Essays in Macroeconomics and Macroeconometrics

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Matthias Meier

aus Frechen

2017

Dekan:	Prof. Dr. Daniel Zimmer
Erstreferent:	Prof. Dr. Christian Bayer
Zweitreferent:	Prof. Dr. Keith Küster
Tag der mündlichen Prüfung:	21. Juni 2017

Acknowledgements

In preparing this dissertation, I received support from many people to whom I am grateful. First of all, I wish to express my sincere gratitude to my main supervisor, Christian Bayer, for providing the environment, guidance, freedom, and resources needed to conduct my research. His genuine and effective feedback shaped both this dissertation and me as a researcher. I also benefited tremendously from Keith Kuester, my second supervisor. I want to thank him for his extraordinary support and invaluable advice especially during the later stages of my dissertation.

I consider myself very fortunate to have collaborated with a number of excellent and inspiring coauthors. I am grateful to Ariel Mecikovsky for sharing the experience of writing the first paper together and for countless discussions that were pivotal for developing this dissertation. I am grateful for the collaboration with Josè Luis (Pepe) Montiel Olea and Bulat Gafarov, from whom I have learned a lot.

Being a student at the Bonn Graduate School of Economics was instrumental in preparing this dissertation and I want to thank all professors and administrators involved. What is more, I made great friends among my co-students who never failed to provide diversion when needed.

I am greatly indebted to my wife, Leyla, who always supported me in preparing this dissertation. Her love was central to carry on during periods of frustration. I am even more grateful to my parents. All that I am I owe to my parents. From early on they encouraged me to find answers to my questions. I am thankful to them for trusting me to make the right choices on my path. This dissertation is dedicated to my parents.

Contents

List of Figures

List of Tables

1

Introduction

This thesis contains five chapters that contribute to macroeconomics and macroeconometrics. The first line of research in this dissertation, chapters 2-4, studies the role of producer heterogeneity for macroeconomics. This dissertation shows how heterogeneity matters for both business cycles and cross-country income differences. The second line of research, chapters 5-6, studies time series models, in particular vector autoregressions. These models are widely used in business cycle research including the work in my first line of research.

Chapter 2 contributes to the business cycle literature. Following the Great Recession, a wide range of business cycle research has emphasized the role of investment. This includes much of the uncertainty literature and the literature on financial frictions. A defining characteristic of investment is time to build, which is the time an investing firm needs to wait for delivery of ordered capital goods. In Chapter 2, I examine the role of time to build for business cycles. While existing business cycle models assume constant time to build, I document that time to build is volatile and largest during recessions. Motivated by this finding, I develop a heterogeneous firms general equilibrium model in which time to build fluctuates exogenously. In the model, investment is partially irreversible. The longer time to build, the less frequently firms invest, and the less firm investment reflects firm productivity. As a result, an increase in time to build worsens the allocation of capital across firms and decreases aggregate productivity. In the calibrated model, a shock increasing time to build by one month lowers investment by 2% and output by 0.5%. Structural vector autoregressions corroborate the quantitative importance of time to build shocks.

The next chapter studies business cycles as well. Chapter 3, which is joint work with Ariel Mecikovsky, contributes to the uncertainty literature. Various measures of uncertainty are countercyclical and there is ample evidence that uncertainty shocks are contractionary. In addition, a large number of structural macroeconomic models have been proposed to study various transmission channels of uncertainty shocks. In Chapter 3, we ask which of the proposed channels are empirically important for the transmission of uncertainty shocks. Exploiting highly disaggregated industry-level data from the US, we examine the empirical relevance of several transmission mechanisms. To this end, we study models with factor adjustment frictions, nominal ridigities, and financial frictions. We provide testable implications of these models based on the interaction between the severity of a friction and the job flows response to uncertainty shocks. Empirically, uncertainty shocks lower job creation and raise job destruction in the aggregate and in more than 80% of industries. We show that these responses are significantly magnified by the severity of financial frictions in a given industry, in line with the model-based findings. We do not find supportive evidence for the other transmission channels.

