Essays in Macroeconomics and Heterogeneity

Christian Martin Rörig

Faculty of Economics

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

St. Catherine's College   September 2021

To my loving parents, grandparents and sisters.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60,000 words.

Christian Martin Rörig
September 2021

Acknowledgements

Essays in Macroeconomics and Heterogeneity

Christian Martin Rörig

Summary

This thesis contains four chapters, addressing two separate but equally prominent research areas in macroeconomics. The first two chapters focus on the effects of firm heterogeneity and firm dynamics on macroeconomic outcomes, particularly how the transmission of macroeconomic shocks is impacted by strategic behaviour and the distribution of financially constrained firms. The last two chapters address the topic of identification in macroeconomics with the last chapter focusing on the measurement of heterogeneous dynamic effects after a macroeconomic policy shock.

The first chapter examines how oligopolistic competition, firm heterogeneity and entry of firms alter the transmission of monetary policy through the lens of a New Keynesian model. The standard textbook New Keynesian model relies on counter-cyclical profits for monetary policy to have an effect as recently pointed out by Broer et al. (2020). This paper first tests this hypothesis by adding oligopolistic competition and static free entry to an otherwise standard NK model. Including search and matching frictions to the labour market, breaks the result of money neutrality despite free entry of firms. Beyond that the model is further augmented to accommodate firm heterogeneity and dynamic entry to investigate how the market structure affects the propagation of monetary policy shocks. The findings are threefold: First, through the channel of oligopolistic competition, heterogeneity in firms' productivity leads to pro-cyclical profits for lower levels of wage rigidity compared to the homogeneous case. Second, firm heterogeneity increases the response in aggregate output which is in line with Mongey (2017). Third, dynamic entry is further enhancing this effect, yet the strengthening of competition has a negative impact on firms' profits. Local projections using Compustat firm-level data and Romer & Romer (2004) monetary policy shocks support the predictions of the theoretical model.

The second chapter, which is co-authored with Miguel H. Ferreira and Timo Haber, investigates how distributional properties of financially constrained firms shape macroeconomic outcomes. Using a unique dataset covering the universe of Portuguese firms and their credit situation we revisit the relation between firm size, their financial situation, and sensitivity to the cycle. First, we provide two stylized facts: (1) Financially constrained firms react more to the business cycle and this mechanism is orthogonal to the size channel proposed by Crouzet & Mehrotra (2020). (2) Constrained firms are found across the

entire size distribution, also in the top percentiles, which is in contrast to what standard financial friction models would predict. We then show that ex-ante heterogeneity of firms, a possible explanatory factor, persists over the firms' life cycle and affects constrained and unconstrained firms differently. Incorporating this ex-ante heterogeneity into an otherwise standard financial frictions model simultaneously accounts for the stylized facts, gives rise to large constrained firms, and leads to larger aggregate fluctuations and capital misallocation.

In the third chapter, together with my supervisor Pontus Rendahl, we investigate the effect of saving shocks on business cycle fluctuations. A common modeling tool – and a popular narrative – used to explain the financial crisis of 2008-2010 is a sudden increase in the desire to save. Such "marginal propensity to save" (MPS) shocks can be triggered by, for instance, a rise in uncertainty surrounding the economic climate, and depress interest rates, inflation, and generally cause an economic contraction. This paper uses the long-run properties arising from MPS shocks in both exogenous- and endogenous growth models with sticky prices in order to identify their causal effect on output. We find that time series data from the United States is strongly supportive of the notion that MPS shocks indeed have a causal, and contractionary, effect on economic activity, lending support to the most common approach of studying the financial crisis.

The fourth chapter is co-authored with Adrian Ochs and aims to improve and facilitate the measurement of heterogeneous macroeconomic policy effects by using machine learning techniques. Specifically, the paper proposes a flexible framework to identify state-dependent effects of macroeconomic policies. In the literature it is common to either estimate constant policy effects or introduce state-dependency in a parametric fashion. This, however, demands prior assumptions about the functional form. Our new method allows to identify state-dependent effects and possible interactions in a data-driven way. Specifically, we estimate heterogeneous policy effects using semi-parametric varying-coefficient models in an otherwise standard VAR structure. While keeping a parametric reduced form for interpretability and efficiency, we estimate the coefficients as functions of modifying macroeconomic variables, using random forests as the underlying non-parametric estimator. Simulation studies show that this method correctly identifies multiple states even for relatively small sample sizes. To further illustrate our method, we apply the semi-parametric framework to the historical data set by Ramey & Zubairy (2018) and offer a more granular perspective on the dependence of the fiscal policy efficacy on unemployment and interest rates.

Contents

Contents

Contents

List of figures

List of tables

Chapter 1

Monetary policy with oligopolistic competition, firm heterogeneity and entry

Abstract

The standard textbook New Keynesian model relies on counter-cyclical profits for monetary policy to have an effect as recently pointed out by Broer et al. (2020). This paper first tests this hypothesis by adding oligopolistic competition and static free entry to an otherwise standard NK model. Including search and matching frictions to the labour market, breaks the result of money neutrality despite free entry of firms. Beyond that, the model is further augmented to accommodate firm heterogeneity and dynamic entry to investigate how the market structure affects the propagation of monetary policy shocks. The findings are threefold: First, through the channel of oligopolistic competition, heterogeneity in firms' productivity leads to pro-cyclical profits for lower levels of wage rigidity compared to the homogeneous case. Second, firm heterogeneity increases the response in aggregate output which is in line with Mongey (2017). Third, dynamic entry is further enhancing this effect, yet the strengthening of competition has a negative impact on firms' profits. Local projections using Compustat firm-level data and Romer & Romer (2004) monetary policy shocks support the predictions of the theoretical model.

Keywords: Monetary policy, oligopolistic competition, heterogeneous firms, firm entry.
JEL Codes: E62, E22, E23

1

1.1   Introduction

The standard framework to study the monetary transmission mechanism, which describes how interest rate changes affect the rest of the economy, has been the New Keynesian (NK) model. Counter-cyclical fluctuations in the firms' profits play a crucial role in this process: A decrease in dividends received by households raises labour supply through a wealth effect and the mere presence of profits reduces the offsetting negative income effect of the wage rise. This ultimately leads to an increase in hours worked and subsequently aggregate output. Muting the channel of firms' profits received by households as done in Broer et al. (2020) by introducing capitalists into a two-agent version of the NK model or by allowing for free entry of firms via a zero-profit condition as done in Bilbiie (2017) completely restores the frictionless equilibrium of money neutrality. Hence, profits seem to be an under-appreciated feature of the transmission mechanism of monetary policy. Profits are influenced by the market structure, which in turn is a function of the type of firm competition, concentration and potential entry and exit of firms. Yet, the standard NK framework usually restricts itself to the assumption of static monopolistic competition. This paper contributes to the question of how market structure affects the transmission of monetary policy and vice versa, since monetary policy shocks might alter the endogenous market structure. More specifically, I examine how the different determinants of market structure – competition, market concentration and firm entry – contribute to shaping the aggregate effect of monetary policy.

   This paper relaxes the assumption of monopolistic competition in an otherwise standard NK model with nominal rigidities by introducing oligopolistic competition as in Atkeson & Burstein (2008) and entry of firms. The finding of Bilbiie (2017) that free entry leads to money neutrality even under sticky prices is replicated for the oligopolistic case. As a remedy to restore monetary non-neutrality, I include search and matching frictions into the labour market following Christiano et al. (2016). There, households supply labour inelastically, hence acyclical profits imposed by a zero profit condition do not neutralize the transmission of monetary policy. Search and matching frictions guarantee changes in employment despite households supplying labour inelastically. Beyond that, building on the labour market setup proposed by Christiano et al. (2016), I introduce firm heterogeneity and dynamic entry of firms to the oligopolistic framework in order to investigate the effect of market structure on the propagation of monetary policy shocks. I show that larger firms, the ones with higher productivity levels, respond less to changes in marginal cost relative to smaller, less efficient firms. Hence, the propagation of monetary policy shocks

becomes a function of the endogenous market structure, characterised by the amount of competing firms and concentration within the market. This mechanism gives rise to several findings. The main findings are threefold: First, through the channel of oligopolistic competition, heterogeneity in firms' productivity leads to pro-cyclical profits for lower levels of wage rigidity compared to the homogeneous case. Second, firm heterogeneity increases the response in aggregate output which is in line with Mongey (2017). Third, dynamic and pro-cyclical entry is further enhancing this effect, yet the strengthening of competition has a negative impact on firms' profits.

Empirically, the motivation to include multiple strategically interacting firms into the framework is straightforward, as the majority of product markets are not made up by a sole monopolist. Furthermore, most markets are highly concentrated. As Mongey (2017) points out, the median effective number of firms is only 3.7[1]. Moreover, entry is a non-negligible aspect of most product markets and an important feature to explain business cycle fluctuations as shown in recent studies. Broda & Weinstein (2010) for instance provide evidence that 35% of an increase in aggregate sales stems from newly introduced products. Also Bilbiie et al. (2012) emphasize the importance of new products, including those produced by existing firms, as source of aggregate output fluctuations.

Literature. While the monetary policy literature usually abstracts from firm dynamics, thus relying on monopolistic competition with static product varieties, research on business cycles and on trade occasionally accounts for oligopolistic market structures and entry, yet mostly separated. The underlying project hence combines several strands of literature to investigate the effects of firm dynamics on the transmission mechanism of monetary policy.

Pioneered by Hopenhayn (1992)'s model of industry dynamics, many papers incorporated firms' entry and exit in macroeconomics models to study their effects on real business cycles. Contributions that endogenized firms' entry are for instance Jaimovich & Floetotto (2008), Lewis (2009) and Bilbiie et al. (2012) among others. Bilbiie et al. (2012) showed that variation in the number of producers and products through entry and exit can be an important propagation mechanism for fluctuations. Within their setup they are able explain stylized facts such as the pro-cyclical behaviour of entry and profits[2]. In his seminal paper, Melitz (2003) extended the model of Hopenhayn (1992) to accommodate heterogeneity in firms' productivity. His model captures the difference in firms'

---

[1]A measure of market concentration computed as the inverse of the Herfindahl index

[2]Specifically, their assumption of translog preferences results in counter-cyclical markups with pro-cyclical profits.

responses to being exposed to trade. Trade induces only the more productive firms to export while simultaneously forcing the least productive firms to exit the market. Hence, market shares and profits are reallocated to the more productive firms. This project aims to capture potentially different responses of firms to shocks in the nominal interest rate. Papers studying the effect of entry in the context of monetary policy include Bergin & Corsetti (2008), Bilbiie et al. (2014) and Bilbiie (2017). The latter points out that free entry of firms restores the frictionless equilibrium of money neutrality even under sticky prices. However, all of them rely on the standard assumption of monopolistic competition.

Effects of strategically competing firms have been mainly studied within real business cycle models being agnostic about any effects on the transmission of monetary policy as in Etro & Colciago (2010), Colciago & Etro (2010) and Colciago & Rossi (2015) among others. Etro & Colciago (2010) show in a real business cycle model with endogenous market structures that a temporary positive productivity shock fosters entry and consequently strengthens competition which in turn reduces markups temporarily and increases real wages. They argue that oligopolistic competition creates an intertemporal substitution effect which boosts consumption and employment. Oligopolistic competition in trade models was introduced by Brander (1981) and extended with free-entry by Brander & Krugman (1983). Atkeson & Burstein (2008) renewed the interest in oligopolistic market structures to explain some trade features and more recent contributions in that area include Impullitti et al. (2017) and Impullitti & Licandro (2018) who combine oligopolistic competition and free entry of firms to study the gains from trade.

The first applications in the NK framework have been developed by Faia (2012) and Lewis & Poilly (2012). The most recent study in an NK framework capturing oligopolistic competition and entry stems from Etro & Rossi (2015) who derive a New Keynesian Phillips curve under Calvo staggered pricing and endogenous market structures with Bertrand competition. They find that both strategic interactions and endogenous business creation strengthen nominal rigidities. This reduces the slope of the Phillips curve and consequently amplifies the real effects of monetary policy shocks.

Moreover, the monetary policy literature has traditionally focused on the aggregate effect of policy shocks. Questions on how micro-level heterogeneity and distributional effects change the transmission of monetary policy are of quite recent interest (see e.g. Sterk & Tenreyro (2015), Cloyne et al. (2016), Auclert (2017), Kaplan et al. (2018)). However, these papers have mainly focused on heterogeneity in households not across firms. Among the first to introduce heterogeneity between firms in a NK model are Ottonello & Winberry (2018) who study the role of heterogeneity in firms' financial positions in determining

the investment channel of monetary policy. In a heterogeneous firm NK model they show that firms with high default risk are less responsive to monetary shocks because their marginal cost of external finance is high. Probably closest related to this project is the paper by Mongey (2017) who proposes an equilibrium menu cost model with a continuum of sectors, each consisting of two strategically engaged firms competing in prices and facing heterogeneity in product demand. He shows that compared to a model with monopolistically competitive sectors, the duopoly model features a smaller inflation response to a monetary shock and an output response which is more than twice as large. This project departs from Mongey (2017) in allowing for endogenous entry and selection and a different form of oligopolistic competition.

Outlook. The paper is structured as follows: First, section 1.2 introduces the reader to the baseline model for symmetric firms, followed by a note on static free entry and money non-neutrality in section 1.3. Section 1.4 lays out the extended model to accommodate heterogeneity in firms' productivity and introduces dynamic entry. The calibration of the theoretical model is explained in detail in section 1.5. Section 1.6 analyses the effect of oligopolistic competition, firm heterogeneity and entry on the propagation of monetary policy shocks. Empirical evidence for heterogeneous responses to monetary policy is presented in section 1.7. Finally, section 1.8 concludes.

1.2   Baseline Model

The baseline model is augmenting the standard NK model by oligopolistic competition and search and matching frictions in the labour market as this forms the basis for the monetary policy analysis later on. At first, this model will be compared to a model with oligopolistic competition but with a simplistic competitive labour market[3] for the case of static free entry of firms.

Time is discrete and infinite. The model environment consists of a representative household obtaining utility from consuming the final good, a final good producer using the intermediate goods as input and firms producing the intermediate good. Intermediate producers face nominal rigidities which are modelled as a quadratic adjustment cost following Rotemberg (1982). Within intermediate sectors, firms compete strategically à la Cournot and are symmetric in their marginal cost for now. The labour market is modelled separately and exhibits search and matching frictions following the framework

---

[3]Described in detail in the appendix in section A.1.

of Mortensen & Pissarides (1994). The bargaining process follows the alternating offer bargaining as proposed in Christiano et al. (2016) in order to introduce endogenous wage inertia. A monetary authority sets nominal interest rates and monetary policy shocks are the only disturbance in the model economy.

1.2.1 Households

There is a continuum of infinitely lived identical households. The representative household has a unit measure of labour which it supplies inelastically to the Diamond-Mortensen-Pissarides labour market. A fraction of L_t of the household is employed and receives a real wage w_t. The residual fraction $1 - L_t$ is unemployed receiving unemployment benefits D_t from the government.[4] There is perfect consumption insurance within each household so that each member is provided with the same level of consumption. The household's preferences are consequently the equally weighted average of the preferences of its members. The representative household maximizes the following lifetime utility:

$$\max_{C_t, B_{t+1}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(C_t), \tag{1.1}$$

subject to the nominal budget flow equation of the household

$$C_t P_t + B_{t+1} = B_t(1 + i_t) + w_t P_t L_t + (1 - L_t) P_t D_t + \pi_t, \tag{1.2}$$

where C_t and B_{t+1} denote the consumption choice and risk-free bond purchases by the household in period t. The household takes equilibrium real wages w_t and the equilibrium gross interest rate of $1 + i_t$ as given. Furthermore, π_t denotes nominal firm profits after taxes, P_t denotes the price level of the final good. The first order condition yields the Euler equation of the household:

$$u'(C_t) = \beta \mathbb{E}_t \left[(1 + i_t) \frac{P_t}{P_{t+1}} u'(C_{t+1}) \right]. \tag{1.3}$$

1.2.2 Production

Production is modelled according to the oligopolistic competition framework proposed by Atkeson & Burstein (2008). There is a representative competitive final goods firm which

---

[4]Unemployment benefits are financed by a lump-sum tax which is deducted from the firms' profits before the household receives the residual.

aggregates intermediate goods according to a constant-elasticity of substitution (CES) technology as in Dixit & Stiglitz (1977). Furthermore, there is a continuum of sectors producing an intermediate good. The market structure of each such sector is an oligopoly made up of n identical firms (for now). Within each sector, firms are competing à la Cournot and produce output using labour only. Prices are sticky, as firms have to pay a quadratic adjustment cost when changing prices.

Final Good Producer. The final good Y_t is produced by a competitive firm with a CES production function, taking a continuum of measure one of intermediate goods as input,

$$Y_t = \left(\int_0^1 Y_t(i)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \quad i \in [0,1], \tag{1.4}$$

where $Y_t(i)$ denotes output of sector i and σ is the elasticity of substitution across intermediate goods. Intermediate goods are imperfect substitutes (i.e. $\sigma < \infty$). The final good producer buys the intermediate goods and maximizes profits according to

$$\max_{Y_t(i)} P_t \left(\int_0^1 Y_t(i)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} - \int_0^1 P_t(i) Y_t(i) di, \tag{1.5}$$

where $P_t(i)$ denotes the price for the intermediate good i. Maximizing profits yields the demand function for each intermediate good i

$$Y_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\sigma} Y_t, \tag{1.6}$$

The relative demand for intermediate i is hence a function of its relative price. The price index can be derived using nominal output as

$$P_t Y_t = \left(\int_0^1 P_t(i) Y_t(i) di \right).$$

Plugging in the demand for each variety, we get

$$P_t Y_t = \left(\int_0^1 P_t(i)^{1-\sigma} P_t^{\sigma} Y_t di \right),$$

which boils down to the following expression for the aggregate price level

$$P_t = \left(\int_0^1 P_t(i)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}. \tag{1.7}$$

Intermediate Producers. There are $n_t(i)$ intermediate producers $j \in [1, n_t(i)]$ in each sector i at time t. Sector output is the CES aggregate of all intermediate goods in that sector:

$$Y_t(i) = n_t(i)^{\frac{1}{1-\zeta}} \left(\sum_{j=1}^{n_t(i)} Y_t(i,j)^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}, \tag{1.8}$$

where $Y_t(i,j)$ is the amount produced by firm j in industry i at time t. We assume that goods within a sector are more substitutable than goods across sectors, hence $1 < \sigma < \zeta$. Following Jaimovich & Floetotto (2008), I abstract from a variety effect by multiplying with $n_t(i)^{\frac{1}{1-\zeta}}$. Again, analogous to above, the sectoral price index $P_t(i)$ is given by:

$$P_t(i) = n_t(i)^{\frac{\zeta}{1-\zeta}} \left(\sum_{j=1}^{n_t(i)} P_t(i,j)^{1-\zeta} \right)^{\frac{1}{1-\zeta}},$$

and the inverse demand function for goods within a sector is given by:

$$Y_t(i,j) = \left(\frac{P_t(i,j)}{P_t(i)} \right)^{-\zeta} Y_t(i).$$

Furthermore, each firm produces according to a constant returns to scale production function with hired labour, $h_t(i,j)$, as the sole input elevated by total factor productivity A_t

$$Y_t(i,j) = A_t h_t(i,j). \tag{1.9}$$

Intermediate oligopolists face a common price for hiring workers. As prices set by the intermediate producers are sticky (see below) oligopolists cannot freely adjust prices to maximize profits but they will minimize cost. The minimization problem is subject to the constraint of producing enough to meet the demand for the intermediate good

$$\min_{h_t(i,j)} P_t^h h_t(i,j),$$

subject to

$$A_t h_t(i,j) \geq Y_t(i,j),$$

where P_t^h is the price for hiring one unit of labour $h_t(i,j)$. Taking first order conditions we get that marginal cost, mc_t, equals the cost to hire one unit of labour divided by its productivity

$$mc_t = \frac{P_t^h}{A_t} = \frac{\vartheta_t}{A_t} P_t, \tag{1.10}$$

where ϑ_t is the relative price of hiring one unit of labour to the price of the homogeneous final product ($\vartheta_t = \frac{P_t^h}{P_t}$). Marginal cost is identical across all sectors and firms as we abstract from any heterogeneity in productivity for now.

Intermediate firms' optimization problem. Since firms are finite and not atomistic, they take their impact in sectoral output into account. More specifically, as they compete à la Cournot they internalize the quantities produced by their competitors when optimizing their production plan. Moreover, as changing prices comes with a quadratic adjustment cost in the spirit of Rotemberg (1982), the firms' optimization problem is intertemporal and each firm consequently discounts future profits when optimizing quantities according to a stochastic discount factor $m_{t,t+\tau}$. The profit maximization problem for firm j in sector i can be written as follows

$$\max_{Y_{t+\tau}(i,j)} \mathbb{E}_t \sum_{s=0}^{\infty} m_{t,t+\tau} \left(P_{t+\tau}(i,j) Y_{t+\tau}(i,j) - \frac{\vartheta_{t+\tau}}{A_{t+\tau}} P_{t+\tau} Y_{t+\tau}(i,j) - \frac{\theta_p}{2} \left(\frac{P_{t+\tau}(i,j)}{P_{t+\tau-1}(i,j)} - 1 \right)^2 P_{t+\tau}(i,j) Y_{t+\tau}(i,j) \right),$$

$$s.t. \qquad P_t(i,j) = \left(\frac{Y_t(i,j)}{Y_t(i)} \right)^{-1/\zeta} P_t(i), \tag{1.11}$$

where θ_p is the Rotemberg coefficient for quadratic price adjustment cost.

The first order condition implies that the optimal sectoral price (firms are symmetric) is a markup over marginal cost[5]

$$P_t(i,j) = \mu_t(i,j) \frac{\vartheta_t}{A_t} P_t, \tag{1.12}$$

where

$$\mu_t(i,j) = \frac{\Theta_t(i,j)}{(\Theta_t(i,j) - 1) \left[1 - \frac{\theta_p}{2} (\pi_t(i) - 1)^2 \right] + \theta_p \pi_t(i) (\pi_t(i) - 1) - \Gamma_t(i)}$$

$$\Theta_t(i,j) = \left[\frac{1}{\zeta} + \left(\frac{1}{\sigma} - \frac{1}{\zeta} \right) s_t(i,j) \right]^{-1} \tag{1.13}$$

$$\Gamma_t(i) = \theta_p \mathbb{E}_t \left[m_{t,t+1} \pi_{t+1}(i)^2 (\pi_{t+1}(i) - 1) \frac{Y_{t+1}(i) n_t}{Y_t(i) n_{t+1}} \right]$$

with $s_t(i,j)$ denoting the market share of firm j in sector i defined as $\left(\frac{P_t(i,j)}{P_t(i)} \right)^{1-\zeta}$. For symmetric firms, the firm specific index j can be omitted and the market share is simply $\frac{1}{n_t(i)}$ with $n_t(i)$ being the number of firms in industry i. $\pi_t(i)$ is the inflation of the sectoral

---

[5]For a detailed derivation please see appendix A.2.

price. For flexible prices ($\theta_p = 0$) and symmetric firms the markup reduces to

$$\mu_t(i) = \frac{\Theta_t(i)}{\Theta_t(i) - 1},\tag{1.14}$$

with

$$\Theta_t(i) = \left[\frac{1}{\zeta} + \left(\frac{1}{\sigma} - \frac{1}{\zeta}\right)\frac{1}{n_t(i)}\right]^{-1}.\tag{1.15}$$

For an increasing number of firms $n_t(i)$ in a sector, the sectoral market structure becomes monopolistic with a markup of $\frac{\zeta}{\zeta-1}$ as products are imperfect substitutes also within each sector. Furthermore, the monopolistic case is also nested if both elasticities of substitution are equal, i.e. $\sigma = \zeta$. If there is a single firm per sector, the markup equals $\frac{\sigma}{\sigma-1}$.

1.2.3 Personnel agency, workers and the labour market

The labour market is modelled separately to the intermediate producers' problem as in Christiano et al. (2016). They adjust the workhorse search and matching framework by Mortensen & Pissarides (1994) to resolve the Shimer (2005) puzzle. Shimer (2005) shows that a standard DMP labour market does not produce realistic volatility of labour market variables following a technological shock. However, Christiano et al. (2016) show that adding hiring cost (Pissarides, 2009) and a bargaining process known as alternating offer bargaining (AOB)[6] produces realistic responses of employment and posted vacancies with respect to what is observed in the data. The reasons to employ this labour market setup are twofold: First, it induces wage rigidity endogenously and solves the Shimer (2005) puzzle making it preferable to a standard Calvo model of wage rigidity. Secondly, it allows us to investigate in more detail how labour market variables react to firm dynamics compared to a simplistic competitive labour market. The following section lays out the DMP labour market model with hiring cost and alternating offer bargaining as in Christiano et al. (2016).

Personnel agencies post vacancies and workers search for employment. The law of aggregate employment, L_t, is given by

$$L_t = (\rho + x_t)L_{t-1},\tag{1.16}$$

where ρ is the probability that a match between a personnel agency and a worker continues to be productive in the next period. Hence, ρL_{t-1} denotes the number of workers who

---

[6]First adoption in this context by Hall & Milgrom (2008).

have been employed in $t-1$ and remain employed in t. Furthermore, $x_t L_{t-1}$ denotes the number of new meetings between firms and workers at the start of period t. Following Christiano et al. (2016) meetings always result in employment, hence x_t can be interpreted as the hiring rate. The number of workers searching for work at the start of each period is the sum of workers who were unemployed in period $t-1$, which is given by $1-L_{t-1}$, and the number of workers who separate from firms at the end of $t-1$, which is given by $(1-\rho)L_{t-1}$. Hence, the probability of finding a job f_t is given by

$$f_t = \frac{x_t L_{t-1}}{1-\rho L_{t-1}}. \tag{1.17}$$

Personnel agencies earn revenues by renting out workers to intermediate goods producers. If a firm wishes to meet a worker in period t it must post a vacancy which costs s in real terms. The vacancy is filled with probability Q_t. In case the vacancy is filled, the agency must pay a fixed cost of κ before being able to bargain with the new worker. The fixed cost of meeting a worker delivers more volatility in the job finding rate and hence a more volatile response of unemployment to cyclical productivity shocks. Interpretations for the fixed cost component can be one-off negotiation costs, administrative costs or training costs (Pissarides, 2009). The value to the firm of each worker match can be expressed as follows

$$J_t = \vartheta_t^p - w_t^p, \tag{1.18}$$

where ϑ_t^p denotes the expected present value of the flow of revenues from hiring out the employees over the duration of the match, given the real price of renting a worker $\vartheta_t = \frac{P_t^h}{P_t}$. Analogously, w_t^p denotes the present value of the real wage, $w_t = \frac{W_t}{P_t}$. Hence, in recursive form the present values can be written as follows

$$\vartheta_t^p = \vartheta_t + \rho E_t m_{t+1} \vartheta_{t+1}^p, \qquad w_t^p = w_t + \rho E_t m_{t+1} w_{t+1}^p, \tag{1.19}$$

where $m_{t,t+1}$ is the stochastic discount factor as discussed above. Free entry of personnel agencies guarantees that in equilibrium, the expected value of a match has to equal its cost

$$Q_t(J_t - \kappa_t) = s_t, \tag{1.20}$$

where κ_t denotes the cost of hiring a worker and s_t denotes the cost of search in form of posting vacancies. The left hand side expresses the expected value of hiring a worker which has to equal the cost to find a match. Let furthermore V_t denote the value of an employed worker which can be expressed as the sum of the expected present value of

wages earned while the match endures and the continuation value when worker and firm separate

$$V_t = w_t^p + A_t, \tag{1.21}$$

and

$$A_t = (1 - \rho)E_t m_{t+1} \left[f_{t+1} V_{t+1} + (1 - f_{t+1})U_{t+1} \right] + \rho E_t m_{t+1} A_{t+1}. \tag{1.22}$$

The variable U_t denotes the value of an unemployed worker, which is simply the sum of unemployment benefits and the continuation value of unemployment

$$U_t = D_t + \tilde{U}_t, \tag{1.23}$$

where again $\tilde{U}_t$ denotes the continuation value of unemployment

$$\tilde{U}_t = E_t m_{t+1} \left[f_{t+1} V_{t+1} + (1 - f_{t+1})U_{t+1} \right]. \tag{1.24}$$

Defining labour market tightness as the number of vacancies posted by firms $v_t L_{t-1}$ relative to all workers looking for jobs

$$\Xi_t = \frac{v_t L_{t-1}}{1 - \rho L_{t-1}}. \tag{1.25}$$

We can write the vacancy filling rate and the job finding rate as follows assuming that they are related to market tightness

$$f_t = \eta_m \Xi_t, \quad Q_t = \eta_m \Xi_t^{-\eta}, \quad \eta_m > 0, 0 < \eta < 1. \tag{1.26}$$

where η_m is the level parameter in the matching function and η the matching function parameter.

1.2.4 Alternating Offer Bargaining

Instead of using a simple Nash bargaining rule where both agents, firms and workers, share the surplus according to their bargaining power, we use a version of the alternating offer bargaining game between firms and workers adopted by Christiano et al. (2016) and initially proposed by Hall & Milgrom (2008). The advantage of AOB is that it creates endogenous inertia in wages without the need to impose it in a Calvo fashion exogenously. As Hall & Milgrom (2008) point out, the alternating offer bargaining process is detaching the real wage from general labour market conditions making it less sensitive. Consequently,

AOB is a further remedy to the Shimer (2005) puzzle as it allows firms to enjoy a larger share of the rent when an expansionary shock hits, therefore increasing their incentive to expand employment and vice versa for a negative shock.

This section describes the process of bargaining and the resulting wage arrangement between firms and workers following the model described in Christiano et al. (2016). After all productive matches L_t have been initialised, each worker engages in bilateral bargaining of the wage rate with the personnel agency. Each worker-firm pair takes the outcome of all other wage discussions that period as given. Furthermore, conditional on remaining matched, each pair has beliefs about future wage arrangements. Specifically, they assume that future wages do not depend on the outcome in the current period.

The model frequency is assumed to be quarterly, yet the bargaining game proceeds across M subperiods within each quarter, where M is even. The firm is the first to make a wage offer at the start of the first subperiod. Optimally, the firm would target a wage as low as possible subject to the worker not rejecting it. Immediately after having received the offer, the other party can either accept or reject it. If the worker rejects he can make an counter-offer in the next period which is an even subperiod. This goes on as long as firm and worker reject their offers, respectively, until the last subperiod M when the worker makes a take-it or leave-it offer. There is also a probability δ that one party leaves the negotiations after having rejected an offer and bargaining breaks down.

More formally, a firm is offering exactly the wage $w_{j,t}$ in subperiod $j < M$ and j being odd which satisfies the following indifference condition:

$$V_{j,t} = \max\{U_{j,t}, \delta U_{j,t} + (1 - \delta)[D_t/M + V_{j+1,t}]\}. \tag{1.27}$$

It is assumed that if the agent is indifferent between accepting and rejecting the offer, he accepts it. The left hand side of equation (24) denotes the value of an agent being employed and receiving the offered wage $w_{j,t}$:

$$V_{j,t} = w_{j,t} + \rho E_t m_{t+1} w_{t+1} + A_t,$$

which is simply the sum of current wage and the present discounted value of future wage earnings conditional on the worker-firm match remaining productive. The right hand side of the indifference equation is the maximum over the worker's options if he chooses to reject the wage offer, namely an outside option $U_{j,t}$ and the worker's disagreement payoff. The latter denotes the value of a worker who rejects the offer but intends to male a counter-offer. The outside option is defined as the share of unemployment benefits

13

in period t after j subperiods and the continuation value of unemployment as defines above:

$$U_{j,t} = \frac{M - j + 1}{M} D_t + \tilde{U}_t.$$

Hence the term $\delta U_{j,t}$ reflects the case that talks break down and the worker remains unemployed and receives proportionate benefits. The second term of the disagreement payoff reflects the case that the worker receives unemployment benefits for the subperiod and then makes a counter-offer. The worker offers the highest possible wage subject to the firm not rejecting it. Hence, the offer satisfies the following indifference condition:

$$J_{j,t} = \max\{0, (1 - \delta)[-\phi_t + J_{j+1,t}]\}, \tag{1.28}$$

where the left hand side is the value of a firm which accepts the offer:

$$J_{j,t} = \frac{M - j + 1}{M} \vartheta_t + \rho E_t m_{t+1} \vartheta_{t+1} - (w_{j,t} + \rho E_t m_{t+1} w_{t+1}), \tag{1.29}$$

where again, it is assumed that a worker produces $1/M$ intermediate goods each subperiod following employment. As before, the right hand side reflects the trade off between the firm's outside option (i.e. zero) and it's disagreement payoff. In contrast to the worker, the firm incurs a real cost ϕ_t in order to make a counter-offer. Hence, if the firm intends to disagree, yet to prepare a counter-offer, it faces the cost ϕ_t and the potential gain $J_{j+1,t}$ of accepting $w_{j+1,t}$.

Under the assumption that each bargaining party prefers a disagreement and subsequent counter-offer to their outside option we can solve the bargaining game backwards. Starting with the last subperiod M, the worker offers the highest wage possible as a take-it-or leave it offer subject to the firm not rejecting it. Hence, the firm will be indifferent between accepting and taking the outside option. Therefore we have that:

$$J_{M,t} = 0.$$

Using that and equation 1.29, we can derive the present value of wage flows $w_{j,t}^P$ which are defined above as the sum of current wage and discounted future wages:

$$w_{M,t}^P = \vartheta/M + \tilde{\vartheta}_t^P.$$

Knowing the present value at subperiod M we can the proceed to calculate the present value at $M - 1$ using the worker's indifference condition as it's the firm's turn to make an

offer. Analogously, w_{M-2}^P can be derived using 1.28 and so on. The final solution to the bargaining problem is simply $w_{1,t}^P$, the wage a firm would offer in the first stage of the game. Christiano et al. (2016) show that the linear indifference curves lead to the following closed-form solution for w_t^P:

$$w_t^P = \frac{1}{\alpha_1 + \alpha_2} [\alpha_1 \vartheta_t^P + \alpha_2 (U_t - A_t) + \alpha_3 \phi_t - \alpha_4 (\vartheta_t - D_t)],$$

where

$$\alpha_1 = 1 - \delta + (1 - \delta)^M,$$
$$\alpha_2 = 1 - (1 - \delta)^M,$$
$$\alpha_3 = \alpha_2 \frac{1 - \delta}{\delta} - \alpha_1,$$
$$\alpha_4 = \frac{1 - \delta}{2 - \delta} \frac{\alpha_2}{M} + 1 - \alpha_2,$$

which are strictly positive. Rearranging the terms above gives us the alternating offer bargaining rule:

$$J_t = \beta_1 (V_t - U_t) - \beta_2 \phi_t + \beta_3 (\vartheta_t - D_t), \tag{1.30}$$

with $\beta_i = \alpha_{i+1} / \alpha_i$ for $i \in 1, 2, 3$.

1.2.5 Monetary authority

Lastly, there is a central bank setting the nominal interest rate following a Taylor rule. The central bank takes into account the deviation from price stability as captured by inflation and the deviation of output from its potential as captured by its steady-state measure:

$$\log(R_t / R_{ss}) = \rho_R \log(R_{t-1} / R_{ss}) + (1 - \rho_R) \left[\theta_\pi \log(\pi_t / \pi_{ss}) + \theta_Y \log(Y_t / Y_{ss}) + v_t \right], \tag{1.31}$$

where π_t is gross inflation of the final good price, θ_π is the weight on inflation in the reaction function. Analogously, θ_Y is the weight on the output gap, measured as deviation from steady state output, Y_{ss}. ρ_R is a smoothing parameter and the policy shock v_t is the only aggregate disturbance in the model. The monetary policy shock has unit variance and zero mean. The shock is a one time shock and not autocorrelated.

1.2.6 Aggregation and market clearing

Imposing price stickiness following Rotemberg (1982) has the advantage that there is no price dispersion between firms or sectors and aggregation is straightforward as outputs and consequently prices are identical for the case of symmetric firms. Hence, we have that $\forall (i,j) \in [0,1] \times [1, n_t] : y_t(i,j) = y_t, n_t(i) = n_t, p_t(i,j) = p_t$. Aggregate hired labour is then given by $H_t = n_t h_t$. Aggregate output is obtained by combining the definition of final good and intermediate CES aggregators and the linear production function of the intermediate good

$$
Y_t = \left(\int_0^1 \left(n_t^{\frac{1}{1-\zeta}} \left(\sum_{j=1}^{n_t} y_t^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}} \right)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} = \left(\int_0^1 (n_t y_t)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}},
$$

$$
= \left(\int_0^1 (n_t A_t h_t)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} = A_t H_t. \tag{1.32}
$$

Analogously the price level equals the individual firm's price

$$
P_t = \left(\int_0^1 \left(n_t^{\frac{\zeta}{1-\zeta}} \left(\sum_{j=1}^{n_t} p_t^{1-\zeta} \right)^{\frac{1}{1-\zeta}} \right)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}} = p_t.
$$

Consequently, from equation 1.13 we get that marginal cost is equal to the inverse of the markup

$$
\vartheta_t / A_t = \frac{1}{\mu_t}. \tag{1.33}
$$

Furthermore, the labour market needs to clear, hence hired labour by the intermediate goods sector H_t has to equal the labour employed through the DMP wholesalers L_t,

$$
H_t = L_t. \tag{1.34}
$$

Finally, market clearing demands that the goods market clears and the real adjustment cost as well as the real costs of searching and meeting employees is paid. Hence, it has to hold that

$$
Y_t = C_t + \frac{\theta_p}{2} (\pi_t - 1)^2 Y_t + (s/Q_t + \kappa) x_t L_{t-1}. \tag{1.35}
$$

1.3   Static free entry and monetary non-neutrality

In a very recent paper, Broer et al. (2020) emphasize an under-appreciated feature of the standard textbook NK model. Namely, that both the counter-cyclical response of profits and their steady-state size play a key role for the response of employment and output to monetary policy shocks. With elastic labour supply and preferences in the King-Plosser-Rebelo class, the deviation of total income from labour income is decisive for the response of labour supply. More specifically, as households receive profits lump-sum such a deviation occurs as profits fall following an increase in goods demand and wages. As this leaves households poorer, it triggers the required increase in labour supply to meet the rise in demand. However, if this channel is muted, there is no aggregate response in aggregate employment and output and hence we are back to the result of money neutrality even under sticky prices. This transmission channel is clearly implausible as Broer et al. (2020) point out and greatly at odds with the pro-cyclical response of profits as observed in the data (Christiano et al., 2005).

Independently, in an environment more closely related to this paper, Bilbiie (2017) finds the same result of monetary neutrality for sticky prices with free entry of firms through a zero-profit condition. The condition's logic is the following: firms enter a sectoral market as long as profits are positive and exit as long as they are negative. A fixed cost of production ensures that in equilibrium there is a finite amount of firms in each industry. Bilbiie (2017) shows that for the standard workhorse model of monopolistic competition by Dixit & Stiglitz (1977) that free entry can replace price flexibility even if prices are fixed. Following an expansionary monetary policy shock, demand for each firm goes up increasing labour demand which in turn raises the real wage. Profits fall and as a consequence, under the assumption of free entry, firms exit. Aggregate quantities will not change but the distribution of production is different. There will be less firms (extensive margin), each producing more (intensive margin). Both effects perfectly offset each other. Yet again, the basis for this channel are counter-cyclical profits and elastic labour supply.

Including a DMP labour market on the one hand endogenously introduces wage rigidity and on the other hand ensures that the model is not prone to the neutrality result as in Broer et al. (2020) and Bilbiie (2017) as labour is supplied inelastically and search and matching friction determine employment. In order to demonstrate the different responses for the baseline model with elastic labour supply[7] and the baseline model with a DMP labour market as introduced above, we allow for free entry and exit through a zero-profit

---

[7]As defined in the appendix in section A.1.

17
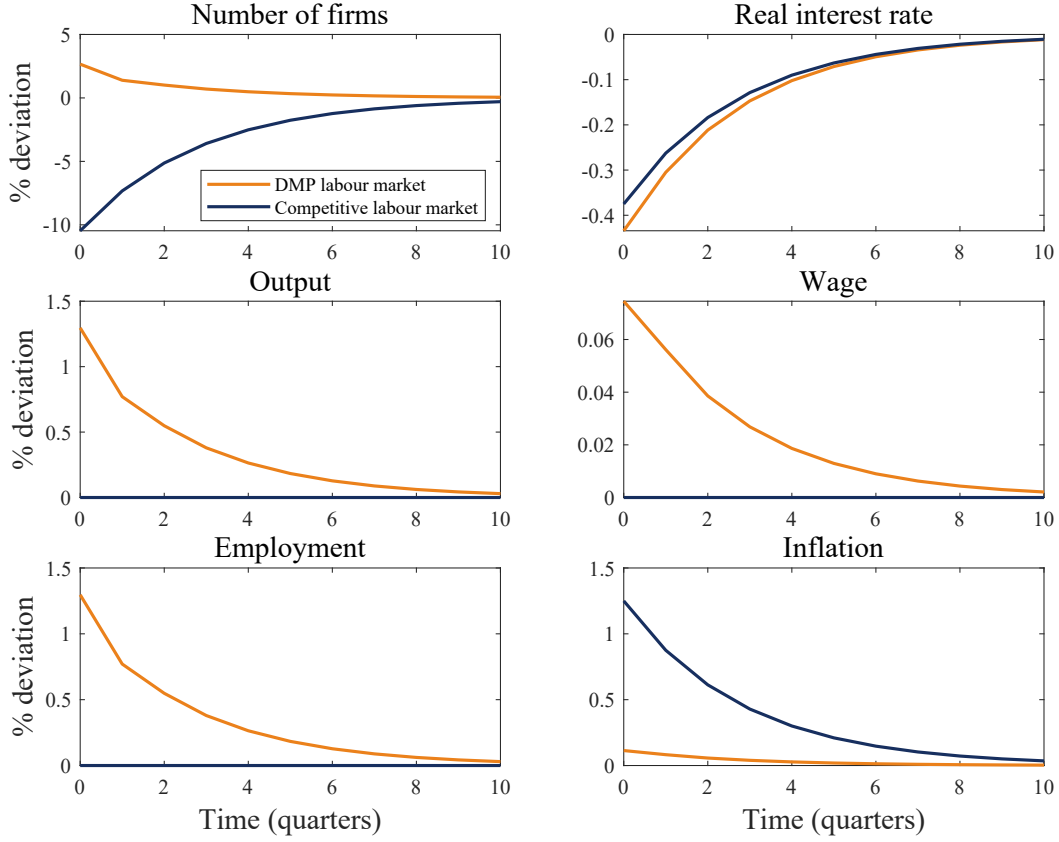
Fig. 1.1 Impulse response functions after an expansionary monetary policy shock.

Notes. Impulse response functions after a one percent shock to the nominal interest rate at $t = 0$. The solid orange line shows the responses of the baseline model with search and matching frictions and a zero-profit condition. The blue line denotes the responses when households supply labour elastically as in the standard textbook NK model (Details in the appendix in section A.1).

condition. Recall that entry is free and entrants use the same production technology as incumbent firms. Consequently, the aggregate zero-profit condition as shown below directly backs out the number of firms in each market:

$$Y_t - \frac{w_t}{A_t} Y_t - \frac{\theta_p}{2} (\pi_t - 1)^2 Y_t - M = 0, \tag{1.36}$$

where M is a fixed cost of production, set exogenously.

Figure 1.1 plots the impulse response functions for both the baseline model with a competitive labour market and with a DMP labour market and alternating offer bargaining. In both settings, the number of firms is determined endogenously through a static zero-profit condition. Both models are calibrated such that the steady-state amount of firms

and the aggregate values of output and employment are the same.[8] In the competitive labour market case, there is no response in aggregate output nor employment but firms exit due to a counter-cyclical response in profits. Furthermore, the absolute change in the number of firms is much larger compared to the DMP labour market scenario as the extensive margin perfectly offsets the intensive margin without leaving any residual response in output. However, if households supply labour inelastically as in the DMP setup, free entry does not neutralize monetary policy with respect to output and employment. Moreover, as the DMP labour market endogenously introduces wage rigidity, free entry leads to pro-cyclical business creation which is consistent with empirical observations (see e.g. Fairlie & Fossen (2018)).

1.4 Generalized model with firm heterogeneity and dynamic entry

This section extends the baseline model by firm heterogeneity and dynamic entry of firms in order to gauge the effect of changes in the market structure and concentration for monetary policy and vice versa.

1.4.1 Intermediate production with heterogeneous firms

There is a fixed number of intermediate producers indexed by j that are organized in a fixed number of heterogeneous productivity categories denoted by s, similar to Andrés & Burriel (2018). Hence, the CES aggregation of the intermediate product can be written as:

$$Y_t = N_t^{\frac{1}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} \left(y_t(j,s) \right)^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}, \tag{1.37}$$

where S is the total number of productivity categories, and N_t^s is the number of firms in productivity class s. Note that the sector index i has been dropped as sectors are symmetric.

The input demand function and the price index associated with the intermediate good producers' optimization are as follows:

$$y_t(i,j,s) = \left(\frac{P_t(i,j,s)}{P_t(i)} \right)^{-\zeta} y_t(i),$$

---

[8]Details on the calibration are covered further down.

$$P_t(i) = N_t^{\frac{\zeta}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} \left(P_t(i,j,s) \right)^{1-\zeta} \right)^{\frac{1}{1-\zeta}},$$

Analogous to equation 1.13, the intermediate firms' profit maximization problem with heterogeneity in productivity can be written as follows:

$$\max_{y_{t+\tau}(i,j,s)} \mathbb{E}_t \sum_{s=0}^{\infty} m_{t,t+\tau} \left(P_{t+\tau}^{j}(i) y_{t+\tau}(i,j,s) - \frac{\vartheta_{t+\tau}}{z(i,s)A_{t+\tau}} P_{t+\tau} y_{t+\tau}(i,j,s) - \frac{\theta_p}{2} \left(\frac{P_{t+\tau}(i,j,s)}{Pt+\tau-1(i,j,s)} - 1 \right)^2 P_{t+\tau}(i,j,s) y_{t+\tau}(i,j,s) \right)$$

$$s.t. \qquad y_t(i,j,s) = \left(\frac{P_t(i,j,s)}{P_t(i)} \right)^{-\zeta} y_t(i) \tag{1.38}$$

where $z(i,s)$ is the idiosyncratic productivity factor of productivity category s in sector i. In order to keep the model tractable it is assumed that there are no shocks to firms' productivity. Furthermore, sectors are symmetric and each sector comprises the same number of firms of each productivity category and the same productivity distribution prevails. The first order condition of the firms' problem gives us an expression for the time-varying markup of firm j of productivity class s which is analogous to above:

$$P_t(i,j,s) = \mu_t(i,j,s) \frac{\vartheta_t}{z(i,s)A_t} P_t, \tag{1.39}$$

where

$$\mu_t(i,j,s) = \frac{\Theta_t(i,j)}{(\Theta_t(i,j)-1)\left[1-\frac{\theta_p}{2}(\pi_t(i)-1)^2\right] + \theta_p \pi_t(i)(\pi_t(i)-1) - \Gamma_t(i)},$$

$$\Theta_t(i,j,s) = \left[\frac{1}{\zeta} + \left(\frac{1}{\sigma} - \frac{1}{\zeta} \right) s_t(i,j,s) \right]^{-1}, \tag{1.40}$$

$$\Gamma_t(i) = \theta_p \mathbb{E}_t \left[m_{t,t+1} \pi_{t+1}(i)^2 (\pi_{t+1}(i)-1) \frac{y_{t+1}(i,j,s)}{y_t(i,j,s)} \right],$$

The expression is equivalent to the baseline model, except here markups differ between firms within a sector with respect to their individual level of productivity and corresponding market share.

Proposition 1 *A firms' markup over its marginal cost increases with its market share. The standard monopolistic case is nested in the model for $s_t(i,j,s) = 1$.*

Log-linearising the markup of firm j around its steady state gives a firm-specific augmented Phillips curve similar to (Benigno & Faia, 2010; Bilbiie et al., 2014; Guilloux-Nefussi, 2016). Although, the concept of the Phillips curve looks at dynamics of inflation at the

aggregate level, we aim to understand how the individual firm behaviour influences the aggregate results. The Phillips curve for a firm in productivity class s is given as

$$\pi_t(s) = \frac{\Theta_{ss}(s) - 1}{\theta_p}\left(\frac{\hat{\vartheta}_t}{z(s)\hat{A}_t} + \hat{\mu}_t(s)\right) + \beta\mathbb{E}_t\pi_{t+1}(s). \tag{1.41}$$

The Phillips curve for the individual firm shows that the sensitivity of price adjustments to a change in marginal cost is decreasing in a firms' steady-state market power. Recall, that $\Theta_{ss}(s)$ is lower, the larger a firm is. Since the steady-state market shares vary for different producers depending on their relative productivity, the aggregate response depends on the market structure.

Proposition 2 *Large firms, the ones with higher productivity levels, respond less to changes in marginal cost as their Phillips curve is flatter.*

1.4.2 Dynamic entry of firms

Entry is dynamic as in Andrés & Burriel (2018). The number of operating firms is endogenous and as mentioned above, firms can have different exogenous steady state levels of productivity. This will consequently lead to different firm sizes. All potential entrants are of the lowest productivity level. Yet, a fraction of firms can transition to the nearest group size each period. More specifically, with probability p_u firms become more productive and subsequently larger and with probability p_l firms shrink in size, respectively. Furthermore, a fraction δ of firms independent of their productivity level stop producing each period. The law of motion for the number of firms in each productivity or size class can therefore be written as

$$N_{t+1}^s = p_u(1-\delta)N_t^{s-1} - p_l(1-\delta)N_t^{s+1} + (1 - p_u - p_l)(1-\delta)N_t^s, \tag{1.42}$$

where the right hand side is the sum of those surviving firms which either increase in size, decrease in size or stay in the same productivity bin s. The overall number of operating firms can be written as the sum of surviving firms across all productivity classes and new entrants

$$N_{t+1} = (1-\delta)N^E + \sum_{s=1}^{S}(1-\delta)N_t^s. \tag{1.43}$$

The exit shock δ and the transition probabilities are assumed to be exogenous while entry is endogenous. Furthermore, there is a mutual fund of firm profits which pays a dividend to the representative household equal to the total amount of nominal profits of all firms

producing in that period. The representative household can choose to buy shares s_t in the mutual fund of all the firms in the economy. The price of a share at time t is equal to the discounted present value of future profits streams averaged across firms which equals the average value of producing firms v_t. Hence, accommodating firm entry, the households budget constraint is given by:

$$C_t P_t + B_{t+1} + \sum_{s=0}^{S} v_t^s N_t^s s_t^s = B_t(1+i_t) + w_t P_t L_t + (1-L_t)P_t D_t$$

$$+ (1-\delta) \sum_{s=0}^{S} s_{t-1}^s N_{t-1}^s (p_u(\pi_t^{s+1} + v_t^{s+1}) + p_l(\pi_t^{s-1} + v_t^{s-1}) + (1-p_u-p_l)(\pi_t^s + v_t^s))$$

where s_t^s denotes the amount of shares and π_t^s denotes the profit streams of firms of productivity class s at period t. The latter part of the right hand side hence, denotes the sum of profits and remaining present values after transition of all surviving firms in the economy at that period. Households maximize over consumption, bonds purchases and investment in firms. First order conditions with respect to consumption and bonds give the Euler equation as above. Optimal investment in firms of each productivity class s has to satisfy:

$$v_t^s = \mathbb{E}_t \left[m_{t,t+1} \left(p_u(\pi_{t+1}^{s+1} + v_{t+1}^{s+1}) + p_l(\pi_{t+}^{s-1} + v_{t+1}^{s-1}) + (1-p_u-p_l)(\pi_{t+1}^s + v_{t+1}^s) \right) \right] \tag{1.44}$$
$$\text{for } s \in \{1,2,...,S\}$$

where $m_{t,t+1} = \beta \frac{u'(C_{t+1})}{u'(C_t)}(1-\delta)$ is the stochastic discount factor.

Each period there is a continuum of potential entrants to the market of producing firms. Entry is subject to frictions. Firms have to pay an exogenous sunk entry cost f^E. Furthermore, firms remain idle for one period before starting to produce. Yet, new entrants face the same structural shocks δ as incumbent firms. By assumption entrants are of the lowest productivity class and can become more productive with probability p_u. Free entry ensures that firms will enter as long as their expected gain from producing equals the entry cost:

$$v_t^E = f^E, \tag{1.45}$$

where the sunk cost is identical across all sectors i and constant over time.

1.4.3 Aggregation and market clearing

Aggregate output is obtained by combining the definition of final good and intermediate CES aggregators and the linear production function of the intermediate good but now aggregating over all productivity groups s:

$$
Y_t = \left(\int_0^1 \left(N_t^{\frac{1}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} y_t^{j,s \frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}} \right)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}
$$

$$
= \left(\int_0^1 \left(N_t^{\frac{1}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} (z^s h_t^{j,s})^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}} \right)^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}
$$

$$
= N_t^{\frac{1}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} (z^s h_t^{j,s})^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}}
$$

where in the third step we used the assumption that sectors are symmetric. For perfect substitutes, i.e. $\zeta \to \infty$, the expression boils down to

$$
Y_t = \sum_{s=1}^{S} \sum_{j=1}^{N_t^s} (z^s h_t^{j,s}), \tag{1.46}
$$

which is the sum of hired labour enhanced by the group specific productivity parameter z^s over all firms j and productivity classes s. As within each productivity class, firms are symmetric, the expression reduces further to $\sum_{s=1}^{S} N_t^s z^s h_t^s$. Total output can also be written using the average productivity across all firms in the economy, $Y_t = \bar{z} H_t$, where $\bar{z}$ is defined as follows for the case of perfect substitutes

$$
\bar{z} = \frac{\sum_{s=1}^{S} N_t^s z^s h_t^s}{\sum_{s=1}^{S} N_t^s h_t^s}. \tag{1.47}
$$

Prices must satisfy the following condition

$$
1 = \sum_{s=1}^{S} \frac{N_t^s}{N_t} \left(\frac{P_t^s}{P_t} \right)^{1-\zeta}. \tag{1.48}
$$

Table 1.1 Non-estimated parameters and calibrated values

Parameter	Value	Description	Reference
Panel A: Parameters			
β	0.9926	Discount factor	Standard
γ	-1	Risk aversion	Standard
α	0.8	Taylor rule smoothing parameter	Standard
θ_π	1.5	Taylor rule inflation coefficient	Standard
θ_y	0.05	Taylor rule output gap coefficient	Standard
θ_p	5	Rotemberg coefficient of price adjustment cost	Standard
A	1	Economy wide TFP	Standard
σ	5	CES parameter across sectors	Atkeson & Burstein (2008)
ζ	10	CES parameter within sectors	Atkeson & Burstein (2008)
ρ	0.9	Job survival probability	Christiano et al. (2016)
M	60	Max. bargaining rounds per quarter	Christiano et al. (2016) adj.
δ	0.002	Probability of talks breaking down	Christiano et al. (2016) adj.
z^1	1	TFP of less productive firms	Fernández & López (2014)
z^2	1.5	TFP of efficient firms	Fernández & López (2014)
δ	2.5%	Business death rate	Jaimovich & Floetotto (2008)
p_u	1%	Probability of becoming more efficient	Own assumption
p_l	2.5%	Probability of becoming less efficient	Own assumption
Panel B: Steady state values			
Q	0.70	Vacancy filling rate	Christiano et al. (2016)
u	0.05	Unemployment rate	Christiano et al. (2016)
D/w	0.4	Unemployment benefits relative to wage earnings	Christiano et al. (2016)
κ/Y	1%	Hiring cost/output	Christiano et al. (2016)
s/Y	0.05%	Vacancy cost/output	Christiano et al. (2016)
$N^1/(N^1+N^2)$	0.80	Share of small firms	Andrés & Burriel (2018)

Market clearing has to take into account the overall amount of sunk entry cost paid at each period

$$Y_t = C_t + \frac{\theta_p}{2}(\pi_t - 1)^2 Y_t + (s/Q_t + \kappa)x_t L_{t-1} + N^E f^E, \tag{1.49}$$

and the labour market has to clear

$$\sum_{s=1}^{S} N_t^s h_t^s = H_t = L_t. \tag{1.50}$$

1.5 Calibration

Table 1.1 documents the chosen parameter values and targeted steady state values for the calibration of the model. The parameters have been selected either using consensus values from the literature or were calibrated to reproduce some moments in the data. Starting with the non-estimated parameters, we use a discount factor of 0.9926 which is consistent with an annualized real interest rate of 3%. The household's utility function

exhibits constant relative risk aversion with $\gamma = -1$ which boils down to the nested case of log-utility. The Taylor rule with interest rate smoothing is specified according to standard values from the literature. Following the oligopolistic framework by Atkeson & Burstein (2008), elasticity of substitution across sectors (σ) has to be lower than within sectors (ζ). The chosen parameters lead to a steady-state markup of 1.25 for a single firm and a 10% markup for an infinite amount of firms within a sector. Recall that also within each sector, intermediate goods are defined as imperfect substitutes. The Rotemberg coefficient for price adjustment cost is chosen such that for a steady state markup of 1.25, the implied frequency of price changes equals a Calvo probability for a price change of 0.75.