Chapter 4 contributes to the literature on macroeconomic development. Development accounting shows that a large share of the cross-country income differences remains unexplained by differences in physical and human capital intensity, which suggests an important role for aggregate total factor productivity. One explanation for these differences in total factor productivity builds on to the observation that firm productivity tends to be more dispersed in less developed economies. The idea is that resources are misallocated across producers and reallocating them would boost aggregate productivity. In Chapter 4, Ariel Mecikovsky, Christian Bayer, and I ask whether differences in micro-level factor productivities should be understood as a result of frictions in technology choice. Using plant and firm-level data from Chile, Colombia, Germany, and Indonesia, we document that the bulk of all productivity differences is persistent even within industries and related to highly persistent differences in the capital-labor ratio. This suggests a cost of adjusting this ratio. In fact, a model with such friction in technology choice can explain our findings. At the same time, the loss in productive efficiency from this friction is modest in the sense that eliminating it would increase aggregate productivity by 3-5%.

Chapter 5 and 6 primarily contribute to time series econometrics. Time series econometrics is at the core of applied macroeconomic research and structural vector autoregressions are among the most widely used models. A reduced-form vector autoregressive (VAR) model is a multivariate time series model that is based on the notion that every model variable depends on its own lags as well as the lags of every other model variable. A structural VAR model combines the reduced-form VAR model with additional theoretical restrictions to identify uncorrelated structural shocks. If these restrictions are sufficiently strong, they point-identify a unique structural VAR model. Imposing weaker restrictions only yields set-identification. Based on a joint project with Bulat Gafarov and Josè Luis Montiel Olea, Chapter 5 and 6 examine set-identified structural VAR models. Chapter 5 studies models that impose equality and/or inequality restrictions on a single shock, e.g. a monetary policy shock. The paper proposes a computationally convenient algorithm to evaluate the smallest and largest feasible value of the structural impulse response, e.g. the response of GDP one year after a monetary policy shock. We further show under which conditions these values are directionally differentiable and propose delta-method inference for the set-identified structural impulse response. We apply our method to set-identify the effect of unconventional monetary policy shocks.

In Chapter 6 we study models that impose equality and/or inequality restrictions on multiple shocks. The projection region is the collection of structural impulse responses compatible with the vectors of reduced-form parameters contained in a Wald ellipsoid. We show that the projection region has both frequentist coverage and robust Bayesian credibility. To address the possibility that projection inference is conservative, we propose a feasible calibration algorithm, which achieves exact robust Bayesian credibility of the desired credibility level, and, additionally, exact frequentist coverage under differentiability assumptions.

2

Time to Build and the Business Cycle

2.1 Introduction

Capital goods are complex and manufactured to the specific needs of an investing firm. For example, an assembly line consists of many elements that need to fit together; think of conveyor belts, robotic arms working along these belts, and the concrete foundation that supports the machines. Further, an assembly line needs to fit the specific good it produces. The complexity and specificity of capital cause a time gap between the order of capital goods and their delivery. This time gap is commonly referred to as time to build and is assumed constant in modern business cycle theory.[1] My paper first documents substantial variation in time to build, with peak values in recessions. Second, I ask whether exogenous fluctuations in time to build are of first-order importance for business cycles.

To address this question, I develop a dynamic stochastic general equilibrium model. Firms in my model face persistent shocks to their own productivity. Their investment is partially irreversible. The market for capital goods is characterized by search frictions, which imply time to build. Variations in time to build immediately result from changes in this friction. Calibrating the model to US manufacturing data, I find that time to build fluctuations are quantitatively important. A one month increase in time to build lowers investment by 2% and output by 0.5%.

[1] While Kydland and Prescott (1982) assume four quarters time to build, the standard assumption in real business cycle models quickly shifted to one quarter, see Prescott (1986) for example.

A lengthening in time to build is contractionary. This is due to two channels. First, later delivery of outstanding investment orders, as follows from longer time to build, mechanically reduces contemporaneous investment and thus production. Second, and this channel is both novel and quantitatively central, longer time to build worsens the allocation of capital across firms. While the efficient allocation dictates that more productive firms use more capital, longer time to build weakens the alignment between capital and productivity. At the core of the mechanism, later delivery of an investment order affects the ex-ante productivity forecasts for the periods the investment good is used as well as the associated forecast uncertainties. In turn, firms invest less frequently and, if they invest, their investment reacts less to their contemporaneous productivity. A lengthening of time to build therefore means capital is less well aligned with firms' productivity, meaning aggregate productivity is lower and so are output, investment, and consumption.