The DMP labour market assumes a match survival rate $\rho = 0.9$ which is in line with the values used in Christiano et al. (2016) and Walsh (2003). Furthermore, the amount of sub-periods over which an alternating offer bargaining process could extend is set to $M = 60$ which roughly equals the number of business day in a quarter as pointed out by Christiano et al. (2016). The value for the chance of a sudden breakup in wage negotiations is equal to the bayesian estimate of 0.2 percent by Christiano et al. (2016). In order to increase endogenous wage rigidity induced by the alternating bargaining process, the breakup probability and the number of subperiods are reduced. The model parameters η_m and ϕ are chosen, so that, conditional on the other parameter, the model produces a steady state unemployment rate of 5% and a steady state vacancy filling rate of $Q = 0.7$ which is consistent with Den Haan et al. (2000) and Ravenna & Walsh (2008). Unemployment benefits amount to 40% of prior wage earnings. The variable cost for posting vacancies κ and for meeting potential employees is calibrated such that the total amount of each cost position is 1% of the model economy's total output.

Without loss of generality and following Andrés & Burriel (2018), the model is calibrated for two productivity classes $s = 1, 2$. Hence, there is a number of firms with a low productivity level z^1 and a number of more efficient firms with productivity $z^2 > z^1$ in each sector. Following the estimates by Fernández & López (2014), large firms' TFP is almost two times the productivity of small firms. In the simple model without firm heterogeneity, firms' productivity is set equal to the average productivity of the heterogeneous case. The number of firms are calibrated to reproduce the stylized facts such that most of the firms are small (roughly 80%) but that the majority of workers is employed by large firms. The latter is a result of the productivity differential between heterogeneous firms. As in the model with entry, the number of firms in each productivity class is endogenous, entry cost and transition probabilities are calibrated in order to replicate those facts.

Following Jaimovich & Floetotto (2008) the rate of exogenous business exit is $\delta = 2.5\%$ which determines the ratio of entrants in steady state and reflects empirical observations.

1.6 Monetary policy analysis

This section analyses the effect of a monetary policy shock. In order to gauge the effect of market structure on monetary policy and vice versa, we first compare the baseline model with homogeneous firms to the generalized model with heterogeneous firms yet without entry. Both models are calibrated such that average productivity is identical and they produce the same steady state results of aggregate variables. Second, in a separate analysis, the model is extended by dynamic entry, in order to disentangle the effect of firm heterogeneity and dynamic entry on the transmission of monetary policy. Again, in order to make the results comparable, the same steady states are targeted for both specifications.

1.6.1 Effect of firm heterogeneity on monetary policy

As shown above, we expect that large firms respond less to an increase in marginal cost than less productive firms. Figure 1.2 plots the impulse response functions of selected variables for an expansionary monetary policy shock. The plots shown compare the responses of a model with homogeneous firms, in blue, to the model with heterogeneous firms, in red, for two different levels of wage rigidity. The solid lines denote a regime with less sticky wages than denoted by the dotted lines.

The upper left panel plots the response in aggregate profits for the different model specifications. Clearly, imposing firm heterogeneity, profits respond more for any given value of wage rigidity. Moreover, there are levels of wage rigidity for which a model with homogeneous firms produces counter-cyclical profits, while an environment with heterogeneity in productivity achieves pro-cyclical profits. Hence, cyclicality in aggregate profits appears to depend on the market structure, specifically on the market concentration. The higher the market concentration in a sector, the more pro-cyclical becomes the response in aggregate profits. The reason is that heterogeneous firms face a common wage, yet following a monetary shock large firms will change their prices by less increasing their demand. If the change in demand is sufficiently large, this leads to pro-cyclical profits for the larger firms. As their steady state share of profits is also larger, aggregate profits will ultimately be also pro-cyclical.

Compared to the different responses in aggregate profits, the effect of firm heterogeneity appears to be rather small for the other aggregates plotted. Generally, for a higher level of wage rigidity employment and output react more strongly as firms enjoy a larger share of rent which increases their incentive to expand employment. Heterogeneity in firms' productivity levels amplifies this effect. Large firms, the ones which are more efficient, raise their prices by less than small firms in response to an increase in marginal cost. Given that their market share is higher, aggregate inflation increases also by less than compared to a setting with homogeneous firms. Consequently, imposing firm heterogeneity, firms face on average a higher demand and increase their output by more. This in turn, reallocates labour to the more productive firms, boosting the response in output. As a result to this reallocation of employees, the labour share decreases. Although, the effect is small, the competition channel might contribute to the widely observed trend of a decline in labour shares (Barkai, 2016; Caballero et al., 2017; De Loecker & Eeckhout, 2017). Intuitively, the reallocation shift is larger for higher wage flexibility as this widens the gap between firms' marginal cost and therefore the productivity advantage of large firms compared to small firms becomes more relevant. While the aggregate effect of oligopolistic competition is rather small, heterogeneous firms clearly respond differently to a monetary policy shock as shown in figure 1.3. The figure plots the impulse response functions for selected firm-level variables. For the same two regimes of wage rigidity it shows the responses for a less productive firm versus a more efficient firm. As already indicated above, profits of the large firm will respond pro-cyclical given a minimum level of wage rigidity. However, as small firms lose market share to larger firms, they can face counter-cyclical profits. The average profit response will ultimately depend on how many firms of each type compete within a sector and how large their endogenous steady-state market shares are. The response in prices is as expected less strong for more productive firms, however, the difference is not large. Recall, that by assumption, the elasticity of substitution is bounded by the elasticity across and within sectors, $\Theta \in [\sigma, \zeta]$. Consequently, sales and subsequently market shares respond more for the dominant firms, which in turn reduces the market power of smaller firms. The competition channel hence, shifts labour from less productive firms to more efficient firms. As already indicated above, the shift of market shares and analogously employment from the small firms to the large firms is stronger if wages are less rigid.
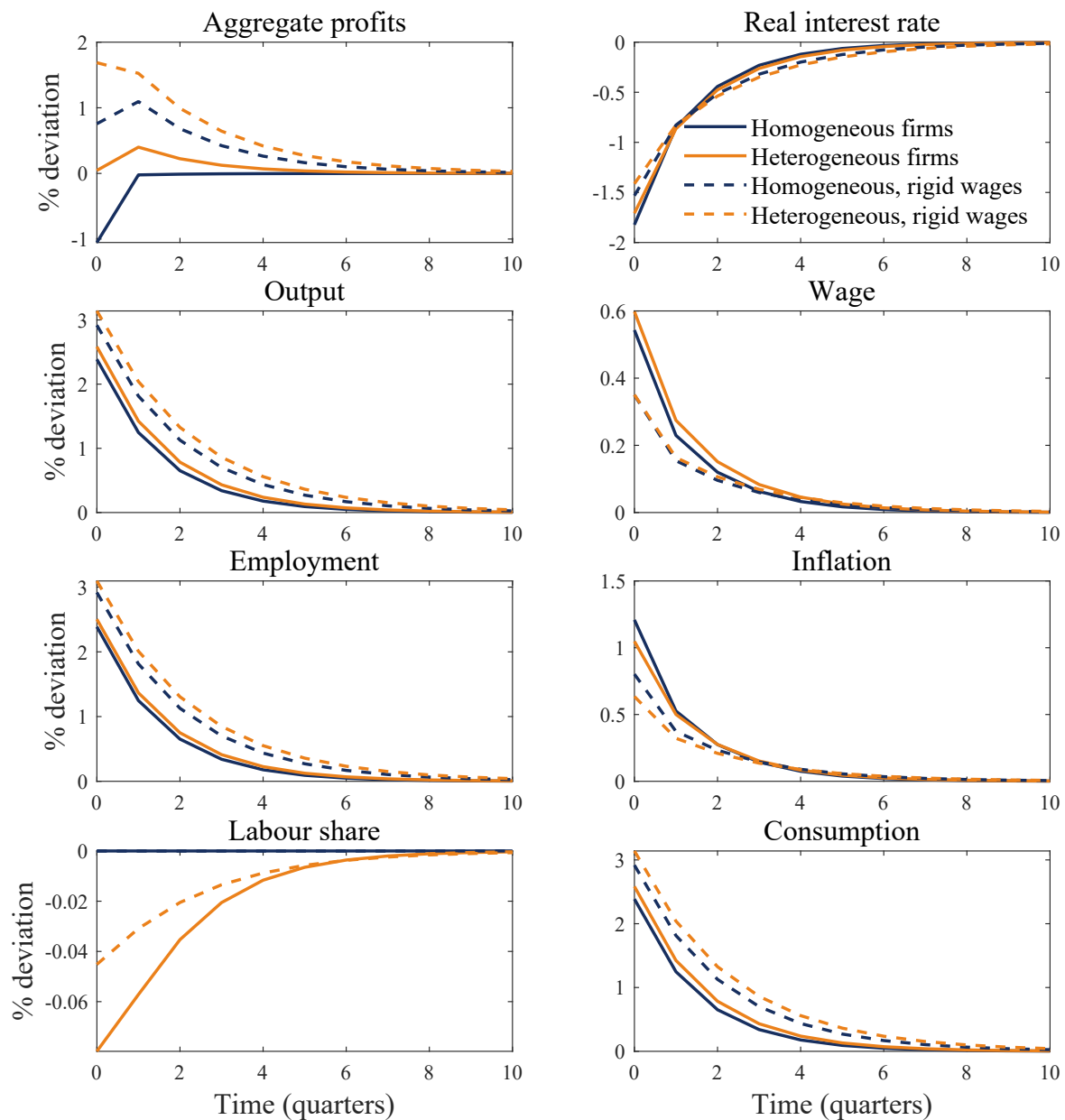
Fig. 1.2 Effect of firm heterogeneity on monetary policy response.

Notes. Impulse response functions after an expansionary one percent shock to the nominal interest rate at $t = 0$. The solid blue line shows the responses of the baseline model without firm heterogeneity. The red line denotes the responses when firms are heterogeneous in productivity. The dashed lines show responses for an increased level of wage rigidity by decreasing the probability of talks breaking down.
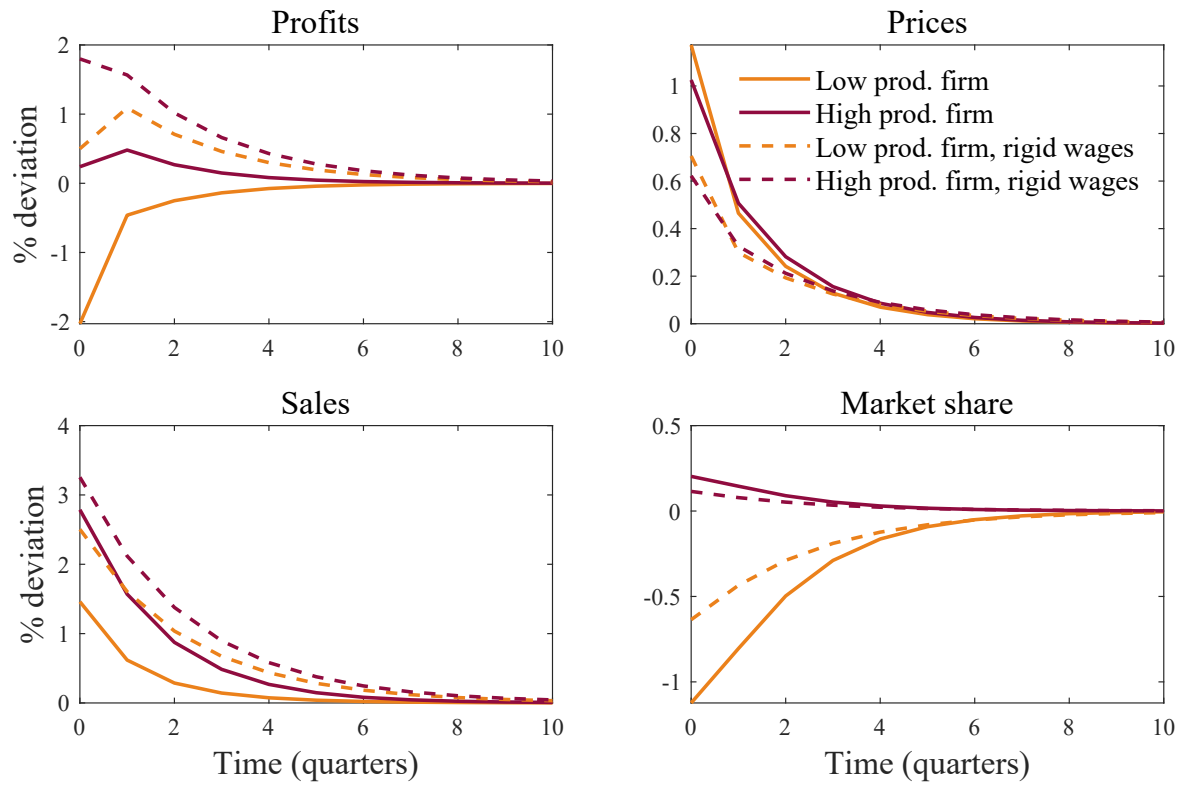
Fig. 1.3 Productivity-specific impulse responses

Notes. Impulse response functions after an expansionary one percent shock to the nominal interest rate at $t = 0$. The solid lines indicate a regime with less rigid wages, whereas the dotted lines denote a regime with higher wage rigidity.

1.6.2 Effect of dynamic entry on monetary policy

Figure 1.4 plots the corresponding impulse response functions for the case of dynamic entry compared to a model with firm heterogeneity but without entry or exit of firms. Two models of dynamic entry are considered, one with symmetric firms and one with firm heterogeneity. Again, all models are calibrated such that they are identical in their steady state values of the main aggregate variables of interest. The most striking result is that dynamic entry seems to strengthen the propagation of a monetary policy shock on output. Pro-cyclical business creation increases the demand for workers which leads to an increase in vacancy postings as shown by the lower left panel. This increases labour market tightness according to equation 1.25 which in turn decreases the vacancy filling rate following equation 1.26. The decrease is due to the concavity of the Cobb-Douglas matching function. Hence, the amount of matches increases by less than the job postings as the number of unemployed workers does not change instantly. Consequently, the hiring rate, shown in the lower right panel, increases and employment rises according to the law of motion defined in equation 1.16. This translates into an increase in output which is higher than in the case of no entry even for homogeneous firms. Due to the assumption that entrants have to wait one period before they produce, the responses of inflation and the real interest rate, respectively, differ significantly only after one quarter. Although, firms raise their prices similarly across models initially, new entrants and consequently a strengthening in oligopolistic competition induces downward pressure on firms' prices. This causes the response in inflation to drop as soon as entrants become productive.

Imposing firm heterogeneity enhances the response of output even further. A larger productivity differential raises the incentive of entering the market as expected profits are higher. Consequently, the immediate response in entrants is much higher for the case of firm heterogeneity as shown in the upper left panel. However, the lower the transition probability to the higher productivity level, the less pronounced is the effect. If the response in the number of entrants is strong enough, this could even lead to a decrease of inflation following an expansionary monetary policy shock after new firms become productive. Hence, pro-cyclical entry and subsequently enhanced competition might contribute to the price puzzle[9] often observed in the data.

Figure 1.5 illustrates the effects of firm heterogeneity and entry at the firm-level in response to an expansionary monetary policy shock. Again, large firms reset their prices by less than small firms and therefore, face a higher demand. Yet, the initial increase in

---

[9]Refers to a rise (fall) in the aggregate price level in response to a contractionary (expansionary) monetary policy shock.
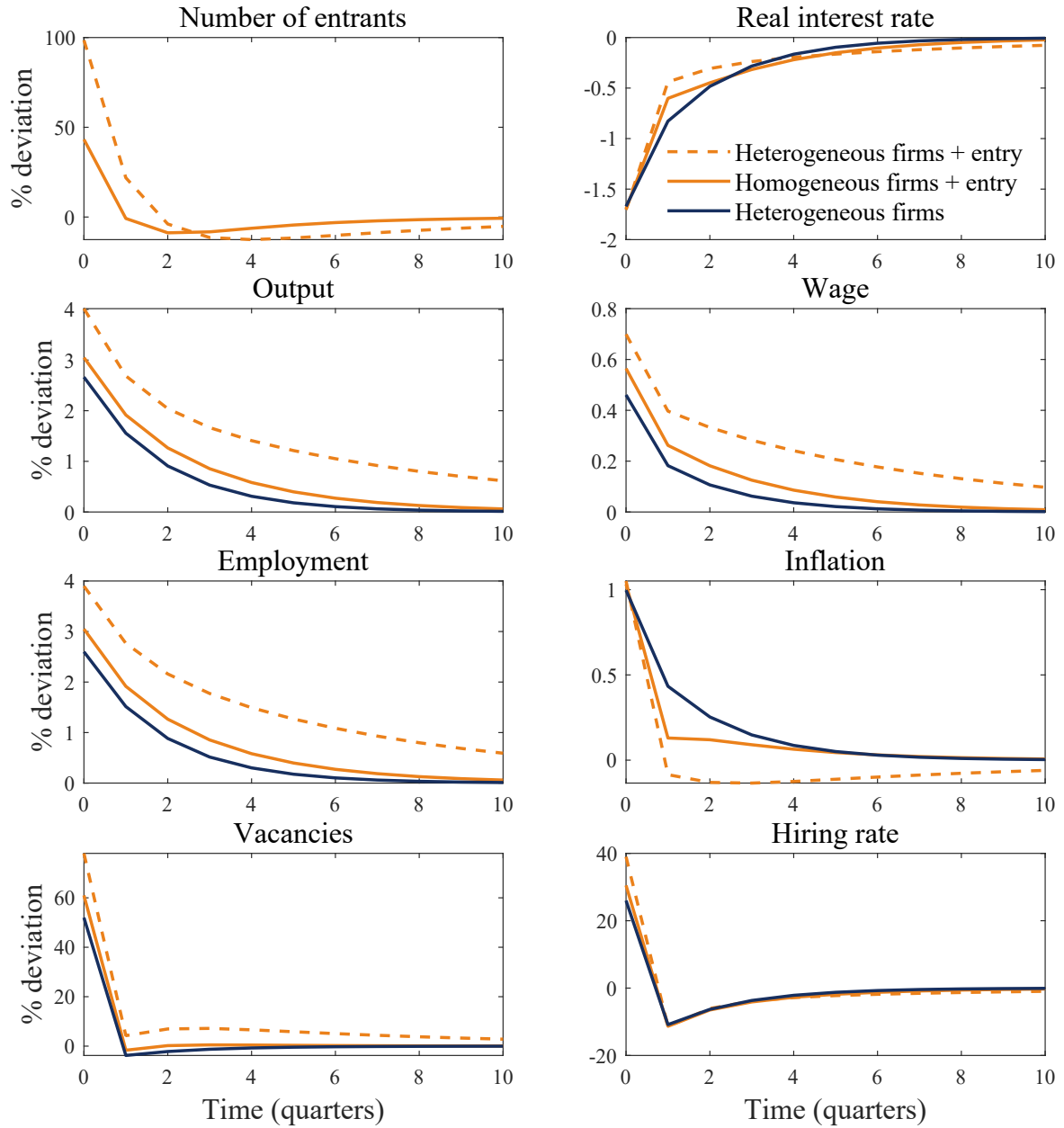
Fig. 1.4 Effect of entry on monetary policy response

Notes. Impulse response functions after an expansionary one percent shock to the nominal interest rate at $t = 0$. The solid red line shows the responses of the model with firm heterogeneity but without entry. The dashed green line denotes the responses when firms are homogeneous in productivity and there is entry. The solid green line shows responses for heterogeneity and entry.
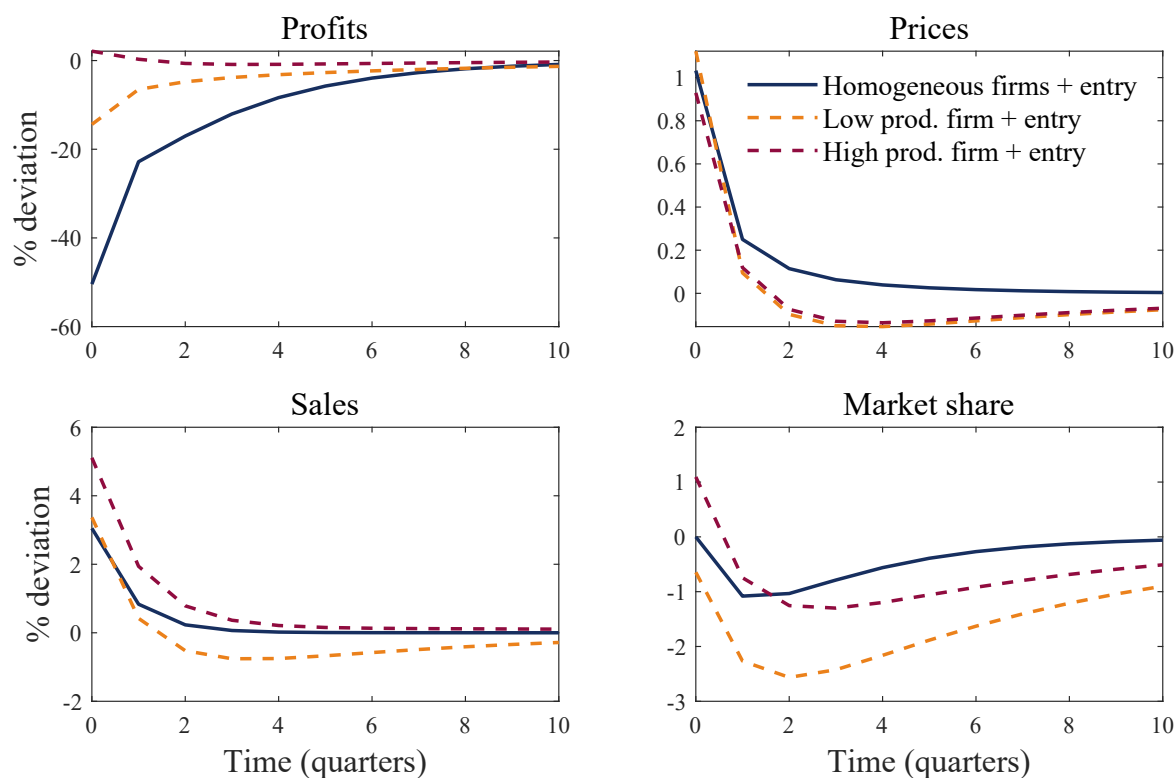
Fig. 1.5 Productivity-specific impulse responses with entry

Notes. Impulse response functions after an expansionary one percent shock to the nominal interest rate at $t = 0$. The solid green line shows the responses of the model with homogeneous firms and entry. The dashed lines show responses of the efficient firms and the less productive firms, respectively.

market share for large incumbents is followed by a decrease due to entry of new firms. Furthermore, entry has a negative impact on firms' profits. The increase in labour demand raises the response in wages, hence decreasing firms' profits. Consequently, in a setting with entry, aggregate profits are becoming counter-cyclical for the same parametrization a model without entry produced pro-cyclical profits.

1.7 Heterogeneous responses to monetary policy: Empirical evidence

This section provides empirical evidence on how the response of firms' sales and operating income after a monetary policy shock varies across firm size as a proxy for productivity. Section 1.7.1 describes the data sources. Section 1.7.2 describes the econometric framework used in order to estimate impulse response functions of firms' sales and operating income to a measure of monetary policy shocks. Section 1.7.3 presents the empirical results and section 1.7.4 checks whether the results are robust to different estimation specifications.

1.7.1 Data description

The empirical specification combines measures of monetary policy shocks and quarterly firm-level data.

Monetary policy shocks. An updated series of Romer & Romer (2004) (henceforth RR) monetary policy shocks is obtained from Wieland & Yang (2016). Their measure of monetary policy shocks is defined as the intended federal funds rate, inferred by quantitative and narrative records, and orthogonalised to internal forecasts prior to the meeting date of the committee. The updated quarterly time series spans the time frame from 1969q1 to 2007q4. The time series stops thereafter due to the binding zero lower bound. I define the time series of monetary policy shocks such that a positive value indicates an expansionary shock which makes the interpretation of regression coefficients more intuitive.

Firm-level data. Firm-level data is drawn from Compustat, a panel for publicly listed U.S. firms. The advantage of using Compustat are threefold: Firstly it offers quarterly data, a high enough frequency to study effects of monetary policy. Secondly, the panel starts in 1961 offering plenty of observations to study within-firm variation. Lastly, it contains

a rich set of balance-sheet information. The main drawback is that it excludes private, smaller firms.

The analysis focuses on manufacturing firms corresponding to sub-sectors 311-339 according to the North American Industry Classification System (NAICS). The sample contains active as well as inactive firm in order to avoid any survival bias. Firm size is approximated using information on the total amount of assets per firm (ATQ). Following Crouzet & Mehrotra (2020), the cut-off level for large firms is defined as the 99th percentile of the distribution of assets.[10] The main variables of interest in correspondence to the theoretical model predictions are firms' sales (SALEQ) and inventory (INVTQ) proxying the firms' output as well as operating income before depreciation (OIBDQ) as a measure for firm profits.

1.7.2   Estimation framework

This section shortly outlines the empirical framework used to estimate potential differences in the response of firms to a monetary policy shock relative to their size. The estimation framework is analogous to the local projection method proposed by Jordà (2005). Impulse response functions to a monetary policy shock are estimated using the following specification:

$$\Delta y_{i,t,t+h} = \alpha_i + \alpha_t + \beta_h \epsilon_t^M \mathbb{1}_{\{i \in [99,100]\}} + \sum_{l \in L} (\gamma_l + \delta_l \epsilon_t^M) \mathbb{1}_{\{i \in \mathscr{L}\}} + \Gamma' Z_{i,t-1} + \mu_{i,t} \qquad (1.51)$$

where y denotes the variable of interest in logs, i indexes the firm, t is the quarterly date and h is the horizon at which the impulse response is estimated. Firm-level and time fixed effects are denoted by α_i and α_t, respectively. The main coefficient of interest is β_h, measuring the semi-elasticity in y, h quarters after a standardized monetary policy shock ϵ_t^M relative to the benchmark group of the bottom 99% firms with respect to their size.[11] $\mathscr{L}$ refers to the three digit NAICS sub-sector specification and by introducing sub-sector dummies and interactions with the monetary policy shock, I control for sub-sector specific intercepts and slopes. $Z_{i,t-1}$ is a vector of additional firm controls such as leverage and sales growth lagged by one period. Standard errors are clustered in two ways to account for correlation within firms and within quarters.

---

[10]Results are pretty robust when instead setting the cutoff at the 90th percentile or the 95th percentile.

[11]Recall that the monetary policy shocks are defined such that a positive sign indicates an expansionary shock.

1.7.3 Empirical results

Figure 1.6 plots the estimated semi-elasticity differential β_h for sales, inventory and operating income over the horizon of eight quarters. The estimated responses seem to be mostly in line with the predictions of the theoretical model. The largest firms, which are assumed to be the most productive firms, seem to increase their output significantly more than smaller firms after an expansionary monetary policy shock as suggested by the response in sales and inventory. In fact, the top firms' output grows by roughly 7% more, two quarters after a one standard deviation monetary policy shock. Conditional on firm size being a good proxy for firm productivity the model showed exactly that. The model showed that the increase in households demand is matched by a proportionally larger increase in the sales of large firms compared to small firms. However, empirical estimates are likely to be biased as they include foreign sales, whereas the model is considering monetary shocks in a closed economy. Assuming that foreign sales do not change following a national (U.S.) monetary policy shock estimates are measured conservatively as this would induce a downward bias.[12]

Although, operating income also increases on average more for large firms directly after an expansionary monetary policy shock, the estimates are not significant in the baseline specification. It is very likely that firm profits exhibit more idiosyncratic within firm variance than sales or inventory as they are also affected by the cost side which would increase standard errors. Furthermore, in the baseline specification we compare the top 1% of manufacturing firms with the bottom 99% across all sectors. Clearly, this is an oversimplified and too aggregated view of the oligopolistic competition firms are facing. In order to address this issue, the following subsection also estimates more granular specifications.

1.7.4 Robustness

In order to check the robustness of the results, the strategy is twofold: First, a more granular definition of size indicators is used. Above, the percentiles for firm size were determined using the entire universe of manufacturing firms. However, as sectors are unlikely to be symmetric in the distribution of firms' total asset holdings, the effect is mostly explained by competition across sectors opposed to within sectors as proposed by the model. Hence,

---

[12]Yet, it is likely that foreign sales would increase following an expansionary monetary shock depreciating the national currency which would reduce this bias and depending on which effect is larger could eventually induce an upward bias. Unfortunately, Compustat firm-level data on sales does not distinguish between national and international sales to control for any bias.

Sales

Inventory
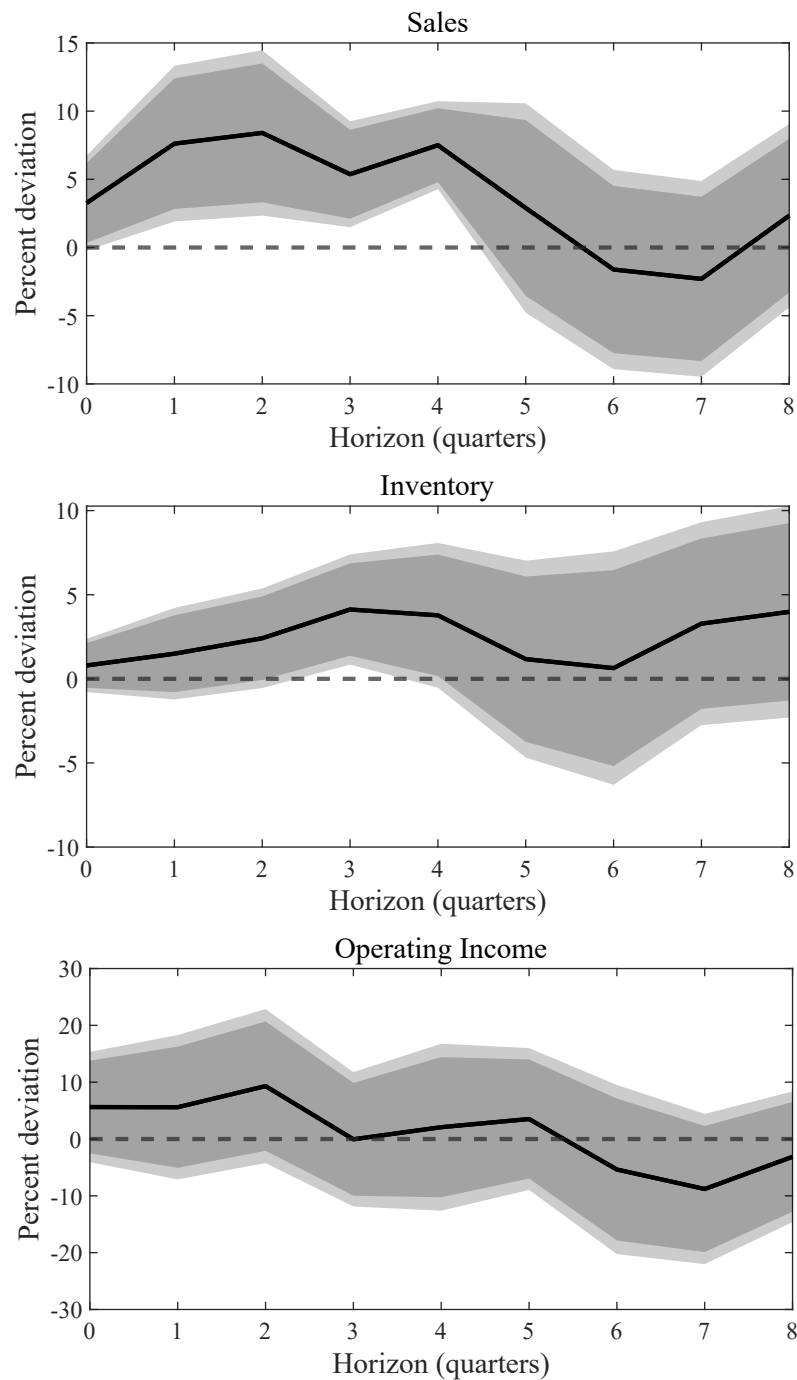
Operating Income

Fig. 1.6 Estimated sales and income response for top 1% firms vs. the rest

Note: Estimated response of sales and operating income for the top 1% firms in size relative to the bottom 99% to Romer & Romer (2004) monetary policy shocks using local projections. The shocks are normalized such that a positive sign denotes an expansionary shock. Standard errors are clustered at the firm-level and over time. The dark and light shaded areas represent the 90% and 95% confidence interval, respectively.

Table 1.2 Local projection results

Estimation specification	Quarter (h)				
	0	1	2	3	4
	Panel A: Sales projections				
Baseline	2.79	7.08***	7.93***	5.60***	7.01***
NAICS sub sectors	2.22	3.12	3.89*	1.41	5.35*
Gertler & Karadi (2015) (GK) shocks	6.35*	5.67**	4.47**	4.08*	5.70
NAICS sub sectors and GK shocks	1.78**	2.19*	1.7	1.16	1.37
	Panel B: Operating income projections				
Baseline	3.28	2.22	7.03	-2.07	-2.14
NAICS sub sectors	2.48	1.97	10.07**	8.17**	10.33**
Gertler & Karadi (2015) (GK) shocks	7.33	4.64	4.17	5.05	6.77
NAICS sub sectors and GK shocks	1.32	1.60	0.95	-0.28	1.18

Notes. Local projection coefficients of interaction between size indicator and standardized monetary policy shock measure in percent. Controlled for firm, industry and time fixed effects. Standard errors are robust to firm heteroskedasticity and autocorrelation. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

an alternative size indicator is calculated at three-digit NAICS sub-sector level. Second, an alternative proxy for monetary policy shocks is used. Gertler & Karadi (2015) (henceforth GK) offer a high-frequency approach of identifying shocks to the federal funds rate.

Table 1.2 lists the results of the alternative specifications over an horizon of four quarters. Defining size indicators at the three-digit NAICS level results in slightly lower and less significant estimates for the sales differential. In contrast, estimates for the operating income differential over time, are larger and significant two quarters after the monetary policy shock.

When using GK's monetary policy shock series instead of RR's, the sales response estimates are qualitatively in line and again slightly lower when using sub-sector size cutoffs. However, while the estimates for operating income show the correct sign for all but one horizon, they are not significant at standard levels. Overall, the data suggests that large firms' sales react more strongly relative to the response of small firms in the aftermath of a monetary policy shock. This is in line with the predictions of the theoretical model above. However, while the point estimates of the response in operating income are mostly in line with the model predictions, they do not appear to be robust.

1.8   Conclusion

This paper presents a New Keynesian model with strategically interacting firms and search and matching frictions in the labour market. In a setting with firm heterogeneity, more

productive firms exhibit a weaker response in their pricing after a change in marginal costs. Hence, oligopolistic competition induces higher price rigidity for large firms relative to small firms. As a consequence, the pro-cyclical response in firms' sales is larger for more productive firms as they face a higher (lower) demand after an expansionary (contractionary) monetary policy shock. Local projections using Compustat firm-level data and different measures of monetary policy shocks support those predictions and appear to be robust. In turn, this leads to a pro-cyclical response in large firms' market and employment shares. As large firms are more productive, this shift also reduces the aggregate labour share following an expansionary monetary policy shock. Hence, aggregate output responds more than in the case of homogeneous firms due to re-allocating labour from small to large, more productive, firms. This is in line with a very recent paper by Baqaee et al. (2021) who provide evidence for a pro-cyclical response in aggregate TFP and re-allocations to high-markup firms. Furthermore, I show that cyclicality in profits ultimately depends on the productivity differential between firms. More specifically, under heterogeneity, less rigid wages are necessary to obtain a pro-cyclical response in firm profits. Allowing for dynamic entry strengthens the response in output due to an increase in labour demand and enhances the effect of firm heterogeneity.

References

Andrés, J. & Burriel, P. (2018). Inflation and optimal monetary policy in a model with firm heterogeneity and bertrand competition. *European Economic Review*, 103, 18–38.

Atkeson, A. & Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5), 1998–2031.

Auclert, A. (2017). *Monetary policy and the redistribution channel*. Technical report, National Bureau of Economic Research.

Baqaee, D., Farhi, E., & Sangani, K. (2021). *The Supply-Side Effects of Monetary Policy*. Technical report, National Bureau of Economic Research.

Barkai, S. (2016). Declining labor and capital shares. *Stigler Center for the Study of the Economy and the State New Working Paper Series*, 2.

Benigno, P. & Faia, E. (2010). *Globalization, pass-through and inflation dynamic*. Technical report, National Bureau of Economic Research.

Bergin, P. R. & Corsetti, G. (2008). The extensive margin and monetary policy. *Journal of Monetary Economics*, 55(7), 1222–1237.

Bilbiie, F. O. (2017). Monetary neutrality with sticky prices and free entry.

Bilbiie, F. O., Fujiwara, I., & Ghironi, F. (2014). Optimal monetary policy with endogenous entry and product variety. *Journal of Monetary Economics*, 64, 1–20.

Bilbiie, F. O., Ghironi, F., & Melitz, M. J. (2012). Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, 120(2), 304–345.

Brander, J. & Krugman, P. (1983). A reciprocal dumping model of international trade. *Journal of international economics*, 15(3-4), 313–321.

Brander, J. A. (1981). Intra-industry trade in identical commodities. *Journal of international Economics*, 11(1), 1–14.

Broda, C. & Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *American Economic Review*, 100(3), 691–723.

Broer, T., Harbo Hansen, N.-J., Krusell, P., & Öberg, E. (2020). The new keynesian transmission mechanism: A heterogeneous-agent perspective. *The Review of Economic Studies*, 87(1), 77–101.

Caballero, R. J., Farhi, E., & Gourinchas, P.-O. (2017). Rents, technical change, and risk premia accounting for secular trends in interest rates, returns on capital, earning yields, and factor shares. *American Economic Review*, 107(5), 614–20.

References

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1), 1–45.

Christiano, L. J., Eichenbaum, M. S., & Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84(4), 1523–1569.

Cloyne, J., Ferreira, C., & Surico, P. (2016). Monetary policy when households have debt: new evidence on the transmission mechanism.

Colciago, A. & Etro, F. (2010). Real business cycles with cournot competition and endogenous entry. *Journal of Macroeconomics*, 32(4), 1101–1117.

Colciago, A. & Rossi, L. (2015). Firm dynamics, endogenous markups, and the labor share of income. *Macroeconomic Dynamics*, 19(6), 1309–1331.

Crouzet, N. & Mehrotra, N. (2020). Small and Large Firms over the Business Cycle. *American Economic Review*. Conditionally accepted.

De Loecker, J. & Eeckhout, J. (2017). *The rise of market power and the macroeconomic implications*. Technical report, National Bureau of Economic Research.

Den Haan, W. J., Ramey, G., & Watson, J. (2000). Job destruction and propagation of shocks. *American economic review*, 90(3), 482–498.

Dixit, A. K. & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3), 297–308.

Etro, F. & Colciago, A. (2010). Endogenous market structures and the business cycle. *The Economic Journal*, 120(549), 1201–1233.

Etro, F. & Rossi, L. (2015). New-keynesian phillips curve with bertrand competition and endogenous entry. *Journal of Economic Dynamics and Control*, 51, 318–340.

Faia, E. (2012). Oligopolistic competition and optimal monetary policy. *Journal of Economic Dynamics and Control*, 36(11), 1760–1774.

Fairlie, R. W. & Fossen, F. M. (2018). Opportunity versus necessity entrepreneurship: Two components of business creation.

Fernández, C. & López, P. (2014). Explaining small firms tfp gap: An analysis with italian, spanish and french firm-level data. *Banco de España, Working Paper, forthcoming*.

Gertler, M. & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1), 44–76.

Guilloux-Nefussi, S. (2016). Globalization, market structure and inflation dynamics.

Hall, R. E. & Milgrom, P. R. (2008). The limited influence of unemployment on the wage bargain. *American economic review*, 98(4), 1653–74.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, (pp. 1127–1150).

Impullitti, G. & Licandro, O. (2018). Trade, firm selection and innovation: The competition channel. *The Economic Journal*, 128(608), 189–229.

Impullitti, G., Licandro, O., & Rendahl, P. (2017). Technology, market structure and the gains from trade.

Jaimovich, N. & Floetotto, M. (2008). Firm dynamics, markup variations, and the business cycle. *Journal of Monetary Economics*, 55(7), 1238–1252.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1), 161–182.

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3), 697–743.

Lewis, V. (2009). Business cycle evidence on firm entry. *Macroeconomic Dynamics*, 13(5), 605–624.

Lewis, V. & Poilly, C. (2012). Firm entry, markups and the monetary transmission mechanism. *Journal of Monetary Economics*, 59(7), 670–685.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695–1725.

Mongey, S. (2017). Market structure and monetary non-neutrality. In *2017 Meeting Papers*, number 184: Society for Economic Dynamics.

Mortensen, D. T. & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The review of economic studies*, 61(3), 397–415.

Ottonello, P. & Winberry, T. (2018). *Financial Heterogeneity and the Investment Channel of Monetary Policy*. Technical report, National Bureau of Economic Research.

Pissarides, C. A. (2009). The unemployment volatility puzzle: Is wage stickiness the answer? *Econometrica*, 77(5), 1339–1369.

Ravenna, F. & Walsh, C. E. (2008). Vacancies, unemployment, and the phillips curve. *European Economic Review*, 52(8), 1494–1521.

Romer, C. D. & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4), 1055–1084.

Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy*, 90(6), 1187–1211.

References

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, 95(1), 25–49.

Sterk, V. & Tenreyro, S. (2015). The transmission of monetary policy through redistributions and durable purchases.

Walsh, C. E. (2003). Labor market search and monetary shocks. *Elements of Dynamic Macroeconomic Analysis*, (pp. 451–486).

Wieland, J. F. & Yang, M.-J. (2016). *Financial dampening*. Technical report, National Bureau of Economic Research.

Chapter 2

Financial factors, firm size and firm potential

Abstract

Using a unique dataset covering the universe of Portuguese firms and their credit situation we revisit the relation between firm size, their financial situation, and sensitivity to the cycle. First, we provide two stylized facts: (1) Financially constrained firms react more to the business cycle and this mechanism is orthogonal to the size channel proposed by Crouzet & Mehrotra (2020). (2) Constrained firms are found across the entire size distribution, also in the top percentiles, which is in contrast to what standard financial friction models would predict. We then show that ex-ante heterogeneity of firms, a possible explanatory factor, persists over the firms' life cycle and affects constrained and unconstrained firms differently. Incorporating this ex-ante heterogeneity into an otherwise standard financial frictions model simultaneously accounts for the stylized facts, gives rise to large constrained firms, and leads to larger aggregate fluctuations and capital misallocation.[1]

Keywords: Firm size, business cycle, financial accelerator.
JEL Codes: E62, E22, E23

---

[1]This chapter is based on joint work with Miguel Ferreira and Timo Haber.

2.1 Introduction

A substantial amount of research in macroeconomics focuses on the propagation of aggregate shocks via financial factors and their relation to individual firm characteristics. In a seminal work on this topic, Gertler & Gilchrist (1994) propose firm size as an effective proxy for financial constraints. Smaller firms are arguably more risky, less liquid and face an elevated external finance premium. Accordingly, smaller firms are more sensitive to aggregate shocks, as they tend to be in a weaker financial position. Revisiting this hypothesis with more granular data, Crouzet & Mehrotra (2020) show that differences in firm cyclicality are not correlated with size, except for when one compares very large firms to the rest. Moreover, firm size remains significant, even when controlling for a number of balance sheet measures of financial conditions. So, contrary to the earlier literature, their analysis suggests that financial factors may not be the deeper source of size-based cyclicality differences.

In this paper, we contribute to this debate by presenting both novel evidence and theoretical insights on the relationship between financial factors, size and sensitivity to the business cycle. Our empirical evidence comes from the Bank of Portugal's confidential credit registry database, matched with bank and firm balance sheet data between 2006 and 2017. Since all financial institutions are required to report both potential and actual credit above € 50, we are able to construct detailed, firm-specific and credit-based measures of financial constraints. To the best of our knowledge, we are the first to utilize this level of empirical data on credit relationships in answering these questions and discipline the choice of our theoretical model. As such, our contribution is twofold.

First, using the firms' detailed credit information we present two stylized facts. First, we establish that financially constrained firms are more sensitive to the business cycle, in line with the financial accelerator theory. This empirical result holds conditional on size and is robust across a number of different specifications. On the one hand, this fact supports a broad financial accelerator theory, i.e. that a firm's financial position matters for aggregate cyclicality. On the other hand, this is also evidence against theories where a firm's financial position and its size coincide, since both are jointly significant in determining firm cyclicality. This conclusion is reinforced by our second stylized fact; both constrained and unconstrained firms can be found across the entire firm size distribution. In fact, size and other firm variables commonly used in the literature as proxies for financial constraints are only weakly correlated to our measures of a firm's financial health. This again suggests that size is not a sufficient proxy for financial frictions and that there exists

a considerable within bin heterogeneity across the firm size distribution, not captured in standard financial frictions models.

In our second contribution, we then aim to reconcile the stylized facts with the existing financial frictions theory by accounting for this aforementioned within bin heterogeneity. In particular, we present empirical evidence demonstrating that ex-ante heterogeneity across firms matters and persists over the firm's life cycle. We then use this as motivation to refine the standard financial frictions model by including a permanent component to the firms' productivity process. This rather simple addition of ex-ante heterogeneity enables the model to match the stylized facts, as it introduces a large and permanent heterogeneity in optimal firm sizes and spells of financial constraints. Finally, we explore how the existence of large constrained firms and ex-ante heterogeneity shape the aggregate response following a perfect foresight shock to total factor productivity. Relative to the standard Khan & Thomas (2013) model with only a transitory productivity component, the aggregate response is stronger due to the cyclicality of large constrained firms in the top percentiles of the size distribution.

Literature. Our work follows a large literature in macroeconomics that has developed and analysed models with heterogeneous firms and financial frictions.

One of the early contributions in this literature by Cooley & Quadrini (2001) shares many features with our current model. They augment an otherwise standard Hopenhayn (1992) model of heterogeneous firms with financial frictions and persistent shocks. In doing so, they are able to match the empirical facts that both smaller firms, conditional on age, and younger firms, conditional on size, are more dynamic (ie. job creation and destruction, growth, volatility of growth and exit are all higher). In similar fashion, Pugsley et al. (2021) highlight the importance of ex-ante heterogeneity in explaining the firm size distribution and the recent decline in firm dynamism.

Another, more recent instance where permanent productivity differences plays a crucial role is Mehrotra & Sergeyev (2020). They argue that financial frictions played a relatively minor role in unemployment increases associated with the Great Recession and that employment was reduced due to shocks that affected unconstrained and constrained firms alike. Khan & Thomas (2013) and Ottonello & Winberry (2018) argue instead for the importance of financial frictions in the propagation of financial and monetary policy shocks, respectively. The seminal paper by Bernanke & Gertler (1989) on the financial accelerator mechanism develops a simple neoclassical model where the condition of the borrowers' balance sheets is a source for output dynamics. Positive business cycle

shocks increase borrowers' net worth, lower agency costs, and thus increase investment, amplifying the business cycles.

Our theoretical contribution emphasizes the importance of permanent productivity differences for matching the observed distribution of constrained firms, conditional on size. We then highlight the importance of matching this distribution in amplifying both productivity and financial shocks.

We also relate to the literature on misallocation of productive resources. While Buera et al. (2011) argue that financial frictions explain a large share of cross country differences in misallocation and consequently in aggregate productivity, Midrigan & Xu (2014) by parametrizing a firm dynamics model with financial frictions, consistent with producer level data, find small losses from misallocation. Other papers such as David et al. (2016), Restuccia & Rogerson (2017), Baqaee & Farhi (2020) and Peters (2020) highlight for other sources of misallocation besides financial frictions, such as markups, imperfect information or even tax code and regulation.

We contribute to this literature by illustrating that matching the distribution of constrained firms, conditional on size, amplifies the misallocation resulting from financial frictions up to three times, when compared to a standard firm dynamics and financial frictions model.

Lastly, we relate to the empirical literature assessing the differences in cyclicality of constrained firms and the debate on how to identify these firms in the data. Gertler & Gilchrist (1994) find empirical evidence for the financial accelerator mechanism. They analyse the cyclical behaviour of small vs. large manufacturing firms and interpret this as evidence for the financial accelerator. Their main assumption is that size is a good proxy for financial constraints. Sharpe (1994) finds a statistically significant relationship between a firm's financial leverage and the cyclicality of its labour force. Employment growth at highly leveraged firms is more sensitive - they are less likely to hoard labour. This cyclicality also holds for the size dimension - confirming Gertler & Gilchrist (1994). Gilchrist & Himmelberg (1995) find that investment still responds to cash flow even after controlling for its role for forecasting future investment opportunities, with the effect being stronger for financially constrained firms.

More recently, Crouzet & Mehrotra (2020) using firm level data, showcase that differences in size-related cyclicality only arise at the very top of the distribution, with the bottom 99.5% of firms having non-significant differences in cyclicality. The authors also state they don't find any evidence of the financial accelerator mechanism.

These results are potentially related to Farre-Mensa & Ljungqvist (2016) findings, who suggest that typical measures of financial constraints are not associated with firms that behave as if they were constrained. Even indices that combine different firm characteristics such as the ones proposed by Kaplan & Zingales (1997), Whited & Wu (2006) and Hadlock & Pierce (2010) do not correlate well with firms that behave as financially constrained. These findings are also supported by Bodnaruk et al. (2015), who use text analysis of the 10-k financial reports to gauge if firms are constrained or not, and find a weak correlation with common constrained measures.

By making use of a very detailed firm level credit data, with information on credit lines still available to the firm and overdue credit, we provide evidence supporting the financial accelerator mechanism. Yet we show that size is a much weaker proxy than anticipated by several papers. In fact, the financial accelerator mechanism is orthogonal to a separate size-related channel of reduced cyclicality for large firms, with the latter being the focus of Crouzet & Mehrotra (2020).

Outlook. The paper is structured as follows. Section 2.2 presents the data we use for the empirical analysis as well as to discipline our theoretical model. We proceed to present the three stylized facts outlined above in Sections 2.3 and 2.4. In Section 2.5 we set out the model to incorporate these facts and in Section 2.6 we discuss model predictions of aggregate effects. Finally, Section 2.7 concludes.

2.2 Data

We draw on a unique combination of datasets that cover the Portuguese economy between 2006 and 2017, all managed by the Bank of Portugal.

First, we use the *Informação Empresarial Simplificada* (IES) Central Balance Sheet Database (CBSD) that is based on annual accounting data of individual firms. Portuguese firms have to fill out mandatory financial statements in order to comply with their statutory obligation. Consequently, this dataset covers the population of virtually all non-financial corporations in Portugal from 2006 onward. We combine this with the Central Credit Register (CCR), that contains monthly[2] information on the actual and potential credit above 50 euros extended to individuals and non-financial corporations, reported by all financial institutions in Portugal. Actual credit includes loans that are truly taken up, such

---

[2]Given that the firm balance sheet data is of yearly frequency, we consider the month in which the balance sheet data was reported. Results were robust to shifting and averaging the monthly credit data.

as mortgages, consumer loans, overdrafts and others. Potential credit encompasses all irrevocable commitments to the subject that have not materialised into actual credit, such as available credit on credit cards, credit lines, pledges granted by participants and other credit facilities.[3] We then merge these two databases on the firm level.

Moreover, we then also add the Monetary Financial Institutions Balance Sheet Database in order to gain information on the balance sheets of banks that extend credit to non-financial institutions. We combine this on a firm level using the bank identifier and the share of loans extended by one firm to arrive at our detailed dataset.

Similar to Buera & Karmakar (2017), who use the same dataset, we restrict the set of firm in this panel dataset to those with at least five consecutive observations and to firms which are in business at the time of reporting. Furthermore, we only consider privately or publicly held firms and drop micro firms, i.e. those with overall credit amounts of less than 10,000 €. Descriptive statistics for the relevant variables can be found in table B.2 and B.3 in Appendix B.3.

2.2.1 Construction of constrained measures

Based on the detailed credit information in the data set, we construct several binary and continuous measures indicating whether a firm is financially constrained. Financial constraints are most commonly conceived as a supply side phenomenon. Firms that could potentially obtain credit in perfect credit markets are unable to do so due to asymmetric information considerations on the supply side. For example, a firm that has a profitable investment project that requires external financing cannot realise it as the bank is not satisfied with the creditworthiness of that firm. This may happen either via the price dimension - an interest rate that is too high - or on the quantity dimension - the credit is denied altogether. We aim to identify constrained firms along the quantity dimension, using the credit information for each firm. Given that credit allowances are changing over time, this provides us with a time-varying and firm-specific measure for being financially constrained. It shall be noted however, that while credit information offers a far more detailed notion of a firm being constrained, relative to financial characteristics such as leverage or liquidity, it is still a proxy. Hence, taking this into account, we consider a wide range of binary and continuous measures for identifying whether a firm is constrained.

---

[3]Further details on the credit information used are documented in appendix B.1.

Binary measures In our baseline definition a firm is credit constrained at time t, if it has no potential credit available at time t:

$$\text{Constrained I} := \mathbb{1}_{\text{Potential credit}_t=0}.$$

As outlined above, potential credit summarizes all the irrevocable commitments by credit-granting institutions. Even though this measure enables an understanding of whether firms have drawn down their credit lines and are thus potentially constrained it also encompasses a lot of noise. One problem might be that specific banks may not grant credit lines. This would show up as potential credit = 0 in the data. But this does by no means make these firms credit constrained. They just simply do not have credit lines, which is quite common. In order to address this issue, we consider three approaches. First, we test the robustness of our results dropping all cases in which potential credit is zero throughout. Second, we estimate the panel regression using firm fixed effects, which will control for those cases by construction. Third, we introduce the following adjusted measure, demanding that potential credit was positive in the previous period $t-1$ and hence, the firm seems to have hit the credit line:

$$\text{Constrained II} := \mathbb{1}_{\text{Potential credit}_t=0 \ \& \ \text{Potential credit}_{t-1}>0}.$$

Another issue when relying only on potential credit, might be that although, firms have exhausted their committed credit line, they could still successfully apply for a short- or long-term loan. In order to cope with this issue, we introduce two further measures. The third measure re-defines those firms as unconstrained, which managed to secure short- or long-term credit, while potential credit was zero:

$$\text{Constrained III} := \mathbb{1}_{\text{Potential credit}_t=0 \ \& \ \Delta\text{Effective credit}_{t+1}<0}.$$

The fourth measure augments the baseline definition by specifically considering those firms as being constrained for which overdue credit is growing:

$$\text{Constrained IV} := \mathbb{1}_{\text{Potential credit}_t=0 \ \& \ \Delta\text{Overdue credit}_{t+1}>0}.$$

The rational behind this definition is that growth in overdue credit is likely a signal for a firm in bad financial shape.

While the measures presented so far are conceptually in the spirit of a firm hitting the credit constraint and thus being strictly constrained, it might also be that a firm is granted credit, yet the amount is not sufficient to finance any planned investment. The following measure aims to capture this notion of potentially constrained firms:

$$\text{Constrained V} := \mathbb{1}_{\text{Total Credit}_{t+1} > \text{Total Credit}_t + \text{Potential Credit}_t}.$$

Continuous measures   In order to account for different levels of severity of financial constraints, we also introduce two continuous measures. The first one is defined as the ratio of potential credit and cash to total liabilities of the firm, thereby accounting for the fact that a firm might have enough cash making it financially unconstrained, despite being credit constrained:

$$\text{Constrained VI} := \frac{\text{Potential credit}_t + \text{Cash}_t}{\text{Liabilities}_t}.$$

Our final measure can be interpreted as the continuous counterpart to the third measure taking into account potential credit and ex-post changes to short- and long-term credits:

$$\text{Constrained VII} := \frac{\text{Potential credit}_t + \Delta\text{Short- and long-term credit}_{t+1}}{\text{Liabilities}_t}.$$

Check the appendix for a more detailed description of the dataset and the constrained measures. Tables B.1, B.2 and B.3 report descriptive statistics. Figures B.1, B.2 and B.3 report the evolution of the share of constrained firms and credit over time.

2.3   Stylized facts

The rich dataset covering the universe of Portuguese firms and including both balance sheet and credit information, allows us to generate novel evidence on the characteristics of constrained firms. This Section presents two stylized empirical facts which form the basis for our theoretical model. First, we revisit the relationship between firm characteristics and the elasticity of firm outcomes with respect to the cycle. Similar to Crouzet & Mehrotra (2020), we find firms' cyclicality to differ only at the top 1% of the size distribution. But, more importantly, we find evidence that supports the financial accelerator mechanism theory, with constrained firms being more cyclical, and this to be orthogonal to the size cyclicality. Second, we proceed by illustrating that size, among other variables commonly

used in the literature as proxies for financial constraints, are weakly correlated to the firm's financial health. In fact, constrained firms can be found over the entire firm distribution.

2.3.1 Size and financial factors matter for cyclicality

This Section aims to test the financial accelerator mechanism empirically. In particular, we want to test whether constrained firms are more cyclical than unconstrained firms. In order to make our results comparable to Crouzet & Mehrotra (2020), we first replicate their results for the universe of Portuguese firms and then augment their estimation strategy, using our set of firm-specific and time-varying measures of financial factors. Following Crouzet & Mehrotra (2020), the specification estimated is:

$$g_{i,t} = \Delta GDP_t + \sum_{j \in \mathscr{J}} (\alpha_j + \beta_j \Delta GDP_t)\mathbf{1}_{i \in \mathscr{S}_t^{(j)}} + (\zeta + \eta \Delta GDP_t)Const_{i,t}$$
$$+ \sum_{l \in \mathscr{L}} (\gamma_l + \delta_l \Delta GDP_t)\mathbf{1}_{i \in \mathscr{L}} + \epsilon_{i,t}, \tag{2.1}$$

where i identifies a firm and t identifies a year. The dependent variable $g_{i,t}$ is the year-on-year log change in turnover. The set $\mathscr{S}_t^{(j)}$ is a jth size group, e.g. all firms above the 90th but below the 99th percentile. Furthermore, ΔGDP_t is the year-on-year growth rate of GDP, and $\mathscr{L}$ is a set of industry dummies. $Const_{i,t}$ refers to the firm-specific variable measuring the strength of financial constraints.

Table 2.1 reports estimates of the semi-elasticity of firm-level growth in turnover to GDP growth. The first column reports estimates for the size groups $j \in \{[90,99],[99,99.5], [99.5,100]\}$ with $[0,90]$ as the reference group. On average, small firms have a semi-elasticity of roughly 2.5, meaning for any percent change in GDP growth, their turnover changes by 2.5%. Although, the coefficient for the size group $[90,99]$ is insignificant, results are consistent with the view that larger firms are less sensitive to aggregate fluctuations which is in line with the findings of Crouzet & Mehrotra (2020) and earlier work. The difference is particularly notable at the very top of the firm distribution.

The second column reports results including the baseline binary measure for being constrained $Const.I$. First, it is worth noting, that the estimation coefficient with respect to size hardly change. This is indicative that the mechanism going through size is orthogonal to any financial accelerator mechanism, and hence size might not be a good proxy as already pointed out by Crouzet & Mehrotra (2020). The coefficient for the interaction between the constrained measure and aggregate fluctuations is significant,

Table 2.1 Cyclicality in turnover conditional on size bins
and measures of financial constraints

	Turnover growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.512***	2.441***	2.440***	2.450***	2.470***	2.450***	2.522***	2.508***
	(0.023)	(0.027)	(0.023)	(0.028)	(0.025)	(0.030)	(0.023)	(0.025)
[90,99] × GDP growth	0.042	0.075	0.069	0.075	0.064	0.066	0.042	0.048
	(0.106)	(0.107)	(0.106)	(0.121)	(0.120)	(0.107)	(0.106)	(0.120)
[99,99.5] × GDP growth	-0.726**	-0.693**	-0.698**	-0.742**	-0.754**	-0.688**	-0.732**	-0.760**
	(0.300)	(0.300)	(0.300)	(0.342)	(0.341)	(0.300)	(0.300)	(0.342)
[99.5,100] × GDP growth	-1.426***	-1.382***	-1.391***	-1.607***	-1.610***	-1.384***	-1.428***	-1.650***
	(0.291)	(0.291)	(0.291)	(0.359)	(0.291)	(0.293)	(0.291)	(0.359)
Const.I × GDP growth		0.121**						
		(0.054)						
Const.II × GDP growth			0.561***					
			(0.135)					
Const.III × GDP growth				0.165**				
				(0.069)				
Const.IV × GDP growth					1.192***			
					(0.250)			
Const.V × GDP growth						0.161***		
						(0.048)		
Const.VI × GDP growth							-0.063**	
							(0.025)	
Const.VII × GDP growth								-0.168***
								(0.030)
Observations	1,326,344	1,326,344	1,326,344	1,161,187	1,161,187	1,326,344	1,324,992	1,160,194
R-squared	0.029	0.030	0.030	0.030	0.032	0.029	0.030	0.029
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of turnover with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

offering support for the financial accelerator mechanism. Constrained firms have a 0.1 higher semi-elasticity relative to unconstrained firms according to the baseline measure.

However, as already pointed out when introducing the different measures for being constrained, the baseline measure might capture firms for which potential credit is zero, but in fact are unconstrained. Hence, the baseline measure offers a lower bound of the increased sensitivity of constrained firms. We therefore consider other binary measures trying to overcome those drawbacks which are reported in columns (3) to (6). Estimation results are supportive of the notion that the baseline measure acts as a lower bound and sensitivity might be up to 10 higher for constrained firms as measured by $Const.IV$.

Estimates for the standardized continuous measures are also in line with the findings so far. The negative sign is due to their definition. The higher their value, the less con-
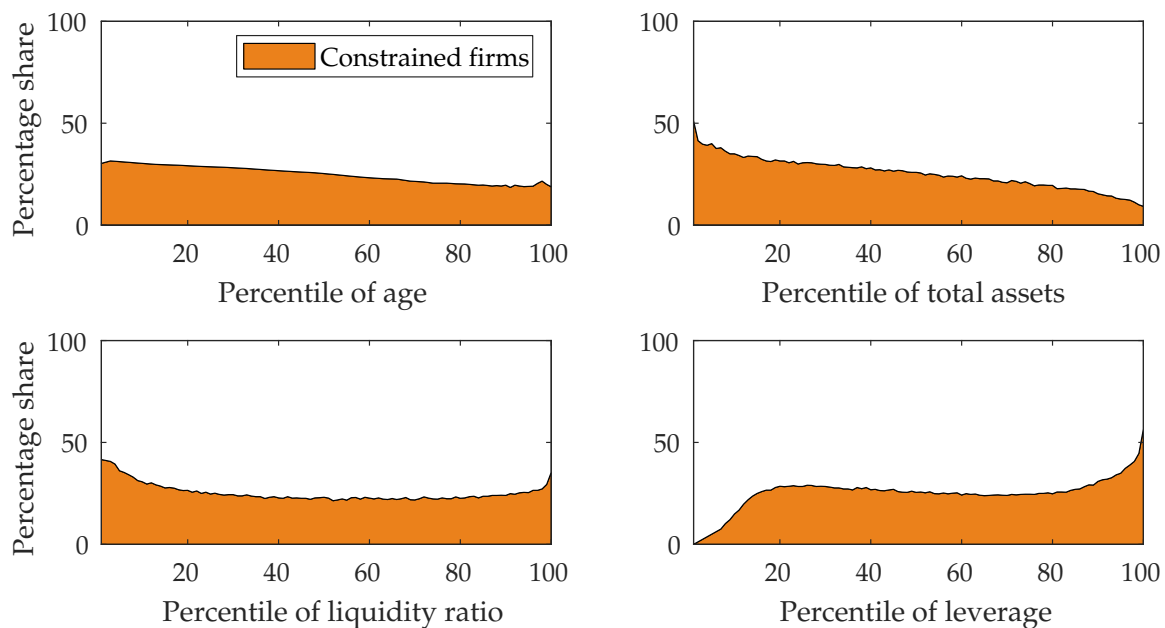
Fig. 2.1 Decomposition of constrained and unconstrained firms
across percentiles of firm variables

Notes. Constrained firms are identified using measure constrained III which classifies firms as constrained if they have exhausted their potential credit and were not granted additional short- or long-term credit in that period as defined in appendix 2.2.1. Figure B.4 - B.6 in the appendix show the analogue plots for related measures.

strained is a firm. Results for sales growth and growth in employees are closely in line, see Tables B.4 and B.5 in Appendix B.3. Besides using different measures, we considered a battery of robustness checks. First, we included time fixed effects to account for broader macroeconomic circumstances. Second, we estimated the specification using firm fixed effects. Third, we estimated the model excluding those firms for which potential credit is zero throughout. Fourth, we controlled for supply effects using aggregated bank data. Estimates were robust across all specifications. Results are reported in tables B.6-B.9 in Appendix B.3.

2.3.2   Constrained firms are found across the firm distribution

Our second stylized fact states that financially constrained firms populate the entire firm distribution along a number of common proxies for financial constraints. Figure 2.1 plots the share of firms that have zero potential credit and declining effective credit (measure III) over percentiles of age, total assets, liquidity ratio and leverage. Evidently, constrained firms can be found in every bin of the firm distribution. This finding is robust across

all binary identifiers for being constrained, with only the overall fraction of constrained firms changing depending on the strictness of the specific measure, as documented in Figures B.4 - B.6 in Appendix B.4.1. While correlations are in line with the existing literature, they are not as strong as existing models would predict. In fact, when running a linear probability model, the probability of being constrained only reduces by about 5% for one standard deviation increase in total assets.[4] Even after accounting for potential attenuation bias, the main conclusion stands: standard firm models typically produce small constrained firms and large unconstrained firms, yet our data does not support this strong dichotomy.

Moreover, even when controlling for a battery of financial variables the explanatory power to predict whether a firm is constrained is relatively low compared to the firms' fixed effects. Hence, existing proxies of financial constraints may be unable to capture this idiosyncrasy, which seems to play a role in credit decisions.

2.4 Firm potential

In contrast to our second stylized fact, standard firm financial frictions models à la Khan & Thomas (2013) predict a very strong correlation between firm size and financial constraints, as firms require a relatively uniform minimum size to become unconstrained. Consequently, one factor that could potentially break this strong correlation are heterogeneous ex-ante conditions for firms. Small firms may be unconstrained as they already reached their potential - i.e. optimal size - while large firms may still be growing and are still constrained. Equally, different potentials would create a dispersion of unconstrained firms across the entire firm size distribution, similar to our second stylized fact. Accordingly, this Section investigates whether such ex-ante heterogeneity exists in our dataset.

We follow Pugsley et al. (2021) by first computing both the standard deviation of log employment by age and the autocorrelation of log employment between age a and $h \leq a$. The results are depicted in Figure 2.2.[5] The left panel showcases that even at early ages the standard deviation of log employment is already relatively high - above 0.8 - which indicates large size differences even at early ages. Also, the standard deviation of log employment is increasing with age. This supports the idea that there exists ex-ante

---

[4]Check Table B.10 in Appendix B.3 for the results of the linear probability model.

[5]To prevent differences across sectors and business cycle conditions from explaining the majority of the standard deviation and autocorrelation, we first control for sector and year fixed effects and then use the residuals of log employment.
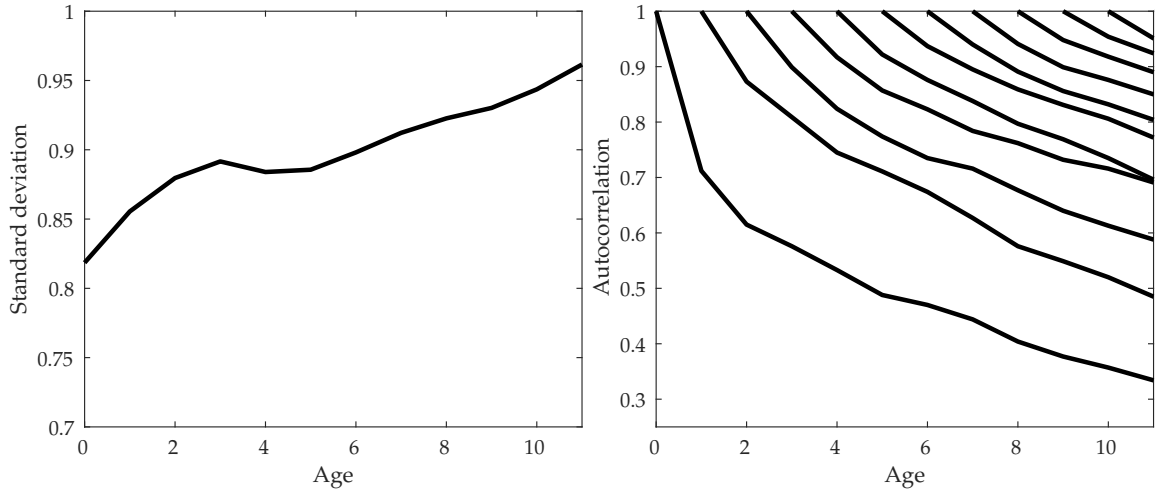
Fig. 2.2 Standard deviation and autocorrelation of log employment by age

Notes. The left panel presents the standard deviation of log employment by age, after controlling for sector and year fixed effects. The right panel presents the autocorrelation of log employment between ages a and $h \leq a$. Across lines h changes, while a changes along the lines.

heterogeneity, as firms at birth are not all equal, and that firms will settle at different levels of employment in the long run, as the standard deviation is increasing with age.

The right panel plot indicates that the long run autocorrelations appear to stabilize at relatively high levels. This may be indicative that ex-ante conditions are persistent and affect the firm even in the long run. This results are in line with the Pugsley et al. (2021) findings, who showcase the importance of ex-ante conditions in explaining the firm size distribution and firm dynamics.

As we are interested in understanding if differences between constrained and unconstrained firms can, partly, be explained by ex-ante heterogeneity, we plot in Figure B.7 in Appendix B.4.2 the standard deviation and autocorrelation of log employment for both constrained and unconstrained firms.[6] It is possible to notice that both these statistics are lower for constrained firms than unconstrained ones. One could have expected the opposite to be true, as constrained firms would potentially have less resources to grow and so their employment tomorrow could have a stronger correlation with employment today. The fact that the autocorrelation is higher across the life-cycle for unconstrained firms may be indicative that they are born closer to their optimal size, when comparing to

---

[6]Here we are using the measure that exclusively takes the amount of potential credit available into account to determine which are the constrained firms. A firm is considered constrained if at age $a - h$ has potential credit equal to zero.

constrained firms. This may then explain, in part, why some young firms are constrained and others are not: the ones born closer to their optimal size have lower investments and do not become constrained, while firms that need to grow to reach the optimal size exhaust their credit lines.

To better understand the importance of ex-ante vs ex-post heterogeneity for constrained and unconstrained firms, we again follow Pugsley et al. (2021) and adopt the statistical model used therein. This model will use the information provided by the autocovariance structure of log employment to capture the importance of both types of heterogeneity.

Consider the following process for employment n by firm i at age a

$$
\underbrace{\ln n_{i,a}}_{\text{log employment}} = \underbrace{u_{i,a} + v_{i,a}}_{\text{Ex-ante component}} + \underbrace{w_{i,a} + z_{i,a}}_{\text{Ex-post component}}, \tag{2.2}
$$

where

$$
\begin{aligned}
u_{i,a} &= \rho_u u_{i,a-1} + \theta_i, & u_{i,-1} &\sim iid\left(\mu_{\tilde{u}}, \sigma_{\tilde{u}}^2\right), & \theta_i &\sim iid\left(\mu_\theta, \sigma_\theta^2\right), & |\rho_u| &\leq 1 \\
v_{i,a} &= \rho_v v_{i,a-1}, & v_{i,-1} &\sim iid\left(\mu_{\tilde{v}}, \sigma_{\tilde{v}}^2\right), & & & |\rho_v| &\leq 1 \\
w_{i,a} &= \rho_w w_{i,a-1} + \varepsilon_{i,a}, & w_{i,-1} &= 0, & \varepsilon_{i,a} &\sim iid\left(0, \sigma_\varepsilon^2\right), & |\rho_w| &\leq 1 \\
z_{i,a} &\sim iid\left(0, \sigma_z^2\right)
\end{aligned}
$$

In this employment process, $u_{i,a} + v_{i,a}$ capture the ex-ante profile, while $w_{i,a} + z_{i,a}$ capture the ex-post one. The ex-ante component is determined by three shocks that are drawn just prior to the birth year, at $a = -1$. The shocks $v_{i,-1}$ and $u_{i,-1}$ represent the initial conditions of the firm, which allow for rich heterogeneity even at birth. θ_i is the permanent component, which will accumulate over the life-cycle at speed ρ_u. In particular, with $\rho_u < 1$, the long-run steady state level of employment will be given by $\frac{\theta_i}{1-\rho_u}$.

This specification will allow for rich heterogeneity not only in terms of optimal size of the firms, depending on the distribution of θ_i, but also in terms of the speed at which firms reach the steady state. As firms start at different points depending on $u_{i,-1}$ and $v_{i,-1}$, each shock with its own persistence parameter, the path from initial to steady state employment will highly differ across firms.

The ex-post component is formed of two different shocks, one i.i.d. shock with expected value of zero, and a persistent one that follows an AR(1) process with i.i.d. inno-

	ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_z
Constrained	0.471	0.867	0.901	0.276	0.741	0.642	0.277	0.184
Unconstrained	0.395	0.729	0.876	0.453	0.677	0.784	0.301	0.190

Table 2.2 Static model parameters for constrained and unconstrained firms.

vations $\epsilon_{i,a}$ and persistence ρ_w. To abstract the ex-post component from affecting the ex-ante one, we set the initial conditions of the persistent shock to $w_{i,-1} = 0$.

To more clearly identify the ex-post and ex-ante contributions one can also derive the formula for the autocovariance, enabling a clear identification of the contribution of both components. The autocovariance formula is given by

$$Cov[\ln n_{i,a}, \ln n_{i,a-j}] = \underbrace{\left(\sum_{k=0}^{a} \rho_u^k \right)\left(\sum_{k=0}^{a-j} \rho_u^k \right)\sigma_\theta^2 + \rho_u^{2(a+1)-j}\sigma_{\hat{u}}^2 + \rho_v^{2(a+1)-j}\sigma_{\hat{v}}^2}_{\text{Ex-ante component}}$$

$$+ \underbrace{\sigma_\epsilon^2 \rho_w^j \sum_{k=0}^{a-j} \rho_w^{2k} + \sigma_z^2 \mathbf{1}_{j=0}}_{\text{Ex-post component}}$$

Derivation of the autocovariance formula is presented in Appendix B.6. The autocovariance is a function of variance and persistence parameters of both ex-ante and ex-post shocks. We calibrate the model for constrained and unconstrained firms by minimizing the sum of squared differences between the model and empirical autocovariance.[7] Table 2.2 presents the parameters resulting from the calibration strategy.[8] Two key parameters of the model are ρ_u and σ_θ, as, together, they imply that steady state heterogeneity exists. The point estimates imply a standard deviation of steady state employment for constrained firms of 0.52 and of 0.67 for unconstrained ones. This, again, highlights that there are differences between both types of firms that come from ex-ante conditions.

Figure 2.3 quantifies the importance of the ex-ante component for the variance of both constrained and unconstrained firms. For both types of firms, the ex-ante component contribution is above 80% at birth. Differences between both types of firms start to arise after year 3, with the ex-ante component explaining more than 50% of the standard deviation for unconstrained firms in the long run, while for constrained firms it is below

---

[7]Firms are split into constrained and unconstrained according to the measure $ConstrainedI$. In the Appendix B.4.2 we redo the exercise using measure $ConstrainedIV$ instead.

[8]Figure B.9 in Appendix B.4.2 plots the model fit to the data for both types of firms.
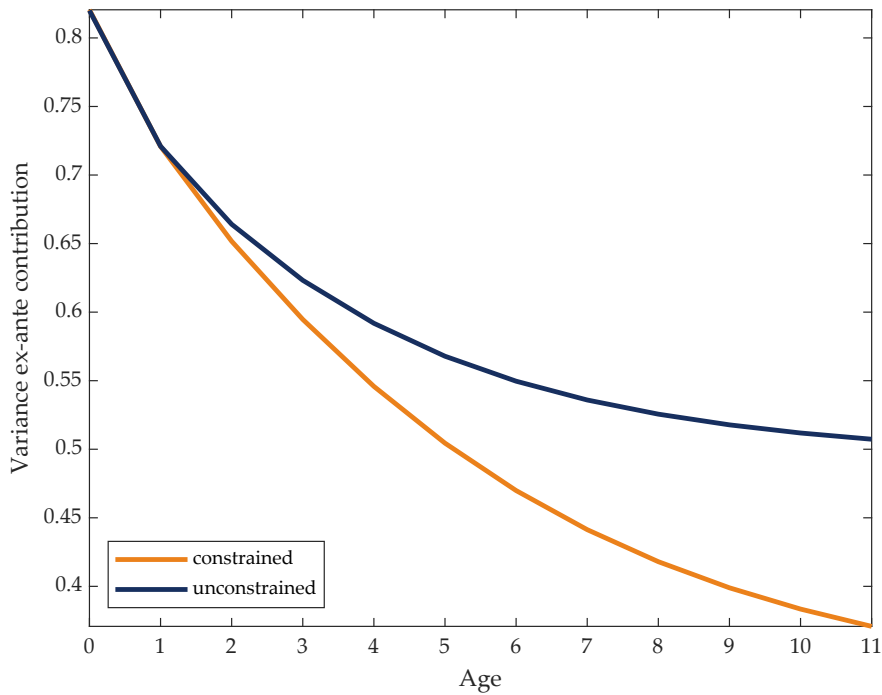
Fig. 2.3 Variance ex-ante contribution.

Notes. Values for constrained firms presented in orange, while blue stands for the unconstrained firms.

40%. As a robustness test we recalibrate the model using $Const.IV$ measure, to guarantee results are not dependent on the selected measure. Table B.12 and Figure B.10 in Appendix B.4.2 present the calibration of the statistical model and fit to data using the measure $Const.IV$. Figure B.11 presents the ex-ante contribution under the alternative calibration, supporting the previous results.

One could have expected the opposite result, as constrained firms would have limited resources to grow and so initial conditions would be more prevalent in explaining the distribution. The fact that the ex-ante contribution is stronger for unconstrained firms may, again, be indicative that this firms are born closer to their optimal size.

All the empirical evidence in this subsection suggests that ex-ante heterogeneity: 1) matters both in the short and in the long-run; 2) more strongly affects unconstrained than constrained firms. Moreover, the standard deviation of steady state level of employment is higher for unconstrained firms. All this evidence may be indicative that unconstrained firms start closer to their steady state level of employment, while firms that still need to grow, exhaust their credit lines to reach their optimal size and so become constrained. To have a better understanding of these facts, we proceed to a general equilibrium, firm dynamics model.

2.5 Model

In this section we present a heterogeneous firms model with financial frictions which aims to reconcile the stylized facts above. We built on Khan & Thomas (2013) and introduce ex-ante heterogeneity through a permanent productivity component which can be interpreted as the firm's business potential. The rationale is that this will break up the strong correlation between size and being financially constrained. Firms with lower permanent productivity will reach their optimal amount of capital earlier and will be unconstrained from then on, while firms which draw a higher permanent component will be constrained much longer as they take longer to grow into their potential.

2.5.1 Households

Households choose consumption, savings and labor supply according to the following maximization problem:

$$V(k) = \max_{c,l,k'} \left\{ U(c,l) + \beta \mathbb{E} V(k') \right\}$$

subject to:

$$k' + c = (1+r)k + \omega l + D$$

.

The first-order conditions for the household problem are standard:

$$U_l(c,l) = \omega U_c(c,l)$$
$$U_c(c,l) = \beta \mathbb{E}\left[(1+r') U_c(c',l') \right].$$

We use the following Greenwood-Hercowitz-Huffman (GHH) utility formulation:

$$U(C,N) = \log(C) + \psi(1-N)$$

Consequently, in the absence of aggregate risk, the first-order conditions are:

$$(1+r) = \frac{1}{\beta}$$
$$\omega = \psi C$$
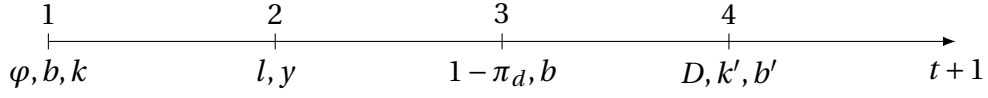
Fig. 2.4 Within period timing of incumbent firm

2.5.2 Production

The production sector features a continuum of firms, indexed by i. Firms are either classified as entrants or incumbents, detailed below.

Incumbents

Incumbent firm i produces according to the following production function:

$$y = \varphi k^{\alpha} l^{\upsilon}, \quad \alpha + \upsilon < 1.$$

where k and l are capital and labor inputs and φ denotes idiosyncratic productivity. Every firm's productivity comprises two components:

$$\ln \varphi_i = w_i + \theta_i,$$

where w_i is an idiosyncratic *transitory* productivity shock, which follows an AR(1) process with persistence ρ_w and variance of innovations σ_ϵ^2. θ_i is the *permanent* productivity component, drawn at birth from a normal distribution with mean μ_θ and variance σ_θ^2

$$\theta_i \sim iid\left(\mu_\theta, \sigma_\theta^2\right)$$
$$w_i' = \rho_w w_i + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \quad \left|\rho_w\right| \leq 1$$

Firm's total profits are given by:

$$\pi = y - \omega l,$$

where ω is the wage per unit of labor.

Figure 2.4 summarizes the within period timing of the incumbent. The firm enters the period with predetermined levels of debt b and capital k and immediately observes its idiosyncratic productivity φ composed of a permanent and transitory component. Next, the firm's labor decision is a static choice that can be found through the firm's first order

condition

$$l(k, \varphi; \omega) = \left(\frac{\upsilon \varphi}{\omega} k^{\alpha} \right)^{\frac{1}{1-\upsilon}}$$

After the production stage, the firm may suffer an exogenous exit shock. The shock happens with probability π_d. So, after the production stage, the value of the firm is given by

$$V^1(x, \varphi) = \pi_d x + (1 - \pi_d) V^2(x, \varphi)$$

If the firm survives the exit shock, at the end of the period it chooses debt b' and capital k' to take to the next period and dividends to distribute this period D to maximize its value

$$V^2(x, \varphi) = \max_{k', b', D} \left[D + \mathbb{E}_{\varphi' | \varphi} \Lambda V^1(x', \varphi') \right] \tag{2.3}$$

$$\text{s.t.:}$$

$$D \equiv x + q b' - k' \geq 0$$

$$b' \leq \xi x$$

$$x' \equiv x(k', b', \varphi') = y(l(k', z'), k', \varphi') - w l(k', \varphi') + (1 - \delta) k' - b'$$

where ξ is the financial parameter that captures the financial frictions in the economy, x is the cash on hand that the firm starts the period today, which is given as the sum of profits plus the value of the non-depreciated capital minus the debt the firm has to pay back. q is the price of the bonds firms issue, with $\frac{1}{q} - 1$ being the equilibrium interest rate in the economy r. Λ is the firm discount factor. As the representative household is the owner of the firm, we assume $\Lambda = \beta$.

Entrants

Entry in this model is exogenous. We assume there is a fixed measure, M_e, of entrants equal to the mass of firms exiting after receiving a death shock. The entrants are assumed to enter with zero debt ($b_0 = 0$) and are log normally distributed over their initial capital k_0 with the mean being anchored at a fraction of the mean of optimal capital levels. The choice of a log normal distribution is motivated by the right skewed distribution of entrants in the data. The initial productivity of each entrant, φ_0, follows the same process as the incumbents productivity. Note that firm entry takes place at the end of a period, and entrants start operating in the next period, given their initial state, (k_0, b_0, φ_0).

2.5.3 Firm Level Decisions

To characterize the firms' decisions we divide the firms into three groups, following Khan and Thomas (2013):

1. **Unconstrained firms.** Firms that can implement the optimal amount of capital and guarantee that in the future they will never be constrained again.

2. **Constrained firms, type 1.** Firms that can implement the optimal amount of capital but not the minimums savings policy that guarantees they will never be constrained again in the future.

3. **Constrained firms, type 2.** Firms that are constrained and cannot implement the optimal amount of capital nor the minimums savings policy.

Unconstrained Firms

This group of firms can implement both the optimal amount of capital and the minimum savings policy that guarantees these firms will never be constrained in the future again. Given the absence of adjustment costs and the stochastic process for φ the optimal amount of capital is given by

$$\max_{k'} -k' + \beta \mathbb{E}_{\varphi'|\varphi} \left[(\pi(k', \varphi') + (1-\delta)k' \right]$$

So the optimal amount of capital solves the following equation

$$\beta \mathbb{E}_{\varphi'|\varphi} \left[\frac{\partial \pi}{\partial k'}(k', \varphi') \right] = 1 + \beta\delta - \beta$$

which is when the expected marginal productivity of capital is equal to the marginal cost of an extra unit. The minimum savings policy these firms implement guarantees they will never be constrained again. It is given by

$$B^*(\varphi_i) = \min_{\varphi_j} \tilde{B}(k^*(\varphi_i), \varphi_j)$$

where $\tilde{B}(k^*(\varphi_i), \varphi_j)$ is the minimum savings that guarantees that going from state φ_i to φ_j the firm is still able to implement the optimal amount of capital. It is given by

$$\tilde{B}(k^*(\varphi_i), \varphi_j) = \pi(k^*(\varphi_i), \varphi_j) + (1-\delta)k^*(\varphi_i) - k^*(\varphi_j) + q\min\left(B^*(\varphi_j), \xi k^*(\varphi_j)\right)$$

Given the optimal amount of capital and the minimum savings policy, the dividends distributed by the unconstrained firms are given by

$$D = x - k^* + qB^*$$

From the dividend constraint $D \geq 0$ we can extract the the minimum threshold for cash-on-hand that guarantees the firm is not constrained

$$\tilde{x} = k^* - qB^*$$

and the firms is constrained if $x \leq \tilde{x}$

Constrained Firms: Type 1

These firms can implement the optimal amount of capital, k^*, but not the optimal savings policy and are therefore partially constrained. As they may still be constrained in future states, they value internal financing more than households value dividends. As a results, for this type of firms, $D = 0$. The amount of debt is given by

$$b' = \frac{(k^* - x)}{q}$$

A firm is type 1 if it can adopt the above amount of debt and capital and at the same time guaranteeing that it does not default in the next period.

Constrained Firms: Type 2

Strictly constrained firms can not implement the optimal amount of capital. Those firms utilize all their borrowing capacity as their marginal value of net worth is greater than unity. Hence, their savings policy is simply

$$b' = \xi x,$$

and their maximum possible investment is consequently

$$k' = x + \xi x < k^*,$$

which is strictly smaller than their optimal level of capital k^*.

2.5.4   Model predictions

The way in which firms respond to different types of shocks will ultimately depend on whether they have reached their optimal amount of capital or whether they are still growing. Hence, in what follows, we refer to firms which can implement their optimal capital level as being unconstrained and otherwise as constrained. Consequently, type 1 constrained firms are considered unconstrained as they can implement the optimal amount of capital and their investment policy is the same as for unconstrained firms if shocks are relatively small.[9] To gain more intuition on the respective investment elasticities to aggregate shocks and the role of ex-ante heterogeneity, we consider a slightly simplified version of the model as outlined in appendix B.5 which results in proposition 1. Constrained firms will only respond more to an aggregate productivity shock if either their marginal product of capital is large enough, i.e. they are far from their potential, or the aggregate shock is quickly fading (ρ is close to 0) which gives unconstrained firms barely any incentive to adjust their capital amount. In fact, the elasticity of unconstrained firms is independent of their potential. The marginal product of capital of constrained firms is higher, the higher their potential and the farther they are from reaching their potential.

Proposition 1 *Constrained firms are more elastic to an aggregate TFP shock than unconstrained firms, absent any cyclically in the constraint, if*

$$mpk > \rho \frac{\alpha}{1 - \alpha} \frac{1}{1 + q_t \xi}.$$

Proof: Proof is provided in the appendix.

Hence, the overall aggregate response of output and capital depends on the distribution of constrained firms across the firms size distribution. Furthermore, the financial accelerator mechanism will only be present in the model economy, if proposition 1 holds on average. However, as already pointed out by Crouzet & Mehrotra (2020), even for low values of ρ, constrained firms will respond less than unconstrained firms and a cyclical collateral constraint is necessary to increase the elasticity of constrained firms. In our discussion about aggregate implications further down, we consider the case of a temporary aggregate shock to total factor productivity (TFP) and a credit shock as a negative shock to borrowing conditions, separately.

---

[9]Large shocks could make the constraint bind again and they would become strictly constrained.

Table 2.3 Parameter values

Parameter	Description	Value		Source
β	Discount factor	0.96		Khan & Thomas (2013)
α	Returns on capital	0.30		Khan & Thomas (2013)
η	Returns on labor	0.60		Khan & Thomas (2013)
δ	Depreciation rate	0.065		Khan & Thomas (2013)
ψ	Labour preference	2.15		Khan & Thomas (2013)
π_d	Exogenous probability of exit	0.02		Data
μ_θ	Average of permanent productivity	0		Normalized
μ_w	Average of transitory shock	0		Normalized
		Trans + perm shock	Transitory shock	
ξ	Collateral constraint	0.57	0.57	Internally calibrated
σ_θ	Standard deviation of permanent productivity	0.16	-	Internally calibrated
ρ_w	Persistence of transitory shock	0.07	0.47	Internally calibrated
σ_w	Standard deviation of transitory shock	0.09	0.09	Internally calibrated
μ_{ke}	Relative size of entrants	0.01	0.09	Internally calibrated
σ_{ke}	Standard deviation of entrants	0.11	0.12	Internally calibrated

Notes. The calibration was done on the full model, i.e. including a transitory and a permanent component of productivity.

2.5.5 Solving and calibrating the model

Solution Method. The model that we set up is similar to Khan & Thomas (2013) in many respects and so naturally we follow their solution algorithm for optimal firm policies. In particular, as outlined in subsection 2.5.3, one can categorize firms into constrained, potentially constrained and unconstrained firms according using the two cash-on-hand thresholds together with current productivity.

To solve for the general equilibrium, we approximate the firm distribution over a fixed grid of net worth using the histogram method proposed by Young (2010). This method has two advantages over a Monte Carlo simulation. For one, it is not prone to sampling bias which might arise from Monte Carlo simulations for specific conditional moments where the law of large numbers is not met. And as a consequence, it is considerably faster than a Monte Carlo approach. The steady state solution is then given at the wage which is leading to a clearance of the goods market.[10] Given the equilibrium wage, we also conduct a Monte Carlo simulation to study the firms' policy responses to aggregate shocks in partial equilibrium.

Calibration. For most of the parameters, which are unrelated to distributions in the model, we follow Khan & Thomas (2013). The set of parameters chosen is documented in the upper part of table 2.3. The discount factor, β, is set to yield an average annual real interest rate of 4%. The production parameters, η and α, imply a labour share of 60% and

---

[10]Market clearing interest rates are given by $1/\beta$.

Table 2.4 Calibrated model fit

Moment	Data	Model trans. + perm. shock	Model transitory shock
Percentage of constrained firms	0.23	0.24	0.24
Share of constrained firms in bottom 20%	0.33	0.29	0.94
Size of 90th-percentile vs. median	9.44	9.65	9.75
Size of constrained firms 90th-percentile vs. median	7.35	7.59	2.48
Size of unconstrained firms 90th-percentile vs. median	9.67	9.27	4.90
Asset share of constrained firms	0.12	0.13	0.01
Share of constrained firms in top 10% vs. bottom 20%	0.36	0.35	0
Percentage of constrained firms in top 1%	0.09	0.06	0

Notes. All moment conditions were equally weighted when minimizing the percentage deviation from the empirical target values. All constrained firms momments are calculated using measure $Const.III$.

capital share of 30%, respectively. Leisure preferences imply that households work one third of their available time.

Firm exit rates in the data are heterogeneous and tend to be lower for larger and older firms. In order to account for that without introducing a size based exit rate schedule, we compute a size weighted average exit rate. When not accounting for lower exit rates among performing firms, small firms with high potential are likely to drop out prior to reaching their optimal amount of capital.[11]

The mean productivity levels for the permanent and transitory component, μ_θ and μ_w, are normalized such that when transforming it to a log-normal distribution, the average productivity component equals one.[12] The rest of the distribution parameters, as well as the maximum borrowing capacity parameter, ξ, are calibrated using the simulated method of moments (SMM). The values presented in table 2.3 minimize the distance between a set of empirical unconditional and conditional moments of the firm distribution, listed in table 2.4, and their model counterparts.

Table 2.4 compares the fit of a model with just a transitory productivity component to a model including both, a transitory and a permanent productivity component. Both models were separately calibrated to find the best match to the data.

When calibrating the model with just a transitory shock to firms' productivity, we use the same ξ as the one internally calibrated in the two shock model and place extra weight

---

[11]The model can still fit the data reasonably well for higher exit rates and far better than a model with just a transitory shock component, yet it gets harder to match the skewness of the firm size distribution as firms with high potential and a long growth path are proportionally more likely to exit before they reach their full size.

[12]Note that the mean of a log-normal distribution is affected not only by the location parameters but also the scale parameter. We adjust it accordingly, such that for any scale parameter, $\mu = 0$ yields an average productivity of 1, when transformed to a log-normal.

on the unconditional share of constrained firms and size moments. Two reasons motivate us to restrict this calibration: 1) when comparing the aggregate responses across the two models we want the elasticity of constrained firms to be comparable and, as established in Proposition 1, the elasticity of these firms depends directly on ξ. Equally, this allows us to simulate the effect of an identical financial shock in both models; 2) the model with just a transitory shock, cannot match the conditional moments such as the relative size of constrained and unconstrained firms, or the share of constrained by size. [13] Given this, we specifically target the fraction of constrained firms and moments of the unconditional size distribution, giving less weight to those moments the model with just a transitory shock could not fit in the first place. This ensures that the underlying size distribution is the same across both models and different predictions are down to differences in the the distribution of constrained firms over the firm size distribution.

As documented in the far right column of table 2.4, the model with just transitory shocks to firms' productivity is not able to match the data well. While it matches the fraction of constrained firms and the unconditional size distribution, it is unable to generate large constrained firms and small unconstrained firms. Hence, with constrained firms concentrated at the bottom of the size distribution, a standard model with just a transitory shock drastically underestimates the asset share of constrained firms. In contrast, when accounting for ex-ante heterogeneity by including a permanent component, and thereby breaking the strong link between size and financial conditions, the model can match the data remarkably well, as documented in the second column of table 2.4.

2.6   Discussion

In this section we start by discussing how the model fits the two stylized facts presented in Section 2.3 and then explore the impacts of accounting for different permanent productivity components across the distribution of firms.

We first start by illustrating that in a standard heterogeneous firms model with only a transitory productivity shock, the model cannot account for the independent effects of size and financial constraints on the investment response to a shock. When adding the permanent productivity shock to a standard model, we can now replicate the empirical

---

[13]In Table B.11 in Appendix B.4.3 we report the calibration of the one shock model without any restrictions, with a free ξ and equal weight for all moments. As can be seen, the model is not able to match the conditional moments and the share of constrained firms across the firm distribution cannot match the empirically observed one, as illustrated in Figure B.12.

Table 2.5 Regression with model data.

	Output growth			
	Model		Model	
	trans. + perm. shock		transitory shock	
	(1)	(2)	(3)	(4)
ΔTFP_t	2.37	2.40	2.29	2.36
$[90,100] \times \Delta TFP_t$	-0.55	-0.55	0.25	0.15
Const. $\times \Delta TFP_t$		0.08		0.18

Notes. Each regression also includes firm fixed effects and all variables separately, however, they were dropped to put focus on the orthogonality of the semi-elasticities w.r.t to size and being constrained in line with table 2.1.

findings from Section 2.3.1, with both size and financial frictions playing an important, yet orthogonal role in explaining the firms' response to an aggregate productivity shock.

Second, we show that, while a model that incorporates both transitory and permanent components of the productivity process can generate constrained firms across the entire distribution, similar to what Figure 2.1 suggests, a model with only a transitory component fails to account for this.

We then proceed to assess the implications of accounting for large constrained firms when faced with an aggregate productivity shock and a financial shock, respectively. Furthermore, we compare the degree of misallocation implied by the model including a permanent productivity component to the standard model with just a transitory component.

2.6.1   Replicating the stylized facts

Size and financial constraints

In Section 2.3.1 we showcase how size and financial frictions both affect firms' response to aggregate shocks and how the two channels appear to be orthogonal to each other. We now check how our benchmark model performs in replicating this stylized fact, and compare it to a similar model without the permanent component of the idiosyncratic productivity process.

To do this, we first run the firm simulation including aggregate productivity shocks. For now, aggregate shocks are assumed to have no persistence. This way, unconstrained firms are completely acyclical which follows from proposition 1 and hence constrained firms are more cyclical. For higher levels of persistence this is only true for a sufficient

degree of cyclicality in the collateral constraint itself. Using the model generated data, we replicate the fixed effects regression from the empirical analysis as follows

$$\Delta \ln y_{i,t} = \Delta TFP_t + (\alpha_j + \beta_j \Delta TFP_t)\mathbf{1}_{i \in S_{[90,100]}} + (\zeta + \eta \Delta TFP_t)\mathbf{1}_{constrained} + \alpha_i + \epsilon_{i,t}$$

where $\Delta \ln y_{i,t}$ is firm i growth rate of output from period t to period $t+1$, ΔTFP_t is the aggregate TFP shock in period t, $\mathbf{1}_{i \in S_{[90,100]}}$ is a dummy for the firm's net worth being in the top decile of the firm distribution, $\mathbf{1}_{constrained}$ is a dummy variable equal to one if the firm is type 2 constrained firm in period t, α_i captures the firm's fixed effect due to heterogeneity in the permanent productivity component and $\epsilon_{i,t}$ is the residual.[14]

Results are presented in Table 2.5. As expected and outlined in proposition 1, large firms react less due to decreasing returns to scale and constrained firms are more elastic since TFP shocks are i.i.d. However, when considering a standard model with just a transitory idiosyncratic productivity component, the coefficient estimating the semi-elasticity of large firms changes quite drastically when controlling for financial conditions. This is due to the high correlation of firm size and financial conditions in this model. When incorporating a permanent productivity component to the model the estimated semi-elasticity of the firms in the upper size decile hardly changes. This suggests that firm size and financial conditions are almost orthogonal[15] to each other, which is in line with the findings in table 2.1.[16]

Constrained firms across the distribution

In Section 2.3.2 we highlight that constrained firms are found across the entire distribution of firms. As illustrated in Figure 2.1, even at the top of the distribution in terms of size, close to 10% of the firms are financially constrained.

Figure 2.5 compares the model generated share of constrained firms across the size distribution with its empirical equivalent. When not allowing for ex-ante heterogeneity between firms, the model can still produce the same overall share of constrained firms, yet the distribution is completely off. Using only a transitory component, the model can

---

[14]We use the type two constrained firms in the model as these are the hard constrained firms and more comparable to the ones with zero potential credit in the data. Although, we equally test the regression accounting for both type 1 and type 2 constrained firms, and results are robust to it.

[15]In fact, the correlation between firms size and being constrained is only -0.07, while it is roughly -0.90 in a model with only a transitory component.

[16]The positive size coefficient in the one shock model is opposite to the empirical evidence. This has to do with selection effects, as the largest firms in the one shock model are the unconstrained with more savings but not necessarily with the largest amount of capital, being the firms with the largest elasticity.

(a) Transitory + permanent shock   (b) Transitory shock
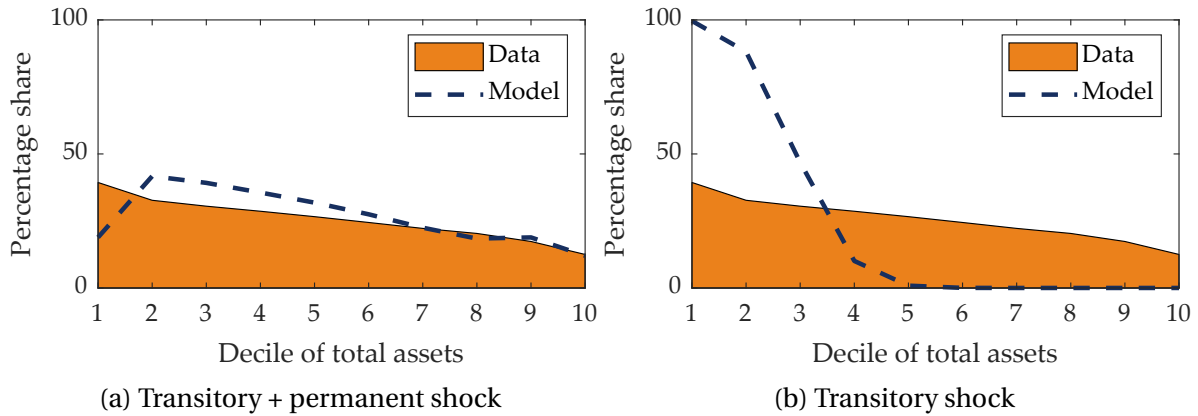
Fig. 2.5 Share of constrained firms across the distribution.

Notes. Constrained firms are identified using measure constrained III which classifies firms as constrained if they have exhausted their potential credit and were not granted additional short- or long-term credit in that period.

neither generate small unconstrained firms nor large constrained firms as depicted in the right panel of figure 2.5. On the other hand, the model with a transitory and a permanent productivity component generates small unconstrained firms, as well as, large constrained firms and is also able to match the untargeted deciles of the empirical distribution quite well.

This is explained by the fact that we have larger firms that are still growing to reach their steady state capital and that are still constrained. At the same time, the model with the two components, accounts for a larger share of small firms that are born at or close to their steady state level of capital. This justifies the lower share of small constrained firms than in the model with only the transitory shock, and also more in line with the data.

Figure B.13 in the appendix offers a different perspective, plotting the density distribution of constrained and unconstrained firms. It is possible to observe that while in the two shock model case the distributions overlap, in the one shock case they are completely separated, with the model only generating small constrained firms and large unconstrained ones.

2.6.2   Aggregate effects

Aggregate productivity shock

We now proceed to assess the aggregate implications of accounting for constrained firms across the entire firm size distribution. First, we consider an unexpected and temporary
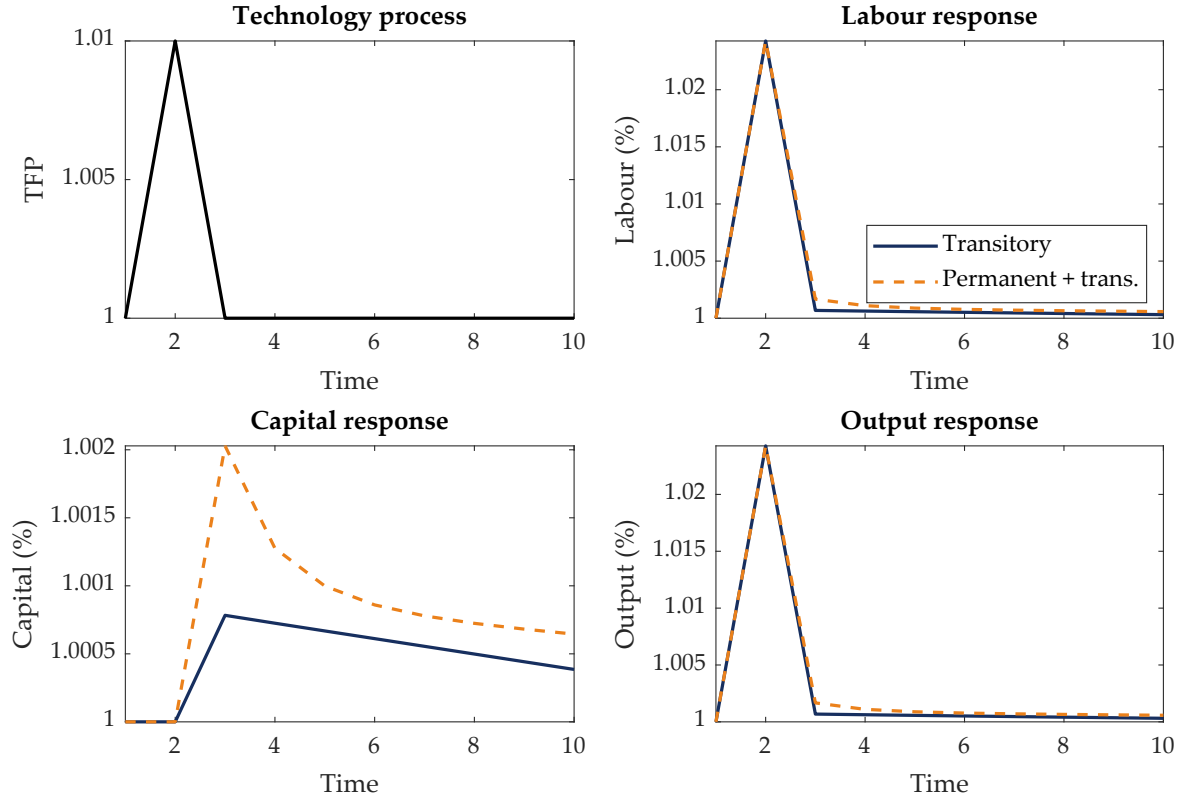
Fig. 2.6 IRFs to the aggregate productivity shock.

Notes. Lines indicate the partial equilibrium response to a shock to overall TFP in the upper left panel, with wages fixed at their steady state level.

1% increase in total factor productivity (TFP) as depicted in the upper left panel of figure 2.6. In a direct response to the shock, firms employ more labour for any predetermined level of capital. While unconstrained firms do not increase their investment in capital due to the transitory nature of the shock, constrained firms leverage their increased net worth to borrow more. This explains why the lagged response in capital is much smaller than the response in labour, as only constrained firms react for the shown case of $\rho = 0$, which is only 23% of all firms in this calibration.

When comparing the two models, we can observe that the aggregate investment response is higher in a model with a permanent and a transitory productivity component. This is simply due to the fact that the asset share of constrained firms is substantially larger than in the transitory shock model. That is a direct consequence of the differences in the distribution of constrained firms, as highlighted in figure 2.5. In fact, figure 2.7, which shows the capital elasticity over the size distribution, illustrates the key difference between both models quite well. For unconstrained firms, as already pointed out, the elasticity
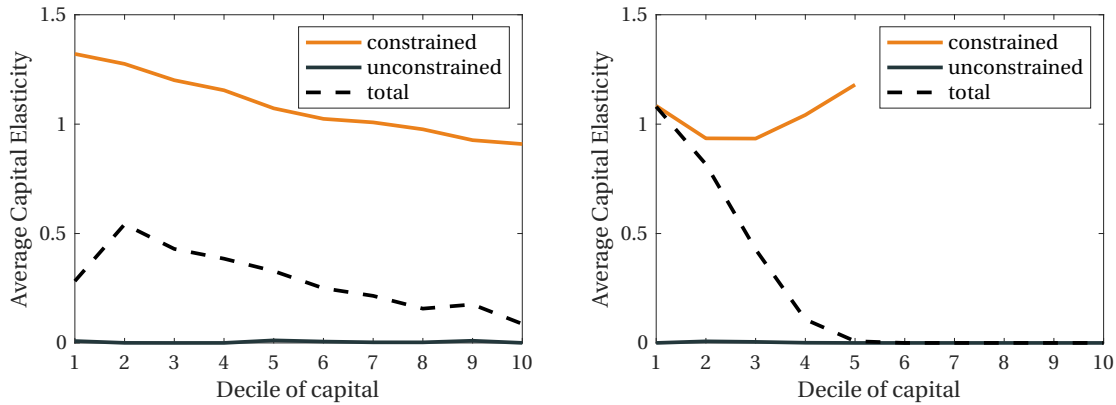
Fig. 2.7 Conditional elasticity over the capital distribution

Notes. Please note that the average capital elasticity of constrained firms has the distribution of constrained firms over the size distribution as its domain. Hence, the orange line in the plot on the right ends at the cutoff after which we only observe unconstrained firms in a model with only a transitory shock.

is zero and for constrained firms, the elasticity is decreasing with size due to decreasing returns to scale. The dashed line is indicating the unconditional average elasticity per decile bin. In a model with just a transitory shock, the overall elasticity is high for small constrained firms but drops to zero at some size cutoff after which all firms become unconstrained. When including a permanent component and thereby generating small unconstrained firms and large constrained firms, the average capital elasticity for small firms is lower than in the one component model but stays above zero for top quantiles of the size distribution. Hence, the capital weighted average elasticity is much higher in the model with a permanent and transitory component, and thus leading to a stronger aggregate capital response.

However, given the small magnitude of the capital response relative to the response in labour, the difference is barely showing up in aggregate output. The effect would become stronger if the borrowing constraint was cyclical or the fraction of constrained firms in the economy was higher.[17]

Financial shock

In a next exercise we consider the effect of a sudden unexpected increase in the severity of financial frictions. We assume a drop in the maximum borrowing capacity of 50%.[18]

---

[17]One should also note that the difference between the models would vanish and eventually flip if the TFP shock gets more persistent and unconstrained firms become more cyclical, as shown in proposition 1.

[18]Khan & Thomas (2013) simulated a 88 percentage point drop in ξ. However, in their calibration the initial level of ξ is 1.38. In our calibration ξ is 0.56, hence a 50% drop equals a 28 percentage point drop in maximum borrowing allowances.
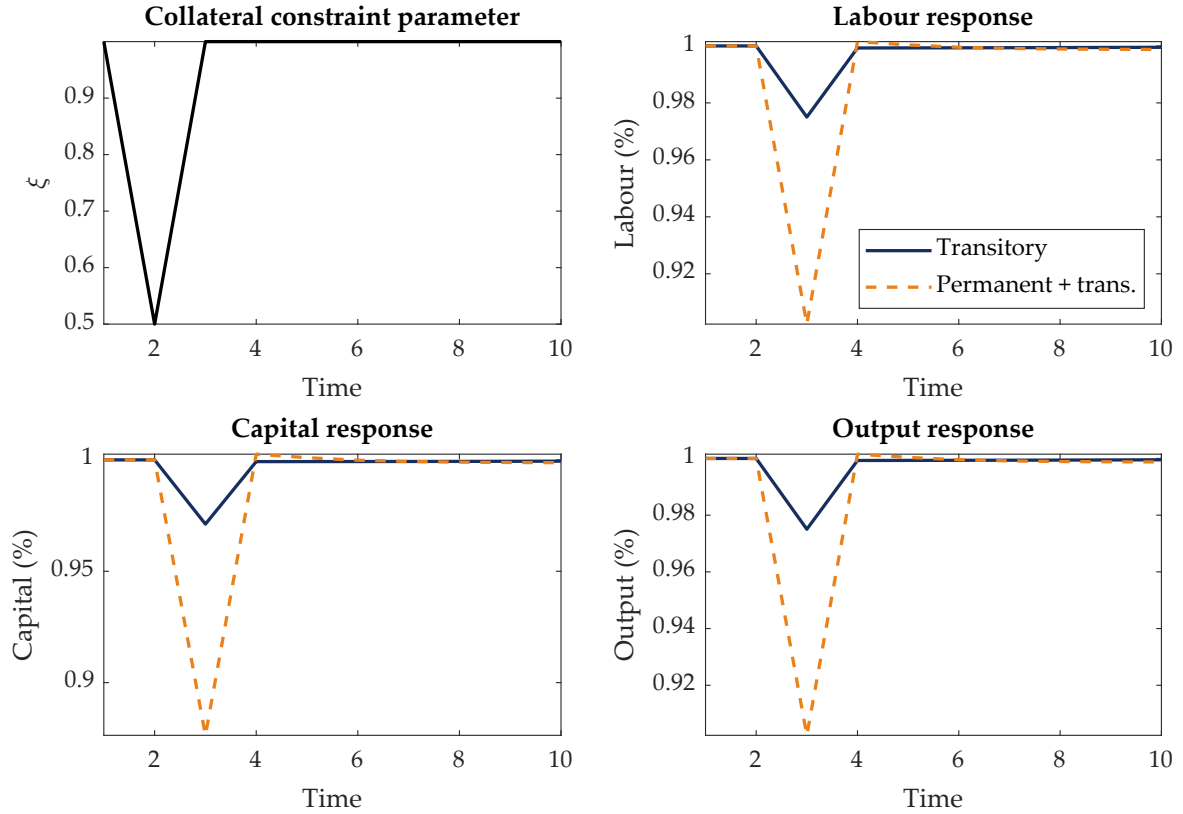
Fig. 2.8 IRFs to a financial shock.

Notes. Lines indicate the partial equilibrium response to a shock to ξ in the upper left panel, with wages fixed at their steady state level.

Again, given the sudden and transitory nature of the financial shock, we assume wages to be fixed at the general equilibrium level before the shock hits.[19]

Figure 2.8 shows the responses to the credit shock depicted in the upper left panel. Since the firm's capital stock is pre-determined, there is no direct impact in period $t = 2$, when the financial shock hits. However, the lower maximum borrowing capacity affects constrained firms in their investment decision, while unconstrained firms remain unaffected by the shock as they finance their investment completely internally.

The resulting aggregate effect of constrained firms having to reduce their investment depends heavily on the distribution of these constrained firms along the firm size distribution. In a model with only transitory productivity shocks, all constrained firms will be concentrated at the lower end of the size distribution. When further accounting for the skewness in the firm size distribution, the capital and asset share of these constrained

---

[19]General equilibrium results for this exercise lead to the same qualitative conclusions, but we prefer the partial equilibrium analysis to isolate the effect coming from the differences in the distribution of constrained firms.

Table 2.6 Deviations from frictionless economy

	Deviations from 1st best	
	trans. + perm. shock	transitory shock
Consumption	0.950	0.982
Capital	0.895	0.952
Output	0.939	0.976
Employment	0.989	0.994
MPK deviation	0.016	0.010
MPK stdev	0.048	0.034

Notes. Reported values are relative to the models without any financial frictions, i.e. when setting the collateral constraint parameter ξ to a sufficiently large value that firms can directly implement their optimal amount of capital.

firms becomes marginal. Hence, despite the drastic shock to financing conditions, the aggregate responses in production factors and ultimately output is relatively minor.

However, when accounting for large constrained firms by introducing ex-ante heterogeneity via a permanent productivity component, aggregate effects get massively amplified simply due to the higher capital share of constrained firms. The quantitative magnitude of the effect clearly depends on the fraction of firms identified as being constrained by the different binary measures ranging from 36% (No potential credit) to as low as 4% (No potential credit and increasing overdue credit) of all firms. Yet, since all measures are suggestive of the notion that constrained firms exist along the entire firm size distribution, a model with just a transitory productivity component could drastically underestimate the aggregate effects of a credit shock.

Capital misallocation

Besides the amplification of macroeconomic shocks, what does a more realistic distribution of financially constrained firms imply for the degree of misallocation in the economy? Pugsley et al. (2021) show that when accounting for ex-ante heterogeneity, the economy exhibits a stronger degree of misallocation. While our results in table 2.6 are qualitatively in line, they suggest that the degree of misallocation can be substantially larger when accounting for the skewness in the firms' size and capital distribution, as well as, large constrained firms.

The stark result is mainly driven by firms with a high draw of permanent productivity but little initial capital. Two modelling assumptions lead to this. First, we assume independence of entry conditions and the firm's potential. Second, the firm's potential is

manifested and observed when the firm enters which is quite a strong assumption and hence results should be rather carefully interpreted as an upper bound of the level of misallocation.