To measure time to build, I use the US Census M3 survey of manufacturing firms. The Census provides publicly available time series for order backlog and shipment in the non-defense equipment goods sector since 1968. These time series allow me to estimate time to build as the time span new capital good orders remain unfilled in capital producers' order books. I document that time to build exhibits substantial variation. It fluctuates between three and nine months. Time to build tends to be largest at the end of recession periods.[2]

The model I develop is a real business cycle model. Households consume and supply labor. The model distinguishes between firms that supply capital and firms that demand capital. On the capital demand side, there are firms that produce consumption goods combining labor with specific capital. To invest in specific capital, they need to hire an engineering firm that devises a blueprint for the investment project. Using the blueprint, the engineering firm searches for a capital supplier to produce the required capital good. Production takes place when engineering firm and capital supplier are matched and goods are delivered at the end of the period. Shocks to the matching technology cause fluctuations in time to build. These shocks may be seen as shortcut for changes in the capital supply network, which make it more difficult

[2] This paper is not the first to document the countercyclicality of the backlog ratio, see, e.g. Nalewaik and Pinto (2015). To the best of my knowledge, however, my paper is the first to relate fluctuations in the backlog ratio to time to build in the context of modern business cycle models.

to supply the required components. The model further features lumpy firm-level investment in line with the micro-level evidence on capital adjustment. The lumpiness arises because investment is partially irreversible.

To evaluate the quantitative importance of shocks to the matching technology, I calibrate the model to US data. The strategy is to jointly target moments of the investment rate distribution and aggregate fluctuations in time to build. In the calibrated model, shocks to the matching technology that raise time to build by one month cause a sharp 2% drop in investment and a more gradual drop in output that peaks after six quarters at 0.5%. I show that the direct effects of later delivery explain the short-term responses while increased capital misallocation explain nearly all of the persistent responses. Misallocation endogenously lowers measured aggregate productivity. Using the calibrated model, I back out a time series of shocks to the matching technology that explain the measured time to build fluctuations. The model predicts that these shocks account for a third of the decline in output and investment during the early 1990s recession and the Great Recession.

To solve the model, I build on the algorithms in Campbell (1998), Reiter (2009) and Winberry (2016a). The conceptual idea is to combine global projection with local perturbation solution methods. Compared to Winberry (2016a), the model in this paper is computationally more involved because the idiosyncratic state additionally consists of outstanding capital good orders. Hence, I show that the algorithms can be applied to solve more involved firm heterogeneity models.

To reassess the results of my business cycle model, I use time series techniques to investigate the importance of time to build shocks. In particular, I fit an eight-variable vector autoregression (VAR) including macroeconomic aggregates, prices, and time to build. To be conservative, I restrict the identified shocks to matching technology to contemporaneously only affect time to build. The restriction also implies that all other shocks may affect time to build contemporaneously.[3]

The results of the structural VAR corroborate the quantitative findings of my business cycle model. I find that adverse shocks to matching technology significantly and persistently lower GDP, investment, and consumption. The identified shock explains more than 20% of the forecast error variance of GDP and consumption. The impulse

[3] The results are robust to the alternative restriction that only shocks to matching technology affect all variables in the VAR contemporaneously.

response functions (IRFs) of output and investment are of similar magnitude as the IRFs in the business cycle model. Moreover, the forecast error variance of time to build explained by the identified shock is almost 50% and the identified shocks are uncorrelated with conventional direct measures of business cycle shocks (e.g., productivity, monetary policy, and tax shocks). This lends support to my business cycle model's assumption of exogenous shocks directly affecting time to build. I further show that my results are robust to relaxing the equality restrictions of the structural VAR by flexible elasticity bounds, using the methods suggested in Gafarov, Meier, and Montiel Olea (2016).

Related literature

This paper contributes to several literatures. First, this paper contributes to the literature studying the macroeconomic implications of lumpy investment. There is ample evidence for investment lumpiness, see Doms and Dunne (1998), and structural explanations are investment irreversibilities or fixed costs of capital adjustment, see Cooper and Haltiwanger (2006). Recent work has investigated the macroeconomic implications of capital adjustment costs for the response to aggregate productivity shocks, see Khan and Thomas (2008) and Bachmann and Bayer (2013), and, for the response to uncertainty shocks, see Bloom (2009), Bachmann and Bayer (2013), and Bloom et al. (2014). In my model, the interaction between time to build and investment irreversibilities is key for the transmission of shocks to the matching technology. The transmission mechanism shares the real options effect prominent in the uncertainty literature, albeit without inducing the volatility effect that higher uncertainty eventually realizes and leads to reversals and overshooting, see Bloom (2009). To the extent that longer time to build increases the effective forecast uncertainty, this paper also contributes to the endogenous uncertainty literature, see Bachmann and Moscarini (2011) and Fajgelbaum et al. (2014).