Figures B.15 and B.16 in Appendix B.4.3 illustrate the mechanism explaining the larger capital misallocation. Figure B.15 showcases the density distribution of the MPKs in both models. It is possible to observe that in the two shock model there is a much higher dispersion of MPKs and average. Figure B.4.3 illustrates that the existence of large constrained firms also contributes to having larger MPK at the top of the distribution.

2.7 Conclusion

In this paper we begin by documenting the role of financial frictions for firms' cyclicality and how this channel is independent of the size effect highlighted in recent contributions to the literature. Empirically we show that at any point of the firm distribution there are both constrained and unconstrained firms, which explains the weak correlation between size and financial frictions. We conclude the empirical section by showing that ex-ante conditions matter and point estimates are suggestive of the notion that they affect constrained and unconstrained firms differently.

This motivates us to build a standard firm dynamic model, adding a permanent productivity component. We showcase that by adding this extra component to the productivity process helps us match the distribution of constrained firms across the size distribution, breaking the typical strong correlation between financial constraints and size, generating a large mass of small unconstrained and large constrained firms.

We conclude the paper by illustrating that this mechanism has implications for aggregate responses to shocks. We find aggregate capital and output to respond slightly more to a productivity shock in a model that accounts for ex-ante firm heterogeneity than a model where idiosyncratic productivity is purely driven by a transitory component. This is due to large constrained firms which a model with just a transitory component is unable to generate. Clearly, the existence of large constrained firms amplifies the effects of a financial shock as we show. Furthermore, similar to Pugsley et al. (2021), we find a higher level of misallocation due to financial frictions when accounting for ex-ante heterogeneity driven by small firms with high potential.

References

Baqaee, D. R. & Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1), 105–163.

Bernanke, B. & Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *The American Economic Review*, 79(1), 14–31.

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, 50(4), 623–646.

Buera, F. J., Kaboski, J. P., & Shin, Y. (2011). Finance and development: A tale of two sectors. *American economic review*, 101(5), 1964–2002.

Buera, F. J. & Karmakar, S. (2017). Real effects of financial distress: the role of heterogeneity.

Cooley, T. F. & Quadrini, V. (2001). Financial Markets and Firm Dynamics. *American Economic Review*, 91(5), 1286–1310.

Crouzet, N. & Mehrotra, N. (2020). Small and Large Firms over the Business Cycle. *American Economic Review*. Conditionally accepted.

David, J. M., Hopenhayn, H. A., & Venkateswaran, V. (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics*, 131(2), 943–1005.

Farre-Mensa, J. & Ljungqvist, A. (2016). Do measures of financial constraints measure financial constraints? *The Review of Financial Studies*, 29(2), 271–308.

Gertler, M. & Gilchrist, S. (1994). Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms. *The Quarterly Journal of Economics*, 109(2), 309–340.

Gilchrist, S. & Himmelberg, C. P. (1995). Evidence on the role of cash flow for investment. *Journal of Monetary Economics*, 36(3), 541–572.

Hadlock, C. J. & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the kz index. *The Review of Financial Studies*, 23(5), 1909–1940.

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5), 1127–1150.

Kaplan, S. N. & Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The quarterly journal of economics*, 112(1), 169–215.

Khan, A. & Thomas, J. K. (2013). Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, 121(6), 1055–1107.

Mehrotra, N. & Sergeyev, D. (2020). Financial shocks, firm credit and the great recession. *Journal of Monetary Economics*.

Midrigan, V. & Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American economic review*, 104(2), 422–58.

Ottonello, P. & Winberry, T. (2018). *Financial heterogeneity and the investment channel of monetary policy*. Technical report, National Bureau of Economic Research.

Peters, M. (2020). Heterogeneous markups, growth, and endogenous misallocation. *Econometrica*, 88(5), 2037–2073.

Pugsley, B., Sedláček, P., & Sterk, V. (2021). The nature of firm growth. *American Economic Review*, 111(2), 547–79.

Restuccia, D. & Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3), 151–74.

Sharpe, S. (1994). Financial market imperfections, firm leverage, and the cyclicality of employment. *American Economic Review*, 84(4), 1060–74.

Whited, T. M. & Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies*, 19(2), 531–559.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1), 36–41.

Chapter 3

Saving Shocks and Business Cycle Fluctuations

Abstract

A common modeling tool – and a popular narrative – used to explain the financial crisis of 2008-2010 is a sudden increase in the desire to save. Such "marginal propensity to save" (MPS) shocks can be triggered by, for instance, a rise in uncertainty surrounding the economic climate, and depress interest rates, inflation, and generally cause an economic contraction. This paper uses the long-run properties arising from MPS shocks in both exogenous- and endogenous growth models with sticky prices in order to identify their causal effect on output. We find that time series data from the United States is strongly supportive of the notion that MPS shocks indeed have a causal, and contractionary, effect on economic activity, lending support to the most common approach of studying the financial crisis.[1]

---

[1]This chapter is based on joint work with Pontus Rendahl.

3.1   Introduction

There is growing consensus on a particular narrative about what triggered the financial crisis of 2008-2010. This narrative is essentially based on three concurring factors: First, housing prices started to decline in the summer of 2007. Second, the financial system was heavily invested in housing-related assets and mortgage-backed securities of low quality and third, the shadow banking system was highly leveraged in housing assets and highly vulnerable to bank runs (Christiano et al., 2017). The fall in housing prices led to a decline in asset values which forced the shadow banking system to sell them immediately, reinforcing the decline and damaging the whole banking system. The subsequent credit crunch made this decrease in asset values even more severe. The reduction in household wealth due to the interplay of these factors and the rise in uncertainty triggered households to spend less and save more. This increase in the saving rate is depicted in the left panel of figure 3.1[2]. While the savings rate has been declining for decades, there is a clear structural break around the time of the financial crisis sending the saving rate on an upward trend. As households cut back on spending, firms reduced investments and hiring due to falling sales. Hence, it can be argued that the increase in the marginal propensity to save (MPS) reinforced the effect of the three factors above, pushing the economy deeper into recession. Consequently, a common modelling tool to explain the financial crisis has been the introduction of MPS shocks as in Eggertsson et al. (2003) and Christiano et al. (2011).

Building on this narrative, this paper investigates the causal link between saving shocks and business cycle fluctuations. The right panel of figure 3.1 illustrates the need for identification as there is no clear relationship between personal savings and output growth. Moreover, identification approaches using short-run restrictions are hard to justify given the simultaneity issue between output and the saving rate, with the saving rate being defined as savings over income. Therefore, we use the long-run properties arising from MPS shocks in both exogenous- and endogenous growth models with sticky prices in order to identify their causal effect on output. Standard textbook growth models as the Solow model or endogenous versions such as the AK model and Schumpeterian growth models would predict an increase in growth following an increase in the saving rate.

---

[2]The plot excludes the period of the Covid crisis as an outlier for legibility reasons. Since this paper was written before the Covid crisis and to make sure results are not driven by the spike in saving rates, baseline results also exclude the Covid period, but we do include it for robustness checks.
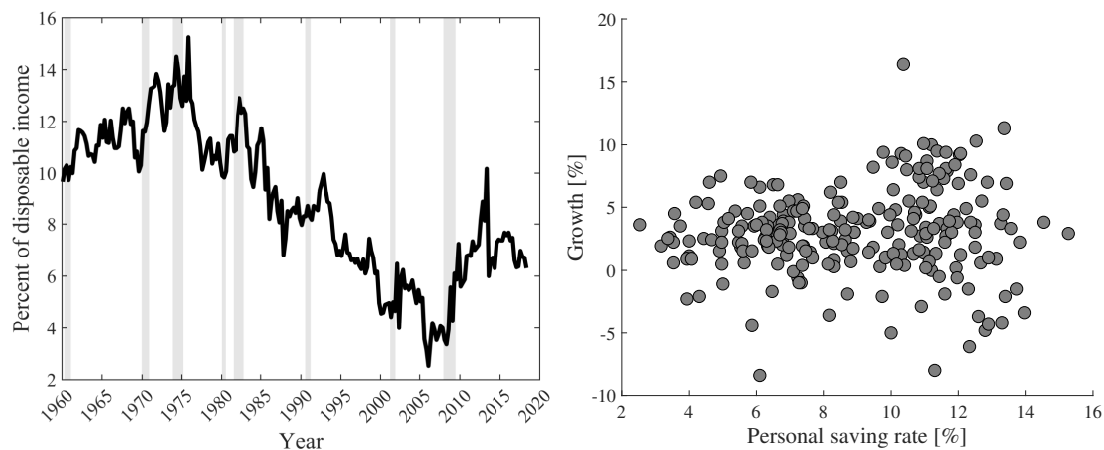
Fig. 3.1 Personal saving rate and growth

Notes. The left panel depicts the personal saving rate in the U.S. from Q1 1960 to Q2 2018. Grey shaded areas indicate the starting and ending dates for recession periods, as determined by the National Bureau of Economic Research (NBER). The right panel depicts the relationship between growth and changes in the saving rate.

However, they only offer a long-run perspective. When taking into account price rigidities, a rise in the saving rate can have a contractionary effect on output in the short-run[3].

Using time series data from the United States, our empirical findings are threefold: First, MPS shocks indeed have a causal, and contractionary, effect on economic activity, lending support to the most common approach of studying the financial crisis. Secondly, we find evidence of the paradox-of-thrift, namely that a rise in the saving rate can lead to a decrease in total savings as output contracts. This corroborates the narrative of the financial crisis being ultimately the response to a negative shock to the demands for goods all across the board and underlines the claim by Christiano et al. (2017) to reconsider the theory of the paradox-of-thrift in modern macroeconomics. Thirdly, combining long- and short-run restrictions we find evidence that saving shocks can have a negative impact orthogonal to uncertainty shocks. In fact, uncertainty shocks appear to be transmitted almost entirely via changes in the saving rate.

Literature. The link between uncertainty surrounding the economic climate and precautionary savings is well documented in theoretical and empirical contributions to the

---

[3]Guerrieri & Lorenzoni (2017) show that an increase in precautionary savings can generate an output drop even if prices are flexible in a model with incomplete markets and borrowing constraints. Under sticky prices, the output drop will be larger if the economy is in a liquidity trap.

literature[4]. Greater uncertainty increases the incentive for households to save in order to protect themselves against the higher likelihood of adverse economic outcomes in the future. Theoretical contributions on the topic of uncertainty and precautionary savings include Leland (1978), Skinner (1988), Zeldes (1989), Deaton (1989), Caballero (1991) and Carroll et al. (1992) among others. Empirical work studying the importance of precautionary savings at the household level include Carroll (1997), Engen & Gruber (2001) and Gourinchas & Parker (2002). The effect of uncertainty on business cycles is of more recent interest in response to the financial crisis. Gourio (2012) for instance shows that an increase in uncertainty modeled as a rise in the perceived probability of disaster leads to a collapse of investment and output. When inferring the probability of disaster from observed asset prices, the variation in this probability measure can explain a significant fraction of business cycle dynamics, especially for the duration of the financial crisis. Related to this, Christiano et al. (2014) find that fluctuations in risk are the most important shock driving business cycle fluctuations. Furthermore, Basu & Bundick (2017) argue that higher uncertainty has even more negative effects if monetary policy can no longer perform its usual stabilizing function because of the zero lower bound and consequently, increased uncertainty about the future played a key role in worsening the financial crisis. In a recent paper, Bloom et al. (2018) find theoretical evidence that uncertainty shocks can generate drops in gross domestic product of up to 2.5%.

Outlook. The rest of the paper is structured as follows. First, we present a version of the quality ladder model by Aghion & Howitt (1992) with sticky prices. An exogenous growth model is nested if we shut down the R&D channel. Section 3.3 introduces the empirical approach used to identify saving rate shocks and their effect on output in the data. Results for U.S. time series data are presented in section 3.4 and section 3.5 concludes.

3.2   Theoretical model

In this section we present the equilibrium conditions of a growth model with sticky prices that are utilized to derive the long-run restrictions imposed in the empirical model. We consider the simplest version of a Schumpeterian growth model where endogenous growth is induced through improvements in the quality of goods as introduced by Aghion & Howitt (1992) and Grossman & Helpman (1991). The main idea of these so-called quality-ladder models is that refinements of existing intermediate goods and techniques

---

[4]See Lugilde et al. (2019) for a comprehensive review on the empirical literature on precautionary savings.

increase their productivity in producing the final output. Quality improvements occur stochastically and at different rates across intermediate good sectors. An important feature of these models is that, when a product is improved, it tends to displace the old one given its higher productivity and substitutability in the production process of the final good. Hence, successful research along the quality dimension eliminates the monopoly rents of the preceding intermediate product. This process of endogenous growth is commonly referred to as "creative destruction". An exogenous version of the growth model is nested when shutting down the option to invest into R&D.

The model departs from the textbook quality ladder model as laid out in Barro & Martin (2004) by assuming that prices are sticky. This allows us to identify short run effects on output following a shock in the households' MPS.

3.2.1   Setup

Household.   There is an infinitely lived representative household which aims to maximise its expected present value of utility over its lifetime using a discount factor β. The household obtains utility from consumption according to CRRA preferences and disutility from supplying labour which are additively separable

$$U_t = \mathbb{E}_t \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\gamma} - 1}{1 - \gamma} - \psi \frac{L_t^{1+\eta}}{1 + \eta} \right). \tag{3.1}$$

The household maximizes equation (1) subject to its budget constraint choosing optimal consumption, bond purchases and how much labour to supply

$$P_t C_t + B_t = P_t w_t L_t + r_t B_{t-1} + \Pi_t. \tag{3.2}$$

First order conditions with respect to consumption and bond holdings lead to the intertemporal Euler equation of the household

$$U_{C_t} = \beta \mathbb{E}_t \left[r_{t+1} \frac{P_t}{P_{t+1}} U_{C_{t+1}} \right]. \tag{3.3}$$

Shocks in the marginal propensity to save are modelled as shocks to the discount factor β. Following a rise in β it is optimal for the household to reduce its contemporaneous consumption in order to save for future periods. The first order condition with respect to

labour leads to the intratemporal labour choice

$$\psi L_t^\eta = C_t^{-\gamma} w_t. \tag{3.4}$$

Final goods sector. In the perfectly competitive final goods sector, the final good Y is produced using technology A, labour L and a continuum of varieties of imperfectly substitutable intermediate goods. Intermediate goods are augmented by their quality rung q^{k_i} and combined according to a constant elasticity of substitution (CES) aggregator.

$$Y_t = A_t L_t^{1/\sigma} \int_{i=0}^1 \left(q^{k_{i,t}} y_{i,t} \right)^{\frac{\sigma-1}{\sigma}} di, \tag{3.5}$$

where $y_{i,t}$ is the amount of sector i's intermediate input used at the final production stage, and q^{k_i} denotes its quality rung. Elasticity of substitution between intermediate goods σ is larger than unity to allow for endogenous growth through quality improvements. Increases in the level of technology A can be viewed as a source of exogenous growth. For the growth accounting exercise further below it is convenient to define the economy's aggregate quality index as the weighted sum of quality levels associated with each sector's intermediate input

$$Q_t = \int_{i=0}^1 \left(q^{k_{i,t}} \right)^{\sigma-1} di. \tag{3.6}$$

Growing Q_t through successful research raises the quality of final consumption as much as it raises its quantity. From the representative firm's profits which are given as follows

$$\Pi_t^Y = P_t Y_t - \int_0^1 p_{i,t} y_{i,t} di - w_t P_t L_t, \tag{3.7}$$

we can derive the firm's demand for labour and the optimal production of the intermediate good y_i, respectively,

$$\frac{1}{\sigma} \frac{Y_t}{L_t} = w_t, \tag{3.8}$$

and

$$y_{i,t} = \left(\frac{p_{i,t}}{P_t} \right)^{-\sigma} \left(\frac{\sigma-1}{\sigma} A_t \right)^\sigma L_t q^{(\sigma-1)k_{i,t}}. \tag{3.9}$$

Intermediate goods producers. There is a continuum of intermediate firms producing the intermediate input using output as the sole factor of production. They produce according to the latest technology acquired from the research sector. Innovations are assumed to be drastic such that the intermediate monopolist is unconstrained by potential

competition from previous patents. As intermediate goods are imperfect substitutes for the final good, producers compete in an environment of monopolistic competition and choose their optimal price taking into account the sector's demand function. However, price setting comes with a quadratic adjustment cost following Rotemberg (1982). Due to nominal rigidities the firms face an intertemporal pricing problem

$$\max_{p_{i,t+s}} \mathbb{E}_t \left[\sum_{s=0}^{\infty} m_{t,t+s} \left(p_{i,t+s} - P_{t+s} - \frac{\theta}{2} \left(\frac{p_{i,t+s}}{p_{i,t+s+1}} - 1 \right)^2 p_{i,t+s} \right) y_{i,t+s} \right], \qquad (3.10)$$

where $m_{t,t+s}$ is a stochastic discount factor capturing the general discount factor β, the marginal rate of substitution between future and present consumption and the survival rate of the firm. The survival rate is defined as $(1 - \mu)$, where μ is the probability of a jump on the quality ladder, hence the likelihood for creative destruction in this sector. The pricing problem is symmetric across sectors as the optimal price and the likelihood for being replaced is independent of the rung on the quality ladder as we will see below. This results in the following mark-up expression for the optimal price

$$p_{i,t}^* = \psi_t P_t, \qquad (3.11)$$

with

$$\psi_t = \frac{\sigma}{(\sigma - 1)(1 - \frac{\theta}{2}(\pi_t - 1)^2) + \theta(\pi_t - 1)\pi_t - \theta \mathbb{E}_t \left[m_{t,t+1}(\pi_{t+1})^2(\pi_{t+1} - 1) \frac{y_{i,t+1}}{y_{i,t}} \right]}, \qquad (3.12)$$

where π is the inflation rate. Normalizing the aggregate price level to unity we have that intermediate goods' prices equal the markup. For flexible prices we get the standard markup expression in terms of the elasticity of substitution between intermediate input, $\psi_t = \frac{\sigma}{\sigma - 1}$.

The market value of the intermediate firm using the latest patent is defined as the discounted expected future profits, which can be written recursively as follows

$$V_{k_i} = \left(p_{k_i,t} - P_t - \frac{\theta}{2}(\pi_t - 1)^2 p_{k_i,t} \right) y_{k_i,t} + m_{t,t+1} V_{k_i}' \qquad (3.13)$$

$$= \left(\psi_t - (1 + \frac{\theta}{2}(\pi_t - 1)^2) \right) y_{k_i,t} + m_{t,t+1} V_{k_i}'. \qquad (3.14)$$

85

R&D and patents. For a given quality rung k_i the probability of success depends on the amount of research resources $z_{i,t}$ in form of the final good used

$$\mu_{k_{i,t}} = \phi(k_{i,t}) z_{i,t}, \tag{3.15}$$

where $\phi(k_{i,t})$ captures the fact that for any given level of research, the probability of success decreases in the number of innovations already made in this sector (thus $\phi'(k_{i,t}) < 0$). Following standard specifications of $\phi(k_{i,t})$ in the literature, we define it as follows

$$\phi(k_{i,t}) = \frac{1}{\lambda} (q^{k_i+1})^{(1-\sigma)}, \tag{3.16}$$

where $1/\lambda$ is the productivity of goods invested in research and development. Note that the specification is chosen such as to offset the positive effect of a sector's position on the quality ladder on profits. Recall, that instantaneous profits depend linearly on q^{k_i+1} via $y_{k_{i,t}}$.

Given that intermediate firms make profits, there will be competition for the latest patents. Ultimately, a single firm is willing to pay the equivalent to its present value at market entrance. Free entry into the R&D sector leads to a zero profit condition

$$\phi(k_{i,t}) V_{k_{i,t}} - P_t = 0, \tag{3.17}$$

which implies that expected profits from research per researcher equal the wage. The zero profit condition pins down the optimal level of research and hence the probability of success μ which turns out to be constant over time. Using the expression of a firm's present value producing on the latest quality rung, we can see that the free entry condition is independent of the specific sector as q^{k_i+1} drops out

$$\frac{1}{\lambda} (q^{k_i+1})^{(1-\sigma)} \mathbb{E}_t \sum_{s=0}^{\infty} m_{t,t+s} \left(p_{i,t+s} - P_{t+s} - \frac{\theta}{2} (\pi_{t+s} - 1)^2 p_{i,t+s} \right) \tag{3.18}$$

$$\left(\frac{p_{i,t}}{P_t} \right)^{-\sigma} \left(\frac{\sigma-1}{\sigma} A \right)^{\sigma} L_t q^{(\sigma-1)(k_i+1)} - P_t, \tag{3.19}$$

$$= \frac{1}{\lambda} \mathbb{E}_t \sum_{s=0}^{\infty} m_{t,t+s} \left(p_{i,t+s} - P_{t+s} - \frac{\theta}{2} (\pi_{t+s} - 1)^2 P_{i,t+s} \right) \tag{3.20}$$

$$\left(\frac{p_{i,t}}{P_t} \right)^{-\sigma} \left(\frac{\sigma-1}{\sigma} A \right)^{\sigma} L_t - P_t = 0. \tag{3.21}$$

Monetary authority. Lastly, there is a central bank setting the nominal interest rate following a Taylor rule. The central bank takes into account the deviation from price stability as captured by inflation and the deviation of output from its potential as captured by its steady-state measure:

$$\log(R_t/R_{ss}) = \rho_R \log(R_{t-1}/R_{ss}) + (1 - \rho_R)[\theta_\pi \log(\pi_t/\pi_{ss}) + \theta_Y \log(Y_t/Y_{ss})], \qquad (3.22)$$

where π_t is gross inflation of the final good price, θ_π is the weight on inflation in the reaction function. Analogously, θ_Y is the weight on the output gap, measured as deviation from steady state output, Y_{ss}. ρ_R is a smoothing parameter.

3.2.2 Model predictions

Table 3.1 Non-estimated parameters and calibrated values

Parameter	Value	Description	Source
		Panel A: Parameters	
β	0.994	Discount factor	Basu & Bundick (2017)
γ	1	Risk aversion	Log utility
η	2	Frisch elasticity	Basu & Bundick (2017)
ψ	0.95	Weight on leisure preference	Basu & Bundick (2017)
ρ_r	0.8	Taylor rule smoothing parameter	Standard
θ_π	1.5	Taylor rule inflation coefficient	Basu & Bundick (2017)
θ_y	0.2	Taylor rule output gap coefficient	Basu & Bundick (2017)
θ_p	100	Rotemberg coefficient of price adjustment cost	Basu & Bundick (2017)
A	1	Economy wide TFP	Normalized
Q	1	Quality index	Normalized
σ	6	CES parameter	Basu & Bundick (2017)
		Panel B: Steady state values	
g_Y	0.5%	Quarterly growth rate of final good	
Y	1	Output of final good	
ψ	1.2	Markup	

Both, the endogenous and exogenous growth model are calibrated such that they result in the same quarterly growth rate of 0.5%. The discount factor is chosen such that it corresponds to a quarterly interest rate of 0.5% which roughly amounts to a 2% annual interest rate. The only shock in this model is a shock to the saving behaviour of households. A shock in the MPS is modeled as a one percent shock to the discount factor. The constant technological progress A in case all growth comes through quality refinements and the

constant quality index Q in the case of exogenous growth are set to one such that final output is the same in both models for the same amount of labour utilized. The constant elasticity of substitution between intermediate inputs is chosen in order to obtain a steady state markup of 20%. Based on this markup, the Rotemberg (1982) coefficient of price adjustment is chosen to mimic a Calvo frequency re-optimizing firms of around 25% each quarter following Keen & Wang (2007)[5]. In the following analysis, this is compared to a model with a flexible-price calibration ($\theta_p = 0$).
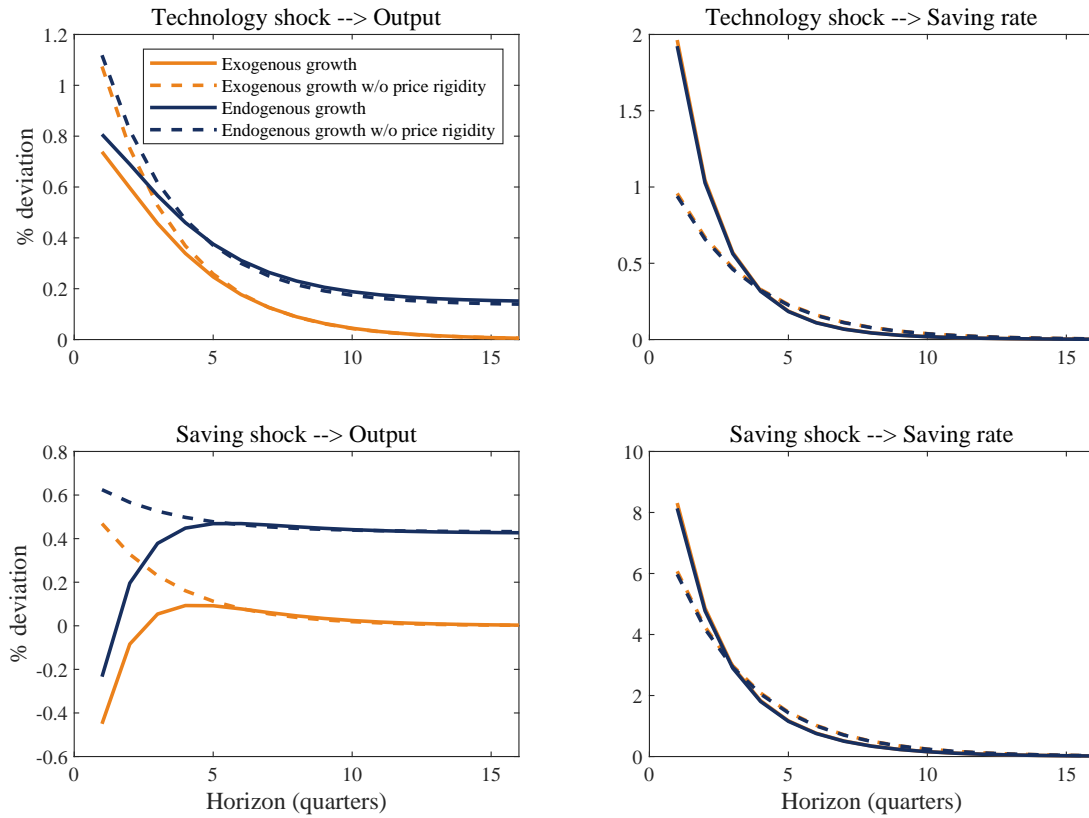
Fig. 3.2 Impulse response functions for the exogenous and endogenous growth model.

Notes.

Figure 3.2 plots the impulse responses of the endogenous and exogenous growth model for both, the sticky-price and flexible-price calibration in case of a temporary shock[6]. Following a productivity shock, output increases for all specifications. In the

---

[5]This is also broadly in line with the calibration in Basu & Bundick (2017) who use a slightly higher quadratic adjustment cost due to their inflation target of 2%.

[6]Impulse responses for a permanent shock to the discount factor are shown in figure C.3 in the appendix

presence of price stickiness the response is lower in magnitude, due to the quadratic adjustment cost faced by intermediate good producers. While, for the exogenous growth model, output goes back to its steady state level, extra savings lead to more R&D activity in the endogenous growth model, ultimately resulting in more growth in the quality index and hence output. In case of a shock to the discount factor, the response depends on whether prices are flexible or rigid. In a flexible-price regime, prices adjust to compensate for the lack in demand and higher savings are increasing output, both in the exogenous and endogenous growth model. If prices are sticky however, the demand shock has a contractionary effect on output. Firms face lower demand, but are unable to flexibly change their prices, ultimately having to reduce production to still maximize profits. Eventually, output will go back to its steady state level in the exogenous growth model, while it reaches a higher steady state in the endogenous growth model due to innovations in the R&D sector.

Those model predictions allow us to derive several identification restrictions for our empirical strategy to estimate the effect of saving shocks on the business cycle. These are lined out in the next section.

3.3   Empirical strategy

This sections introduces the bivariate[7] structural VAR specifications and long-run exclusion restrictions used to identify saving shocks in the data.[8] The empirical model interprets the variations in differences in (log) output, Δy_t, and the saving rate, sr_t, as originating from two types of exogenous disturbances, namely saving shocks and technology shocks which are by definition orthogonal to each other. The propagation mechanisms of those shocks over time are unspecified. We can formalize this idea by writing it as a structural vector autoregression (SVAR) model:

$$A \begin{bmatrix} \Delta y_t \\ sr_t \end{bmatrix} = B(L) \begin{bmatrix} \Delta y_{t-1} \\ sr_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^s \end{bmatrix}, \tag{3.23}$$

where ϵ_t^y and ϵ_t^s denote structural technology shocks and saving shocks, respectively. The orthogonality assumption combined with the normalization that $\epsilon_t = (\epsilon_t^y, \epsilon_t^s)' \sim (0, I_2)$

---

[7]We also considered a trivariate VAR including the federal funds rate to disentangle MPS shocks from saving shocks induced by monetary policy, yet results were very similar to the bivariate VAR.

[8]For a detailed explanation of SVARs and long-run restrictions please refer to Kilian & Lütkepohl (2017).

implies that $\mathbb{E}[\epsilon_t \epsilon_t'] = I$. Given equation 3.23 and assuming invertability, the reduced form VAR is given by

$$
\begin{bmatrix} \Delta y_t \\ sr_t \end{bmatrix} = \boldsymbol{F}(L) \begin{bmatrix} \Delta y_{t-1} \\ sr_{t-1} \end{bmatrix} + \boldsymbol{Q} \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^s \end{bmatrix}, \tag{3.24}
$$

where $\boldsymbol{F}(L) = \boldsymbol{A}^{-1}\boldsymbol{B}(L)$, and $\boldsymbol{Q} = \boldsymbol{A}^{-1}$. Using the mapping from structural shocks to reduced form errors

$$
u_t = \boldsymbol{Q} \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^s \end{bmatrix}, \tag{3.25}
$$

we can simplify the reduced form VAR to

$$
Y_t = \boldsymbol{F}(L) Y_{t-1} + u_t. \tag{3.26}
$$

Informed by the model predictions, we derive several long-run restrictions for saving shocks, while technology shocks can have a permanent effect as in Gali (1999). The long-run exclusion restrictions are imposed on the moving average representation of the SVAR model, which is given by

$$
Y_\infty = (I - F(L))^{-1} Q \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^s \end{bmatrix} = \Theta \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^s \end{bmatrix}, \tag{3.27}
$$

where Θ is the long-run impact matrix.

In order to identify the structural shocks from the residuals of reduced form estimations of the system of variables, we impose the long-run restrictions implied by the theoretical model following the identification approach originally proposed by Blanchard & Quah (1989). They show that even if the economy is subject to multiple demand shocks, the identification conditions are satisfied as long as only the supply shock has lasting effects on output and the other demand disturbances do not alter the relationship between output and the disturbance of interest, here saving shocks. We argue that this is the case as sudden changes to consumption choices by households are at the very core of the definition of demand disturbances.

Identification restrictions. Based on the theoretical results of temporary and permanent saving shocks in an exogenous and endogenous growth model as shown in the appendix, we derive three exclusion restrictions

9

Exclusion Restriction 1 *In an exogenous growth model, a temporary shock to the marginal propensity to save has no long-term effect on output.*

In this case $Y_t = [\Delta y_t, sr_t]'$ and the effect of any of the structural shocks on Y_t will approach zero as the horizon increases. The effect of a structural shock on the level of real GDP y_t is the cumulative sum of its effects on Δy_t. The long-run cumulative effects are given by the matrix Θ. Hence, the first exclusion restriction corresponds to $\Theta_{12} = 0$. In other words, the restriction demands that a saving shock is such that when cumulating the impulse response function of growth, there will be eventually no effect on real GDP. The other entries of the long-run effect matrix remain unrestricted as supply shocks such as technology shocks affect the level of real GDP in the long run. Moreover, there are no restrictions on the second row of Θ because the cumulative responses of a stationary series are different from zero.

Exclusion Restriction 2 *In an endogenous growth model, a temporary shock to the marginal propensity to save has no long-term effect on output growth.*

In an endogenous growth model, a reduction in contemporaneous consumption in combination with sticky prices will act as a demand shock to output but will also increase investment into R&D as output decreases by less than consumption. Higher investment in R&D will lead to a higher probability of research success and subsequently a temporary rise in TFP growth. Therefore, under the assumption of an endogenous growth model, a temporary saving shock can act as a supply shock, leading to a permanent change in the level of output in the long-run. However, it has no impact on the long-term growth rate. Implementing the second exclusion restriction, the vector of stationary processes is defined as $Y_t = [\Delta^2 y_t, sr_t]'$. The restriction to the long-run cumulative effects matrix is the same as above, i.e. $\Theta_{12} = 0$ such that a saving shock has no long-run effect on the growth rate of output.

Exclusion Restriction 3 *In an endogenous growth model, a permanent shock to the marginal propensity to save has no long-term effect on the change in output growth.*

A permanent saving shock increases the steady state investment in R&D and thereby fosters higher TFP growth in the long-run. While the growth rate will increase to reach

---

[9]Please note that we do not consider the restriction of a permanent shock to the saving rate having no long-run effect on growth as this is true by assumption in an exogenous growth model and is not an outcome of the model.

its higher steady state, it is constant in the long-run. Hence, in case of an endogenous growth model, a permanent positive saving shock leads to a rise in the growth rate but no change in the slope of the growth rate thereafter. This leads us to the third exclusion restriction, where we define the vector of stationary processes as $Y_t = [\Delta^3 y_t, \Delta sr_t]'$. Here, we assume the saving rate to be I(1) such that a temporary increase in the first difference leads to a permanent increase in the level of the saving rate. The restriction to the long-run cumulative effects matrix is the same as above, i.e. $\Theta_{12} = 0$ such that a saving shock has no long-run effect on the slope of the growth rate.

Reduced-form specification.   Consistent estimates of the coefficients of $F(L)$ are obtained as functions of the estimated parameters of a reduced-form VAR. The baseline specification of the reduced-form estimation is as follows

$$Y_t = c_t + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t, \tag{3.28}$$

where Y_t is defined according to the specific exclusion restriction utilized, c_t is a (2×2) matrix containing a constant term and a linear trend.[10] β_i is a (2×1) vector of parameter estimates for the i-th lag of Y_t with a maximum of p lags. For our baseline specification we use eight lags, i.e. two years for quarterly data, but consider alternative choices as robustness checks. The last term u_t denotes the residuals of the reduced-form estimation.

From the reduced form estimates we obtain the structural estimates by following the standard implementation of long-run restrictions. First, we calculate the companion form of the beta estimates F, then we find the restricted long-run impact matrix, D, from a Cholesky decomposition of $LR \sum LR'$, where $\sum$ is the covariance matrix of the residuals and LR is a matrix containing the first two row and column entries of $(I - F)^{-1}$. Using the long-run impact matrix, we can back out the contemporaneous impact matrix as $Q = (I - F)D$.

---

[10]We also included dummies for several potential structural breaks but they did not alter the results in any meaningful way.

3.4   Empirical results and robustness

3.4.1   Baseline results

In this section, we present the results of the SVAR models using quarterly U.S. time series data from 1960-Q1 until 2019-Q4. All data comes from the Federal Reserve Economic Database (FRED). Growth is defined as the seasonally adjusted percentage change in real gross domestic product from the preceding period. We use two measures for the saving rate. The first is the personal saving rate as the ratio between personal savings and disposable income of households. The second measures the aggregate saving rate as the ratio between personal savings and the real gross domestic product. Just by the definition of the saving rate, the simultaneity issue becomes apparent as output directly enters the definition of the saving rate. Hence, any short-run identification strategy would be flawed. Long-run restrictions, on the other hand, aim to disentangle this simultaneity by differentiating between technology shocks shifting output permanently and temporary demand shifts in the form of saving shocks. Using the identified temporary response in output after an increase in the desire to save, the aggregate definition allows us to back out the response in the level of personal savings to a saving shock.

Figures 3.3 to 3.5 show the resulting impulse response functions when estimating the baseline specification with eight lags for the first two long-run exclusion restrictions over the sample from 1960-Q1 to 2019-Q4.[11] These results are obtained for aggregate saving rate. Estimates using the personal saving rate are very similar and are reported in the appendix. Imposing the first exclusion restriction, i.e. that a temporary shock in the saving rate cannot have a long-run effect on output, results in a negative short-term impact on economic activity as depcited in figure 3.3. Hence, a sudden cut back in consumption appears to have a contractionary effect which underlines the hypothesis that increased savings made the financial crisis more severe. Only after roughly two years, economic activity is picking up again slowly approaching the level at which it would be in the absence of a saving shock. Interestingly, when computing the response in total personal savings as the sum of the cumulated impulse response of growth in real GDP and the impulse response in the saving rate, we obtain the paradox-of-thrift result as shown in Figure 3.4. Although, households increase their saving rate, the level of savings decreases as the fall in real GDP dominates the increase in the saving rate.

---

[11]Plots for exclusion restriction 3 are in the appendix. Given that this excl. restriction assumes GDP to be an I(3) process we refrain from giving the result much weight.
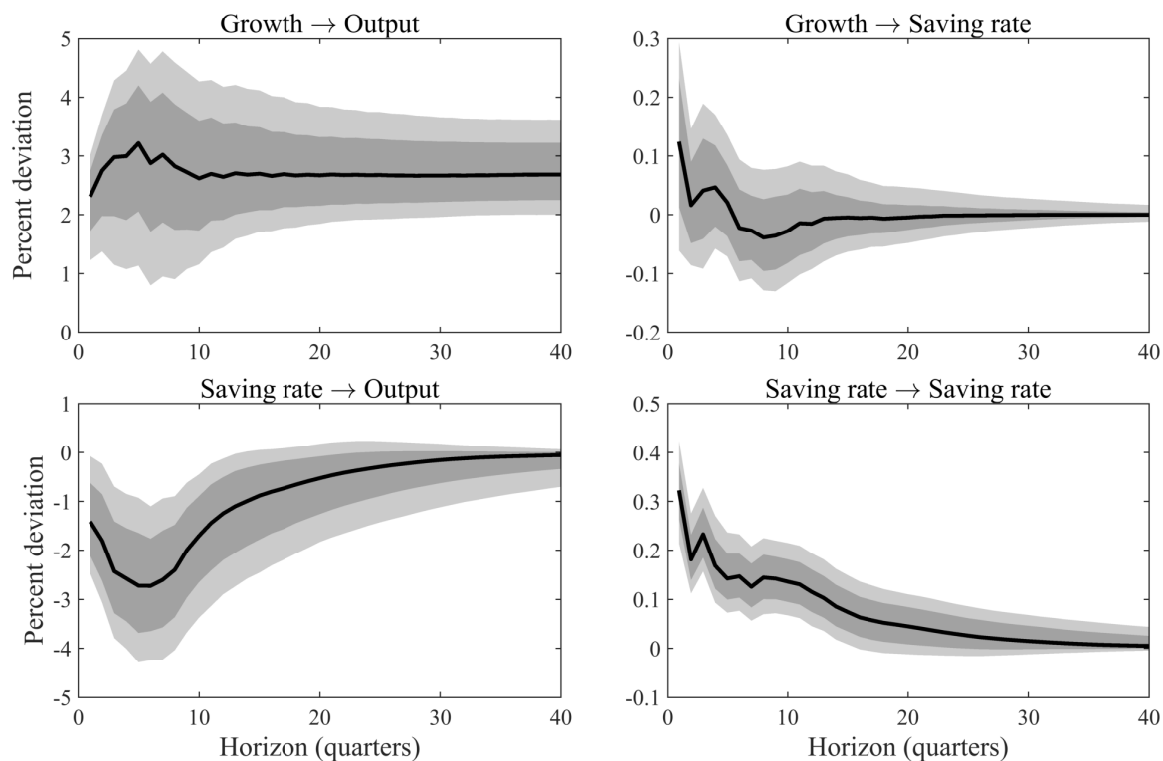
Fig. 3.3 Impulse response functions using exclusion restriction 1 that a temporary shock to the saving rate has no long-run effect on output.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% and 90% confidence bounds.
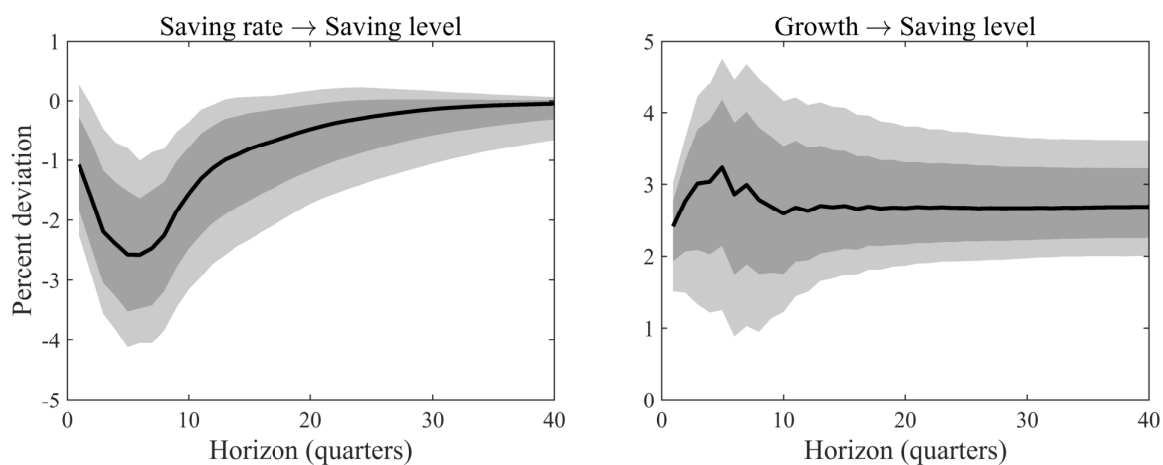
Fig. 3.4 Impulse response function of the level of private savings according to the exclusion restriction in Figure 1.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% and 90% confidence bounds.
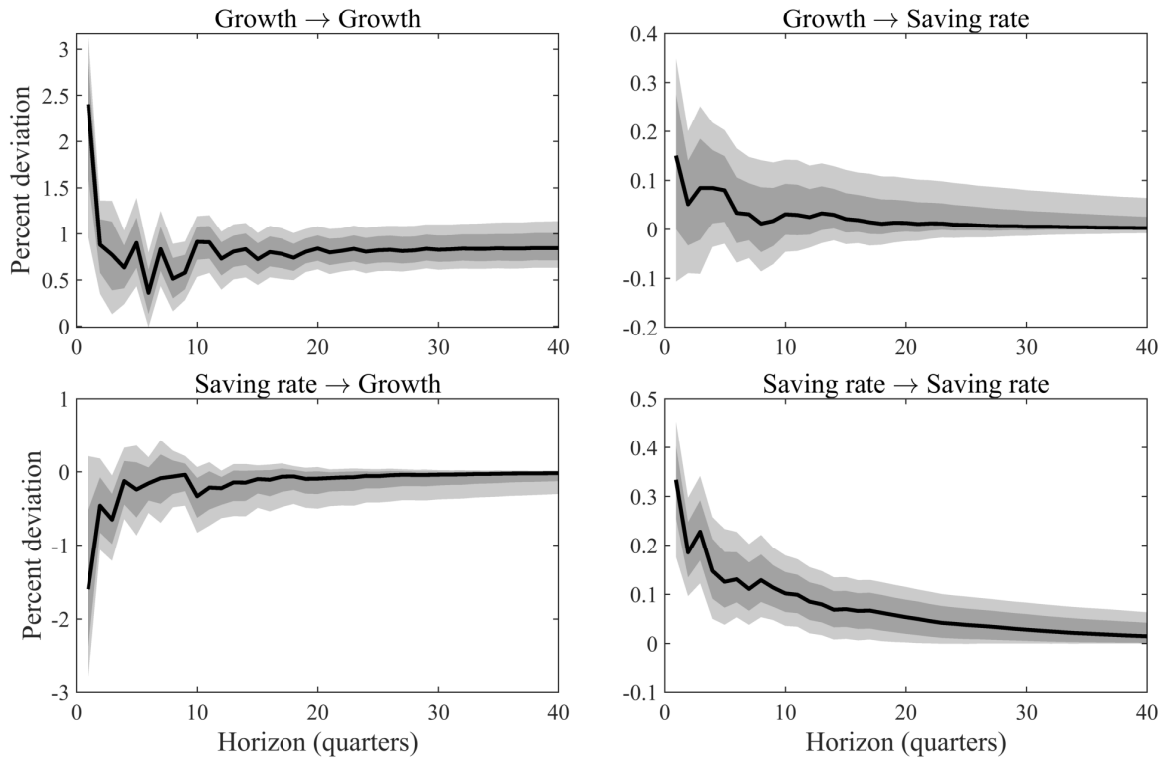
Fig. 3.5 Impulse response functions using exclusion restriction 2 that a temporary shock to the saving rate has no long-run effect on the growth of output.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% and 90% confidence bounds.

Imposing the second exclusion restriction that a temporary saving shock does not have any long-term effect on growth leads to a similar picture as depicted in Figure 3.5. A positive saving shock is followed by a drop in output growth. By definition of the second exclusion restriction, the initial drop reverts back such that growth in real GDP approaches its pre-shock level in the long-run. Furthermore, the identified temporary positive saving shock leads to a permanent decrease in the level of savings which is in line with the paradox-of-thrift argument above. This is due to the fact that while the saving rate returns to its initial level, real GDP will be permanently lower, leading to less total personal savings in the long run.

The third long-run restriction results in a permanently lower level of real GDP growth following a permanent positive saving shock as shown in Figure C.8. However, this identification restriction hinges on real GDP being an I(3) process and the saving rate being of order I(1). Under this assumption, the level of savings is decreasing over time as real GDP is growing slower than its counter-factual in the absence of a saving shock.

3.4.2 Uncertainty and saving shocks

So far, we have identified that shocks to the saving rate with no long-run impact on output or growth, have a contractionary effect on output in the short-term. However, we might be missing a link which is well established in the literature. Uncertainty could act as a latent mediator, driving both, the saving rate up due to precautionary motives and contracting output at the same time. Hence, the question arises, whether we have truly identified the effect of saving rate shocks or rather uncertainty shocks.

In order to tackle this question, we combine long-run and short-run exclusion restrictions in a three-variate SVAR with output, saving rate and proxies for uncertainty. As before, using exclusion restriction 1, we assume that a saving rate shock has no long-term impact on output. An uncertainty shock is defined as having no-long term impact on output but can have an immediate impact on both, output and the saving rate. In order for the system to be just identified, we further assume that saving rate shocks have no contemporaneous impact on uncertainty. The short-run exclusion restriction are in line with the identification strategy proposed by Basu & Bundick (2017). Following the notation by Rubio-Ramirez et al. (2010), equation 3.29 illustrates the identification assumptions imposed on the contemporaneous and long-run impact matrix Q and Θ, respectively.

$$
f(Q,\mathbf{F}) = \begin{bmatrix} Q \\ \Theta \end{bmatrix} = \begin{matrix} \\ \Delta log Y \\ U \\ S \\ \Delta log Y \\ U \\ S \end{matrix} \begin{matrix} \epsilon^y & \epsilon^u & \epsilon^s \\ \begin{bmatrix} x & x & x \\ x & x & 0 \\ x & x & x \\ \hdashline x & 0 & 0 \\ x & x & x \\ x & x & x \end{bmatrix} \end{matrix} \tag{3.29}
$$

As a measure for uncertainty we use the historical newspaper-based economic uncertainty index by Baker et al. (2016) as this spans almost our entire data sample. As a robustness check we also consider the Equity Market Volatility Index (EMV) by Baker et al. (2019) starting in 1985 and the Exchange Volatility Index (VXO) starting in 1986.

Figure 3.6 plots the estimated responses of each shock on output and the saving rate along with 68% and 90% confidence intervals. On impact, a saving rate shock which is now orthogonal to the identified uncertainty shock, still has a contractionary effect on output with a confidence interval of one standard deviation. The effect of uncertainty on output is hardly significant at the 68% level. In fact, when comparing the responses to a

saving rate shock and an orthogonal uncertainty shock, the latter looks relatively similar just lower in magnitude.

To test whether uncertainty shocks are primarily propagated through changes in the saving rate, we mute the response in the saving rate. Specifically, we counter the uncertainty shock with saving rate shocks and thereby offsetting the response in the saving rate. As plotted in the last column of figure 3.6, when muting the saving rate channel, the effect of an uncertainty shock on output becomes insignificant. Hence, it seems that while saving shocks have a contractionary effect independent of uncertainty shocks, uncertainty shocks are mainly transmitted via changes in the saving rate due to precautionary motives.
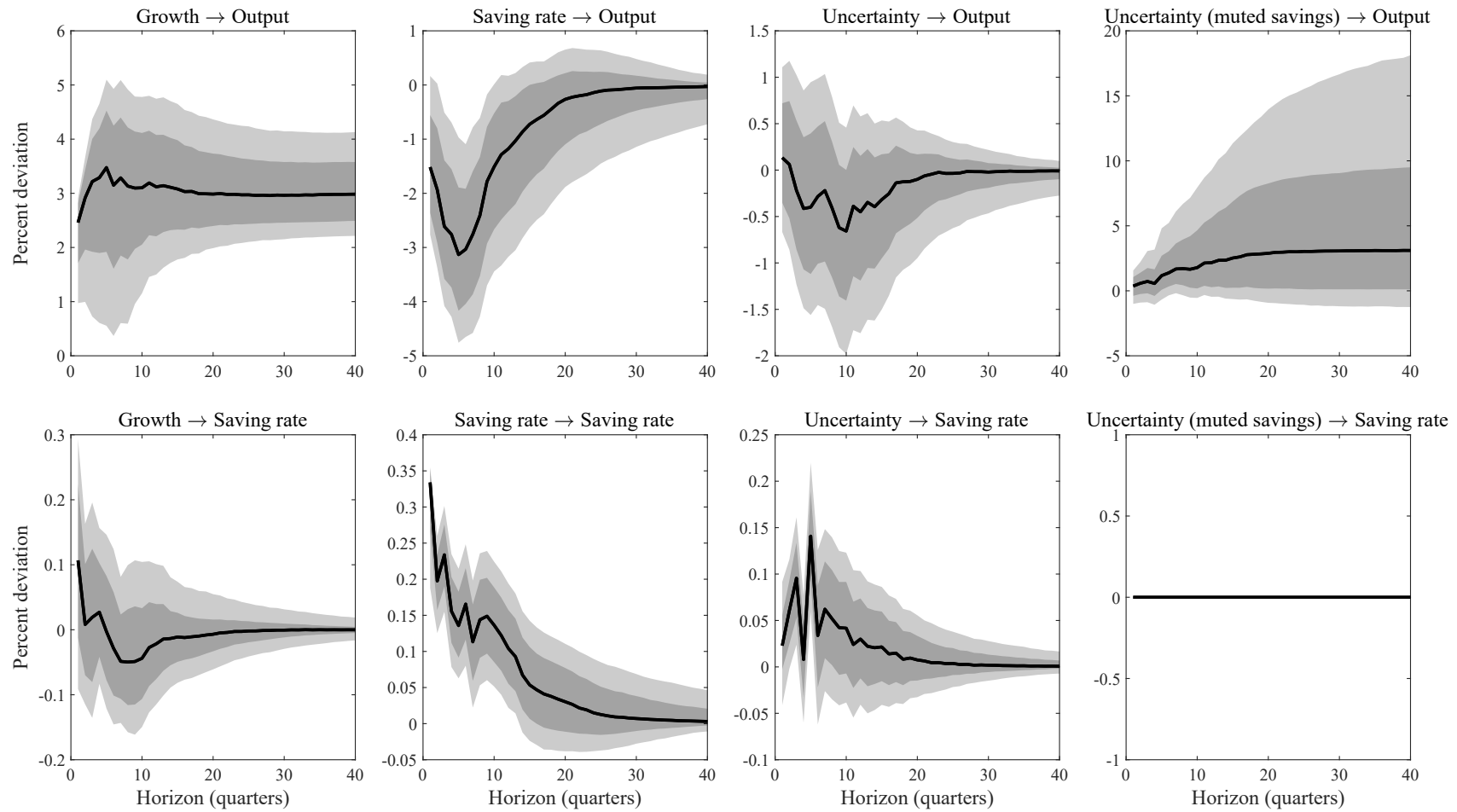
Fig. 3.6 Impulse response functions using long- and short-run exclusion restrictions for uncertainty and saving shocks.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications.

The dark grey transparent area shows the corresponding 68% and 90% confidence bounds.

3.4.3 Robustness checks

Long run identification.   Under the baseline specification, the finding that a positive saving shock can have a contractionary effect on economic activity appeared already robust for three different long run identification restrictions and two definitions of the saving rate, the aggregate saving rate and the personal saving rate in terms of disposable household income. The rest of this section examines the robustness of this finding for alternative choices of lags, trends and time periods. Table 3.2 lists the mean cumulative percentage impulse responses in real GDP after a one percent positive shock in the aggregate saving rate for alternative specifications of the respective structural VAR models[12]. The specification details are defined at the top of the table. The baseline specification uses eight lags, a linear time trend and the sample from 1960 to end of 2019, excluding the Covid crisis, and is included in the fourth column of table 3.2. Bootstrapped standard errors are reported in parentheses beneath the mean estimates.

Across all specifications we observe a negative cumulative effect in real GDP growth after eight quarters following a positive MPS shock. By definition of the exclusion restrictions, the cumulative response in real GDP growth is likely to be lower for the second and third long-run restriction as they do not impose any return of the output level in the long-run. For most of the cases, the negative response is significant for a confidence interval of at least one standard deviation. While reducing the lag length to four yields a weaker response, higher lag lengths in general reduce the estimation bias in VARs as shown by Plagborg-Møller & Wolf (2021). Using twelve lags, results are very close to the baseline estimates confirming that we capture enough flexibility by including eight lags. When choosing a quadratic trend instead of a linear trends, the response is only slightly weaker but still robust. When including the period of the Covid crisis until the end of 2020, results using exclusion restriction 1 are broadly in line, while exclusion restriction 2 predicts a much stronger response. Also when excluding the Financial Crisis or periods before 1990, results are broadly robust.

---

[12]Robustness checks for the saving rate definition in terms of disposable household income are in table C.1 in the appendix

Table 3.2 Cumulative percentage response in real GDP growth for alternative estimation specifications

Identification	At horizon h	Lags			Trend		Time		
		4	8	12	linear	quadratic	1960Q1-2020Q4	1960Q1-2006Q4	1990Q1-2019Q4
Exclusion restriction 1	1	-0.93	-1.50	-1.52	-1.50	-1.24	-0.97	-0.61	-1.23
		(0.70)	(0.73)	(0.69)	(0.73)	(0.74)	(1.06)	(1.00)	(0.44)
	2	-1.15	-1.89	-1.93	-1.89	-1.58	-2.27	-1.00	-1.45
		(0.83)	(0.85)	(0.80)	(0.85)	(0.85)	(0.81)	(1.13)	(0.56)
	4	-1.82	-2.61	-2.68	-2.61	-2.26	-2.60	-1.55	-2.00
		(0.91)	(0.92)	(0.88)	(0.92)	(0.90)	(0.96)	(1.21)	(0.62)
	8	-1.36	-2.38	-2.59	-2.38	-2.05	-2.18	-1.65	-1.63
		(0.79)	(0.91)	(0.86)	(0.91)	(0.85)	(1.08)	(1.07)	(0.68)
Exclusion restriction 2	1	-0.86	-1.39	-1.28	-1.39	-1.29	-0.89	-1.68	-0.94
		(0.78)	(0.79)	(0.76)	(0.79)	(0.75)	(0.84)	(0.91)	(0.51)
	2	-1.13	-1.80	-1.67	-1.80	-1.65	-2.60	-2.29	-1.12
		(1.09)	(1.06)	(0.99)	(1.06)	(1.01)	(0.80)	(1.68)	(0.71)
	4	-1.97	-2.62	-2.41	-2.62	-2.39	-4.12	-3.16	-1.69
		(1.63)	(1.48)	(1.24)	(1.48)	(1.39)	(1.19)	(1.68)	(0.95)
	8	-3.16	-3.06	-2.41	-3.06	-2.67	-6.325	-4.02	-1.53
		(2.62)	(2.12)	(1.86)	(2.12)	(1.98)	(1.82)	(2.56)	(1.38)
Exclusion restriction 3	1	-0.72	-0.67	-1.07	-0.67	-0.73	-2.58	-0.91	-0.25
		(0.34)	(0.42)	(0.48)	(0.42)	(0.41)	(1.26)	(0.53)	(0.41)
	2	-1.08	-0.88	-1.46	-0.88	-0.95	-1.96	-1.36	-0.16
		(0.55)	(0.64)	(0.68)	(0.64)	(0.62)	(1.46)	(0.77)	(0.64)
	4	-2.11	-1.64	-2.36	-1.64	-1.74	-2.34	-2.23	-0.65
		(0.95)	(1.05)	(1.11)	(1.05)	(1.01)	(2.69)	(1.27)	(1.01)
	8	-3.85	-2.60	-3.25	-2.60	-2.75	-4.80	-3.96	-0.63
		(1.77)	(1.85)	(1.92)	(1.85)	(1.77)	(5.44)	(2.34)	(1.67)

Note: Standard errors are parentheses; $'$ $p < 0.32$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3 Cumulative percentage response in real GDP growth
for different uncertainty measures

		Uncertainty measure		
Response	Horizon	Historical Uncertainty Index	EVM Index	VXO Index
Saving rate → GDP	1	-1.09	-1.21	-1.06
		(0.72)	(0.82)	(0.74)
	2	-1.23	-1.42	-1.20
		(0.88)	(1.06)	(0.91)
	4	-1.93	-2.09	-1.89
		(1.00)	(1.24)	(1.05)
	8	-1.78	-1.90	-1.75
		(1.19)	(1.39)	(1.24)
Uncertainty → GDP (muted savings)	1	-0.19	-0.11	-0.16
		(0.70)	(0.96)	(0.71)
	2	-0.28	0.32	-0.25
		(0.94)	(1.40)	(0.94)
	4	0.62	0.65	0.62
		(1.69)	(1.78)	(1.66)
	8	2.78	1.30	2.74
		(3.10)	(1.92)	(3.16)

Note: Standard errors are parentheses; $'$ $p < 0.32$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Long and short run identification. Combining long and short run exclusion restrictions, we find that saving shocks lead to a decline in output orthogonal to uncertainty shocks. This result appears robust for several different proxies for uncertainty as documented in table 3.3. Corresponding impulse response functions are reported in the appendix, see figures C.12 - C.15. Saving shocks are followed by a negative response in output across all uncertainty measures. In fact, even the magnitude of the response at different horizons is pretty robust for all proxies considered, although their data availability varies substantially.

Uncertainty shocks on the other hand have no significant effect on real GDP when shutting down any response in the saving rate suggesting that precautionary savings are the main transmission channel of uncertainty shocks.

SVAR-IV approach. As another robustness check, we use monetary policy surprises as an external instrument to identify the effect of structural saving shocks. We use the shock series by Gertler & Karadi (2015), who, using high-frequency financial data, obtain an external instrument for monetary policy. In estimating the SVAR-IV, we follow Plagborg-Møller & Wolf (2021) and order the external instrument first in an otherwise recursive
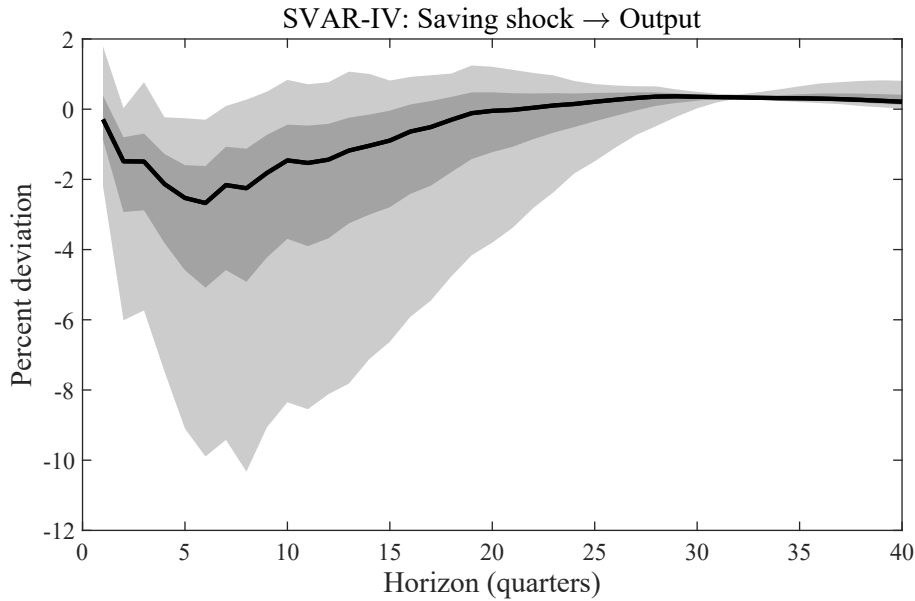
Fig. 3.7 Impulse response function of real GDP after a saving rate shock instrumented by high-frequency monetary policy surprises by Gertler & Karadi (2015)

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The dark grey transparent areas denote the corresponding 68% and 90% confidence bands.

VAR[13]. Identification is then based on the validity and relevance of the instrument. The instrument is valid if it is uncorrelated to the structural shocks to real GDP, i.e. $\mathbb{E}[z_t \epsilon_t^y] = 0$. Given the high-frequency identification approach used by Gertler & Karadi (2015) there is consensus that the instrument is exogenous to shocks in output and hence can be used to identify e.g. effects of monetary policy. We argue that the relevance assumption holds as changes in consumption and thereby savings are one of the main transmission channels of monetary policy (see Holm et al. (2021) among others). Furthermore, if it was rather unrelated and therefore at best a weak instrument we would hardly get any significant results. Figure 3.7 depicts the estimated impulse response in real GDP after a shock in the saving rate induced by a monetary policy surprise[14]. As before, we find evidence that saving shocks have a significant and contractionary effect on economic activity.

---

[13]Also referred to as 'internal instrument' recursive VAR identification and similar to the VARX approach by Paul (2020)

[14]The plot is virtually the same for both saving rate definitions, see figure C.16 in the appendix using the saving rate in terms of disposable household income.

3.5   Conclusion

This paper uses the long-run properties arising from MPS shocks in both exogenous- and endogenous growth models with sticky prices in order to identify their causal effect on output. We find that time series data from the United States is strongly supportive of the notion that MPS shocks indeed have a causal, and contractionary, effect on economic activity, lending support to the most common approach of studying the financial crisis. Furthermore, we find evidence of a paradox-of-thrift, namely that a rise in the saving rate can lead to a decrease in total savings as output contracts. This corroborates the narrative of the financial crisis being ultimately the response to a negative shock to the demands for goods all across the board and underlines the claim by Christiano et al. (2017) to reconsider the theory of the paradox-of-thrift in modern macroeconomics. Last but not least, we provide evidence that saving shocks can have a negative impact orthogonal to uncertainty shocks. In fact, uncertainty shocks appear to be transmitted almost entirely via changes in the saving rate. Our findings are robust to a battery of various robustness checks.

References

Aghion, P. & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica (1986-1998)*, 60(2), 323.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4), 1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., & Kost, K. J. (2019). *Policy news and stock market volatility*. Technical report, National Bureau of Economic Research.

Barro, R. J. & Martin, S.-i. (2004). *Economic Growth*. MIT press.

Basu, S. & Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3), 937–958.

Blanchard, O. J. & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 79(4), 655–673.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., & Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, 86(3), 1031–1065.

Caballero, R. J. (1991). Earnings uncertainty and aggregate wealth accumulation. *The American Economic Review*, (pp. 859–871).

Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly journal of economics*, 112(1), 1–55.

Carroll, C. D., Hall, R. E., & Zeldes, S. P. (1992). The buffer-stock theory of saving: Some macroeconomic evidence. *Brookings papers on economic activity*, 1992(2), 61–156.

Christiano, L., Eichenbaum, M., & Rebelo, S. (2011). When is the government spending multiplier large? *Journal of Political Economy*, 119(1), 78–121.

Christiano, L. J. et al. (2017). *The great recession: A macroeconomic earthquake*. Technical report, Federal Reserve Bank of Minneapolis.

Christiano, L. J., Motto, R., & Rostagno, M. (2014). Risk shocks. *American Economic Review*, 104(1), 27–65.

Deaton, A. (1989). *Saving and liquidity constraints*. Technical report, National Bureau of Economic Research.

Eggertsson, G. B. et al. (2003). Zero bound on interest rates and optimal monetary policy. *Brookings papers on economic activity*, 2003(1), 139–233.

Engen, E. M. & Gruber, J. (2001). Unemployment insurance and precautionary saving. *Journal of monetary Economics*, 47(3), 545–579.

Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American economic review*, 89(1), 249–271.

Gertler, M. & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1), 44–76.

Gourinchas, P.-O. & Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, 70(1), 47–89.

Gourio, F. (2012). Disaster risk and business cycles. *American Economic Review*, 102(6), 2734–66.

Grossman, G. M. & Helpman, E. (1991). Quality ladders in the theory of growth. *The Review of economic studies*, 58(1), 43–61.

Guerrieri, V. & Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. *The Quarterly Journal of Economics*, 132(3), 1427–1467.

Holm, M. B., Paul, P., & Tischbirek, A. (2021). The transmission of monetary policy under the microscope.

Keen, B. & Wang, Y. (2007). What is a realistic value for price adjustment costs in new keynesian models? *Applied Economics Letters*, 14(11), 789–793.

Kilian, L. & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Leland, H. E. (1978). Saving and uncertainty: The precautionary demand for saving. In *Uncertainty in Economics* (pp. 127–139). Elsevier.

Lugilde, A., Bande, R., & Riveiro, D. (2019). Precautionary saving: a review of the empirical literature. *Journal of Economic Surveys*, 33(2), 481–515.

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics*, 102(4), 690–704.

Plagborg-Møller, M. & Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2), 955–980.

Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy*, 90(6), 1187–1211.

Rubio-Ramirez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2), 665–696.

Skinner, J. (1988). Risky income, life cycle consumption, and precautionary savings. *Journal of Monetary Economics*, 22(2), 237–255.

References

Zeldes, S. P. (1989). Optimal consumption with stochastic income: Deviations from certainty equivalence. *The Quarterly Journal of Economics, 104*(2), 275–298.

Chapter 4

Heterogeneous Macroeconomic Policy Effects: A Varying-Coefficient VAR

Abstract

This chapter proposes a flexible framework to identify state-dependent effects of macroeconomic policies. In the literature, it is common to either estimate constant policy effects or introduce state-dependency in a parametric fashion. This, however, demands prior assumptions about the functional form. Our new method allows to identify state-dependent effects and possible interactions in a data-driven way. Specifically, we estimate heterogeneous policy effects using semi-parametric varying-coefficient models in an otherwise standard VAR structure. While keeping a parametric reduced form for interpretability and efficiency, we estimate the coefficients as functions of modifying macroeconomic variables, using random forests as the underlying non-parametric estimator. Simulation studies show that this method correctly identifies multiple states even for relatively small sample sizes. To further illustrate our method, we apply the semi-parametric framework to the historical data set by Ramey & Zubairy (2018) and offer a more granular perspective on the dependence of the fiscal policy efficacy on unemployment and interest rates.[1]

Keywords: Varying-Coefficient Model, VAR, Random Forest, Heterogeneous Policy Effects, Fiscal Multiplier
JEL Codes: C32, E62

---

[1]This chapter is based on joint work with Adrian Ochs.

107

4.1 Introduction

The inherent endogeneity of the macroeconomy plagues the empirical study of policy effects. More often than not, any attempt to directly estimate the effect of the policy variable, x, on the response variable, y, will result in biased coefficient estimates since y might have given rise to the implementation of policy, x, but x affects y. Hence, a large part of the macroeconomic literature has focused on resolving this simultaneity issue by imposing short-run, long-run or sign restrictions. Or, more recently, constructing estimates of unanticipated policy shocks using them as external instruments. Particularly the use of external instruments has increased the confidence in coefficient estimates that no longer rely on sometimes hard to defend identification schemes (e.g. Wolf, 2020).

The compelling research on identification strategies to overcome endogenity concerns allows us to shift our attention to a hitherto often ignored source of estimation bias: model misspecification. With some exceptions, most empirical contributions on estimating macroeconomic policy effects assume a linear and constant effect structure. In contrast, many theoretical contributions taught us that effects could be non-linear due to asymmetric effects (i.e. an interest rate increase might not have the same effect as a decrease) and state-dependence (i.e. the economic environment as well as time). For instance, as Nakamura & Steinsson (2018) point out, monetary and fiscal policy effects may differ depending on the level of slack and openness of the economy. Consequently, even with a cleanly identified policy shock, estimating a linear model will yield biased estimates of the possibly state-dependent policy response.

To reduce the potential for model misspecification bias, we propose a flexible framework to identify state-dependent and asymmetric effects of macroeconomic policies. The approach uses a semi-parametric varying-coefficient estimator in an otherwise standard vector autoregression (VAR). Hence, we call this new approach a varying-coefficient VAR or VC-VAR. While keeping a parametric reduced form for interpretability and more efficient estimation, we estimate the (dynamic) policy effect non-parametrically by using random forests. Random forests allow for an entirely data-driven estimation of potentially heterogeneous policy effects, which can depend on the time or state of the economy. The proposed method could also be interpreted as a machine-learning augmented version of a threshold VAR (TVAR) model with the random forest predicting the splits and interactions of states.

The suggested approach to estimate VC-VARs can therefore be summarised in three steps. Firstly, determine the variables that may alter the VAR coefficient estimates, we call

them moderators following the varying-coefficient literature. Secondly, define the VAR coefficients as functions of the set of moderator variables. Thirdly, estimate the VC-VAR using random forests.

In simulation studies, we show the advantages of the proposed semi-parametric estimation method. The VC-VAR:

1. Detects any state-dependence and interactions between different moderator variables in a dynamic system.

2. Captures non-linear and asymmetric effects, depending on the magnitude and the sign of the policy shock.

3. Performs well in identifying heterogeneous policy responses even for small sample sizes typical for the macroeconomic context. Thus, offering a preferred bias-variance trade-off if the data generating process exhibits state-dependence.

4. Handles many moderator variables and detects the relevant moderator variables even when some moderators are correlated.

Furthermore, we apply the VC-VAR to the historic data-set by Ramey & Zubairy (2018) (RZ) to empirically validate the new method. RZ study the state-dependence of the fiscal multiplier for high or low unemployment and Zero Lower Bound (ZLB) periods. We are able to replicate their results by employing their pre-defined dummy variables for high or low unemployment and ZLB as moderator variables. Furthermore, we show that the newly proposed method yields again similar results when using the federal funds and unemployment rate directly as moderators. Hence, the VC-VAR method does not require the definition of dummy variables to detect state-dependence.

The empirical exercise also allows for several extensions of RZ's (2018) study. Firstly, using the federal funds rate and the unemployment rate as moderator variables offers a more granular perspective on the dependence of the fiscal multiplier on interest and unemployment rates and allows to analyse their interactions. This reveals that it is the unemployment rate that drives the state-dependence of fiscal policy effects and not the ZLB. We find that periods of ZLB mostly coincide with periods of high unemployment. Hence, not taking account of the interaction of unemployment and the ZLB would falsely indicate that fiscal spending was more effective in periods of ZLB. Secondly, using the instrument for the government spending shock as a moderator variable itself, we show that the asymmetric effect of fiscal shocks is small (i.e. positive or negative shocks have similar effects).

Literature.   This paper contributes to several strands of the literature. First, we provide a general method to estimate state-dependent policy effects. Recently, both theoretical (Canzoneri et al., 2016; Eichenbaum et al., 2018; Farhi & Werning, 2016; Rendahl, 2016) and empirical (Ascari & Haber, 2019; Auerbach & Gorodnichenko, 2012, 2013; Fazzari et al., 2015; Paul, 2020; Ramey & Zubairy, 2018) work has shed light on the importance of considering state-dependent and non-linear effects of macroeconomic policies. However, this has required to pre-specify any time- and state-dependence. More specifically, one has to construct dummies based on some discretionary thresholds and decide on the functional form of their interaction with the policy shock. This paper uses machine learning methods to automate this process and aims to shed more light on the high-dimensionality of policy effects. Using an underlying non-parametric estimator, we are agnostic about the functional form of any state-dependence.

Second, the new estimation method makes VARs more flexible, thereby reducing model-specific bias. Recent research on the differences of local projections (LP) and SVARs has described the choice between the two methodologies as a trade-off between variance and bias in applied research[2] (Li et al., 2021). While our approach is also concerned with misspecification bias, the focus is on a different source for this bias. Here, the bias stems from state-dependent coefficients. For Li et al. (2021), the model misspecification arises when the true data generating process (DGP) is a vector autoregressive moving average (VARMA) model, but the econometrician specified a VAR or when not selecting the correct lag length. LP's repeated estimation makes it more flexible for the latter misspecification but at the cost of increased variance. However, this does not hold for misspecifying a non-linear model as a linear one. Neither LP nor VARs can mitigate the bias from such misspecification. Nevertheless, the one-time estimation of the VAR model makes coefficient estimates with our method less volatile than LPs for small samples, which we usually observe in macroeconomics. Hence, we present our method in the context of VARs. This might make VARs more attractive when suspecting state-dependence and choosing between the two methodologies, LPs and VARs. Similarly, Ramey & Zubairy (2018) have argued that LPs allow for more straightforward inclusions of non-linearities. Our new estimation method makes this process simple and automatic for the VAR case, focusing the choice between LPs and VARs back on statistical properties rather than the simplicity of implementation.

---

[2]When estimated in population with infinite lag length, both methods produce the same IRFs (Plagborg-Møller & Wolf, 2019)

Third, we follow a recent strand of research exploring how to utilize advances from the machine learning literature for macroeconomic analysis (Duarte, 2018; Goulet Coulombe, 2020; Gu et al., 2020; Joseph, 2019). This paper is closest to Goulet Coulombe (2020) who also proposes a varying-coefficient model in macroeconomics. However, the main difference is that Goulet Coulombe (2020) aims to improve macroeconomic predictions. We are more interested in identifying potential state-dependence in macroeconomic policies using a high-dimensional semi-parametric VAR framework. Hastie & Tibshirani (1993) first introduced varying coefficient models. These models are linear in their regressors, but their coefficients are estimated by generalized additive models to allow them to vary as a function of other factors. Wang & Hastie (2014), for instance, use regression trees as the base learner to identify varying effects. The advantage of using regression trees over other non-parametric techniques, such as kernel-based estimates, is that they can handle many potentially mixed-type variables.

Outlook. The rest of the paper is structured as follows. Section 2 gives an overview of the varying-coefficient model and motivates the use of random forests as the non-parametric coefficient estimator. Furthermore, section 2 explains how to estimate high-dimensional impulse response functions using the varying-coefficient model. In section 3, we conduct several simulation studies showcasing the performance of the method, while section 4 applies the method to the recent paper by Ramey & Zubairy (2018), estimating state-dependent fiscal multipliers. Finally, section 5 concludes.

4.2   Methodology

This section first motivates the use of trees as the underlying non-parametric estimator for capturing regime shifts in the economy. Next, we introduce a varying-coefficient model using trees as base learners and provide computational details on how to estimate state-dependent coefficients non-parametrically. Furthermore, we show how this estimator can be used in the context of structural vector autoregressions (SVAR) to obtain high-dimensional impulse-response functions.

4.2.1   Macroeconomic state-dependencies as trees

Random forests as non-parametric estimators have increased in popularity as they allow for complex non-linearities and asymmetric effects that can handle high-dimensional

data and are, compared to other machine learning algorithms, robust as they demand little to no tuning.

Full sample

High unemployment

Low unemployment

ZLB

No ZLB

$\epsilon_t^g \geq 0$

$\epsilon_t^g < 0$

$y_t = \rho_1 y_{t-1} + \phi_1 \epsilon_t^g + \epsilon_t^y$ $y_t = \rho_2 y_{t-1} + \phi_2 \epsilon_t^g + \epsilon_t^y$ $y_t = \rho_3 y_{t-1} + \phi_3 \epsilon_t^g + \epsilon_t^y$ $y_t = \rho_4 y_{t-1} + \phi_4 \epsilon_t^g + \epsilon_t^y$
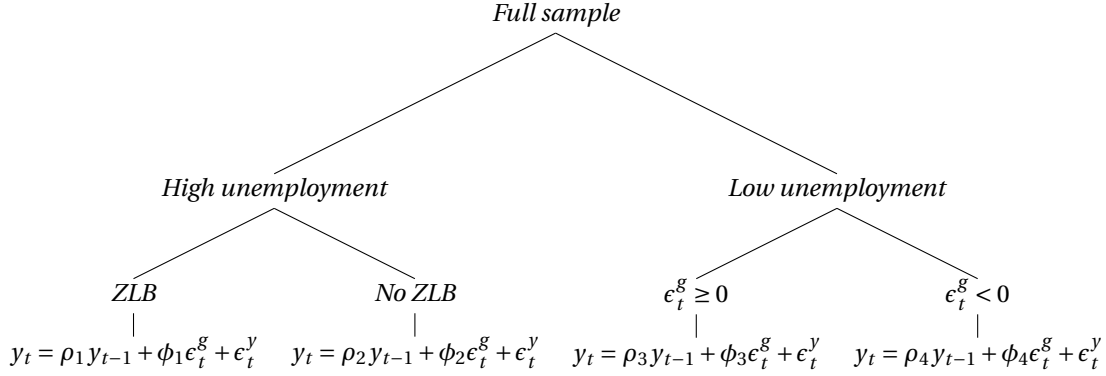
Fig. 4.1 Illustrative tree capturing state-dependent policy effects

Moreover, random forests offer a natural way to represent the macroeconomic state-space and state-dependence in macroeconomic time-series (see also Goulet Coulombe (2020)). Anticipating our empirical application, figure 4.1 shows an illustrative tree capturing potential state-dependence of fiscal policy efficacy as in Ramey & Zubairy (2018). Prior to estimation, a researcher studying fiscal policy might assume that the effect of government spending shocks denoted by ϵ_t^g could depend on the level of unemployment as suggested by Rendahl (2016) and found by Ramey & Zubairy (2018). This is depicted in the first split of figure 4.1 into the "high unemployment" and "low unemployment" branch. Furthermore, for high unemployment, it might matter whether the economy is already close to the zero lower bound (ZLB) or not. If this were the case, the tree-based estimator would split the "high unemployment" branch into two further branches differentiating between "ZLB" and "no ZLB". Additionally, the output response might be asymmetric in the "low unemployment" state, depending on the sign of the policy shock. This would create a further split of the "low unemployment" branch into "$\epsilon_t^g \geq 0$" and "$\epsilon_t^g \leq 0$". This step-by-step illustration shows how macroeconomic state-dependencies can be depicted by trees.

For parametric estimation approaches, the assumptions depicted in figure 4.1 pose a high degree of model uncertainty. To avoid model specification bias, one has to correctly pre-define the different states and cut-off levels, as well as their interactions. Using tree-based learners in our semi-parametric approach, we reduce the problem of model uncertainty to the choice of which effect modifiers to include (e.g. unemployment and ZLB), while the random forest estimates the underlying splits as illustrated in figure 4.1.

4.2.2   Varying Coefficient Model

This subsection introduces a varying coefficient model which employs tree-based learners to augment the estimation of a parametric baseline specification in search of potential state-dependence.[3]

Consider the following simple model structure:

$$y_t = \beta_t \epsilon_t + \boldsymbol{x}'_t \gamma_t + u_t,$$
$$\beta_t = f(\Omega_t),$$

where y_t is some macroeconomic outcome variable of interest, ϵ_t some exogenous policy shock, $\boldsymbol{x}_t$ a vector of controls, and β_t is a high-dimensional coefficient, which is estimated non-parametrically as a function of the state space of the economy Ω_t. We will assume throughout that, ϵ_t, $\boldsymbol{x}_t$ and Ω_t, are exogenous. In this setup, the relationship between the outcome, y_t and some policy shock, ϵ_t may vary depending on the state of the economy captured by Ω_t. In principle, Ω_t can contain many modifying variables and it may overlap with other exogenous regressors in the low-dimensional linear regression.

Tree-based varying coefficient model.   In the definition of the tree-based estimator, we follow Wang & Hastie (2014)[4]. The tree-based method aims to approximate $\beta(\Omega_t)$ by a piece-wise constant function. The idea is to partition the state space Ω_t following a certain optimization criterion and to approximate $\beta(\Omega_t)$ in each partition by a constant vector. Let $C_m{}_{m=1}^M$ denote a partition of the state space $\mathbb{R}^q$ satisfying $C_m \cap C_{m'} = \emptyset$ for any $m \neq m'$, and $\cup_{m=1}^M C_m = \mathbb{R}^q$, where M denotes the number of partitions. Then, the tree-based varying coefficient estimator can be written as

$$y_t = \sum_{m=1}^M \beta_m I_{(\Omega_t \in C_m)} \epsilon_t + \boldsymbol{x}'_t \gamma_t + u_t, \tag{4.1}$$

where $I_{(\bullet)}$ denotes the indicator function with $I_{(c)} = 1$ if event c is true and zero otherwise. The implied heterogeneous policy effect is thus,

$$\beta(\Omega_t) = \sum_{m=1}^M \beta_m I_{(\Omega_t \in C_m)}.$$

---

[3]Another advantage of using trees as base learners over non-parametric kernel smoothing or spline-based methods is that they can handle high-dimensional data of mixed types.

[4]Wang & Hastie (2014) use a tree-based estimator as a base learner for boosting, while we will bag multiple tree-based estimates to obtain a more robust forest-based estimate.

Consequently, a constant OLS estimate of β can be interpreted as the average policy effect over all partitions

$$\beta_{OLS} = \sum_{m=1}^{M} \frac{n_m}{N} \beta_m I_{(\Omega_t \in C_m)}.$$

where n_m and N denote the observations in partition m and the overall sample size, respectively. The set partitions C_m are also referred to as terminal nodes or leaf nodes of the tree, describing the various estimated states of the policy effect. However, the number of nodes M, the particular partitions $\{C_m\}_{m=1}^{M}$ as well as the values for β_m are unknown and require simultaneous estimation. The number of partitions is usually tuned using some criteria evaluating the out-of-sample fit. The latter is crucial in order to avoid over-fitting. For instance, picture the case of a tree that is partitioned entirely based on the in-sample fit. The best and, in fact perfect fit would be reached by having N partitions, each representing a single observation. There are several methods to avoid over-fitting, with cross-validation being the most common one. During cross-validation, the sample is randomly split into a training and validation set, with the first being used for estimating the model and the latter for testing the model's performance out-of-sample. We use k-fold cross-validation where the sample is split into k folds such that k-1 folds are the training set and one fold is the test set. This is repeated k times for each fold.

Having decided on the number of partitions, which is usually done automatically by most implementations, the remaining problem is to find the optimal partition sets and coefficient estimates corresponding to those splits. Given the parametric nature of the governing function, we minimize according to the least squares criterion

$$\left(\hat{C}_m, \hat{\beta}_m\right) = \underset{(C_m, \beta_m)}{\text{argmin}} \sum_{t=1}^{T} \left(y_t - \sum_{m=1}^{M} \beta_m I_{(\Omega_t \in C_m)} \epsilon_t - \boldsymbol{x}_t' \gamma_t \right)^2 \tag{4.2}$$

In the above specification, the estimation of β_m is nested in that of the partitions. Hence, within each partition, β_m is simply the least squares estimator on the corresponding sub-sample. Therefore, minimizing equation (4.2) boils down to finding the optimal set of partitions $\{C_m\}_{m=1}^{M}$.

$$\hat{C}_m = \underset{C_m}{\text{argmin}} \sum_{t=1}^{T} \sum_{m=1}^{M} \left(y_t - \beta_m(C_m) I_{(\Omega_t \in C_m)} \epsilon_t - \boldsymbol{x}_t' \gamma_t \right)^2 \tag{4.3}$$

Forest-based varying coefficient model. The forest-based estimator represents a collection of many tree-based estimates following the random forest ensemble algorithm by Breiman (2001). The idea is simply to reduce the variance of a single tree's estimate by

averaging over multiple trees. More specifically, k trees are estimated based on k independent identically distributed random samples of the data set. With each tree casting a prediction for β_m, the random forest isdictor is formed by taking the average over all trees. This procedure is commonly referred to as bootstrap averaging or bagging as introduced in Breiman (1996). Additionally, Breiman (2001) proposes to use a random subset of state space variables in Ω_t, which improves the accuracy by minimizing the correlation between state space variables while at the same time keeping the strength of their signal. Breiman (2001) shows that random forest are more robust and accurate compared to single trees. Given their preferred characteristics, we will use forest-based estimates throughout this paper.

The underlying optimization problem remains the same as in equation (4.3) for a single tree as a base learner with only slight changes to the tree producing algorithm now using random subsets of the state space at every split. The forest-based estimate of the varying coefficient is then simply the average over all these trees:

$$\beta_{RF}(\Omega_t) = \frac{1}{K} \sum_{k=1}^{K} \sum_{m_k=1}^{M_k} \beta_{m_k} I_{(\Omega_t \in C_{m_k})}.$$

Computational details. We rely (for now) on the implementation of a coefficient-wise varying coefficient regression model by Buergin & Ritschard (2017) readily available in their R-package `vcrpart`. While this section will give a short overview of how the estimation algorithm works, we refer for further details to the original paper. Algorithm 1 provides a formal summary of how to estimate a varying coefficient model using a random forest as its underlying non-parametric estimator.

The Random Forest-VCM function in Algorithm 1 is a wrapper function growing numerous trees based on bootstrapped samples of the original data. Coefficient predictions are then obtained by averaging over all tree predictions leading to reduced variance and hence more robust estimates.

The individual trees are estimated in the Randomized Tree-VCM function in Algorithm 1. The main optimization objective is to minimize the negative log likelihood of the linear parametric function. Given some minimal node size, i.e. the number of observations which are represented by each terminal node of the tree, and a minimum log-likelihood improvement required for each split, the algorithm searches for the optimal partition over the value space of the modifying variables for each coefficient at each node. This is done until no further split satisfies the minimum number of observations in each terminal node, or the log-likelihood improvement is smaller than the minimum requested.

Algorithm 1 Estimation of varying coefficient model using a random forest following
Buergin & Ritschard (2017)

Parameters:	T	number of trees in forest, e.g., T = 100
	N_0	minimum node size, e.g., $N_0 = 30$
	$D_{\min}$	minimum $-2\cdot$ log-likelihood reduction, e.g., $D_{\min} = 2$
	$P_{\max}$	maximum levels of pruned tree, e.g., $P_{\max} = 3$

function RANDOM FOREST-VCM(S, Ω)

    $H \leftarrow \emptyset$ ▷ Initialize Forest

    for trees in $t = 1$ to T **do**

        $S_t \leftarrow$ A bootstrap sample from the dataset S

        $h_t \leftarrow$ Randomized Tree-VCM(S_t, Ω)

        $H \leftarrow H \cup h_t$

    end for

    return H

end function   ▷ Coefficient predictions by averaging over all trees in H

function RANDOMIZED TREE-VCM(S, Ω)

    Initialize $\mathscr{C}_{k1} \leftarrow \Omega_{k1} \times \ldots \times \Omega_{kL}$ and $M_k \leftarrow 1$ for all partitions $k = 1, \ldots, K$.

    repeat

        Compute $\hat{l}^{\widehat{\mathscr{M}}} = \max_{\beta,\gamma} l^{\widehat{\mathscr{M}}}(\beta, \gamma)$ of the current model using the latest partitions

$$\widehat{\mathscr{M}} : y_i = \mathbf{x}'_i \boldsymbol{\gamma} + \sum_{k=1}^{K} \sum_{m=1}^{M_k} I_{(\Omega_{ik} \in \mathscr{C}_{km})} x_{ik} \beta_{km}$$

    using all observations $i \in S$.

        for partitions $k = 1$ to K **do**

            for nodes $m = 1$ to M_k and randomized subset of moderator variables $\tilde{l} = 1$ to $\tilde{L}$ **do**

                for all unique candidate split Δ_{kmlj}, in $\{\omega_{kli} : \Omega_{ik} \in \mathscr{C}_{km}\}$ that divides $\mathscr{C}_{km}$ into two

                nodes $\{\mathscr{C}_{kmlj1}, \mathscr{C}_{kmlj2}\}$ and satisfies $\min_s \sum_i I_{(\Omega_{ik} \in \mathscr{C}_{kmljs})} \geq N_0$ **do**

                Using only the observations $\{i : \Omega_{ik} \in \mathscr{C}_{km}\}$ of the node $\mathscr{C}_{km}$, compute

                $\hat{\ell}^{\widehat{\mathscr{M}}_{kmlj}} = \max_{\{\beta_1,\beta_2\}} \ell^{\widehat{\mathscr{M}_{kml}}}(\beta_1, \beta_2)$ of the approximate search model

$$\widehat{\mathscr{M}}_{kmlj} : y_i^{(s)} = \hat{y}_i + \sum_{s=1}^{2} I_{(\Omega_{ik} \in \mathscr{C}_{kmljs})} x_{ik} \beta_s$$

                and compute the training error reduction $D_{kmlj} = -2\hat{\ell}^{\widehat{\mathscr{M}}} + 2\hat{\ell}^{\widehat{\mathscr{M}}_{kml}}_{\{i:\Omega_{ik} \in \mathscr{C}_{km}\}}$,

                where $\hat{\ell}^{\widehat{\mathscr{M}}}_{\{i:\Omega_{ik} \in \mathscr{C}_{km}\}}$ is the subtotal of $\hat{\ell}^{\widehat{\mathscr{M}}}$ for the observations of $\mathscr{C}_{km}$.

            end for

            end for

        end for

        Split node $\mathscr{C}_{k'm'}$ by $\Delta_{k'm'l'j'}$ where $D_{k'm'l'j'} = \max D_{kmlj}$ and increase $M_{k'} \leftarrow M_{k'} + 1$.

    until no candidate split satisfies N_0 or $D_{k'm'l'j'} < D_{\min}$

    return Return the Tree-VCM

end function

Similar to bootstrapping, growing individual trees with subsampling of combinations of coefficients, nodes, and moderators as split candidates reduces the variance. The subsampling decorrelates the individual trees since they have different split candidates. The decorrelation of the individual trees makes the averaging over all trees more effective in reducing the variance. Nevertheless, the number of combinations as split candidates is a tuning parameter and in some cases, no subsampling might yield the best out of sample fit. We find that this is mostly the case when we have few coefficients and moderators.

4.2.3   High-dimensional impulse response functions

This section shows how to utilize the semi-parametric estimator outlined above to estimate state-dependent effects in a vector autoregression (VAR) framework and subsequently how to obtain state-dependent IRFs (For a brief summary of the step-by-step estimation of state-dependent IRFs, see Algorithm 2 below.).

The key difference to the usual way of computing SVARs is that we impose the identifying restrictions directly on the structural equations rather than the reduced form covariance matrix. We first show that these approaches are equivalent and then clarify why direct estimation is beneficial when estimating state-dependent IRFs.

To illustrate our approach, assume a bivariate VAR(1) model and that there exists an internal instrument, x_t, such that the system can be ordered recursively and is just identified. Our focus lies on the flexible semi-parametric estimation of SVARs and the construction of state dependent IRFs given a valid identification[5]. For this reason, we assume throughout that our identifying restrictions coincide with the true data generating process (DGP). This is without any loss of generality since also any fully parameterized estimation of SVARs hinges on the validity of its identifying assumptions. Let us further assume that the internal instrument, x_t has a state dependent contemporaneous effect on the outcome variable, y_t. The structural model takes the following form.

$$\mathbf{A}Y_t = \mathbf{B}Y_{t-1} + \epsilon_t$$

$$\begin{bmatrix} a_{11} & 0 \\ a_{21}(\Omega_t) & a_{22} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^x \\ \epsilon_t^y \end{bmatrix} \tag{4.4}$$

---

[5]For information on how to implement the usual identification techniques with our varying coefficient approach, please see the appendix.

with the mean and variance of the structural errors normalized to zero and one, respectively and no correlation between the errors such that $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[\epsilon_t \epsilon_t'] = I_n$, where I_n is the identity matrix. The state dependence of the effect of the instrument, x_t on y_t is indicated by writing a_{21} as a function of the state, Ω_t. Let the state be defined as $\Omega = \{s_t, r_t\}$ and note that the state could include many more external variables that may overlap with other exogenous regressors. This implies that the coefficient a_{21} might vary according to different values for s_t and r_t.

Furthermore, we assume A to be a non-singular matrix enabling us to obtain the reduced form (4.5) by inverting A:

$$Y_t = \mathbf{F} Y_{t-1} + u_t$$
$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21}(\Omega_t) & f_{22}(\Omega_t) \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix} \tag{4.5}$$

where $\mathbf{A}^{-1} \equiv \mathbf{Q}$, $\mathbf{F} = \mathbf{Q} * \mathbf{B}$ and $u_t = \mathbf{Q}\epsilon_t$.

Due to the state dependent effect, $a_{21}(\Omega_t)$ of the internal instrument on the outcome variable, using a linear estimator would misspecify the system and hence, lead to biased IRFs. This is the case for both, the structural and the reduced form as a_{21}, f_{21} and f_{22} are all a function of the state, Ω_t. Hence, whenever state dependence is present, the usual SVAR identification procedure yields biased estimates.

If we are interested in the effect of a structural shock of x_t on y_t, we can also directly estimate the ratio of a_{21} and a_{22}. This is, given the restrictions, we can rewrite system (4.4) as

$$x_t = \frac{b_{11}}{a_{11}} x_{t-1} + \frac{b_{12}}{a_{11}} y_{t-1} + \frac{1}{a_{11}} \epsilon_t^x, \tag{4.6}$$

where we define $\frac{1}{a_{11}} \epsilon_t^x \equiv u_t^x$ and

$$\begin{aligned} y_t &= -\frac{a_{21}(\Omega_t)}{a_{22}} x_t + \frac{b_{21}}{a_{22}} x_{t-1} + \frac{b_{22}}{a_{22}} y_{t-1} + \frac{1}{a_{22}} \epsilon_t^y \\ &= -\frac{a_{21}(\Omega_t)}{a_{22}} u_t^x + \left(\frac{b_{21}}{a_{22}} - \frac{a_{21}(\Omega_t)}{a_{22}} \frac{b_{11}}{a_{11}} \right) x_{t-1} + \left(\frac{b_{22}}{a_{22}} - \frac{a_{21}(\Omega_t)}{a_{22}} \frac{b_{12}}{a_{11}} \right) y_{t-1} + \frac{1}{a_{22}} \epsilon_t^y. \end{aligned} \tag{4.7}$$

Hence, this allows us to directly estimate the impact effect, $-\frac{a_{21}(\Omega_t)}{a_{22}}$, using a varying coefficient model for y_t,

$$y_t = \beta_1(\Omega_t) x_t + \beta_2(\Omega_t) x_{t-1} + \beta_3(\Omega_t) y_{t-1} + u_t^y. \tag{4.8}$$

Note that β_1 in equation (4.8) is the relative impact effect on y_t of a structural shock that increases x_t by one unit. This is, rather than identifying q_{21}, direct estimation gives the scaled effect $\frac{q_{21}}{q_{11}}$, which can be interpreted as the impact effect of a shock that increases x_t by one unit.[6] Mostly, this is the unit of interest for policy. Nevertheless, to obtain absolute impact effects (i.e. a one standard deviation shock), one can use the standard deviation of the residuals from estimation equation (4.6), which is equivalent to q_{11} and multiply it with β_1 to obtain q_{21}. Note that the standard deviation of the structural shocks is assumed to not change over the different estimated regimes. Hence, the scaling can be done with the residuals from equation (4.6) and (4.8) rather than obtaining them for every state.

The reason for using the direct estimation approach now becomes apparent. Instead of estimating a fully parameterized version of equation (4.8), we can apply the semi-parametric estimation method outlined above to obtain state-dependent estimates of β_1 in one step.[7] Using the more conventional approach and imposing the restrictions on the covariance matrix of the reduced form errors would necessitate estimating the reduced form for every regime and obtain the covariance matrix for every estimation. This would lead to efficiency losses and potentially unstable results when the amount of observations for a regime is small.

It is straightforward to produce IRFs in a simple model where only one coefficient varies between two values. For the example above, assume that a_{21} takes on a different value for $s_t > s^*$ and $s_t <= s^*$. Then, we can obtain values for the impact effect for β_1 and similarly for f_{21} and f_{22} when $s_t > s^*$ and $s_t <= s^*$ and compute the two IRFs.

However, we usually do not know a priori which variables are state-dependent and whether there are interactions within the variables that determine the state dependence. For this reason, in practice, we estimate (4.5) fully flexibly to get the state dependent dynamic matrix, F, as:

$$Y_t = \mathbf{F}(\Omega_t)Y_{t-1} + u_t \tag{4.9}$$

and similarly the structural equation (4.8) to obtain the relative impact, $\beta_1(\Omega_t)$:

$$y_t = \beta_1(\Omega_t)x_t + \beta_2(\Omega_t)x_{t-1} + \beta_3(\Omega_t)y_{t-1} + u_t$$

---

[6]To see this more clearly, note that $\begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22}} \begin{bmatrix} a_{22} & 0 \\ -a_{21} & a_{11} \end{bmatrix} = \begin{bmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{bmatrix}$.

[7]Paul (2020) includes instruments for the structural shock of interest as an exogenous variable in a VAR to more easily estimate state-dependent effects. Our direct estimation approach is also valid when there is no external instrument available.

Estimating the SVAR model fully flexibly complicates the computation of IRFs. Instead of displaying all possible IRFs under all combinations of the varying coefficients, we propose to predict the IRFs for different values of the state variables. That is, after estimating the model, one can define several states of interest, e.g.: $\Omega^* = \{s_t = s^*, r_t = r^*\}$, $\Omega^{**} = \{s_t = s^{**}, r_t = r^{**}\}$ and $\Omega^{***} = \{s_t = s^{***}, r_t = r^{***}\}$ to then compute the IRFs for each of these states. The state-dependent IRF more generally then becomes:

$$
\widehat{IRF}_{t+h}(\Delta, \Omega_t) = \mathbb{E}\left[y_{t+h} \mid \epsilon_{t+\tau} = \left\{ \begin{array}{l} \Delta \quad \text{if } \tau = 0 \\ 0 \text{ if } \tau \in (1, h) \end{array} \right. ; \Omega_t \right]
$$
$$
- \mathbb{E}\left[y_{t+h} \mid \epsilon_{t+\tau} = 0, \forall \tau \in (0, h); \Omega_t \right]
$$

(4.10)

where Δ is the size of the shock. IRFs are conditional on a specific state, i.e. specific values of the state variables, and we do not assume any regime switching.

We adapt the tree-based estimation to VARs rather than LPs since compared to a dummy approach that explicitly imposes the regime shift on the model, our flexible estimation technique might estimate different regimes at different horizons when using LPs. Additionally, even when the same regimes are found for every horizon, the point estimates might be more erratic due to the repeated estimation procedure and less efficient since we lose observations by iterating over the horizons. Appendix D.1 shows that the confidence intervals become wider when producing state-dependent IRFs with LPs.

Algorithm 2 Estimating high-dimensional IRFs

Step 1 Decide on state variables in set $\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$

Step 2 Estimate reduced form equations semi-parametrically using the set Ω as coefficient modifiers

Step 3 Estimate structural equations semi-parametrically using the set Ω as coefficient modifiers

Step 4 From estimated model in **2** predict lag coefficients for specific state values $\{\omega_1 = \omega_1^*, \omega_2 = \omega_2^*, ..., \omega_k = \omega_k^*\} \in \Omega$ and combine to $F(\omega^*)$.

Step 5 From estimated model in **3** predict contemporaneous impact coefficients for state values $\omega \in \Omega$ and combine to $Q(\omega)$.

Step 6 Get state-dependent IRFs by: $\widehat{IRF}_{t+h}(\Delta, \omega) = F^h(\omega)Q(\omega)\epsilon$, where ϵ is the vector of structural shocks.

4.3   Simulation studies

In this section, we showcase the performance of the VC-VAR in identifying heterogeneous policy responses in the data. We consider several different simulation environments from a simple case of one state-dependent coefficient and two states to multiple varying coefficients and various states. For the purpose of illustrating that the VC-VAR approach offers a more favourable bias-variance trade-off compared to an approach using local projections, we included simulation results for the latter in Appendix D.1 (see Figures D.2 - D.4).

Exogenous structural break.   Let's first consider the following dynamic system with an exogenous change in the effect of the policy instrument.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.3 & t < T/2 \\ -0.1 & t \geq T/2 \end{cases}
\end{aligned}
\tag{4.11}
$$

where i_t is some policy instrument, e.g. the nominal interest rate, and y_t is some outcome variable of interest, e.g. output growth. The policy effect is piece-wise constant and there is a structural break amid the simulated sample. We simulate the system for $T = 500$ and estimate the policy effect using the VC-VAR without any priors on the number of structural breaks, their location or the values in the respective partitions. Since the focus of this exercise is to identify potential state- or time-dependence in the policy response, we assume to know the actual exclusion restrictions when estimating the system of equations, not to introduce any endogeneity bias. This is without loss of generality, given that any estimator would be biased if the exogeneity assumption is violated.

The resulting estimates for β_t are compared to OLS estimates and their distributions are plotted in Figure D.1a. As expected, the OLS estimates are only able to capture the average policy effect[8]. The semi-parametric estimates, however, are able to distinguish between the different states of β_t clearly and are very close to the true values.

Figure 4.2a shows the corresponding impulse response estimates for a shock in the policy instrument. While the parametric OLS estimate exhibits smaller confidence bounds, it is clearly biased as it only captures the average response missing the structural break.

---

[8]Clearly, if the policy effect is in fact constant, OLS is the best linear unbiased estimator (BLUE).

The median state-dependent response of the semi-parametric estimator, on the other hand, is very close to the true impulse responses. In fact, when plotting the partial dependency[9] of the varying coefficient β_t, we can see that it found the exact structural break in the middle of the sample.

Endogenous change. In the next example we consider the policy effect being moderated by previous values of the outcome variable y. While before we simulated a clear structural break, in this case the states can quickly switch endogenously. The piece-wise definition of β_t allows introducing all kinds of non-linearities and asymmetries in the policy response for the forest-based estimator to uncover.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.3 & y_{t-1} < 0 \\ -0.1 & y_{t-1} \geq 0 \end{cases}
\end{aligned}
\tag{4.12}
$$

Figure D.1b plots the resulting distributions of the estimates for β_t. Again, the forest-based estimator clearly identifies two states, while the OLS estimate solely identifies the average policy effect. The partial dependency plot in figure 4.2b shows that the semi-parametric estimator correctly identifies the split between positive and negative lagged values of y_t. Consequently, the estimated state-dependent impulse responses follow the actual ones closely.

Exogenous and endogenous change. Next, we consider both, an exogenous structural break as well as an endogenous change in the policy effect over time.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.1 & y_{t-1} < 0, t < T/2 \\ -0.35 & y_{t-1} \geq 0, t < T/2 \\ -0.5 & y_{t-1} < 0, t \geq T/2 \\ -0.9 & y_{t-1} \geq 0, t \geq T/2 \end{cases}
\end{aligned}
\tag{4.13}
$$

---

[9]The partial dependency is computed by predicting the coefficient for the domain of the modifying variable – here, time.

As we can see in figure D.1c, the semi-parametric approach has no problem in identifying the four different states of β_t and gives a much more granular picture of the true policy effect. Figure 4.2c documents how closely the semi-parametrically estimated impulse response functions track the actual ones for all four states. In contrast, the OLS estimate only informs about the average response over the entire sample. The surrogate tree[10] illustrates how well the underlying random forest identifies the actual splits and state-dependent values of β_t.

Impact and dynamic change. So far, we have only considered cases in which only the impact coefficient β_t varied across states. However, it might also be that the policy rule is state-dependent, affecting the propagation of the shock. Let's consider the following dynamic system, with state-dependence in the impact effect of the policy and a varying policy parameter γ_t.

$$
\begin{aligned}
i_t &= 0.8i_{t-1} + \gamma_t y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.1 & y_{t-1} < 0, t < T/2 \\ -0.35 & y_{t-1} \geq 0, t < T/2 \\ -0.5 & y_{t-1} < 0, t \geq T/2 \\ -0.9 & y_{t-1} \geq 0, t \geq T/2 \end{cases}, \quad \gamma_t = \begin{cases} 0.3 & y_{t-1} < 0 \\ 0.1 & y_{t-1} \geq 0 \end{cases}
\end{aligned}
\tag{4.14}
$$

In general, with more lags and endogenous variables in the VAR structure, it becomes harder to justify why only single coefficients should be state-dependent. Hence, we will usually estimate more than one coefficient of the system semi-parametrically. This shifts the focus of considering state-dependent splits of a single coefficient as we did above to rather predicting the whole system of coefficients for specific state values.

Figure 4.3a plots estimates of the impulse response after a shock to the policy instrument for specific states. While the OLS estimates only offer an aggregated view, the semi-parametric estimates capture the state-dependence quite well. Instead of selecting specific state values to predict the system and subsequently the heterogeneous impulse responses, another way of presenting the functional form of the state-dependence is to compute partial dependency impulse response functions. Partial dependency impulse response functions are computed by varying one state over its domain and fixing all other
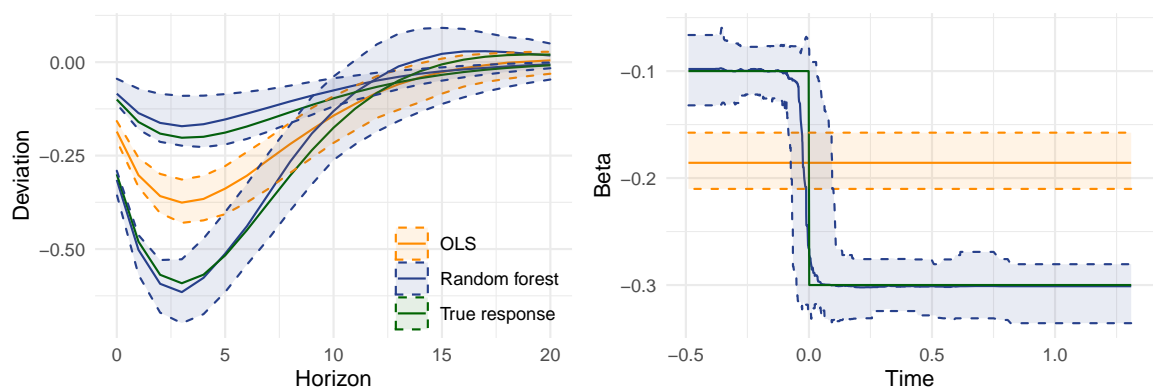
---

[10]Surrogate trees aim to illustrate the resulting splits by random forests with the latter being a collection of many trees. Obtaining a surrogate tree is done by fitting a single tree to the predictions of the random forest.

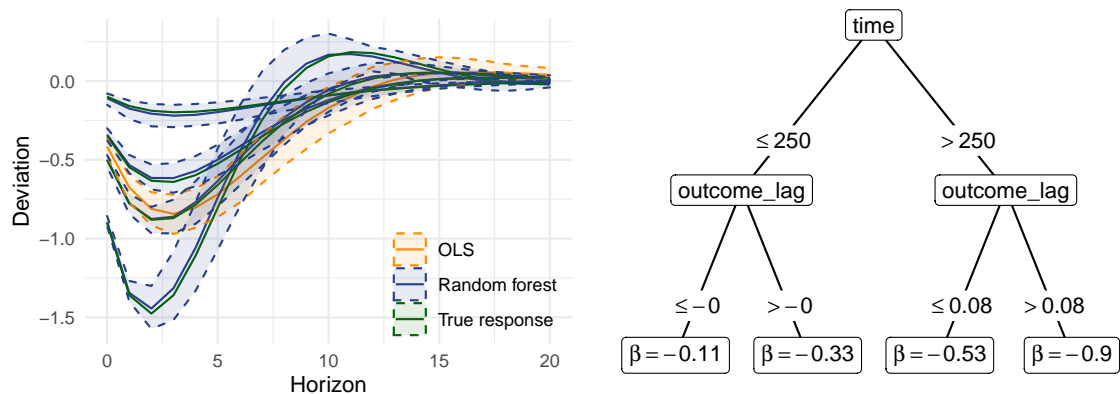Fig. 4.2 IRFs and Surrogate Trees - OLS and Random Forest

(a) Exogenous structural break
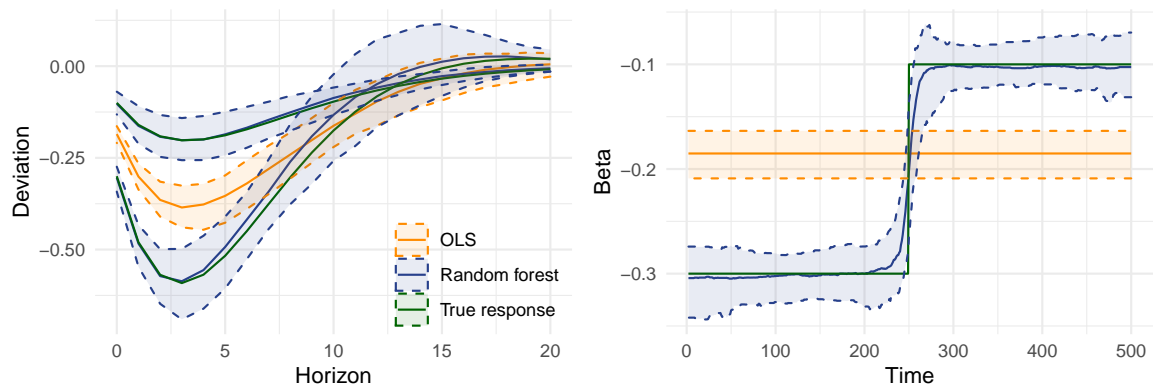
(b) Endogenous change

(c) Endogenous and exogenous change

Note: This figures compares the IRFs using OLS (orange) and Random Forest (blue) in different simulation scenarios. The green lines depict the true response. Simulations are conducted with T = 500.

states at their median values and predicting the impulse response. This leads to impulse response functions with the state of interest as an additional dimension as depicted in figure 4.3b. We can see that the estimated high-dimensional impulse response in blue is very close to the true response in green and the split points are correctly identified.

Many potential moderators. The previous simulation examples used the relevant moderators for the coefficient estimates to test whether our proposed estimator finds the correct splits. However, in practice, we might not know in advance which macroeconomic variables actually impact the policy response. In this example we feed several moderators to our estimator, of which only some are relevant. Furthermore, we allow for some correlation between the relevant and non-relevant moderators. Hence, this subsection shows that the VC-VAR approach can find the relevant moderators among many other, potentially correlated but non-relevant, moderators. We show this using a so-called variable importance measure (VIM) for random forests. Finding the relevant moderators among many candidates is vital in practice when researchers are unsure which moderators might impact coefficient estimates.

We simulate the same data generating process as defined in 4.13. However, instead of just providing the relevant variables time and y_{t-1} as moderators, we include five additional but irrelevant variables to the set of potential moderators. Two of which, m_1 and m_2, are correlated with y_{t-1} and the rest, $m_3 - m_5$, is drawn from a standard normal distribution and thereby pure noise.

We calculate the variable importance measure (VIM) in three steps: (1) Estimate the varying coefficient model using all potential moderators. (2) Randomize the observations within one moderator and obtain the predicted values. (3) Obtain the out-of-sample mean squared error by subtracting the full model's predicted values from the model's predicted values with one moderator's observations randomized.

The idea behind this procedure is the following. If a moderator is important for the prediction, then feeding the observations of that moderator in a randomized order into the model would yield a worse prediction.[11] Similarly, if a moderator is not relevant for prediction, even if we randomize its observations, that should not impact prediction. Hence, whenever the mean squared error increases a lot, we assume the moderator to be important.

---

[11]Please note that if a variable which is already part of the VAR system is also included as a moderator, we only randomize the variable for the prediction of the coefficients of the system in order to isolate its importance for capturing state-dependence, not for the VC-VAR system as a whole.

Fig. 4.3 IRFs and Partial Dependency Plots

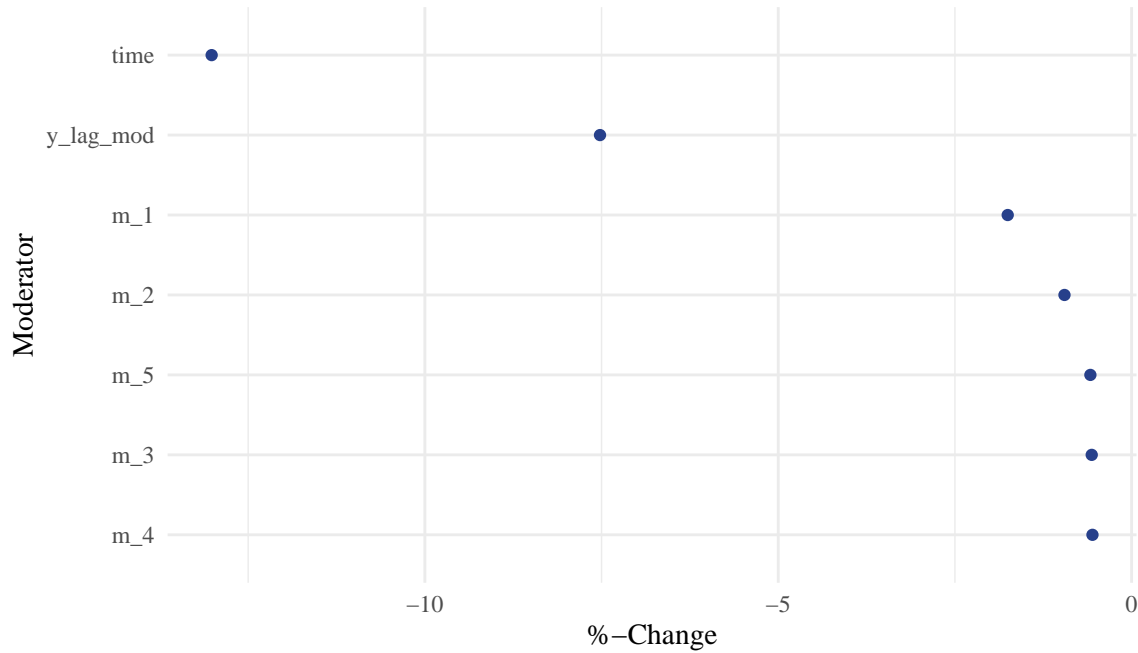(a) Impulse responses for different cases
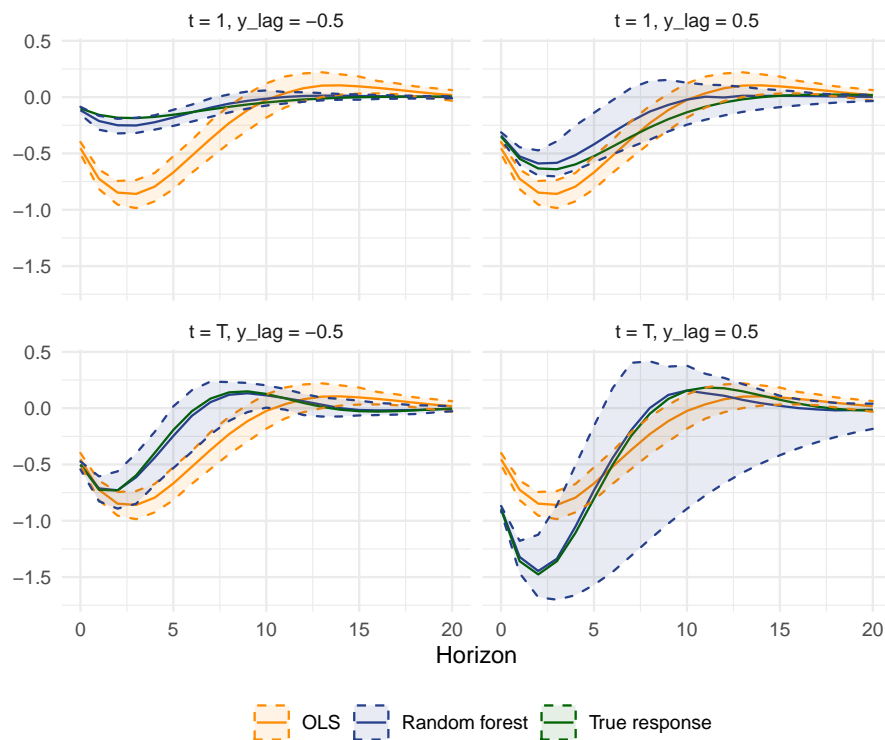
(b) Partial dependency impulse responses
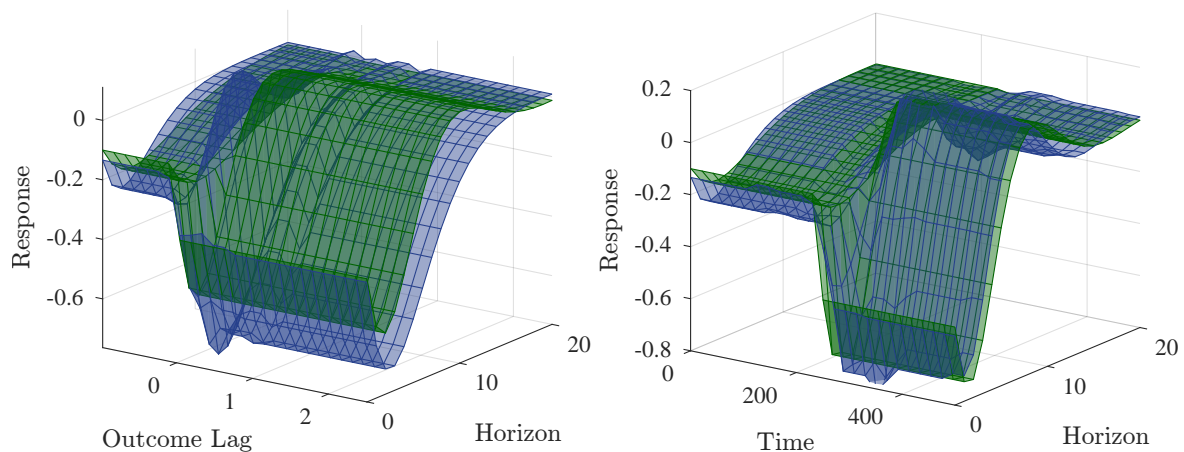
Note: This figures compares the IRFs and their partial dependency using OLS (orange) and Random Forest (blue) for impact and dynamic changes. The green lines depict the true response. Simulations are conducted with T = 500.