Second, my paper relates to recent work studying the interaction between time to build fluctuations and investment irreversibility. Studying time to build for residential housing, Oh and Yoon (2016) document a time series pattern fairly similar to the one for equipment capital goods documented in this paper. In their model, higher uncertainty increases time to build because residential construction occurs in stages

and each stage involves irreversible investment. Kalouptsidi (2014) studies the bulk shipping industry and shows that procyclical fluctuations in time to build dampen the volatility of investment into ships.

Third, in modeling a frictional market for capital goods, I build on the search literature. Since Mortensen and Pissarides (1994) search frictions are popular in labor market models. For capital markets, Kurmann and Petrosky-Nadeau (2007) and Ottonello (2015) show that search frictions amplify business cycle shocks. Tightness on the capital goods market governs the intensive margin of investments, while in my setup search frictions also affect the extensive margin of investment. Shocks to the matching technology in my model build on the labor market search and matching literature, see Krause et al. (2008), Sedláček (2014), and Sedláček (2016) for example.

The remainder of this paper is organized as follows: Section 2.2 documents time to build. Section 2.3 presents the central model mechanism and Section 2.4 develops the quantitative business cycle model. I discuss the calibration in Section 2.5 and results in Section 2.6. Section 2.7 provides the SVAR evidence. Finally, Section 2.8 concludes.

2.2 Does time to build vary over the cycle?

My goal is to estimate time to build using survey data on the order books of capital good producers. I show that time to build exhibits substantial variation between three and nine months with peak values during recessions.

I use US Census data collected in the Manufacturers' Shipments, Inventories, and Orders Survey (M3). The M3 covers two third of manufacturers with annual sales above 500 million USD and some smaller companies to improve industry coverage. M3 participants are selected from the Economic Census and the Annual Survey of Manufacturers and the M3 is benchmarked against these datasets. US quarterly investments are computed by the Bureau of Economic Analysis using the M3.[4]

[4] See *Concepts and Methods of the U.S. National Income and Product Accounts* (2014, ch. 3).

The Census provides publicly available data for shipments and order backlog at the sectoral level. Under the premise of excluding defense goods, I use the sector category *non-defense equipment goods*, which is available at monthly frequency from 1968 through 2015.[5] M3 data satisfies a stock-flow equation for equipment good orders, where O denotes new orders net of cancellations, S shipments, and B the beginning-of-period stock of order backlog[6]

$$B_{t+1} = B_t + O_t - S_t. \tag{2.1}$$

My baseline measure of time to build, also called backlog ratio, is

$$TTB_t \equiv \frac{B_t}{S_t}. \tag{2.2}$$

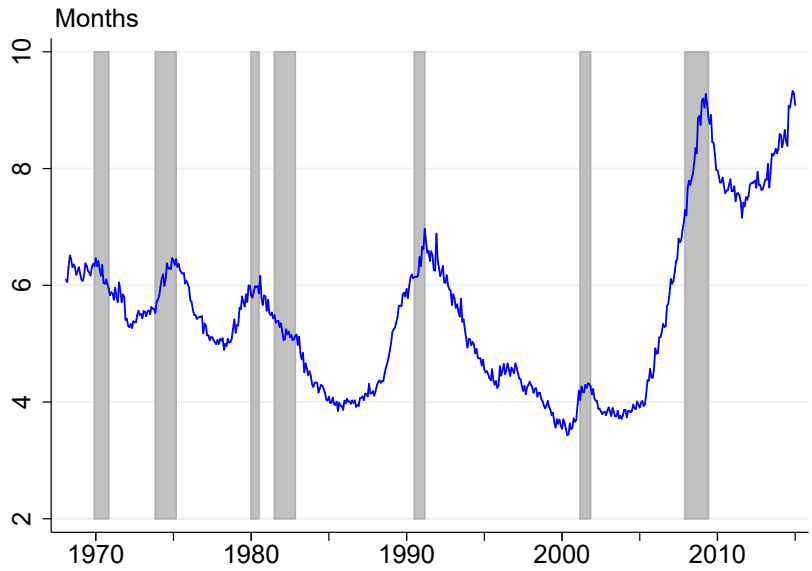
It measures the intensity of flows (shipments) out of the stock of backlogged orders, expressed in months. Figure 2.1 shows the evolution of time to build, which exhibits substantial variation. Time to build tends to start increasing before recessions and peaks at the end of recession periods. In Appendix 2.A.1, I plot the component series of (2.1) over time. The correlation of annualized real GDP growth with log time to build is -0.3. Detrending the slow-moving trend from time to build using the HP filter with a smoothing parameter of 810,000, the correlation increases to -0.4. The finding of a countercyclical backlog ratio coincides with previous studies, see Nalewaik and Pinto (2015) for example.