Fig. 4.4 Variable Importance Measure Simulation

Note: This figure depicts the variable importance measure showing the percentage reduction in the mean squared error when including the moderator compared to its randomized version. The simulation was conducted with 500 observations and the Random Forest was estimated with 100 tress (increasing the number of trees did not change the results much).

Figure 4.4 presents the variable importance measures from our simulated model. Moderators *time* and y_{t-1} are correctly identified as the relevant moderators to explain the underlying state-dependence. For both, randomizing the order of their observations increases the means squared prediction error substantially, while the irrelevant variables have hardly any additional predictive power. Thus, this simulation shows that the estimator can correctly distinguish the relevant moderators from other non-relevant ones.

4.4   Empirical application

This section illustrates the proposed methodology in an applied setting replicating and extending the study by Ramey & Zubairy (2018) (henceforth RZ). The replication exercise reveals that our method yields similar results to conventional IRF estimating techniques. Furthermore, we show that the new estimating methodology allows us to study state-dependence in more detail, extending the study by RZ.

RZ study the effect of government spending shocks on output to investigate the government spending multiplier in low and high employment regimes as well as at the ZLB. To do so, they derive a historical time series of government spending news shocks and use them in local projections as well as TVAR analyses as an instrument for government spending. We focus our comparison on the TVARs since they are more similar to our setting. [12]

To study state-dependence of government spending effects, RZ construct a dummy for slack taking the value of zero whenever unemployment is below 6.5% and one in case it is equal or higher. Similarly, they define a dummy for periods in which the economy was close to the ZLB.

The definition of dummies and thereby pre-definition of states is restrictive. First, deciding on the threshold is usually either an ad-hoc assumption or quite the opposite and the result of trying a battery of different thresholds. Both approaches come with a high degree of model uncertainty. Second, dummies are binary in nature allowing only for two states. Of course, one could include multiple dummies to account for multiple states but then one would have to justify the choice of several thresholds. Our method can be seen as more flexible allowing us to estimate any underlying state-dependence of the VAR non-parametrically as well as implicitly estimating interactions between the states in one estimation step. Studying the interaction of states might be of interest to the researcher since often the states in questions are correlated. For example, it is plausible that high unemployment states are correlated with periods close to the ZLB.

4.4.1 Sanity check: Using Dummies as Modifier Variables

In a first benchmark exercise, we compare the RZ results to estimates obtained from our VC-VAR model when directly feeding in the dummy as the sole effect modifier. Given the pre-defined dummy, the tree-based learners can only either split in line with the dummy or not at all in case the coefficient is state-independent. Hence, by just including a dummy, the semi-parametric estimator offers the same amount of restricted flexibility as a TVAR. Consequently, feeding in the dummy should yield equivalent results to what RZ find.

---

[12]Note that RZ use an LP-IV approach in their main analysis. Nevertheless, LP-IV can be shown to be equivalent to ordering the instrument first in an SVAR (Plagborg-Møller & Wolf (2019)). To obtain the same relative IRFs that the LP-IV approach yields, however, we have to divide by the impact impulse response on government spending. To compute the multiplier, the division by the impact effect on government spending is not necessary since we are interested in the ratio of the two IRFs of government spending and output. Thus, the division by the impact impulse would cancel out.

Fig. 4.5 ZLB Dependent Multiplier with Different Methods
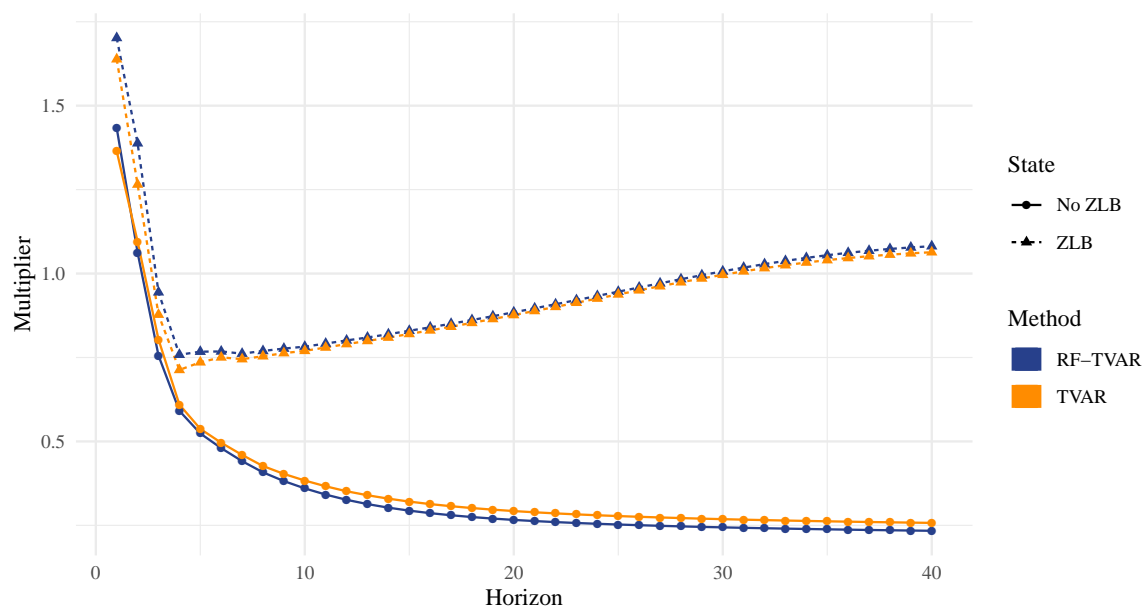
Note: This figure compares the estimated government multipliers for a ZLB and non-ZLB state. The mulitpliers are estimated using a TVAR (orange lines) and a Random Forest (blue lines) with the ZLB dummy directly fed in as a moderator.

Figure 4.5 depicts estimates of the fiscal multiplier in normal times and close to the ZLB. The orange lines correspond to the estimates found by RZ using a TVAR, while the blue lines refer to estimates from the semi-parametric VAR. It can be seen that the state-dependent multipliers converge to very similar values. Some small initial differences are expected since RZ compute the TVARs separately for each state[13], while the semi-parametric VAR makes use of the entire sample.

4.4.2   Semi-Parametric Estimation

In a next step we aim to illustrate the flexibility of the semi-parametric approach. First, in contrast to RZ we do not pre-define any states using dummy variables but directly feed in the continuous variables of the unemployment rate and the interest rate leaving it up to the VC-VAR to find appropriate splits. Second, while RZ estimate separate models looking for potential state-dependence of fiscal policy in regimes of high unemployment and times close to the ZLB, we estimate this simultaneously allowing for potential interaction between regimes. Moreover, we allow all coefficients in the structural VAR estimation

---

[13]According to the replication STATA file `tvar.do` by Ramey & Zubairy (2018).

Fig. 4.6 IRF of Output
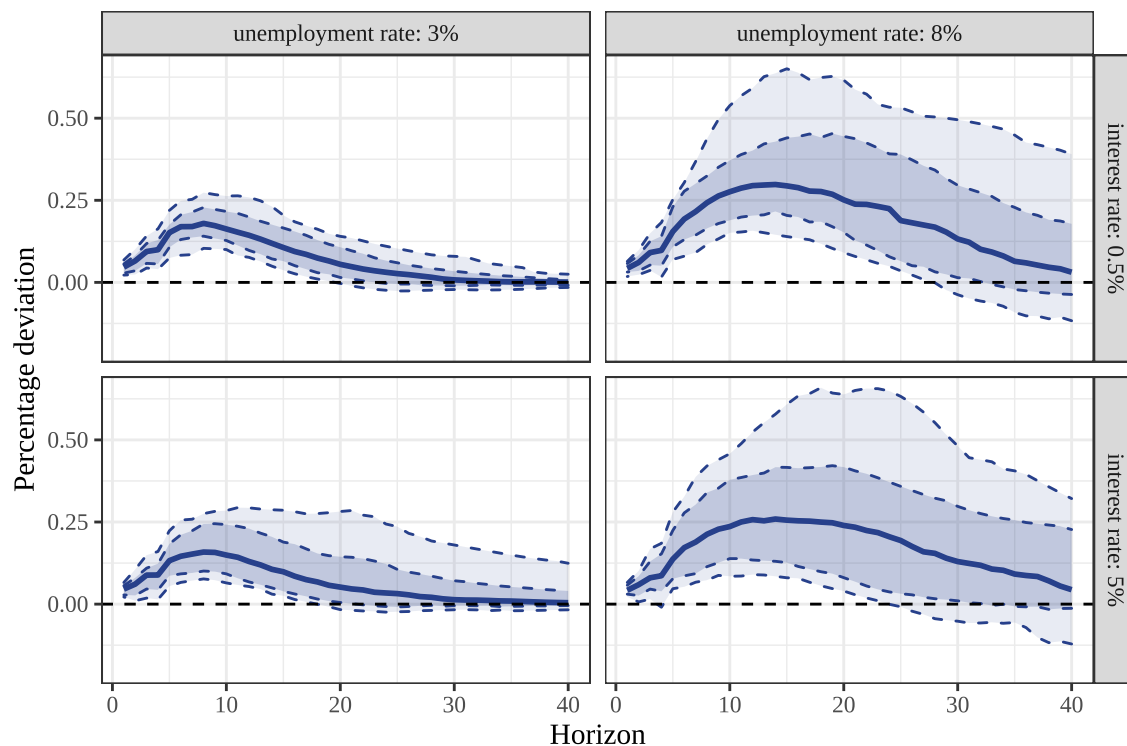
Note: This figure shows the IRFs of output following a fiscal policy shock. The IRFs are plotted for different combinations of low and high unemployment under a ZLB or no ZLB. The shaded areas depict 68% and 90% confidence intervals.

to vary across different states of the unemployment rate, the interest rate, the world war dummy and their potential interactions. The result is an SVAR model estimate which can be predicted for different states. Hence, in order to compute IRFs, we simply predict the entire system of coefficients for a given combination of states. The underlying assumption of these high-dimensional IRFs is that there is no regime switching and the IRF is conditional on a specific state.

Visualizing high-dimensional impulse response functions is not as straightforward as for one-dimensional impulse responses. In the case of a low-dimensional state space, one can plot a battery of impulse response functions for combinations of specific values of the effect modifiers. Figure 4.6, for instance, shows the estimated impulse response functions for two different levels of the unemployment rate and the interest rate when the world war dummy is zero. The dark shaded areas depict the 68% and the lightly shaded areas the 90% confidence interval, while the solid line denotes the point estimate.

Fig. 4.7 Unemployment and Interest Rates
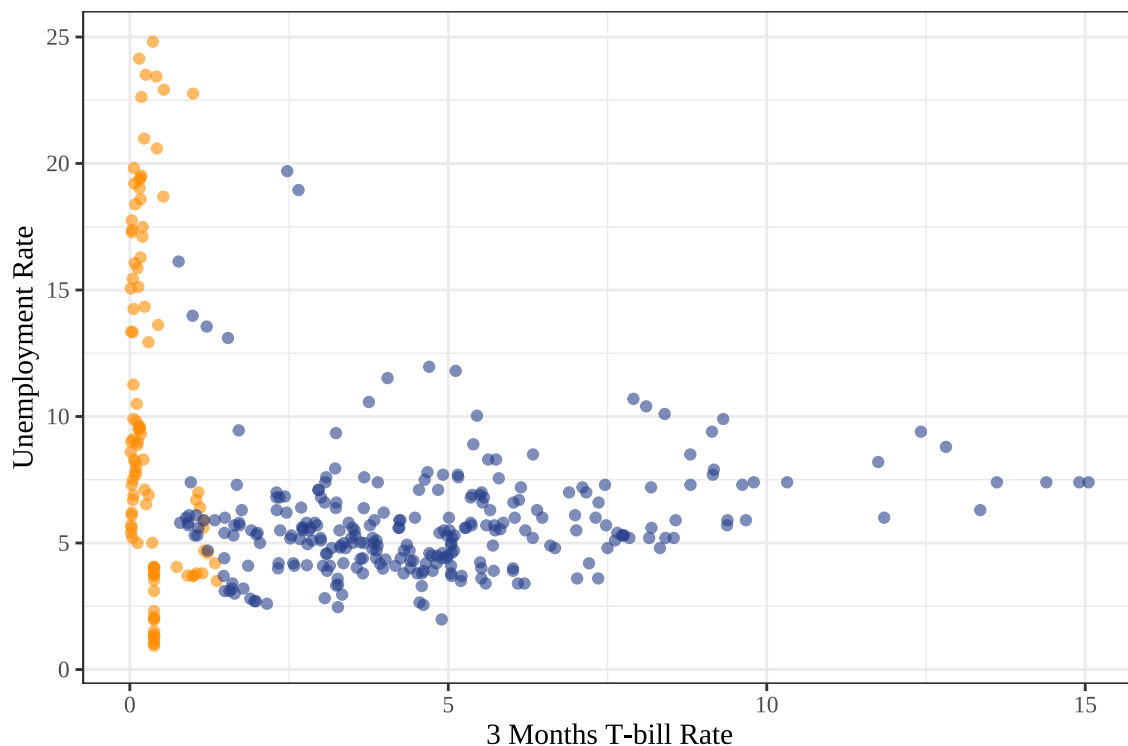
Note: A scatter plot of unemployment and interest rates. The orange dots depict observations during a ZLB period as defined by RZ.

The estimates suggest that fiscal policy is more effective in times of high unemployment. This is in line with the findings by RZ and the theoretical contribution by Rendahl (2016). However, when controlling for unemployment, the efficacy of fiscal policy is not higher close to the zero lower bound as RZ find. The difference might be due to the definition of the zero lower bound dummy in Ramey & Zubairy (2018). RZ define periods of continuous low interest rates and low inflation as binded by the zero lower bound without choosing a specific interest rate threshold. Furthermore, their zero lower bound dummy spans the entire sample, while records of the treasury bill's interest rate only start in 1920. However, more likely is that periods close to the zero lower bound are correlated with periods of high unemployment rates. In fact, this correlation is quite striking as figure 4.7 shows. Points in orange denote RZ's definition of the zero lower bound dummy. Without controlling for unemployment rates, RZ's estimates for the state-dependence of fiscal policy in times of a binding zero lower bound might just pick up the effect of high unemployment due to an omitted variable bias.

Fig. 4.8 High-dimensional IRFs

(a) Partial dependency impulse responses
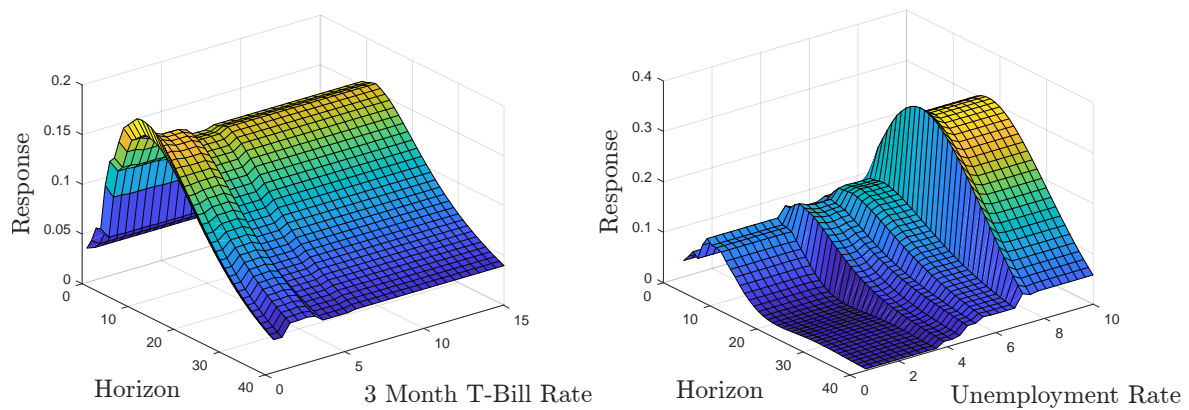
(b) Partial dependency fiscal multiplier
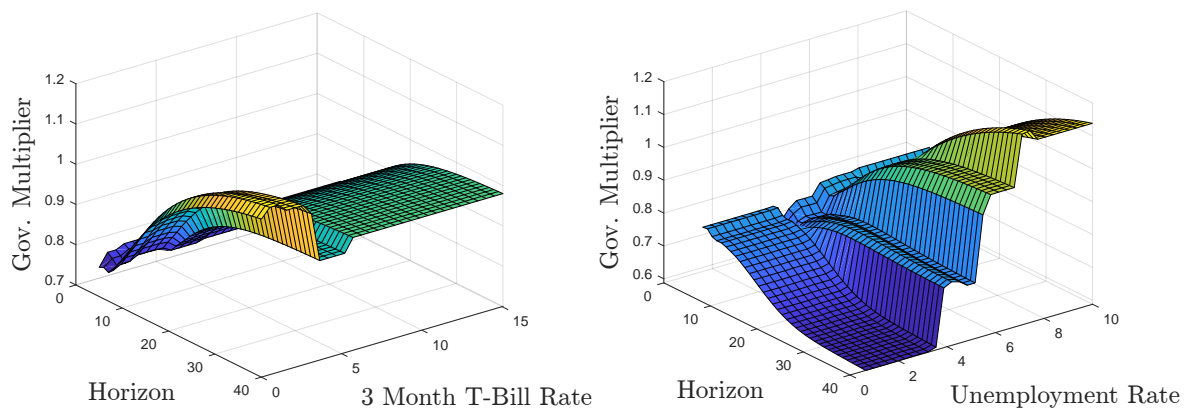
Note: High-dimensional IRFs of output. The top-panel depicts how the moderator variables T-Bill rate and unemployment rate affect the IRF on output. The lower panel analyses the government multiplier.

Another way of visualizing high-dimensional impulse response functions are partial dependency plots. Partial dependency plots are a common tool to make machine learning output easier to interpret. The concept aims to illustrate the marginal effects of individual variables when holding all other variables constant. We make use of this concept and plot partial dependency impulse response functions in figure 4.8a fixing the other variable at its median. The left plot of figure 4.8a shows the partial dependency impulse response for variations in the interest rate of the treasury bill. The impulse response appears constant and only slightly higher for interest rates close to the zero lower bound.

In contrast, the state-dependence of fiscal policy with respect to the unemployment rate is stark. The magnitude of the response of output in case of unemployment rates higher than 8% is more than twice as large as for low unemployment rates. Figure 4.8b relates the high-dimensional response of output to responses in government spending by calculating high-dimensional fiscal multipliers. We follow Ramey & Zubairy (2018) in computing the multiplier as the cumulated response of output over the cumulated response of government spending[14]. While the multiplier is only slightly higher close to the zero lower bound relative to environments of higher interest rates and below unity, the multiplier is increasing for higher unemployment rates and becomes larger than unity.

4.4.3   Estimating Asymmetric Effects

As another illustration of the methodology, we assess potential asymmetric policy effects. To allow for the estimation of asymmetric policy effects, the News shock variable (i.e. RZ's instrument for government spending) will now be introduced as a moderator variable. This allows to predict IRFs for different, including negative, values of the shock variable. Note that when we include the shock series as a moderator variable, it has to be excluded when estimating the reduced or structural form equation for the news shock series. Otherwise, the variable would appear on both, the left- and right-hand side of the regression equation.

Figure 4.9 depicts the government multiplier for various shock values. The multiplier is lower for negative as compared to positive values. Furthermore, it appears that the multiplier does not change much over the negative values of the shock series. However, this might be due to the fact that the news series contains few negative observations, making our estimation less precise.

---

[14]This is the definition Ramey & Zubairy (2018) used to calculate multipliers based on their TVAR estimates.

Fig. 4.9 Fiscal Multiplier for Negative and Positive Gov. Shocks
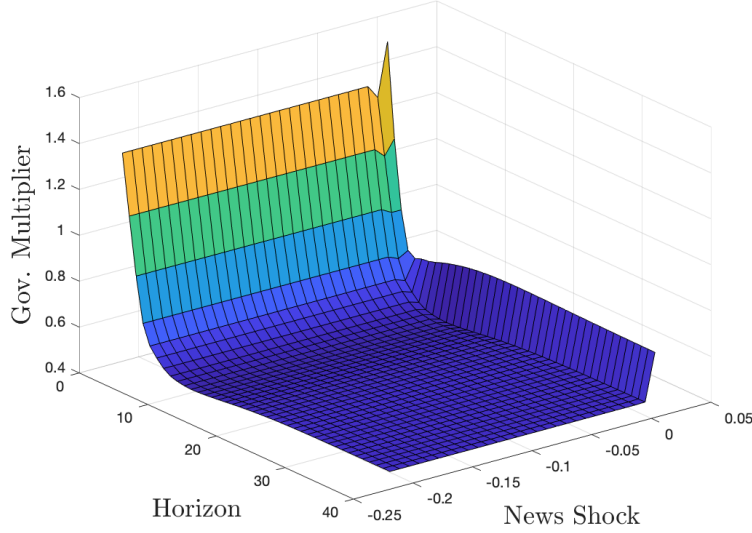
Note: This figure shows the output multiplier of government spending. To test for asymmetric effects of positive or negative shocks, the News shock series is a moderator variable.

4.5   Conclusion

This paper proposes a semi-parametric framework for estimating high-dimensional SVARs and impulse response functions. While a parametric shell keeps the estimation efficient and more interpretable, tree-based learners allow for non-parametric variation in coefficients depending on potential modifiers. Using tree-based learners has two main advantages: 1) it allows us to detect any potential state-dependence without specifying any functional form or cutoff thresholds, 2) forests can cope with a high-dimensional state space allowing us to control for many potential effect modifiers. Simulation studies confirm that the semi-parametric framework correctly identifies state-dependent variation in coefficients even for macroeconomic samples sizes. When applying the estimator to the historical data set by Ramey & Zubairy (2018) we confirm their findings that fiscal policy is more effective in times of high unemployment. However, we only find slight differences in the efficacy of fiscal policy close to the zero lower bound relative to normal times. In fact, RZ's results seem to be driven by the correlation between low-interest rates and unemployment.

References

Ascari, G. & Haber, T. (2019). *Sticky prices and the transmission mechanism of monetary policy: A minimal test of New Keynesian models*. Technical report.

Auerbach, A. J. & Gorodnichenko, Y. (2012). Fiscal multipliers in recession and expansion. In *Fiscal policy after the financial crisis* (pp. 63–98). University of Chicago Press.

Auerbach, A. J. & Gorodnichenko, Y. (2013). Output spillovers from fiscal policy. *American Economic Review*, 103(3), 141–46.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.

Buergin, R. A. & Ritschard, G. (2017). Coefficient-wise tree-based varying coefficient regression with vcrpart. *Journal of Statistical Software*, 80(6), 1–33.

Canzoneri, M., Collard, F., Dellas, H., & Diba, B. (2016). Fiscal multipliers in recessions. *The Economic Journal*, 126(590), 75–108.

Duarte, V. (2018). Machine learning for continuous-time economics. *Available at SSRN 3012602*.

Eichenbaum, M., Rebelo, S., & Wong, A. (2018). *State dependent effects of monetary policy: The refinancing channel*. Technical report, National Bureau of Economic Research.

Farhi, E. & Werning, I. (2016). Fiscal multipliers: Liquidity traps and currency unions. In *Handbook of macroeconomics*, volume 2 (pp. 2417–2492). Elsevier.

Fazzari, S. M., Morley, J., & Panovska, I. (2015). State-dependent effects of fiscal policy. *Studies in Nonlinear Dynamics & Econometrics*, 19(3), 285–315.

Goulet Coulombe, P. (2020). The macroeconomy as a random forest. *Available at SSRN 3633110*.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–779.

Joseph, A. (2019). Shapley regressions: A framework for statistical inference on machine learning models. *arXiv preprint arXiv:1903.04209*.

Li, D., Plagborg-Møller, M., & Wolf, C. K. (2021). Local Projections vs. VARs: Lessons From Thousands of DGPs. (pp. 44).

Nakamura, E. & Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3), 59–86.

References

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics*, 102(4), 690–704.

Plagborg-Møller, M. & Wolf, C. K. (2019). Local projections and vars estimate the same impulse responses. *Unpublished paper: Department of Economics, Princeton University*, 1.

Ramey, V. A. & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy*, 126(2), 850–901.

Rendahl, P. (2016). Fiscal policy in an unemployment crisis. *The Review of Economic Studies*, 83(3), 1189–1224.

Wang, J. C. & Hastie, T. (2014). Boosted varying-coefficient regression models for product demand prediction. *Journal of Computational and Graphical Statistics*, 23(2), 361–382.

Wolf, C. K. (2020). SVAR (Mis)Identification and the Real Effects of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, 12, 1–32.

Appendix A

Appendix for Chapter 1

A.1 Competitive labour market

As in the standard NK model, there is a representative household obtaining utility from consumption and disutility from supplying labour, which are additively separable. The problem is constrained by the nominal budget flow equation of the household

$$
\max_{C_t, B_{t+1}, L_t} E_0 \Big[\sum_{t=0}^{\infty} \beta^t \left(u(C_t) - v(L_t) \right) \Big],
$$

s.t.

$$
C_t P_t + B_{t+1} = B_t (1 + i_t) + w_t P_t L_t + \pi_t,
$$

where w_t denotes real wages and π_t denotes the firm profits received. The FOCs are

Euler equation:
$$
u'(C_t) = \beta E_t \Big[(1 + i_t) \frac{P_t}{P_{t+1}} u'(C_{t+1}) \Big], \tag{A.1}
$$

Labour supply:
$$
v'(L_t) = u'(C_t) w_t, \tag{A.2}
$$

with
$$
v(L_t) = \psi \frac{L_t^{1+\eta}}{1+\eta} \tag{A.3}
$$

where η is the Frisch elasticity of labour supply and ψ the weight of disutility of labour. In the calibration for figure 1, I assume $\eta = 1.5$ and $\psi = 1$. The labour demand is coming from the intermediate good producers' profits maximization.

A.2 Endogenous markup derivation

Derivation of markup for $\zeta \to \infty$[1]:

$$\max_{Y_{t+s}^j(i)} \sum_{s=0}^{\infty} E_t \left[Q_{t,t+s} \left(P_{t+s}(i) Y_{t+s}^j(i) - \frac{w_{t+s}}{A_{t+s}} P_{t+s} Y_{t+s}^j(i) - \frac{\theta_p}{2} \left(\frac{P_{t+s}(i)}{P_{t+s-1}(i)} - 1 \right)^2 P_{t+s}(i) Y_{t+s}^j(i) \right) \right]$$

$$s.t. \qquad P_t(i) = \left(\frac{Y_t(i)}{Y_t} \right)^{-1/\sigma} P_t = \left(\frac{Y_t^j(i) + Y_t^{-j}(i)}{Y_t} \right)^{-1/\sigma} P_t$$

FOC:

$$P_t(i) - \frac{1}{\sigma} \left(\frac{n Y_t^j(i)}{Y_t} \right)^{-1/\sigma - 1} P_t Y_t^j(i) \frac{1}{Y_t} - \frac{w_t}{A_t} P_t - \theta \left(\frac{P_t(i)}{P_{t-1}(i)} - 1 \right) (-1/\sigma) \left(\frac{n Y_t^j(i)}{Y_t} \right)^{-1/\sigma - 1} P_t Y_t^j(i) \frac{1}{Y_t} \frac{1}{P_{t-1}}$$

$$- \frac{\theta_p}{2} \left(\frac{P_t(i)}{P_{t-1}(i)} - 1 \right)^2 P_t(i) + \frac{\theta_p}{2} \left(\frac{P_t(i)}{P_{t-1}(i)} - 1 \right)^2 \frac{1}{\sigma} \left(\frac{n Y_t^j(i)}{Y_t} \right)^{-1/\sigma - 1} P_t Y_t^j(i) \frac{1}{Y_t}$$

$$- \theta_p E_t \left[Q_{t,t+1} \left(\frac{P_{t+1}(i)}{P_t(i)} - 1 \right) \right] P_{t+1}(i) Y_{t+1}^j(i) \frac{P_{t+1}}{P_t^2(i)} \frac{1}{\sigma} \left(\frac{n Y_t^j(i)}{Y_t} \right)^{-1/\sigma - 1} P_t \frac{1}{Y_t} = 0$$

Re-substituting the expression for $P_t(i)$ and re-writing $\frac{P_t(i)}{P_{t-1}(i)}$ as π_t, we can simplify the expression as follows:

$$P_t(i) - \frac{1}{\sigma} \frac{1}{n_t} P_t(i) - \frac{w_t}{A_t} P_t - \frac{1}{\sigma} \frac{1}{n_t} \theta_p (\pi_t(i) - 1) \pi_t(i)$$

$$- \frac{\theta_p}{2} (\pi_t(i) - 1)^2 P_t(i) + \frac{1}{\sigma} \frac{1}{n_t} \frac{\theta_p}{2} (\pi_t(i) - 1)^2 P_t(i)$$

$$- \theta_p E_t \left[Q_{t,t+1} (\pi_{t+1} - 1) \pi_t^2 \frac{Y_{t+1}^j}{Y_t^j} \right] P_t(i) = 0$$

Solving for $P_t(i)$ we get:

$$P_t(i) = \mu_t(i) \frac{w_t}{A_t} P_t$$

$$\mu_t(i) = \frac{\sigma n_t(i)}{(\sigma n_t(i) - 1) \left[1 - \frac{\theta_p}{2} (\pi_t(i) - 1)^2 \right] + \theta_p \pi_t(i) (\pi_t(i) - 1) - \Gamma_t(i)}$$

$$\Gamma_t(i) = \theta_p E_t \left[Q_{t,t+1} \pi_{t+1}(i)^2 (\pi_{t+1}(i) - 1) \frac{Y_{t+1}(i) n_t}{Y_t(i) n_{t+1}} \right]$$

---

[1]For $\zeta < \infty$ analogously, just note that market share $s^j(i)_t \left(\frac{P_t^j(i)}{P_t(i)} \right)^{1-\zeta}$

A.3   Equilibrium conditions

Combining everything gives us the set of equations in order to solve for the steady state:

Household's problem:

$$u'(C_t) = \beta E_t \left[(1 + i_t) \frac{P_t}{P_{t+1}} u'(C_{t+1}) \right] \tag{A.4}$$

Monetary authority:

$$ln(R_t/R_{ss}) = \rho_R ln(R_{t-1}/R_{ss}) + (1 - \rho_R)[\theta_\pi ln(\pi_t/\pi_{ss}) + \theta_Y ln(Y_t/Y_{ss}) + v_t], \tag{A.5}$$

from firms' problem:

$$P_t^{j,s}(i) = \mu_t^{j,s}(i) \frac{\vartheta_t}{z^s(i) A_t} P_t, \text{ for s } \in \{1,2,...,S\} \tag{A.6}$$

where

$$
\begin{aligned}
\mu_t^{s,j}(i) &= \frac{\Theta_t^j(i)}{(\Theta_t^j(i) - 1)\left[1 - \frac{\theta_p}{2}(\pi_t(i) - 1)^2\right] + \theta_p \pi_t(i)(\pi_t(i) - 1) - \Gamma_t(i)}, \\
\Theta_t^{s,j}(i) &= \left[\frac{1}{\zeta} + (\frac{1}{\sigma} - \frac{1}{\zeta}) s_t^{s,j}(i)\right]^{-1}, \\
\Gamma_t(i) &= \theta_p \mathbb{E}_t\left[m_{t,t+1} \pi_{t+1}(i)^2 (\pi_{t+1}(i) - 1) \frac{y_{t+1}^{s,j}}{y_t^{s,j}} \right],
\end{aligned}
\tag{A.7}
$$

Dynamic entry of firms:

$$v_t^s = \mathbb{E}_t\left[m_{t,t+1}\left(p_u(\pi_{t+1}^{s+1} + v_{t+1}^{s+1}) + p_l(\pi_{t+}^{s-1} + v_{t+1}^{s-1}) + (1 - p_u - p_l)(\pi_{t+1}^s + v_{t+1}^s)\right)\right], \atop \text{for } s \in \{1,2,...,S\} \tag{A.8}$$

and

$$v_t^E = f^E \tag{A.9}$$

DMP labour market with AOB:

$$L_t = (\rho + x_t)L_{t-1} \tag{A.10}$$

$$Q_t(J_t - \kappa) = s \tag{A.11}$$

$$J_t = \vartheta_t - \bar{w}_t + \rho E_t m_{t+1} J_{t+1} \tag{A.12}$$

$$V_t = \bar{w}_t + E_t m_{t+1} \left[\rho V_{t+1} + (1-\rho)(f_{t+1} V_{t+1} + (1-f_{t+1}) U_{t+1})) \right] \tag{A.13}$$

$$U_t = D + E_t m_{t+1} \left[f_{t+1} V_{t+1} + (1-f_{t+1}) U_{t+1} \right] \tag{A.14}$$

$$J_t = \beta_1 (V_t - U_t) - \beta_2 \phi_t + \beta_3 (\vartheta_t - D_t), \tag{A.15}$$

with $\beta_i = \alpha_{i+1}/\alpha_i$ for $i \in 1,2,3$

$$\alpha_1 = 1 - \delta + (1-\delta)^M,$$

$$\alpha_2 = 1 - (1-\delta)^M,$$

$$\alpha_3 = \alpha_2 \frac{1-\delta}{\delta} - \alpha_1,$$

$$\alpha_4 = \frac{1-\delta}{2-\delta} \frac{\alpha_2}{M} + 1 - \alpha_2,$$

$$Q_t = \eta_m \Xi_t^{-\eta} \tag{A.16}$$

$$\Xi_t = \frac{v_t L_{t-1}}{1 - \rho L_{t-1}} \tag{A.17}$$

$$Q_t = \frac{x_t}{v_t} \tag{A.18}$$

$$f_t = \frac{x_t L_{t-1}}{1 - \rho L_{t-1}} \tag{A.19}$$

Market clearing:

$$Y_t = N_t^{\frac{1}{1-\zeta}} \left(\sum_{s=1}^{S} \sum_{j=1}^{N_t^s} (z^s h_t^{j,s})^{\frac{\zeta-1}{\zeta}} \right)^{\frac{\zeta}{\zeta-1}} \tag{A.20}$$

$$1 = \sum_{s=1}^{S} \frac{N_t^s}{N_t} \left(\frac{P_t^s}{P_t} \right)^{1-\zeta} \tag{A.21}$$

$$Y_t = C_t + \frac{\theta_p}{2} (\pi_t - 1)^2 Y_t + (s/Q_t + \kappa) x_t L_{t-1} + N^E f^E \tag{A.22}$$

$$H_t = L_t \tag{A.23}$$

A.4   Additional Figures
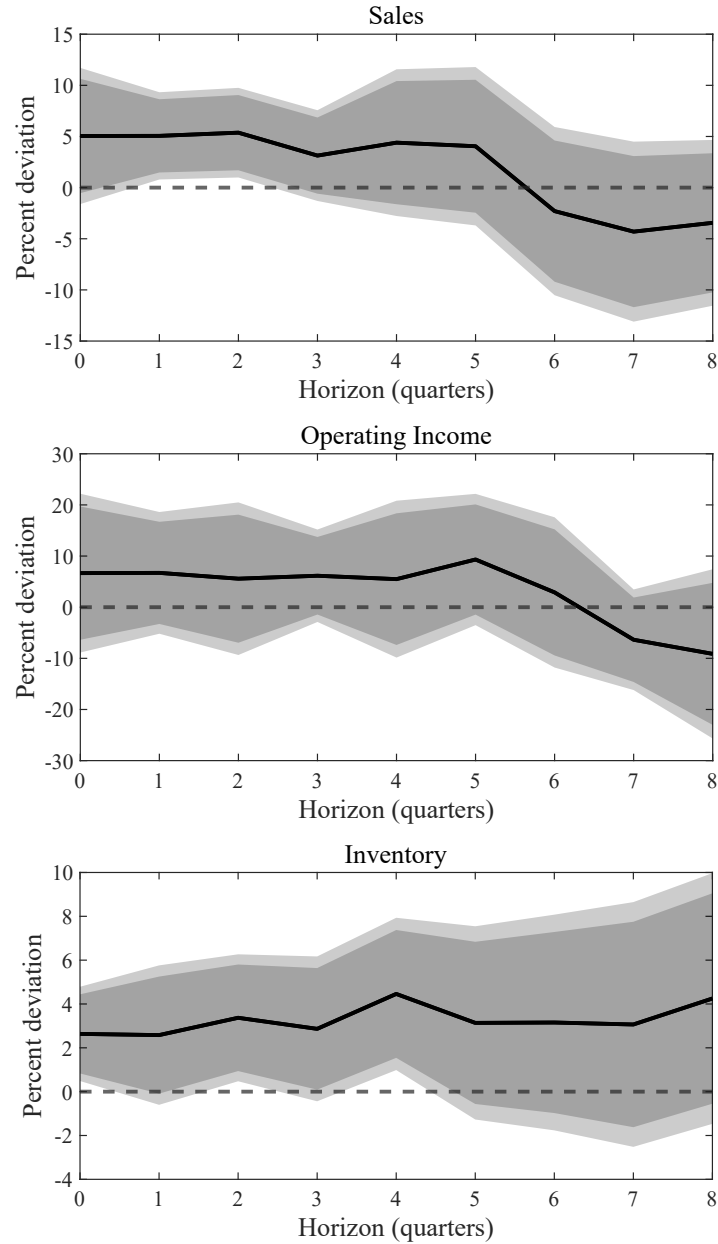
Fig. A.1 Estimated sales and income response for top 1% firms vs. the rest

Note: Estimated response of sales and operating income for the top 1% firms in size relative to the bottom 99% to GK monetary policy shocks using local projections. The shocks are normalized such that a positive sign denotes an expansionary shock. Standard errors are clustered at the firm-level and over time. The dark and light shaded areas represent the 90% and 95% confidence interval, respectively.

Appendix B

Appendix for Chapter 2

B.1   Variable definitions

Central Credit Responsibility Database (Central de Responsabilidades de Crédito)

Identifier (tina)   Anonymized tax identification number.

Global Credit (valor_global)   is the sum of regular credit and potential credit, representing the total available credit that a firm accesses.

Regular Credit (valor_efectivo)   is credit effectively used in a regular situation, i.e., without payment delays as defined in the respective contract. Examples of effective responsibilities are:

- Loans for the acquisition of financial instruments (shares, bonds, etc.);
- Discount and other credits secured by effects;
- Overdrafts on bank accounts;
- Leasing and factoring;
- Used amounts of credit cards.

Potential Credit (valor_potencial)   represents irrevocable commitments of the participating entities. Banco de Portugal requires all credit-granting institutions to report to the CCR their outstanding loan exposure by instrument of all irrevocable credit obligations. Examples of potential responsibilities are:

- Unused amounts of credit cards;

- Lines of credit;

- Guarantees provided by participating entities;

- Guarantees and guarantees given in favor of the participating entities;

- Any other credit facilities likely to be converted into effective debts.

Overdue Credit (valor_vencido)   All outstanding credit exposures recorded as non-performing (including overdue, written off, renegotiated credit, overdue credit in litigation, and written off credit in litigation) are aggregated to calculate overdue credits. It includes principal, interest and related fees.

Short-term Credit (valor_curto)   Short-term credit is calculated using two different definitions. In the first place, short-term credit is defined based on the term-to-maturity as agreed in the credit contract, denoted by valor_curto_o. Specifically, short-term credit has an original maturity of equal to or less than one year. Before 2009, the CCR dataset did not streamline credit exposure based on the maturity structure. Therefore, for the data before 2009, the short-term credit is defined as the aggregation of commercial credit, discount funding, and other short-term funding, which are short-term funding by their nature. In the second place, short-term credit is defined based on residual maturity – the remaining time until the expiration or the repayment of the instrument, denoted by valor_curto_r. Specifically, it is credit with a residual maturity of equal to or less than one year. This variable is only available from 2009 onwards. Potential credit is excluded for both calculations.

Long-term Credit (valor_longo)   Similar to short-term credit, long-term credit is defined based on original and residual maturities. More precisely, long-term credit is credit with an original or residual maturity of more than one year, denoted by valor_longo_o and valor_longo_r, respectively. Long-term credit defined on an original maturity basis (valor_longo_o) for the data before 2009 is the aggregation of total credit excluding commercial credit (type 1), discount funding (type 2), and other short-term funding (type 3). Potential credit is excluded for both calculations.

B.2 Descriptive statistics

Table B.1 reports correlations between all measures. By construction, some correlation coefficients can be relatively low, as some measures represent subsets of others. In fact, correlation coefficients between those binary measures augmenting the baseline measure can be interpreted as the fraction of the subset relative to the baseline definition of being strictly constrained. The correlation coefficients between binary and continuous measures are close to zero, as the variance in the continuous measure is too high to be captured by a binary measure, underlining the importance of including continuous measures in the analysis. The negative sign is simply due to the definition of the continuous measures, as higher values are an indication for being less constrained.

In Table B.2 we report medians for different size related variables for the universe of firms in our panel dataset. Following Crouzet & Mehrotra (2020), we split firms in size bins based on their total asset amount. Furthermore, we document descriptive statistics for the mutually exclusive subsets of constrained and unconstrained firms, based on the baseline measure. We can observe that the median financially constrained firm has less assets, less turnover, is younger and employs less people compared to the median unconstrained firm. Table B.3 reports medians of financial variables for size bins and the subset of constrained and unconstrained firms. Financial variables are expressed in terms of ratios relative to the total amount of assets a firm owns. While the median of smaller firms tends to have a higher leverage ratio, yet also a higher liquidity ratio, there is hardly any difference between the median constrained and the median unconstrained firm. This is suggestive of the fact, that when controlling for size, there are constrained and unconstrained firms across the distribution of financial variables.

Constrained measure	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) $\mathbf{1}_{\text{Potential credit}_t=0}$	1						
(2) $\mathbf{1}_{\text{Pot. credit}_t=0\ \&\ \text{Pot. credit}_{t-1}>0}$	0.31***	1					
(3) $\mathbf{1}_{\text{Pot. credit}_t=0\ \&\ \Delta\text{Effective credit}_{t+1}<0}$	0.82***	0.24***	1				
(4) $\mathbf{1}_{\text{Pot. credit}_t=0\ \&\ \Delta\text{Overdue credit}_{t+1}>0}$	0.26***	0.12***	0.28***	1			
(5) $\mathbf{1}_{\text{Pot. credit}_t=0} + \cdots$	0.73***	0.23***	0.60***	0.19***	1		
(6) $\frac{\text{Pot. credit}_t + \text{Cash}_t}{\text{Liabilities}_t}$	-0.00	-0.00	-0.00	-0.00	-0.00	1	
(7) $\frac{\text{Pot. credit}_t + \Delta\text{Short/Long term credit}_{t+1}}{\text{Liabilities}_t}$	-0.00*	-0.00	-0.00**	-0.00	-0.00	0.33***	1

Table B.1 Correlation between different measures for being constrained

Size group	0 - 90th	90th-99th	99-99.5th	>99.5th	constrained	unconstrained
Assets (€ mio.)	0.25	4.64	23.07	77.09	0.21	0.39
Sales (€ mio.)	0.02	1.01	2.06	0.11	0.00	0.07
Turnover (€ mio.)	0.02	2.50	9.81	15.91	0.12	0.31
Age	12	20	22	21	11	14
Employees	4	18	50	56	2	5

Table B.2 Size and age

Size group	0 - 90th	90th-99th	99-99.5th	>99.5th	constrained	unconstrained
Dividends (€ mio.)	0.00	0.04	0.34	1.28	0.00	0.00
Fixed tangible (€ mio.)	0.04	0.92	4.20	8.23	0.03	0.06
Investment (€ mio.)	-0.00	-0.01	-0.03	-0.02	-0.00	-0.00
Financial investments (€ mio.)	0.00	0.00	0.08	4.77	0.00	0.00
Equity (€ mio.)	0.06	1.43	7.42	26.18	0.04	0.10
Liabilities (€ mio.)	0.18	3.06	15.28	47.67	0.16	0.25
Total income (€ mio.)	0.19	2.63	10.93	25.18	0.12	0.33
EBIT (€ mio.)	0.01	0.12	0.65	2.30	0.00	0.01
Leverage	0.20	0.25	0.21	0.12	0.20	0.20
Liquidity ratio	0.06	0.02	0.01	0.01	0.04	0.05
Potential credit (€ mio.)	0.00	0.10	0.51	1.36	0.00	0.02
Effective credit (€ mio.)	0.04	1.12	4.90	9.98	0.03	0.06
Bank relationships	2	3	4	4	1	2

Table B.3 Financial characteristics

B.3   Additional Tables

Table B.4 Cyclicality in sales (services) conditional on size bins and measures of financial constraints

	Sales growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.305***	2.230***	2.260***	2.261***	2.274***	2.249***	2.283***	2.290***
	(0.042)	(0.048)	(0.042)	(0.052)	(0.047)	(0.053)	(0.042)	(0.047)
[90,99] × GDP growth	0.212	0.257*	0.239	0.325*	0.306*	0.230	0.212	0.294*
	(0.150)	(0.151)	(0.150)	(0.171)	(0.170)	(0.151)	(0.150)	(0.170)
[99,99.5] × GDP growth	-1.042**	-1.000**	-1.016**	-1.389***	-1.412***	-1.015**	-1.047**	-1.399***
	(0.444)	(0.444)	(0.444)	(0.529)	(0.528)	(0.444)	(0.444)	(0.529)
[99.5,100] × GDP growth	-1.669***	-1.614***	-1.635***	-1.795***	-1.789***	-1.640***	-1.681***	-1.839***
	(0.470)	(0.471)	(0.470)	(0.571)	(0.571)	(0.471)	(0.470)	(0.571)
Const.I × GDP growth		0.176*						
		(0.098)						
Const.II × GDP growth			0.310					
			(0.222)					
Const.III × GDP growth				0.204				
				(0.127)				
Const.IV × GDP growth					0.915**			
					(0.430)			
Const.V × GDP growth						0.109		
						(0.085)		
Const.VI × GDP growth							-0.154***	
							(0.055)	
Const.VII × GDP growth								-0.183***
								(0.061)
Observations	805,053	805,053	805,053	709,948	709,948	805,053	804,639	709,661
R-squared	0.023	0.024	0.024	0.024	0.025	0.023	0.023	0.024
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of sales with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.5 Cyclicality in employees conditional on size bins and measures of financial constraints

	Employee growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	1.036***	1.026***	1.010***	1.125***	1.105***	1.004***	1.040***	1.111***
	(0.013)	(0.016)	(0.013)	(0.017)	(0.015)	(0.017)	(0.013)	(0.015)
[90,99] × GDP growth	-0.004	-0.001	0.007	-0.121**	-0.118**	-0.006	-0.008	-0.121**
	(0.046)	(0.047)	(0.046)	(0.052)	(0.052)	(0.047)	(0.046)	(0.052)
[99,99.5] × GDP growth	-0.042	-0.044	-0.033	-0.196	-0.188	-0.034	-0.049	-0.185
	(0.132)	(0.132)	(0.132)	(0.153)	(0.153)	(0.132)	(0.132)	(0.153)
[99.5,100] × GDP growth	-0.330**	-0.330**	-0.318**	-0.398**	-0.385**	-0.316**	-0.341**	-0.401**
	(0.149)	(0.149)	(0.149)	(0.177)	(0.176)	(0.149)	(0.149)	(0.176)
Const.I × GDP growth		-0.026						
		(0.029)						
Const.II × GDP growth			0.129*					
			(0.067)					
Const.III × GDP growth				-0.032				
				(0.036)				
Const.IV × GDP growth					0.306**			
					(0.120)			
Const.V × GDP growth						0.063**		
						(0.026)		
Const.VI × GDP growth							-0.074***	
							(0.013)	
Const.VII × GDP growth								-0.087***
								(0.016)
Observations	1,289,884	1,289,884	1,289,884	1,128,989	1,128,989	1,289,884	1,288,586	1,128,066
R-squared	0.024	0.024	0.025	0.025	0.026	0.025	0.024	0.025
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of employees with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.6 Cyclicality in turnover conditional on size bins and measures of financial
constraints including time fixed effects

	Turnover growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.512***	2.441***	2.440***	2.450***	2.470***	2.450***	2.522***	2.508***
	(0.023)	(0.027)	(0.023)	(0.028)	(0.025)	(0.030)	(0.023)	(0.025)
[90,99] × GDP growth	0.146	0.176*	0.175*	0.168	0.157	0.161	0.125	0.139
	(0.101)	(0.102)	(0.101)	(0.116)	(0.116)	(0.102)	(0.101)	(0.116)
[99,99.5] × GDP growth	-0.732**	-0.702**	-0.700**	-0.714**	-0.725**	-0.709**	-0.764***	-0.734**
	(0.288)	(0.288)	(0.288)	(0.328)	(0.328)	(0.288)	(0.288)	(0.328)
[99.5,100] × GDP growth	-1.608***	-1.574***	-1.571***	-1.750***	-1.745***	-1.581***	-1.643***	-1.792***
	(0.253)	(0.253)	(0.253)	(0.317)	(0.316)	(0.253)	(0.253)	(0.316)
Const.I × GDP growth		0.088*						
		(0.053)						
Const.II × GDP growth			0.643***					
			(0.135)					
Const.III × GDP growth				0.148**				
				(0.067)				
Const.IV × GDP growth					1.516***			
					(0.249)			
Const.V × GDP growth						0.110***		
						(0.047)		
Const.VI × GDP growth							-0.164***	
							(0.024)	
Const.VII × GDP growth								-0.202***
								(0.030)
Observations	1,326,793	1,326,793	1,326,793	1,161,594	1,161,594	1,326,793	1,325,440	1,160,600
R-squared	0.017	0.017	0.018	0.017	0.020	0.017	0.017	0.017
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of turnover with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.7 Cyclicality in turnover conditional on size bins and measures of financial
constraints including firm fixed effects

	Turnover growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.512***	2.441***	2.440***	2.450***	2.470***	2.450***	2.522***	2.508***
	(0.023)	(0.027)	(0.023)	(0.028)	(0.025)	(0.030)	(0.023)	(0.025)
[90,99] × GDP growth	0.045	0.077	0.062	0.110	0.101	0.060	0.031	0.085
	(0.103)	(0.104)	(0.103)	(0.117)	(0.116)	(0.104)	(0.103)	(0.116)
[99,99.5] × GDP growth	-0.788***	-0.756**	-0.772***	-0.683**	-0.688**	-0.764***	-0.811***	-0.703**
	(0.296)	(0.296)	(0.296)	(0.339)	(0.338)	(0.296)	(0.296)	(0.339)
[99.5,100] × GDP growth	-1.493***	-1.456***	-1.474***	-1.744***	-1.739***	-1.468***	-1.516***	-1.779***
	(0.260)	(0.261)	(0.260)	(0.320)	(0.319)	(0.261)	(0.260)	(0.320)
Const.I × GDP growth		0.152***						
		(0.055)						
Const.II × GDP growth			0.363***					
			(0.139)					
Const.III × GDP growth				0.178**				
				(0.070)				
Const.IV × GDP growth					1.774***			
					(0.267)			
Const.V × GDP growth						0.100***		
						(0.049)		
Const.VI × GDP growth							-0.089***	
							(0.024)	
Const.VII × GDP growth								-0.143***
								(0.031)
Observations	1,324,676	1,324,676	1,324,676	1,158,814	1,158,814	1,324,676	1,323,310	1,157,805
R-squared	0.150	0.151	0.151	0.160	0.161	0.150	0.151	0.160
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of turnover with respect to GDP. The first line is based on the
equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity,
and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2.
All specifications contain a constant term and non-interacted indicators. Standard errors are reported in
parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.8 Cyclicality in turnover conditional on size bins and measures of financial constraints excluding firms that have zero potential credit in all periods

	Turnover growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.512***	2.441***	2.440***	2.450***	2.470***	2.450***	2.522***	2.508***
	(0.023)	(0.027)	(0.023)	(0.028)	(0.025)	(0.030)	(0.023)	(0.025)
[90,99] × GDP growth	0.076	0.081	0.113	0.035	0.035	0.075	0.079	0.025
	(0.106)	(0.106)	(0.106)	(0.120)	(0.120)	(0.106)	(0.106)	(0.120)
[99,99.5] × GDP growth	-0.734**	-0.738**	-0.693**	-0.726**	-0.722**	-0.725**	-0.736**	-0.722**
	(0.300)	(0.300)	(0.300)	(0.343)	(0.343)	(0.300)	(0.300)	(0.343)
[99.5,100] × GDP growth	-1.378***	-1.378***	-1.327***	-1.617***	-1.598***	-1.365***	-1.377***	-1.631***
	(0.283)	(0.283)	(0.283)	(0.345)	(0.344)	(0.283)	(0.283)	(0.344)
Const.I × GDP growth		-0.005						
		(0.066)						
Const.II × GDP growth			0.611***					
			(0.135)					
Const.III × GDP growth				0.127				
				(0.086)				
Const.IV × GDP growth					1.383***			
					(0.327)			
Const.V × GDP growth						0.046		
						(0.052)		
Const.VI × GDP growth							-0.051*	
							(0.028)	
Const.VII × GDP growth								-0.164***
								(0.032)
Observations	1,167,541	1,167,541	1,167,541	1,027,839	1,027,839	1,167,541	1,166,458	1,027,039
R-squared	0.031	0.031	0.032	0.031	0.033	0.031	0.031	0.031
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of turnover with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.9 Cyclicality in turnover conditional on size bins and measures of financial constraints including bank controls

	Turnover growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP growth	2.512***	2.441***	2.440***	2.450***	2.470***	2.450***	2.522***	2.508***
	(0.023)	(0.027)	(0.023)	(0.028)	(0.025)	(0.030)	(0.023)	(0.025)
[90,99] × GDP growth	0.095	0.092	0.098	0.118	0.083	0.088	0.064	0.070
	(0.108)	(0.107)	(0.122)	(0.107)	(0.121)	(0.108)	(0.107)	(0.121)
[99,99.5] × GDP growth	-0.685**	-0.688**	-0.744**	-0.675**	-0.758**	-0.677**	-0.723**	-0.764**
	(0.302)	(0.302)	(0.344)	(0.301)	(0.343)	(0.302)	(0.302)	(0.344)
[99.5,100] × GDP growth	-1.355***	-1.365***	-1.572***	-1.328***	-1.582***	-1.356***	-1.403***	-1.621***
	(0.295)	(0.294)	(0.363)	(0.294)	(0.361)	(0.295)	(0.294)	(0.362)
Const.I × GDP growth		0.140**						
		(0.059)						
Const.II × GDP growth			0.540***					
			(0.144)					
Const.III × GDP growth				0.195***				
				(0.075)				
Const.IV × GDP growth					1.274***			
					(0.263)			
Const.V × GDP growth						0.173***		
						(0.050)		
Const.VI × GDP growth							-0.055**	
							(0.026)	
Const.VII × GDP growth								-0.182***
								(0.032)
Observations	1,241,471	1,241,471	1,091,547	1,241,471	1,091,547	1,241,471	1,240,272	1,090,662
R-squared	0.031	0.031	0.030	0.038	0.033	0.030	0.030	0.030
Bank Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE × GDP growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

Notes. Estimates report the semi-elasticity of turnover with respect to GDP. The first line is based on the equivalent regression without interaction fixed effects, as it is otherwise dropped due to multicollinearity, and included as a benchmark. Constrained measures are constructed as documented in Appendix B.2. All specifications contain a constant term and non-interacted indicators. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.10 Linear probability model results

	Constrained binary			
	(1)	(2)	(3)	(4)
Age	-0.05***			
	(0.000)			
Total assets		-0.05***		
		(0.001)		
Leverage			0.03***	
			(0.000)	
Liquidity ratio				0.01***
				(0.000)
Constant	0.36***	0.36***	0.36***	0.36***
Observations	1,765,288	1,765,288	1,764,947	1,764,947
R-squared	0.011	0.000	0.000	0.000

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table B.11 Calibration fit for the 1 shock model with no restrictions

Moment	Data	Model transitory shock
Percentage of constrained firms	0.23	0.20
Share of constrained firms in bottom 20%	0.33	0.85
Size of 90th-percentile vs. median	9.44	9.08
Size of 90-th percentile vs. bottom 20%	30.24	42.65
Size of constrained firms 90th-percentile vs. median	7.35	2.20
Size of unconstrained firms 90th-percentile vs. median	9.67	5.19
Asset share of constrained firms	0.12	0.01
Share of constrained firms in top 10% vs. bottom 20%	0.36	0
Percentage of constrained firms in top 1%	0.09	0

Notes. All moment conditions were equally weighted when minimizing the percentage deviation from the empirical target values.

B.4   Additional Figures

B.4.1   Descriptives
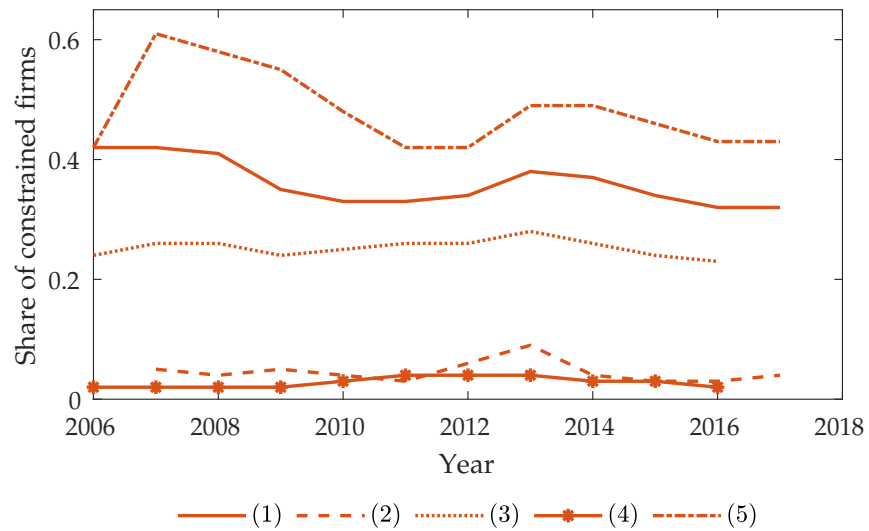
Fig. B.1 Share of constrained firms over time. Measures 1 to 5 as defined in Section 2.2.1.
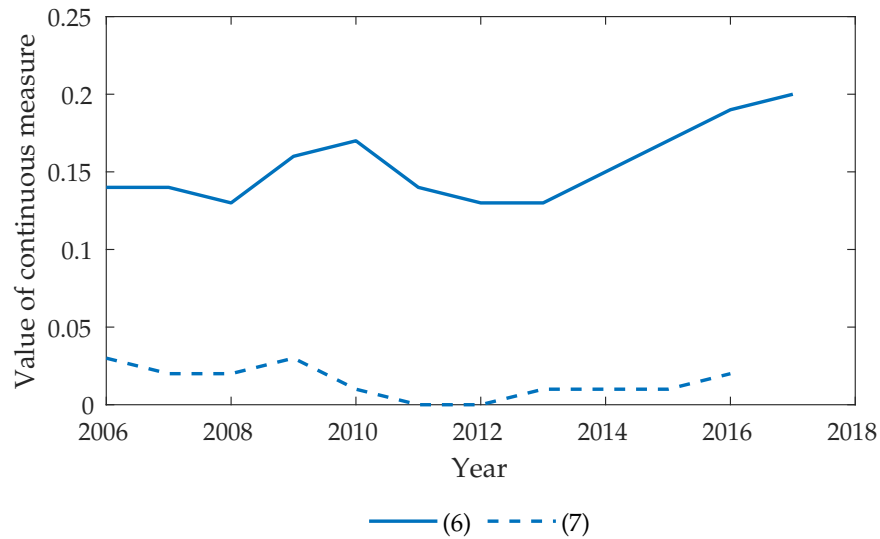
Fig. B.2 Share of constrained firms over time. Measures 6 and 7 as defined in Section 2.2.1.
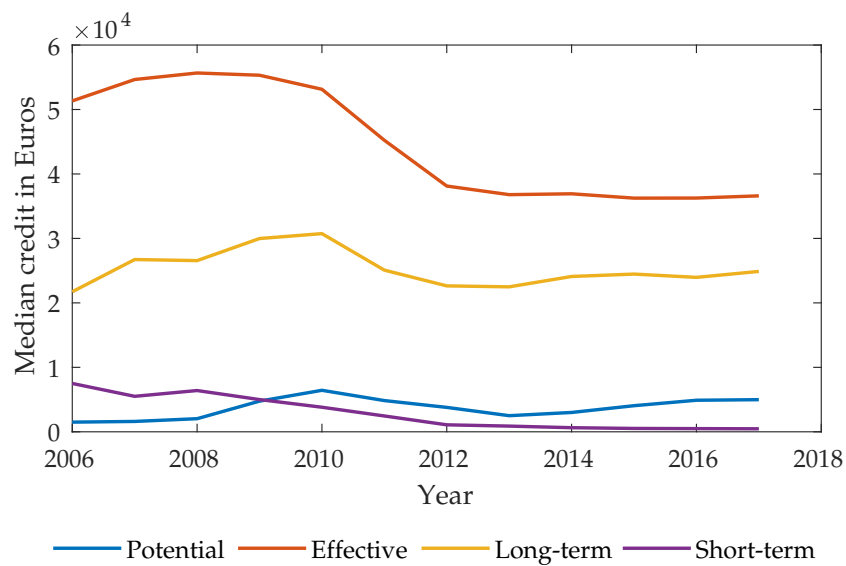
Fig. B.3 Median values for Potential, Effective, Long-term and Short-term credit over time.
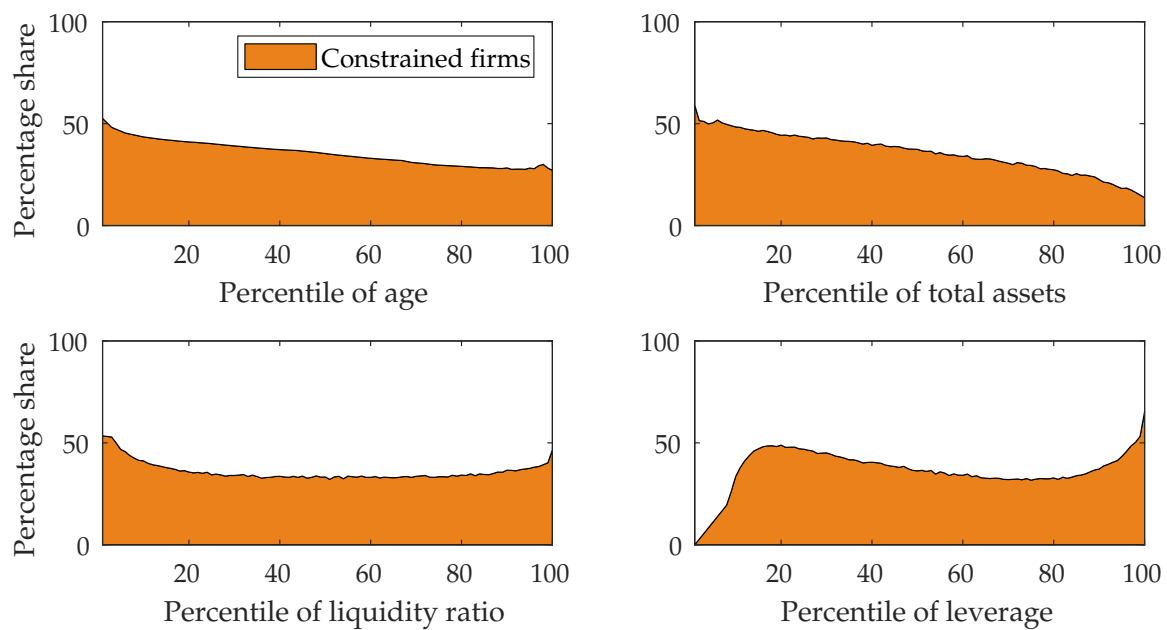
Fig. B.4 Decomposition of constrained and unconstrained firms across percentiles of firm variables using measure Constrained$_1$.
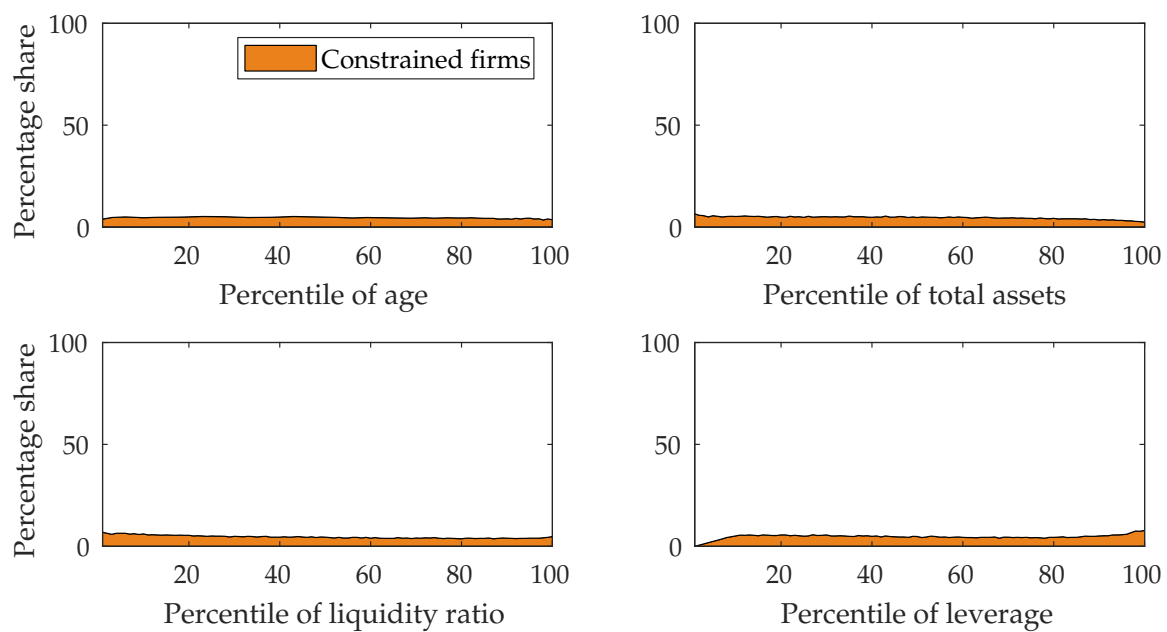
Fig. B.5 Decomposition of constrained and unconstrained firms across percentiles of firm variables using measure Constrained$_2$.
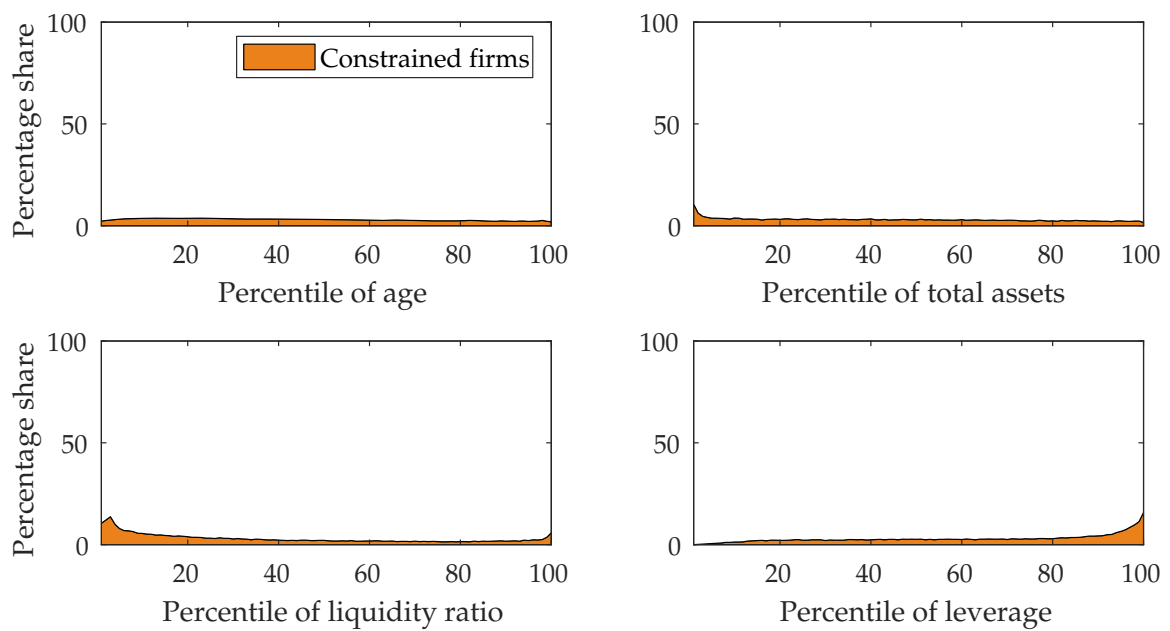
Fig. B.6 Decomposition of constrained and unconstrained firms across percentiles of firm variables using measure Constrained$_4$

B.4.2   Statistical Model
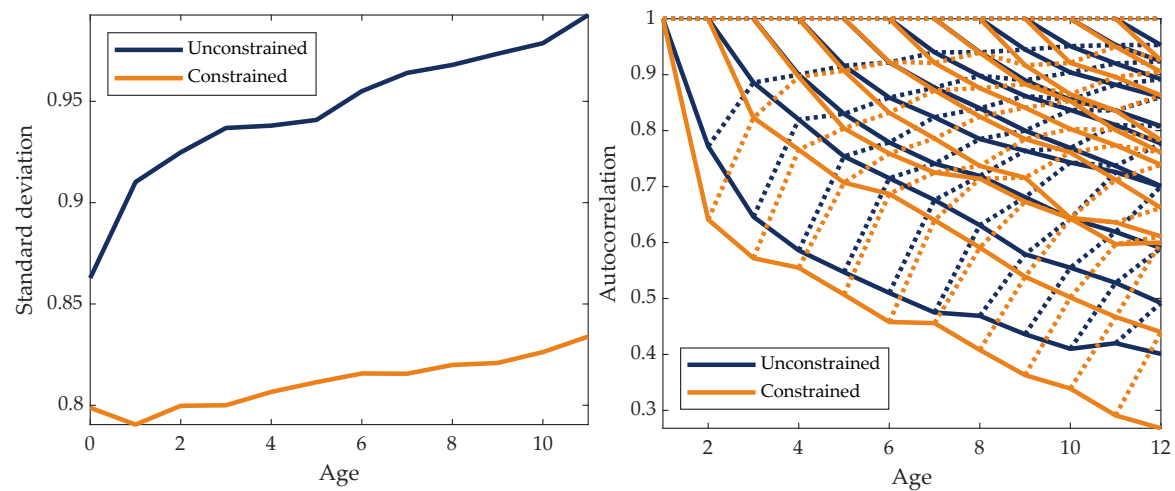
Fig. B.7 Conditional standard deviation and autocorrelation for log employment

Notes. The left panel presents the standard deviation of log employment by age, after controlling for sector and year fixed effects. The right panel presents the autocorrelation of log employment between ages a and $h \leq a$. Across lines h changes, while a changes along the lines. Orange lines stand for constrained firms, while blue lines represent unconstrained firms.
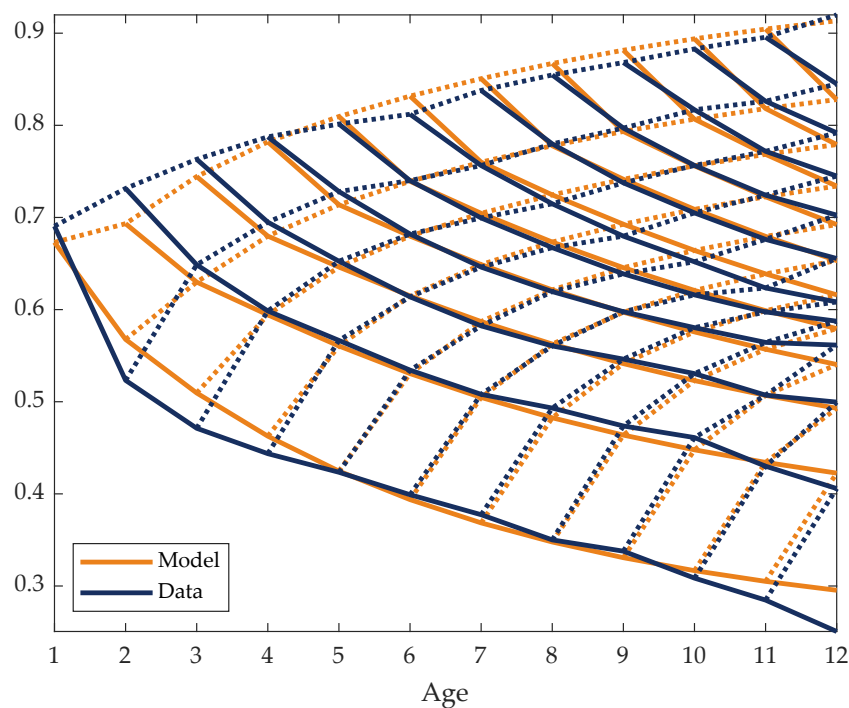
Fig. B.8 Model fit of statistical model for employment process
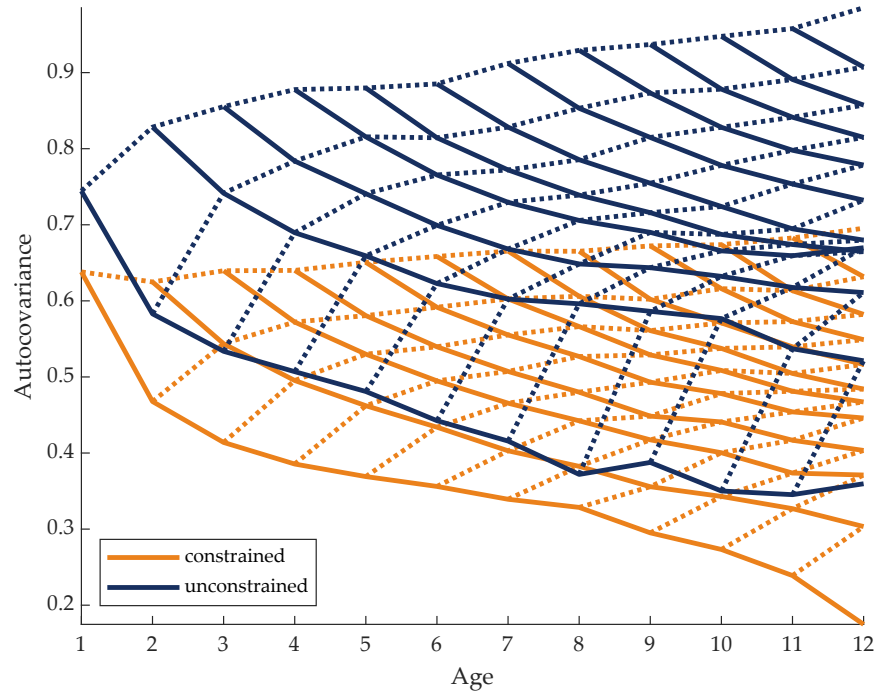
(a) Empirical conditional autocovariances
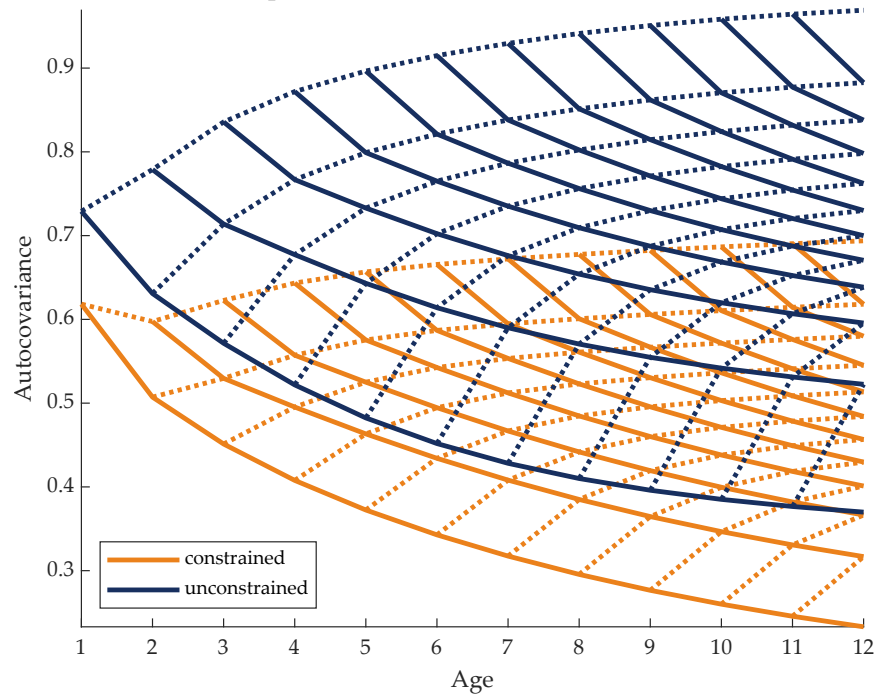
(b) Fitted conditional autocovariances

Fig. B.9 Empirical and model autocovariance for constrained firms (orange) and unconstrained firms (blue) using the measure $Const.I$.

Table B.12 Static model parameters for constrained and unconstrained firms using measure $Const.IV$.

	ρ_u	ρ_v	ρ_w	σ_θ	σ_u	σ_v	σ_ϵ	σ_z
Constrained	0.519	0.873	0.915	0.241	0.501	0.669	0.261	0.202
Unconstrained	0.432	0.765	0.883	0.399	0.755	0.743	0.311	0.175
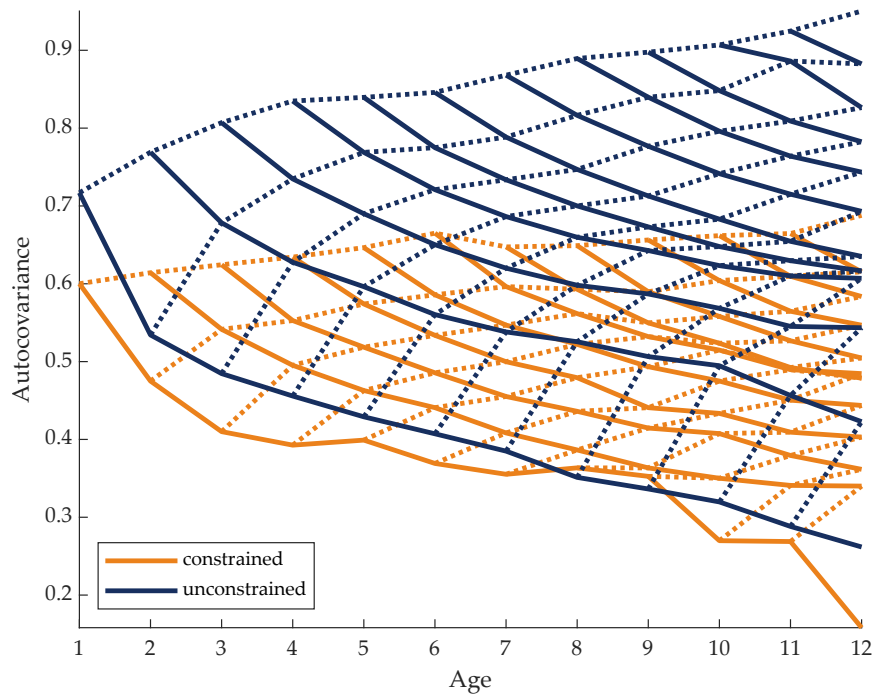
(a) Empirical conditional autocovariances
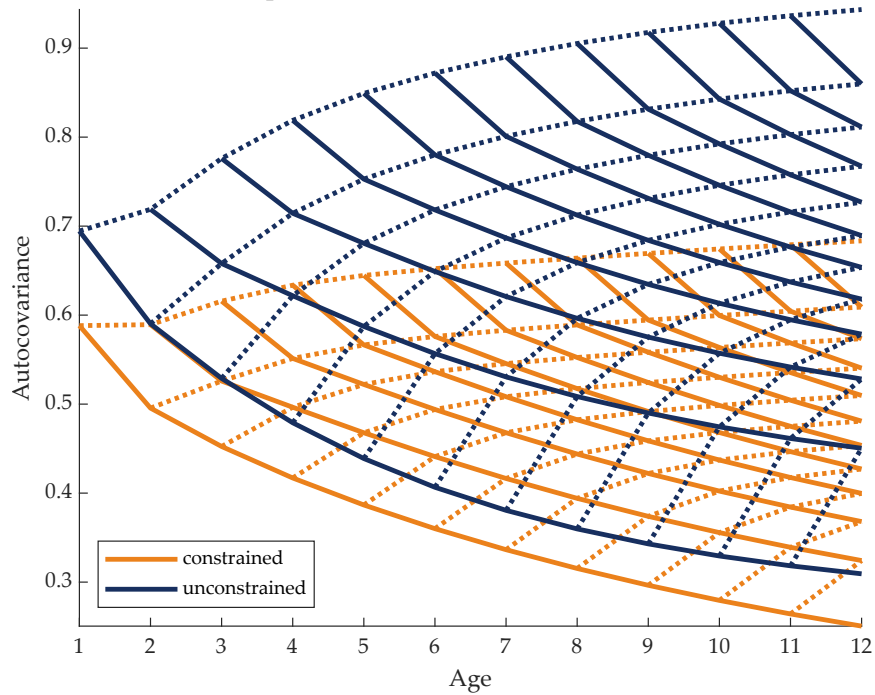
(b) Fitted conditional autocovariances

Fig. B.10 Empirical and model autocovariance for constrained firms (orange) and unconstrained firms (blue) using the measure $Const.IV$.
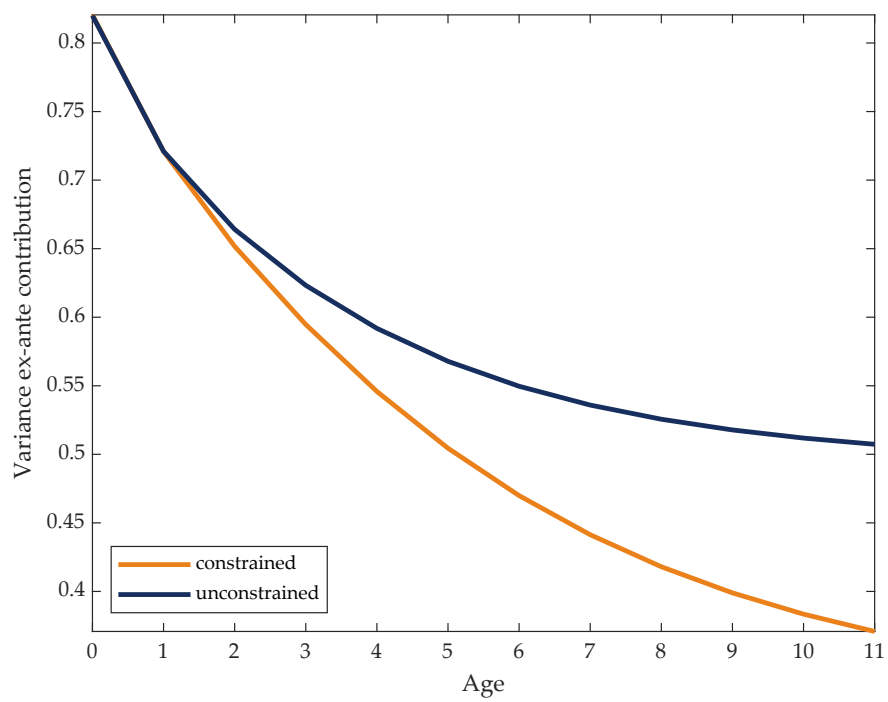
Fig. B.11 Model implied contribution of the ex-ante component to the variance of employment for different ages. Constrained firms are identified using measure $Const.IV$.

B.4.3 Theoretical model
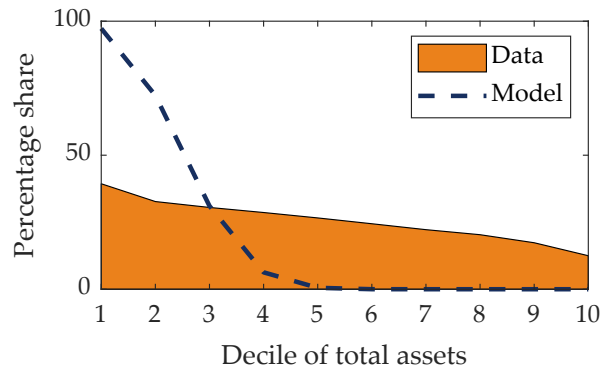
(a) Transitory shock

Fig. B.12 Share of constrained firms across the distribution in the 1 shock model with calibration in Table B.11.
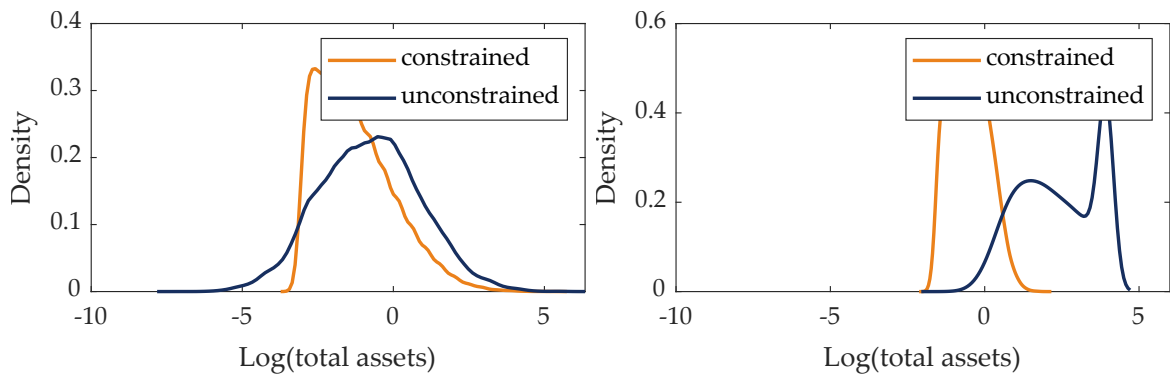
(a) Transitory + permanent shock

(b) Transitory shock

Fig. B.13 Conditional distributions of log of total assets implied by the model

(a) Transitory + permanent shock   (b) Transitory shock
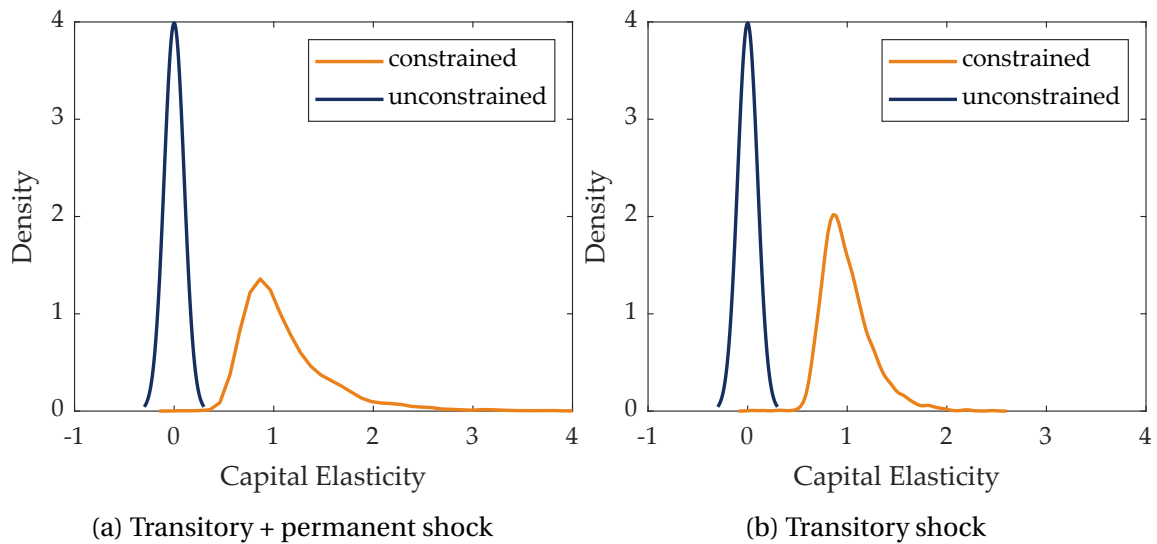
Fig. B.14 Conditional distribution of capital elasticity

(a) Transitory + permanent shock   (b) Transitory shock
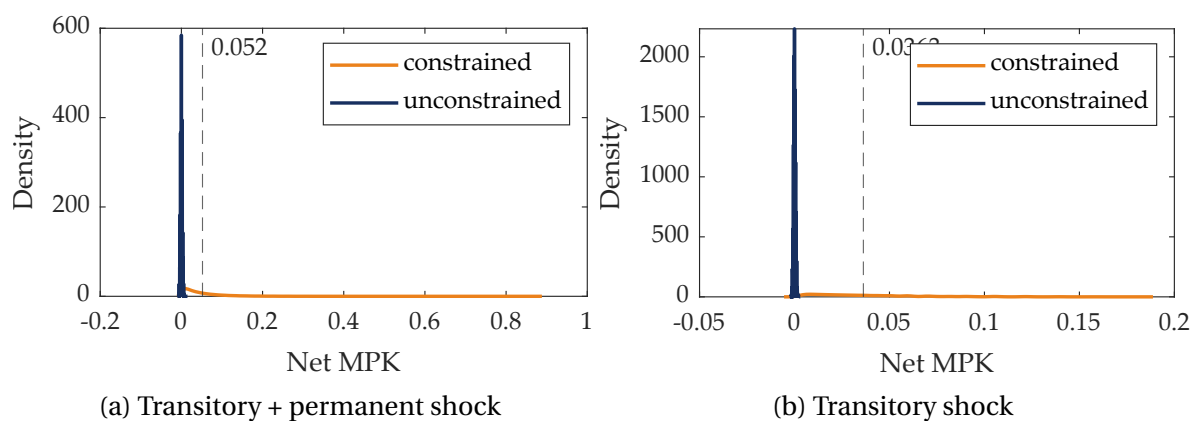
Fig. B.15 Conditional distributions of MPKs

Notes. The dashed line depicts the conditional mean of the marginal product of capital for constrained firms.

(a) Transitory + permanent shock   (b) Transitory shock
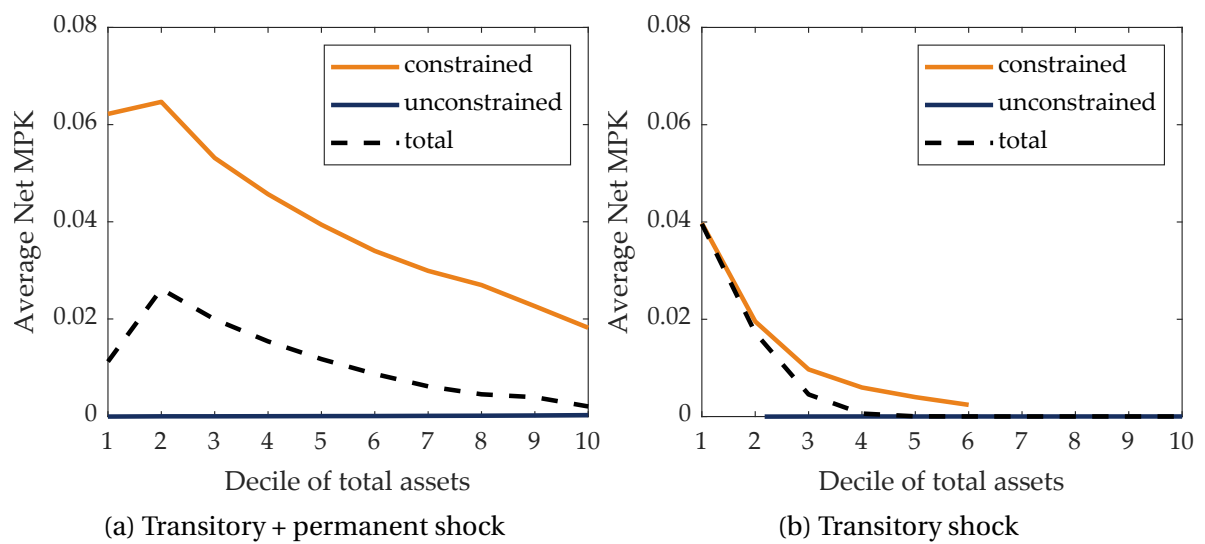
Fig. B.16 Average MPK along total assets distribution

Notes. The line depicts the average MPK of constrained firms per decile bin of total assets along the entire size distribution.

B.5   Simple model - results

Take a very simple model to analyse the impact of heterogeneous productivities on cyclicality. Firms can only invest in physical capital, have permanent productivity and face no uncertainty, except for a stochastic death shock. The problem can be written as:

$$V(k_{t,i}, b_{t,i}, \theta_i) = \pi_d x_{t,i} + (1 - \pi_d)\left(x_{t,i} - k_{t+1,i} + q_t b_{t+1,i} + \beta V(k_{t+1,i}, b_{t+1,i}, \theta_i)\right)$$

subject to

$$x_{t,i} = z_t \theta_i k_{t,i}^\alpha + (1 - \delta) k_{t,i} - b_{t,i}$$

$$\xi x_{t,i} \geq b_{t+1,i}$$

$$k_{t+1,i} \leq x_{t,i} + q_t b_{t+1,i}$$

B.5.1   Unconstrained firms

Steady state growth.   Unconstrained firms optimal capital $k_{t+1,i}^*$ is the solution to:

$$\beta^{-1} = (1 - \delta) + \alpha z_t \theta_i k_{t+1,i}^{*1-\alpha}$$

Hence optimal capital $k_{t+1,i}^*$ is

$$k_{t+1,i}^* = \theta_i^{\frac{1}{1-\alpha}} \left(\frac{\alpha z_{t+1}}{\beta^{-1} - (1 - \delta)}\right)^{\frac{1}{1-\alpha}}$$

where we can choose $z := \left(\frac{\beta^{-1} - (1-\delta)}{\alpha}\right)$ such that, at steady state and for $\theta = 1$, we have that $k_{t+1,i}^* = 1$. In the absence of idiosyncratic shocks and constant total factor productivity z, unconstrained firms are not growing at steady state as they reached their optimal level of capital.

$$g_{uncons} = \frac{k_{t+1,i}^*(\theta_i)}{k_{t,i}^*(\theta_i)} = 1$$

Cyclicality.   Now consider the following setup; at time $t = -1$, $z_t = z$. At time $t = 0$, firms learn the future path of z_t, for $t \geq 0$ will be

$$z_t = z \exp(\rho^t \epsilon)$$

The growth rate then becomes

$$g_{uncons} = \frac{k^*_{t+1,i}(\theta_i)}{k^*_{t,i}(\theta_i)} = \frac{\exp(\frac{\rho}{1-\alpha}\epsilon)\theta_i^{1/(1-\alpha)}}{\theta_i^{1/(1-\alpha)}} = \exp\left(\frac{\rho}{1-\alpha}\epsilon\right)$$

Hence, the elasticity of capital is the same across all unconstrained firms, independent of firm size and firm-specific productivity.

$$\frac{\Delta g_{uncons}}{\Delta\epsilon}|_{\epsilon\approx 0} = \frac{\rho}{1-\alpha}$$

B.5.2 Constrained firms

Steady state growth. Constrained firms invest according to their maximum investment capacity which is capped by the net worth constraint.

$$\begin{aligned}
k_{t+1,i} &= n_{t,i} + q_t b_{t+1,i} \\
&= n_{t,i} + q_t \xi n_{t,i} \\
&= (1 + q_t\xi)(z_t\theta_i k_{t,i}^\alpha + (1-\delta)k_{t,i} - b_{t,i})
\end{aligned}$$

Hence,

$$\begin{aligned}
g_{cons} &= (1 + q_t\xi)(z_t\theta_i k_{t,1}^{\alpha-1} + (1-\delta) - b_{t,i}/k_{t,i}) \\
&= (1 + q_t\xi)(z_t\theta_i k_{t,1}^{\alpha-1} + (1-\delta) - \frac{\xi}{1 + q_{t-1}\xi})
\end{aligned}$$

Due to decreasing returns to scale, the growth rate is affected by the size of the firm, with larger firms growing slower

$$\frac{\Delta g_{cons}}{\Delta k_{t,i}} = (1 + q_t\xi)(\alpha - 1)z_t\theta_i k_{t,1}^{\alpha-2} < 0$$

For firms of the same size, those with a higher permanent productivity component grow quicker

$$\frac{\Delta g_{cons}}{\Delta\theta_i} = (1 + q_t\xi)z_t k_{t,1}^{\alpha-1} > 0$$

Cyclicality   Now consider the same setup as for unconstrained firms; at time $t = -1$, $z_t = z$. At time $t = 0$, firms learn the future path of z_t, for $t \geq 0$ will be

$$z_t = z \exp(\rho^t \epsilon)$$

The growth rate on impact then becomes

$$g_{cons} = (1 + q_t \xi)(z \exp(\rho^0 \epsilon) \theta_i k_{t,1}^{\alpha-1} + (1 - \delta) - \frac{\xi}{1 + q_{t-1}\xi})$$

So, the elasticity of capital with respect to the shock ϵ is decreasing on capital and increasing on the productivity of the firm

$$\frac{\Delta g_{cons}}{\Delta \epsilon}|_{\epsilon \approx 0} = (1 + q_t \xi)(z \theta_i k_{t,1}^{\alpha-1}) = \frac{(1 + q_t \xi)}{\alpha} mpk_i$$

With the derivative of the elasticity with respect to the size and productivity of the firm being negative and positive respectively

$$\frac{\Delta^2 g_{cons}}{\Delta \epsilon \Delta \theta_i}|_{\epsilon \approx 0} = (1 + q_t \xi)(z k_{t,1}^{\alpha-1}) > 0$$

$$\frac{\Delta^2 g_{cons}}{\Delta \epsilon \Delta k_{t,1}}|_{\epsilon \approx 0} = (\alpha - 1)(1 + q_t \xi)(z \theta_i k_{t,1}^{\alpha-2}) < 0$$

When is the elasticity of constrained larger than unconstrained?

$$\frac{\Delta g_{cons}}{\Delta \epsilon}|_{\epsilon \approx 0} > \frac{\Delta g_{uncons}}{\Delta \epsilon}|_{\epsilon \approx 0}$$

This happens when the marginal product of capital of constrained firms is above a given threshold

$$mpk > \rho \frac{\alpha}{1 - \alpha} \frac{1}{1 + q_t \xi}$$

So, two factors will determine which elasticity is larger: (i) the marginal product of capital of constrained firms, which depends on the distribution in terms of both size and productivity. The smaller and the more productive constrained firms are, the higher their elasticity; (ii) the persistence of the aggregate shock. As ρ approaches zero, unconstrained firms will not react to the shock, while the elasticity of constrained firms on impact does not depend on the persistence of the shock.

B.6   Statistical Model Derivation

Write stochastic processes in MA representation:

$$u_{i,t} = \rho_u^{t+1} u_{i,-1} + \sum_{k=0}^{a} \rho_u^k \theta_i$$

$$v_{i,a} = \rho_v^{a+1} v_{i,-1}$$

$$w_{i,a} = \sum_{k=0}^{a} \rho_w^k \varepsilon_{i,a-k} = \sum_{k=0}^{j-1} \rho^k \varepsilon_{i,a-k} + \rho_v^j \sum_{k=0}^{a-j} \rho_v^k \varepsilon_{i,a-j-k} \quad 0 \le j \le a$$

So the level of log employment of firm i at age a is:

$$\ln n_{i,a} = \rho_u^{a+1} u_{i,-1} + \sum \rho_u^k \theta_i + \rho_v^{a+1} v_{i,-1} + \sum_{i=1}^{j-1} \rho^k \varepsilon_{i,a-k} + \rho_v^j \sum_{i=1}^{a-j} \rho_v^k \varepsilon_{i,a-j-k} + z_{i,a}$$

Then the autocovariance of log employment at age a and $a-j$ for $j \ge 0$ is:

$$\begin{aligned}
\mathrm{Cov}\left[\log n_{i,a}, \log n_{i,a-j}\right] &= \left(\sum_{k=0}^{a} \rho_u^k\right) \sigma_\theta^2 \left(\sum_{k=0}^{a-j} \rho_u^k\right) + \rho_u^{a+1} \sigma_{\tilde{u}}^2 \rho_u^{a-j+1} + \rho_v^{a+1} \sigma_{\tilde{v}}^2 \rho_v^{a-j+1} \\
&\quad + \mathrm{Cov}\left[\rho_v^j \sum_{k=0}^{a-j} \rho_v^k \varepsilon_{i,a-j-k}, \sum_{k=0}^{a-j} \rho_v^k \varepsilon_{i,a-j-k}\right] + \mathbf{1}_{\{j=0\}} \sigma_z^2 \\
&= \sigma_\theta^2 \left(\sum_{k=0}^{a} \rho_u^k\right) \left(\sum_{k=0}^{a-j} \rho_u^k\right) + \sigma_{\tilde{u}}^2 \rho_u^{2(a+1)-j} + \sigma_{\tilde{v}}^2 \rho_v^{2(a+1)-j} + \sigma_\varepsilon^2 \rho_w^j \sum_{k=0}^{a-j} \rho_w^{2k} + \mathbf{1}_{\{j=0\}} \sigma_z^2
\end{aligned}$$

Appendix C

Appendix for Chapter 3

C.1 Model derivations

C.1.1 Aggregation and market clearing

Intermediate Production: The final good sector's demand for intermediate production can be found by aggregating equation (8) over all intermediate goods

$$Y^I(t) = \int_{i=0}^{1} \left(\frac{p_i(t)}{P(t)} \right)^{-\sigma} \left(\frac{\sigma - 1}{\sigma} A(t) \right)^{\sigma} L(t) q^{(\sigma-1)k_i} \, di \tag{C.1}$$

$$= \left(\frac{\sigma - 1}{\sigma} A(t) \right)^{\sigma} L(t) \psi^{-\sigma} Q(t). \tag{C.2}$$

Final output: Analogously final output is given by

$$Y(t) = A(t) L(t)^{1/\sigma} \int_{i=0}^{1} \left(q^{k_i(t)} y_i(t) \right)^{\frac{\sigma-1}{\sigma}} \, di \tag{C.3}$$

$$= A(t) L(t)^{1/\sigma} \int_{i=0}^{1} \left(q^{k_i(t)} \left(\frac{p_i(t)}{P(t)} \right)^{-\sigma} \left(\frac{\sigma - 1}{\sigma} A(t) \right)^{\sigma} L(t) q^{(\sigma-1)k_i} \right)^{\frac{\sigma-1}{\sigma}} \, di \tag{C.4}$$

$$= \left(\frac{\sigma - 1}{\sigma} A(t) \right)^{\sigma} L(t) \psi^{1-\sigma} Q(t). \tag{C.5}$$

Remark: Final output is larger than intermediate output as $\psi > 1$, hence $\psi^{1-\sigma} > \psi^{-\sigma}$ for $\sigma > 0$.

Research sector's demand for final good: We get this by rearranging the equation for

research success and integrating over all research sectors:

$$Z(t) = \int_0^1 \frac{\mu}{\phi(k_i(t))} di = \int_0^1 \frac{\mu}{\frac{1}{\lambda} q^{(1-\sigma)(k_i+1)}} di = \lambda \mu q^{(\sigma-1)} Q(t) \tag{C.6}$$

Goods market clearing:

$$Y(t) = C(t) + Y^I(t) + Z(t) + \frac{\theta}{2} (\pi(t) - 1)^2 Y^I(t) \tag{C.7}$$

Labour market clearing:

$$L^d(t) = L^s(t) \tag{C.8}$$

C.1.2   Balanced growth path of the endogenous growth model

In case of the endogenous growth model all growth is coming through growth in the quality index of products $Q(t)$ and we impose a constant total factor productivity A. Recall the definition of the quality index

$$Q(t) = \int_{i=0}^1 \left(q^{k_i}\right)^{\sigma-1} di. \tag{C.9}$$

In each sector i, the term $\left(q^{k_i}\right)^{\sigma-1}$ does not change if no innovation occurs but changes to $\left(q^{k_i+1}\right)^{\sigma-1}$ in the case of successful research. The probability of success is μ as defined above which is independent of time and rung on the quality ladder. Furthermore, we assume the probability of success to be independent across sectors. Hence, we can write the expected change in the quality index as follows

$$E[\Delta Q(t)] = \int_0^1 \mu \left((q^{k_i+1})^{\sigma-1} - (q^{k_i})^{\sigma-1}\right) di \tag{C.10}$$

$$= \mu(q^{\sigma-1} - 1) \int_0^1 \left(q^{k_i}\right)^{\sigma-1} di \tag{C.11}$$

$$= \mu(q^{\sigma-1} - 1)Q(t). \tag{C.12}$$

Hence, we find growth in the quality index as

$$g_Q = \frac{\Delta Q(t)}{Q(t)} = \mu(q^{\sigma-1} - 1), \tag{C.13}$$

where we use the law of large numbers that the expected change is the same as the actual change for the continuum of intermediate firms. Note that for $\sigma = 1$, the growth rate goes to zero as the production of the final good is independent of the intermediate good's quality. Analogously, if $q = 1$ there is no growth through quality improvements as the quality ladder collapses to one rung.