Under two conditions this time to build measure equals the expected waiting time of a new equipment good order: First, the shipment protocol is first-in first-out, i.e. new orders are shipped only after backlogged orders are shipped. Second, shipments are expected to be unchanged in the future. While the first condition is plausibly satisfied, the second one is roughly satisfied given that shipments are highly persistent. In Appendix 2.A.1, I show that an alternative measure of time to build, based on

[5] Notice that for finer disaggregation of the equipment goods sector into two-digit sectors, the distinction of defense and non-defense is not always available.

[6] A new order is defined as a legally binding intention to buy for immediate or future delivery, and the survey does not ask separately for order cancellations. Shipments measure the value of goods delivered in a given period, while order backlog measures the value of orders that have not yet fully passed through the sales account.

Figure 2.1. Time to build



Notes: Time to build is measured as the ratio of order backlog to monthly shipments, for non-defense equipment goods. Shaded, gray areas indicate NBER recession dates.

ex-post shipment realizations, closely resembles my baseline measure. Additionally, I provide the evolution of the individual component series defining the order stock-flow equation.

A caveat of estimating time to build using the M3 is that it excludes structure capital and imported equipment capital, which together account for no more than 35% of total non-residential private fixed investments in the US.[7]

Given the aggregate nature of the data I use, my measure is necessarily one of macroeconomic time to build. If there are cross-sectional differences in time to build, this will be different from the average micro-level time to build. Notice that within the model I develop in Section 2.4, I will recompute the measure of time to build in the exact same way and use that as calibration target.

[7] Out of total private non-residential fixed investment, structure capital constitutes on average 25% over the last 40 years, declining over time with 10% in 2015. Imported equipment capital is on average 10% of total investments, increasing over time with 20% in 2015.

2.3 Firm-level investment and time to build

This section discusses a novel, and quantitatively central, mechanism of my paper. In short, fluctuations in time to build affect how frequently firms invest, and, if they invest, by how much. These changes in the investment policy hamper an efficient reallocation of capital across firms and thereby depress real economic activity.

In general, two key determinants of a firm's investment decision are expected future productivity and uncertainty about future productivity. Higher expected future productivity makes larger investments appear profitable. Higher uncertainty about future productivity may induce the firm to postpone investments if investment is partially irreversible.[8] To understand the specific effects of time to build on a firm's investment decision, it is of central importance that longer time to build shifts the expected usage period of the investment good into the future. Hence, longer time to build affects the expected productivity, and the associated uncertainty, during the usage period.

To illustrate the point, suppose firm productivity follows an AR(1) process

$$x_t = \rho x_{t-1} + \sigma \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1).$$

Conditional on the firm's period zero productivity x_0, the forecast of productivity in period $\tau > 0$ and the associated forecast uncertainty are

$$\hat{x}_\tau = \rho^\tau x_0 \qquad \text{and} \qquad \hat{s}^2_\tau = \sigma^2 \sum_{t=1}^{\tau} \rho^{2(t-1)},$$