The balanced growth paths for the final good, the intermediate good and the investment in research in case of endogenous growth can then be written as

$$\bar{Y}(t) = \frac{Y(t)}{Q(t)} = \left(\frac{\sigma - 1}{\sigma} A\right)^{\sigma} L(t) \psi^{1-\sigma}, \tag{C.14}$$

$$\bar{Y}^I(t) = \frac{Y(t)^I}{Q(t)} = \left(\frac{\sigma - 1}{\sigma} A\right)^{\sigma} L(t) \psi^{-\sigma}, \tag{C.15}$$

$$\bar{Z}(t) = \frac{Z(t)}{Q(t)} = \lambda \mu q^{(\sigma - 1)}, \tag{C.16}$$

and from the goods market clearing condition we get that consumption grows at the same rate g_q and must equal the difference between total final good production $Y(t)$ and the sum of the demands for final goods by the intermediate and research sector as well as the Rotemberg adjustment cost.

$$\frac{C(t)}{Q(t)} = \frac{Y(t)}{Q(t)} - \frac{Y(t)^I}{Q(t)} - \frac{Z(t)}{Q(t)} - \frac{\theta}{2} (\pi(t) - 1)^2 \frac{Y(t)^I}{Q(t)}$$

$$\bar{C}(t) = \bar{Y}(t) - \bar{Y}^I(t) - \bar{Z}(t) - \frac{\theta}{2} (\pi(t) - 1)^2 \bar{Y}^I(t) \tag{C.17}$$

C.2   Balanced growth path of the exogenous growth model

An exogenous version of the growth model is nested when setting $Z(t) = 0$ and shutting down thereby shutting down the R&D channel [make investment productive yet not growth enhancing, like in Solow model]. Under this assumption the growth in the quality index $Q(t)$ is zero as the economy stays on the same quality rung. The only source of growth is then total factor productivity which is imposed exogenously. The growth rate in

the final good and intermediate good is as follows

$$g_Y = g_{Y^I} = \sigma g_A. \tag{C.18}$$

Hence, we can define the balanced growth paths for the final good and the intermediate inputs as

$$\bar{Y}_{ex}(t) = \frac{Y(t)}{A(t)^\sigma} = \left(\frac{\sigma - 1}{\sigma} \right)^\sigma L(t) \psi^{1-\sigma} Q, \tag{C.19}$$

$$\bar{Y}_{ex}^I(t) = \frac{Y^I(t)}{A(t)^\sigma} = \left(\frac{\sigma - 1}{\sigma} \right)^\sigma L(t) \psi^{-\sigma} Q, \tag{C.20}$$

where Q is fixed exogenously. When shutting down the investment channel, the goods market clearing condition simplifies to

$$\bar{C}_{ex}(t) = \bar{Y}_{ex}(t) - \bar{Y}_{ex}^I(t) - \frac{\theta}{2} (\pi(t) - 1)^2 \bar{Y}_{ex}^I(t). \tag{C.21}$$

C.3   Theoretical results
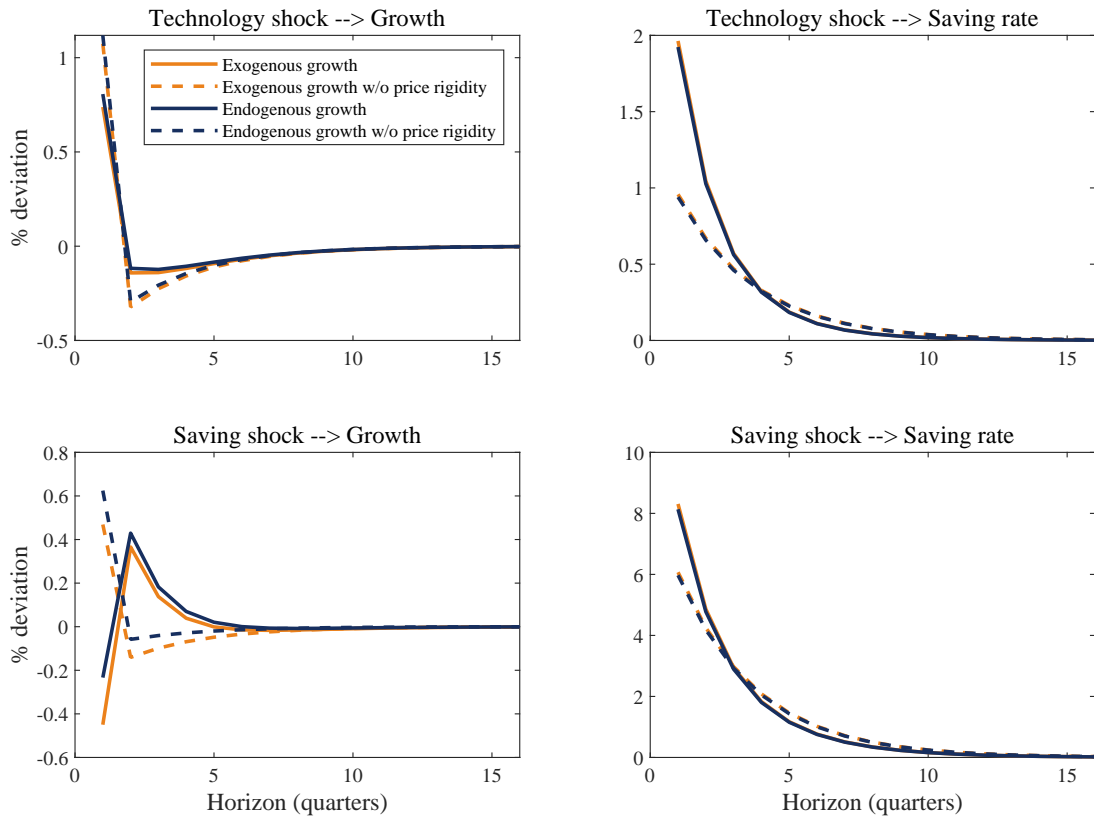
Fig. C.1 Impulse response functions for the exogenous and endogenous growth model presented in case of a temporary shock to the discount factor.
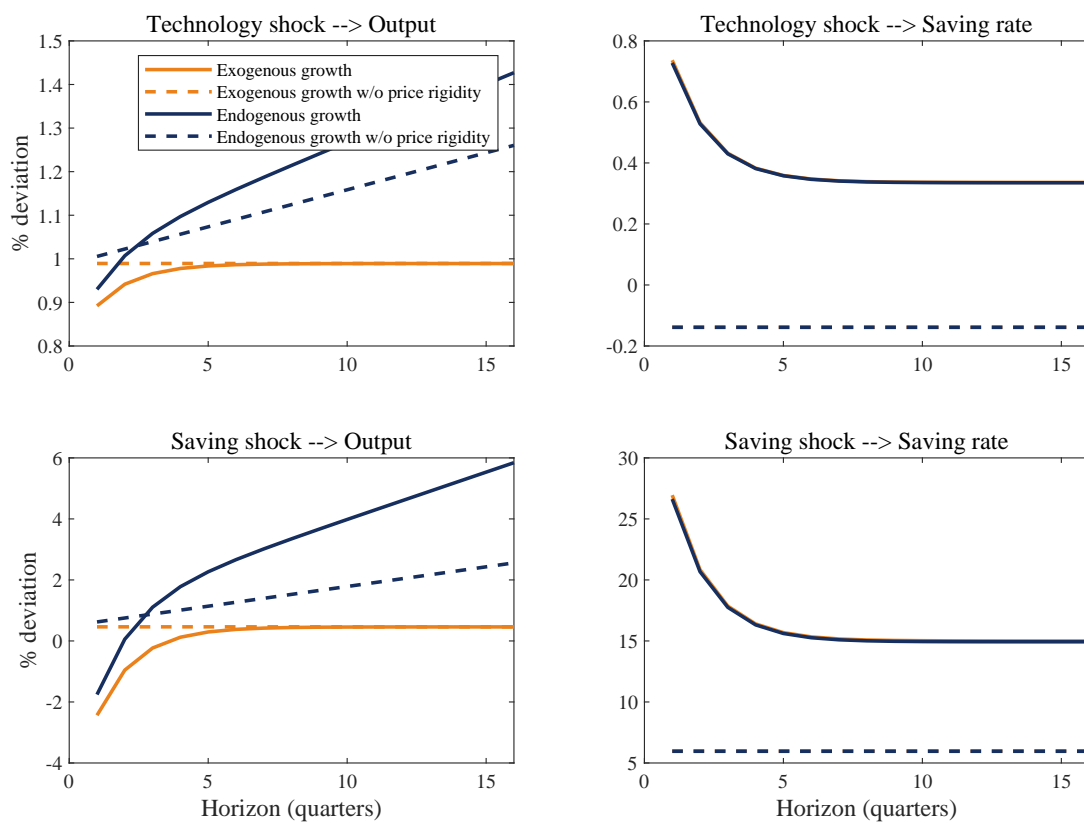
Fig. C.2 Impulse response functions for the exogenous and endogenous growth model presented in case of a permanent shock to the discount factor.
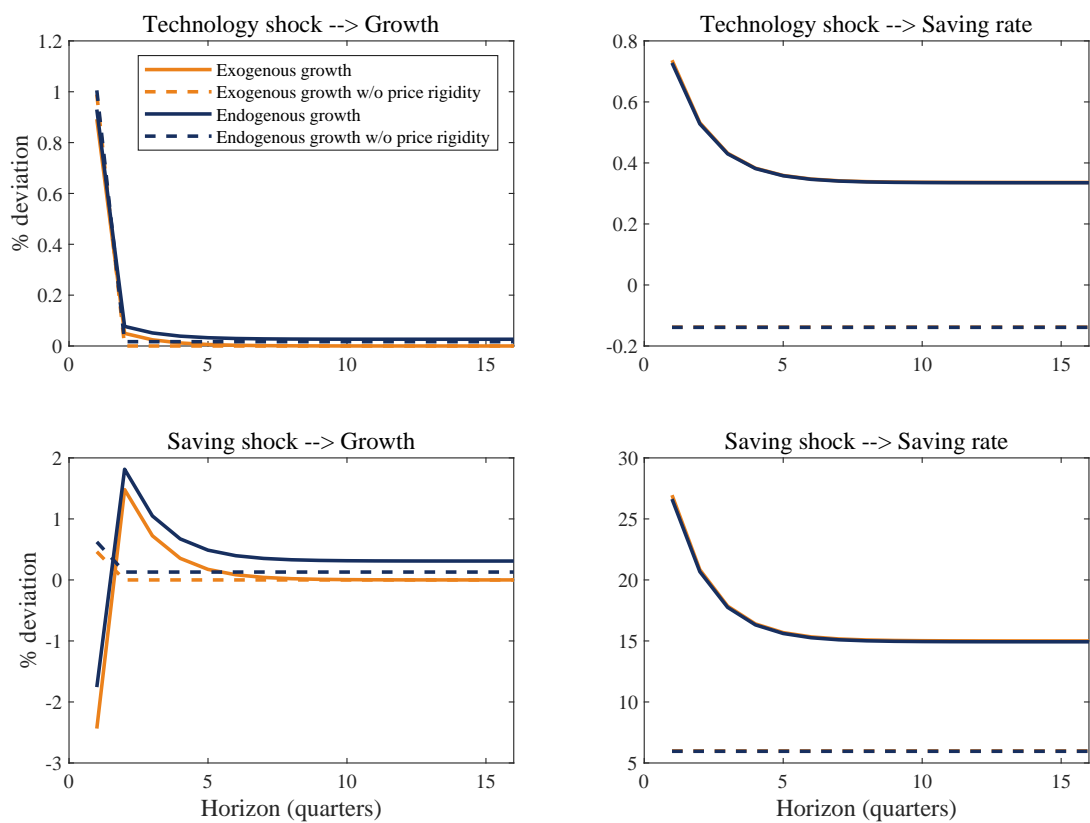
Fig. C.3 Impulse response functions for the exogenous and endogenous growth model presented in case of a permanent shock to the discount factor.
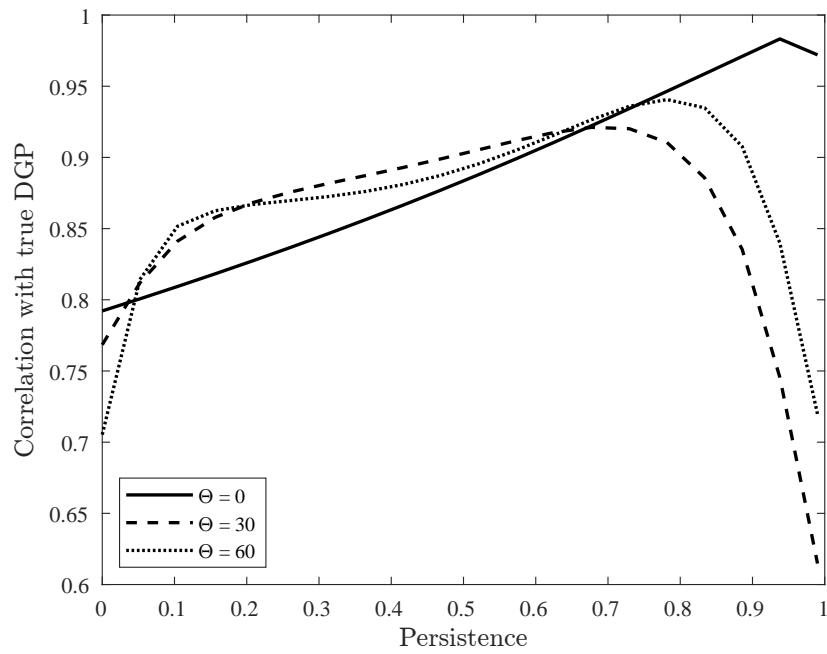
(a) TFP shock
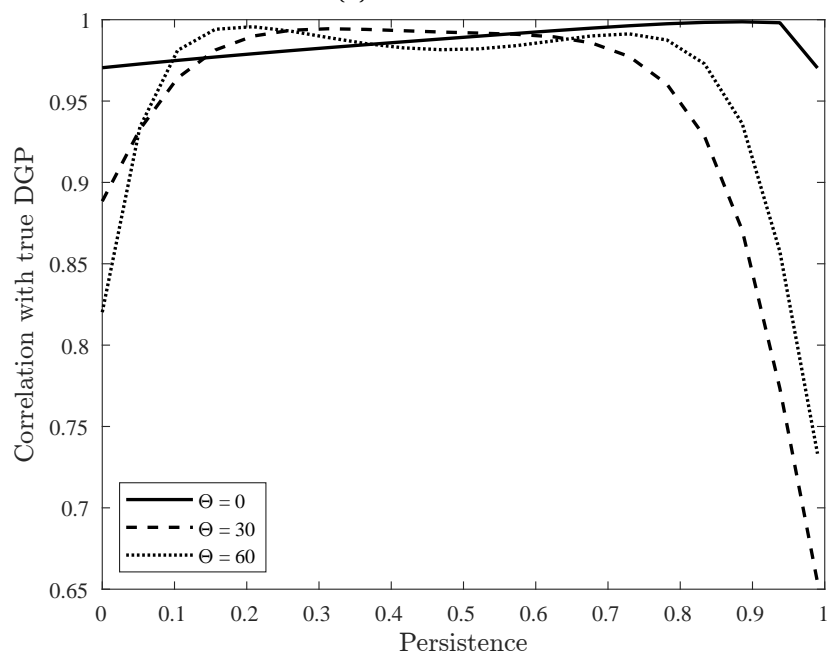
(b) Saving shock

Fig. C.4 Identification of structural shocks in exogenous growth model
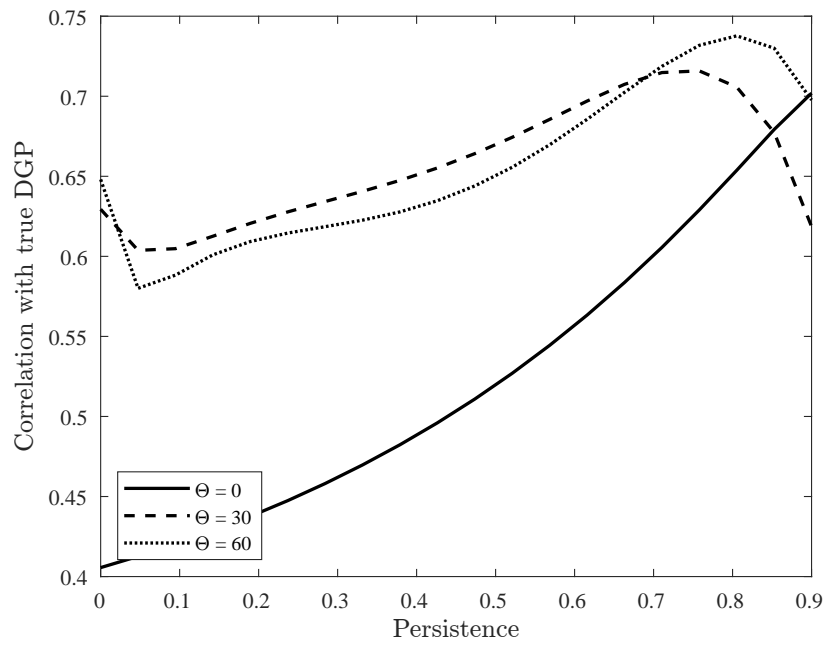
(a) TFP shock
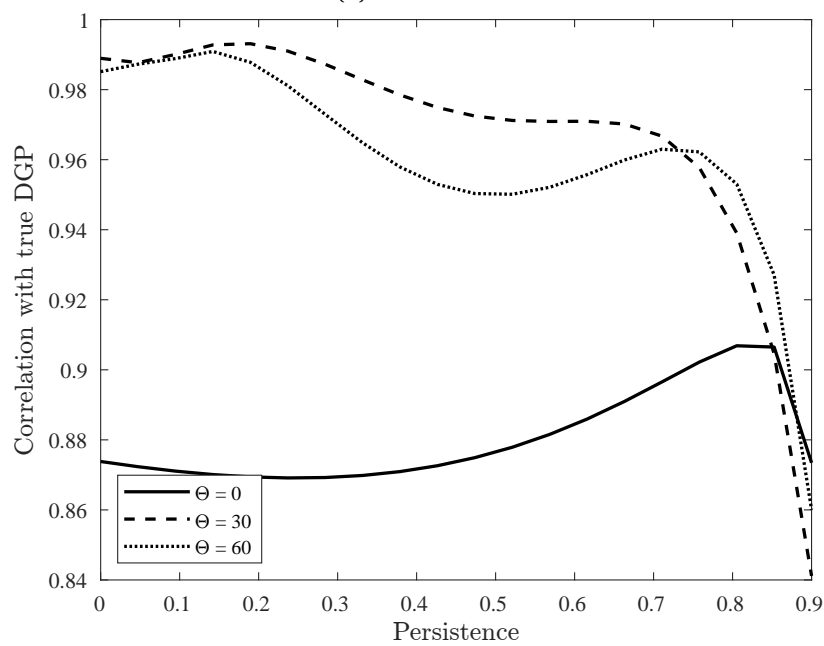
(b) Saving shock

Fig. C.5 Identification of structural shocks in endogenous growth model

C.4 Empirical results
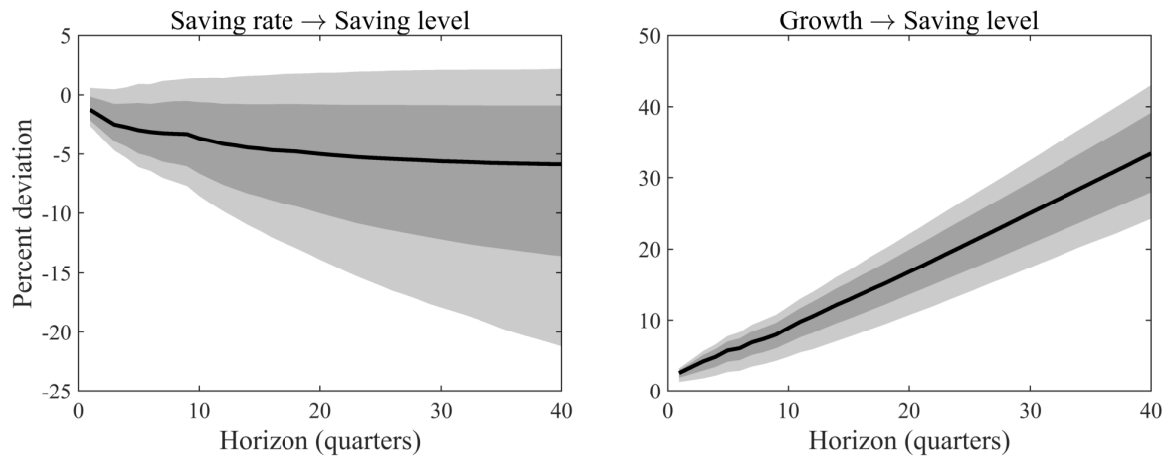
Fig. C.6 Impulse response function of the level of private savings according to the exclusion restriction 2 in Figure 4.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.
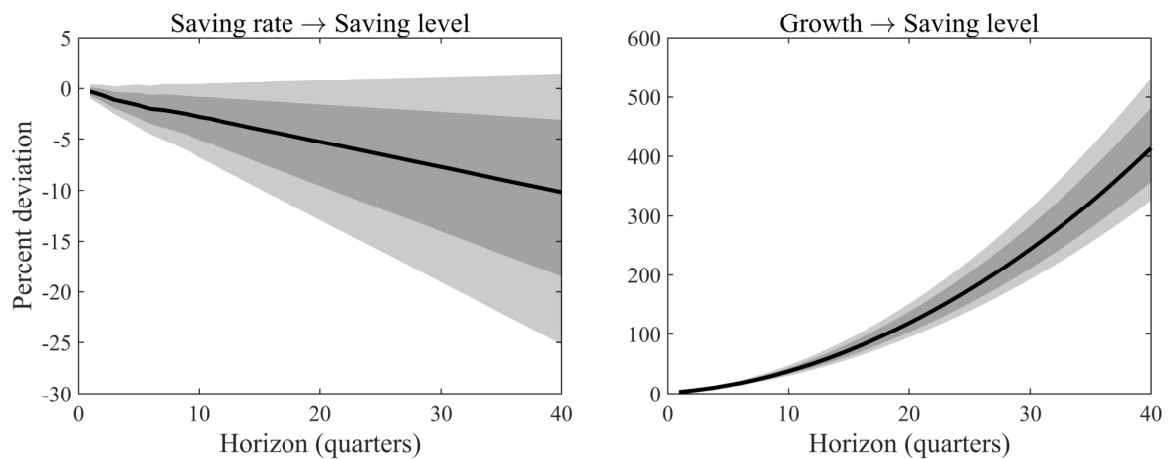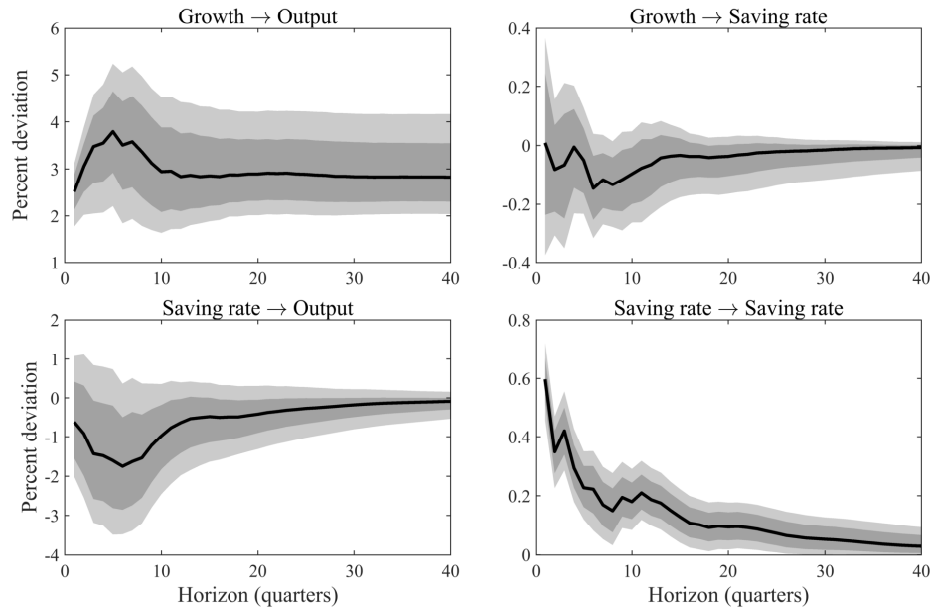

Fig. C.7 Impulse response function of the level of private savings according to the exclusion restriction 3 in Figure 5.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.
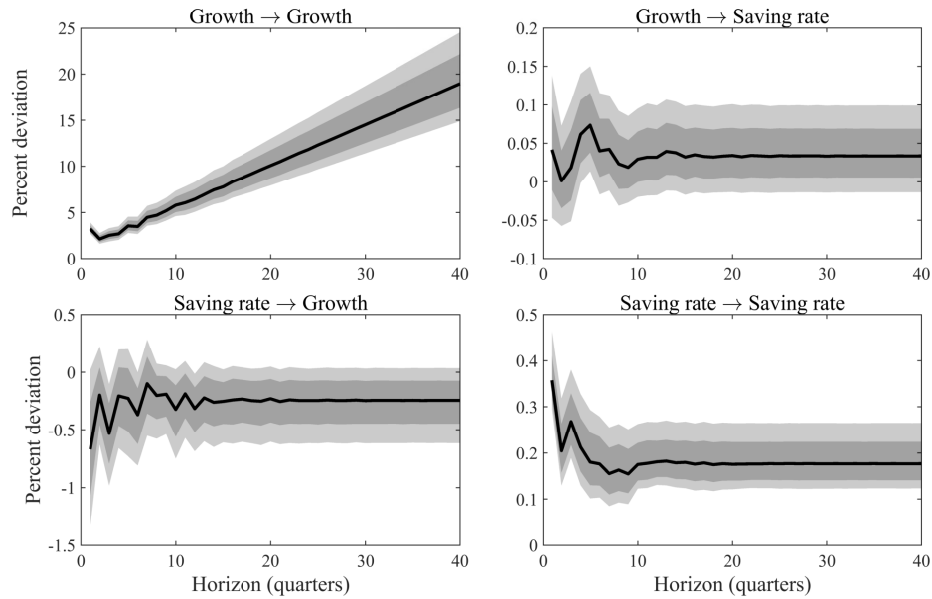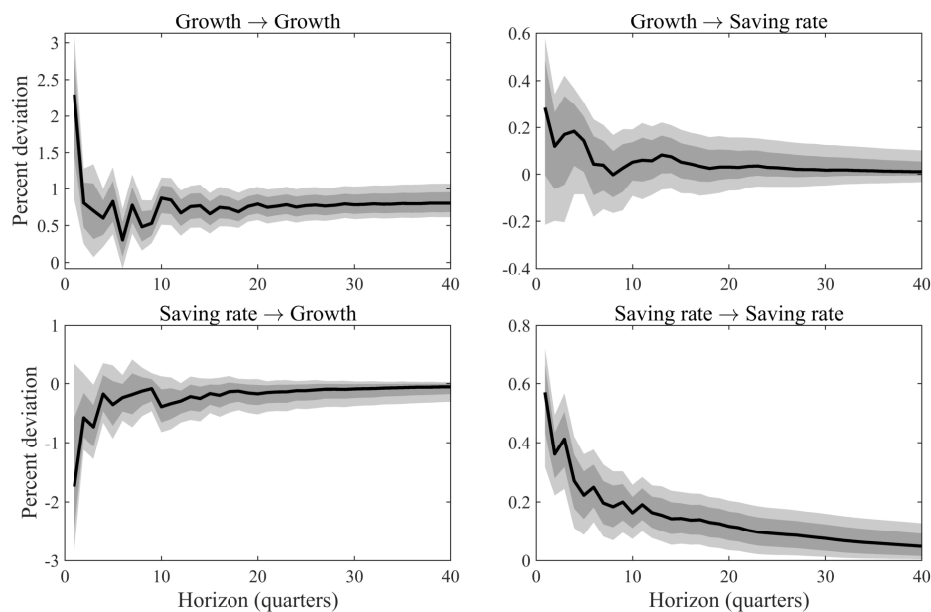
Fig. C.8 Impulse response functions using exclusion restriction 3 that a permanent shock to the saving rate has no long-run effect on the change in growth of output.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.
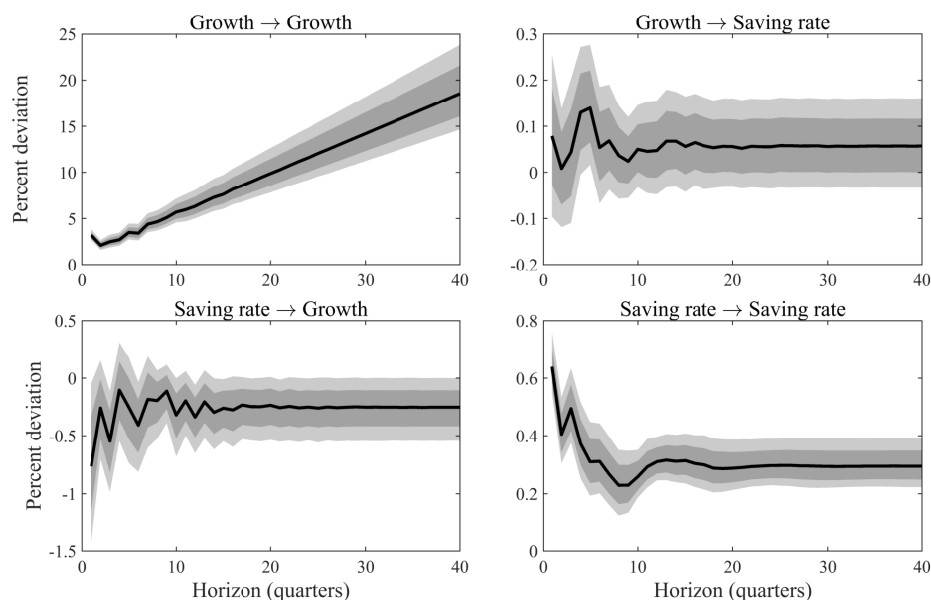

Fig. C.9 Impulse response functions for a temporary shock to the savings rate using exclusion restriction 1. The saving rate is defined as personal savings share of disposable income.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.

Fig. C.10 Impulse response functions for a temporary shock to the savings rate using exclusion restriction 2. The saving rate is defined as personal savings share of disposable income.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.
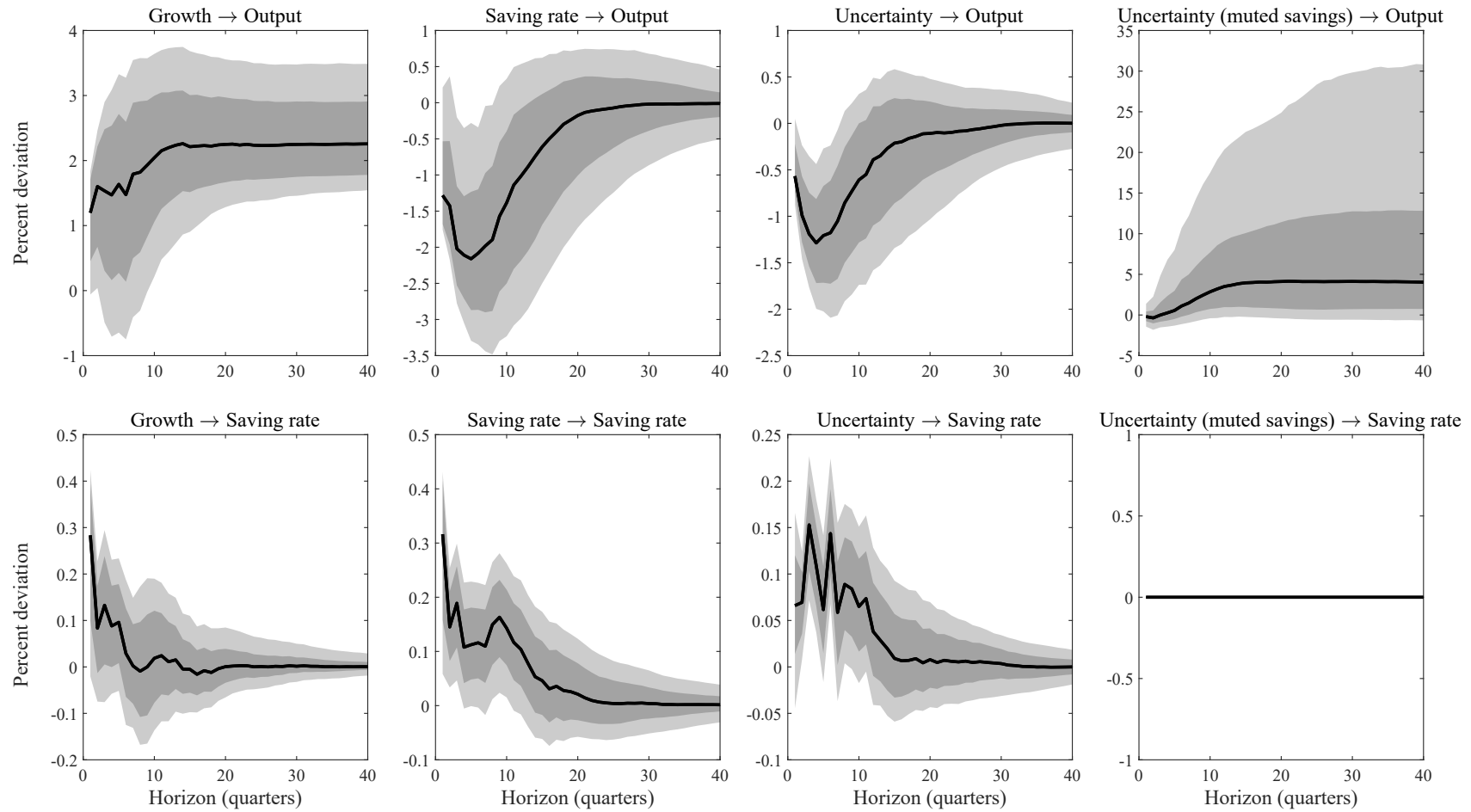

Fig. C.11 Impulse response functions for a permanent shock to the savings rate using exclusion restriction 3. The saving rate is defined as personal savings share of disposable income.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The grey transparent area shows the corresponding 68% confidence bounds.

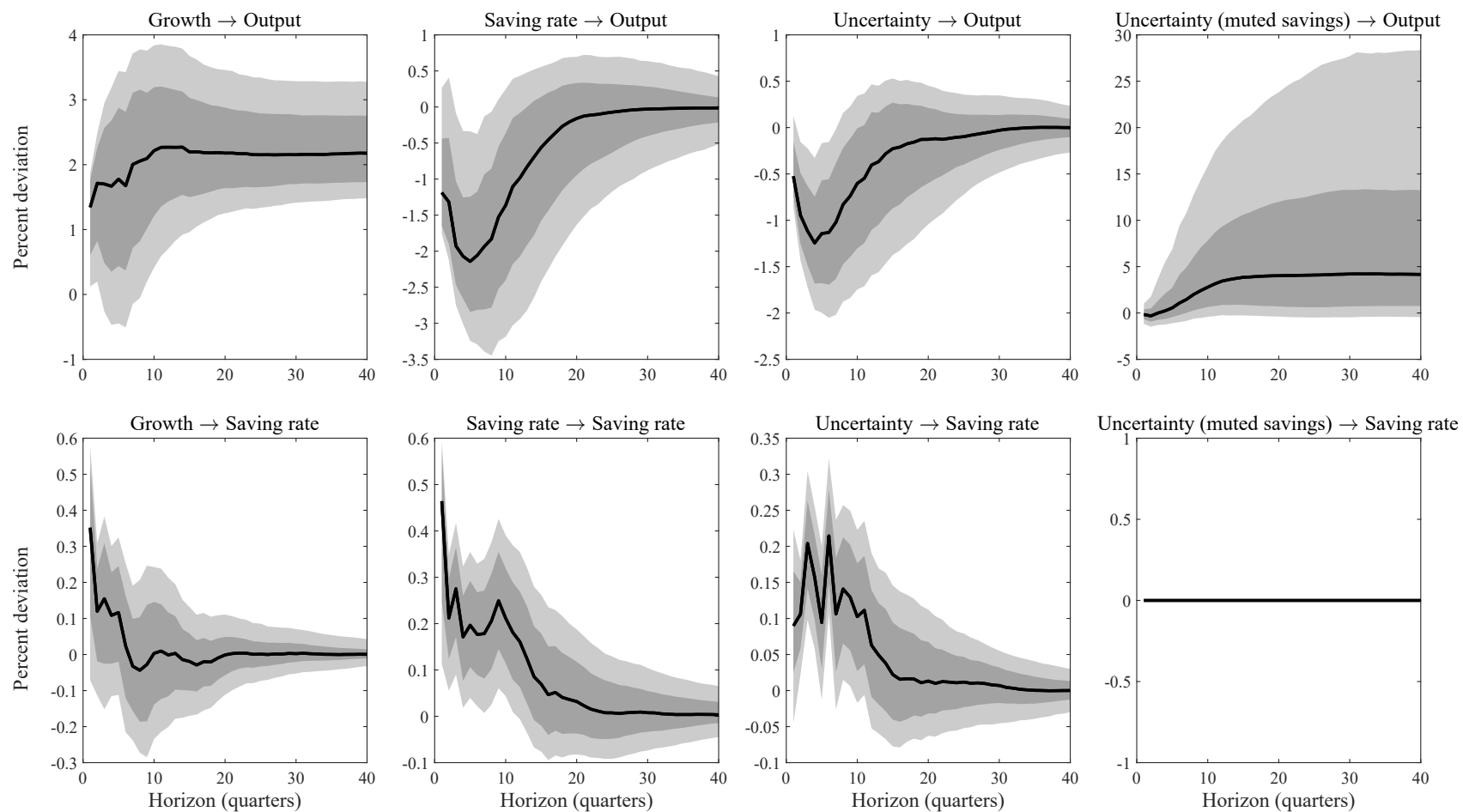Fig. C.12 Impulse response functions using long- and short-run exclusion restrictions for uncertainty and saving shocks.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications.

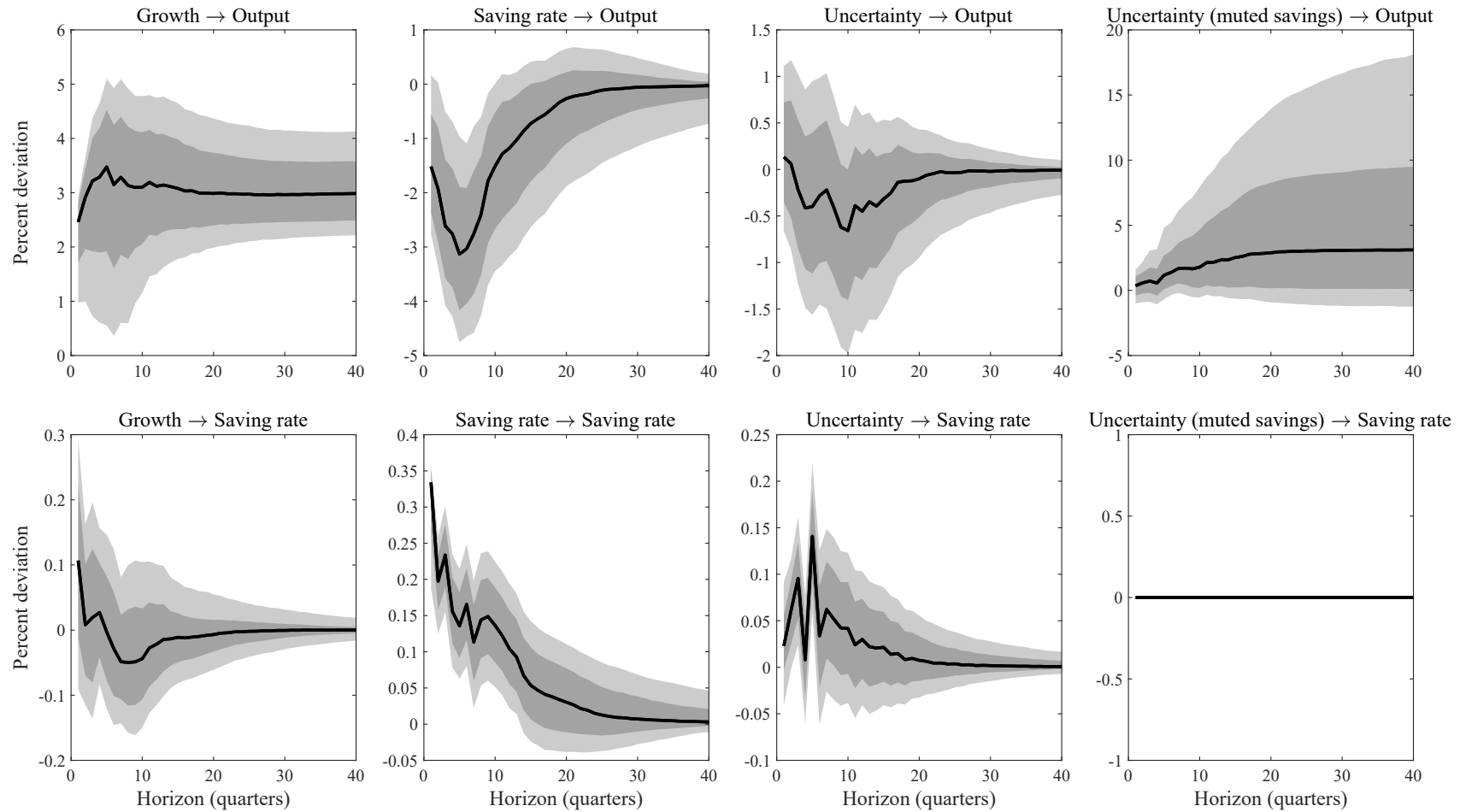The dark grey transparent area shows the corresponding 68% confidence bounds.

Fig. C.13 Impulse response functions using long- and short-run exclusion restrictions for uncertainty and saving shocks.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications.

The dark grey transparent area shows the corresponding 68% confidence bounds.

Fig. C.14 Impulse response functions using long- and short-run exclusion restrictions for uncertainty and saving shocks.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications.

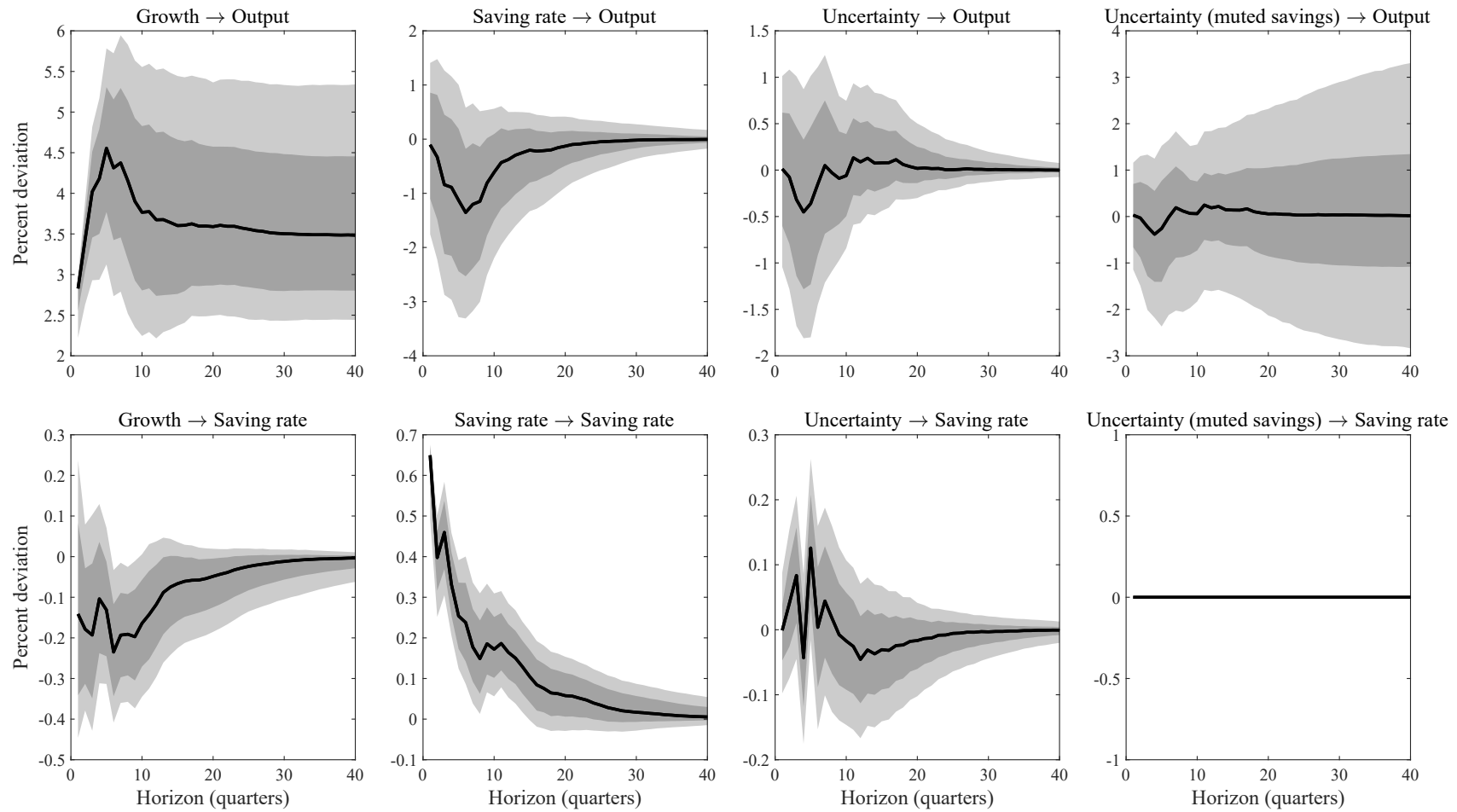The dark grey transparent area shows the corresponding 68% confidence bounds.

Fig. C.15 Impulse response functions using long- and short-run exclusion restrictions for uncertainty and saving shocks.

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications.

The dark grey transparent area shows the corresponding 68% confidence bounds.

Table C.1 Cumulative percentage response in real GDP growth for alternative estimation specifications

| Identification | At horizon h | Lags | | | Trend | | Time | | |
		4	8	12	linear	quadratic	1960Q1-2020Q4	1960Q1-2006Q4	1990Q1-2019Q4
Exclusion restriction 1	1	-0.17	-0.59	-0.39	-1.50	-0.35	-1.48	0.51	-1.16
		(0.81)	(0.94)	(0.99)	(0.73)	(0.93)	(2.42)	(0.83)	(0.46)
	2	-0.27	-0.87	-0.64	-1.89	-0.58	-1.34	0.30	-1.37
		(0.96)	(1.10)	(1.17)	(0.85)	(1.09)	(2.21)	(0.96)	(0.56)
	4	-0.73	-1.42	-1.11	-2.61	-1.09	-1.32	-0.02	-1.93
		(1.06)	(1.21)	(1.34)	(0.92)	(1.19)	(2.29)	(1.02)	(0.62)
	8	-0.58	-1.53	-1.28	-2.38	-1.24	-1.25	-0.44	-1.57
		(0.79)	(1.08)	(1.15)	(0.91)	(1.00)	(2.43)	(0.83)	(0.68)
Exclusion restriction 2	1	-0.67	-1.56	-1.27	-1.39	-1.47	-2.55	-1.56	-0.94
		(0.88)	(0.87)	(0.93)	(0.79)	(0.86)	(1.70)	(0.88)	(0.50)
	2	-0.88	-2.08	-1.68	-1.80	-1.97	-3.19	-2.17	-1.12
		(1.23)	(1.17)	(1.19)	(1.06)	(1.14)	(1.46)	(1.14)	(0.68)
	4	-1.44	-3.00	-2.31	-2.62	-2.82	-4.16	-3.00	-1.69
		(1.84)	(1.63)	(1.63)	(1.48)	(1.59)	(1.82)	(1.59)	(0.91)
	8	-2.33	-3.84	-2.38	-3.06	-3.54	-6.04	-3.89	-1.45
		(2.94)	(2.36)	(2.23)	(2.12)	(2.28)	(2.73)	(2.28)	(1.35)
Exclusion restriction 3	1	-0.74	-0.75	-1.27	-0.67	-0.77	-2.67	-0.81	-0.22
		(0.38)	(0.41)	(0.49)	(0.42)	(0.42)	(1.14)	(0.50)	(0.41)
	2	-1.18	-0.88	-1.77	-0.88	-1.04	-1.89	-1.20	-0.17
		(0.60)	(0.62)	(0.70)	(0.64)	(0.63)	(1.16)	(0.72)	(0.64)
	4	-2.17	-1.64	-2.74	-1.64	-1.71	-2.25	-1.79	-0.70
		(1.00)	(0.98)	(1.13)	(1.05)	(0.99)	(1.83)	(1.14)	(1.01)
	8	-3.96	-2.60	-3.83	-2.60	-2.82	-3.90	-3.16	-0.73
		(1.84)	(1.64)	(1.94)	(1.85)	(1.64)	(3.01)	(1.92)	(1.69)

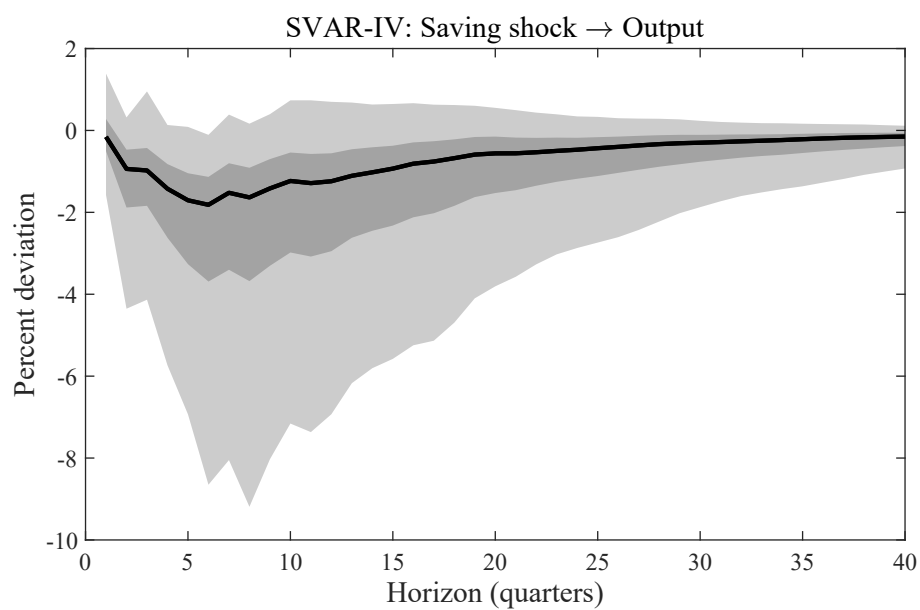Note: Standard errors are parentheses; $'$ $p < 0.32$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Fig. C.16 Impulse response function of real GDP after a saving rate shock instrumented by high-frequency monetary policy surprises by Gertler & Karadi (2015)

Notes. The black solid line shows the median impulse response function for 10,000 bootstrap replications. The dark grey transparent areas denote the corresponding 68% and 90% confidence bands.

Appendix D
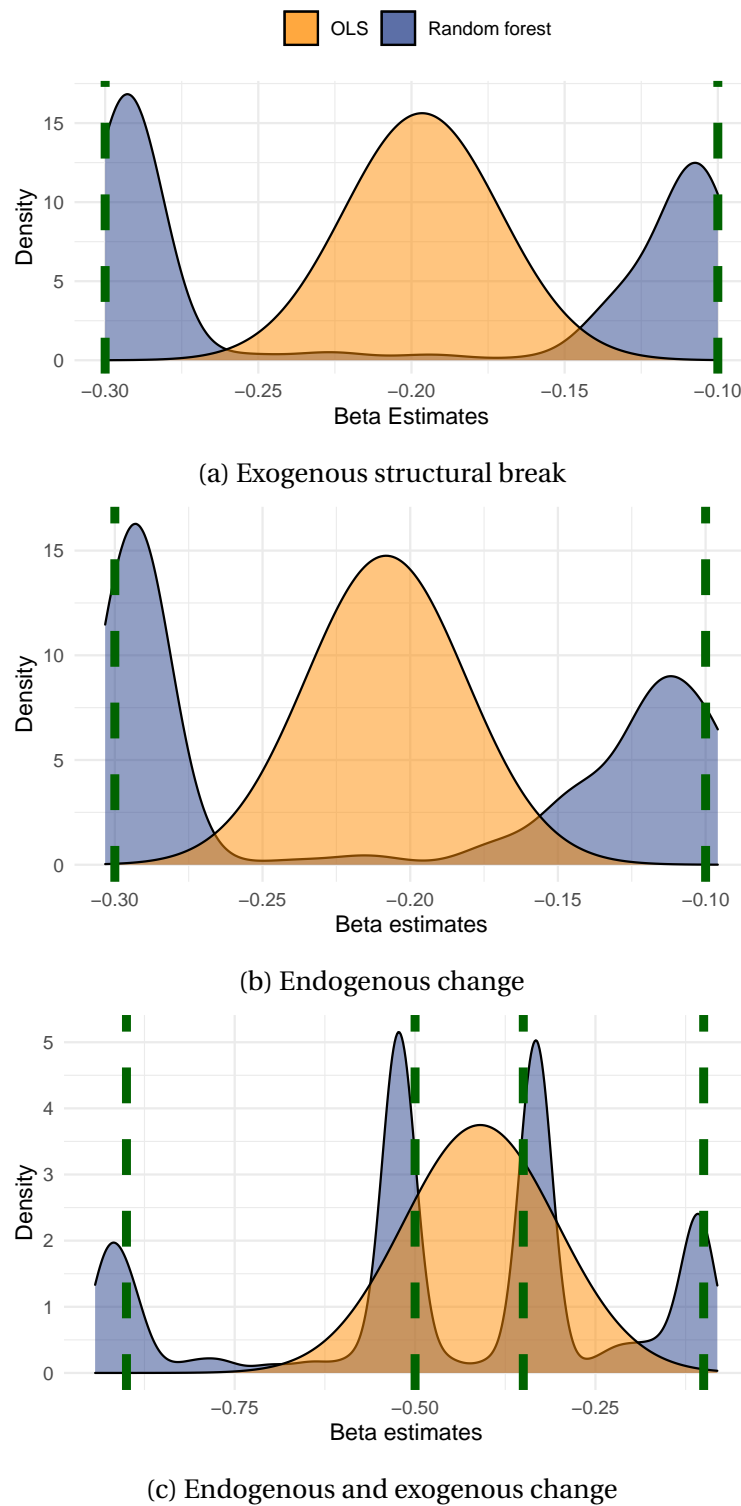
Appendix for Chapter 4

D.1   Simulation studies

(a) Exogenous structural break

(b) Endogenous change

(c) Endogenous and exogenous change

Fig. D.1 Distributions of estimates for varying policy effect for T = 500.
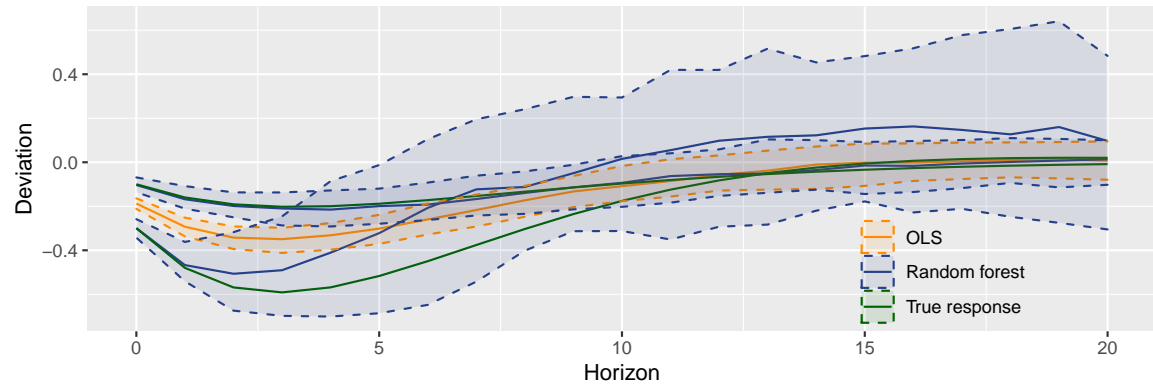
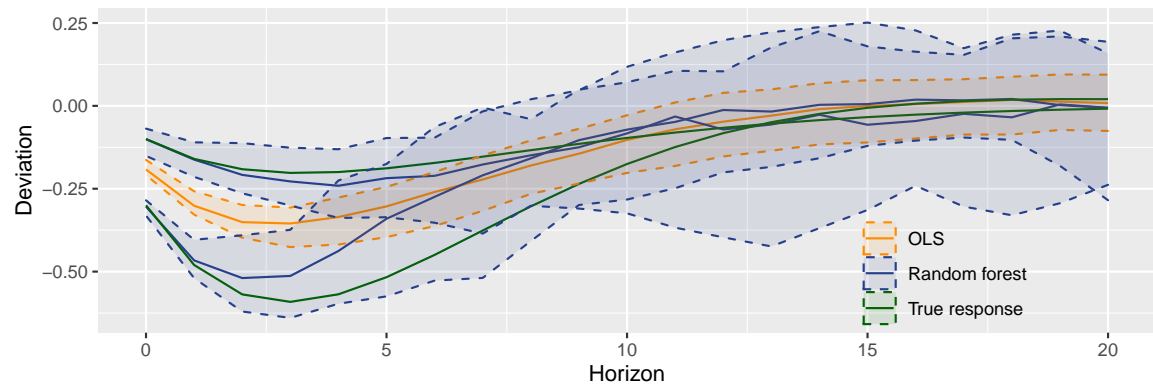Fig. D.2 Impulse response estimates for setting 1 using local projections

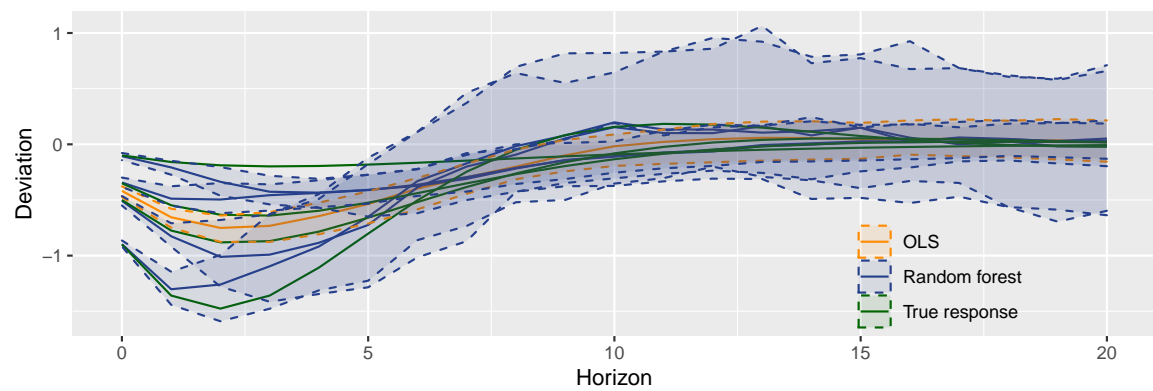Fig. D.3 Impulse response estimates for setting 2 using local projections

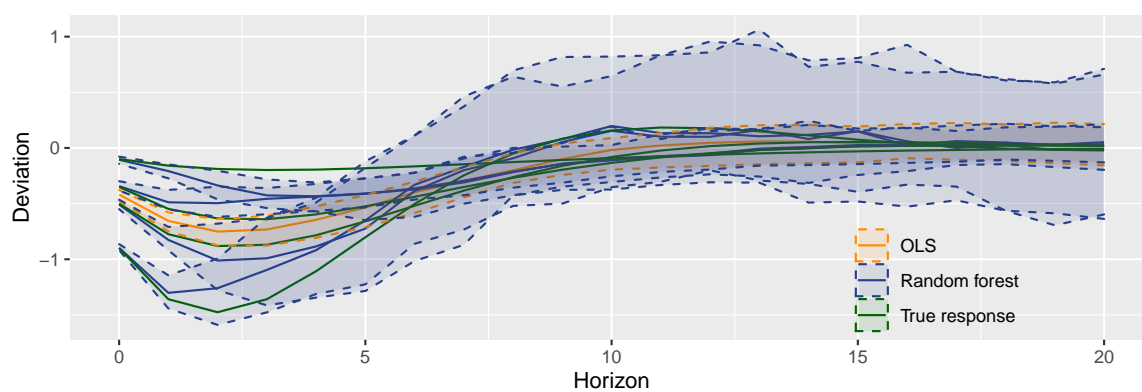Fig. D.4 Impulse response estimates for setting 3 using local projections

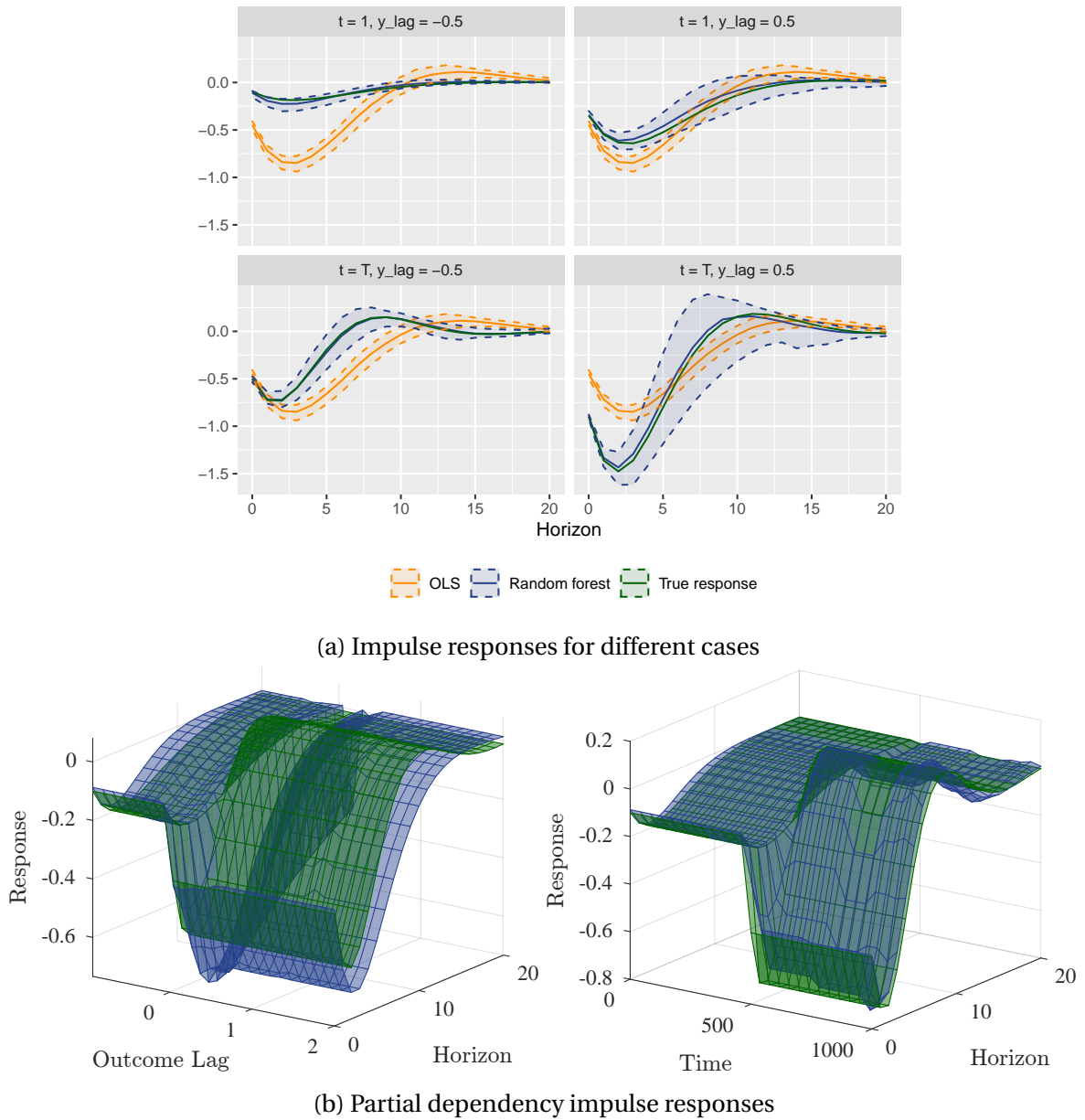Fig. D.5 Impulse response estimates for setting 3 using local projections

(a) Impulse responses for different cases

(b) Partial dependency impulse responses

Fig. D.6 Impulse response functions and their partial dependency for impact and dynamic changes, for T = 1000.