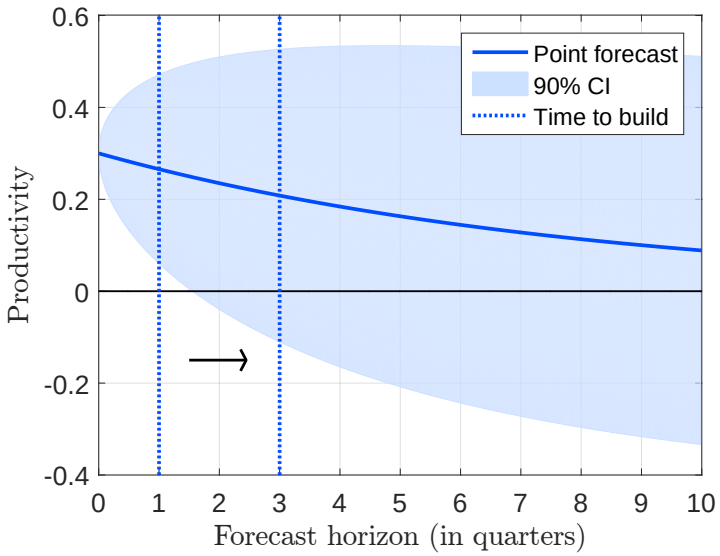
respectively. Consider τ the expected period of investment delivery and $0 < \rho < 1$. Longer time to build, that is larger τ, moves the forecast for productivity after delivery closer to the (zero) long-run mean of productivity and the associated forecast uncertainty increases. Figure 2.2 illustrates the first and second moment effect of an increase in time to build from one to three quarters.[9]

---

[8] Abel and Eberly (1996) show analytically that the inaction range, in which not adjusting capital is optimal, expands in uncertainty when capital is partially irreversible.

[9] Longer time to build increases the *relevant* forecast uncertainty by shifting the *relevant* forecast horizon, which is not captured by empirical estimates of forecast uncertainty as Jurado et al. (2015).

Figure 2.2. Productivity forecasts and time to build



Notes: Starting from an initial productivity level one unconditional standard deviation above zero, the figure plots the productivity forecast, $\hat{x}_\tau$, and its 90% confidence interval (CI), $[x_0 - 1.96\hat{s}_\tau, x_0 + 1.96\hat{s}_\tau]$, per forecast horizon, τ. The arrow illustrates a shift in time to build from one to three quarters, roughly resembling the increase in time to build observed from 2006 to 2009. The figure is based on the parameters for the firm-level productivity process calibrated in Section 2.5.

What are the implications of longer time to build for the firm's investment policy? First, longer time to build reduces the sensitivity of investment to contemporaneous deviations of productivity from its long-run mean. This follows directly from mean-reversion, and I refer to this intensive-margin change in the investment policy as *regression-to-the-mean* effect. Second, higher time to build increases the uncertainty about productivity after delivery. Assuming partial investment irreversibility, the real option value of waiting increases. That is, the firm finds it optimal to tolerate larger deviations of the current capital stock from its optimal size. In turn, the adjustment frequency falls. I refer to this extensive-margin change in the investment policy as *wait-and-see* effect.[10,11]

[10] The wait-and-see effect is also prominent to explain contractionary aggregate effects of exogenous uncertainty shocks, see, e.g., Bloom (2009) and Bachmann and Bayer (2013). In my setup, however, uncertainty is driven by changes in the expected delivery period. The volatility effect, leading to fast reversals as discussed in Bloom (2009), is not present in my setup.

[11] If productivity shocks are permanent, $\rho = 1$, the regression-to-the-mean effect is turned off, while the wait-and-see effect will be strengthened through larger effects on forecast uncertainty.

Increases in time to build have aggregate consequences because the altered investment policy hampers the efficient allocation of capital across firms with different levels of productivity. Intuitively, more of the high productivity firms with low capital stocks do not invest or invest less. Increased capital misallocation endogenously lowers measured aggregate productivity, output, investment, and consumption.

2.4 Modeling cyclical fluctuations in time to build

This section develops a model which extends the basic real business cycle model in two ways. First, producers of consumption goods vary in their productivity and use producer-specific capital. Second, investment in specific capital is partially irreversible and supplied through a frictional capital market giving rise to time to build. Shocks to the matching technology cause fluctuations in time to build.

2.4.1 Households

Households value consumption and leisure. I assume the existence of a representative household with separable preferences

$$U(C_t, L_t) = \frac{C_t^{1-\sigma}}{1-\sigma} - \psi L_t, \tag{2.3}$$

where C_t is consumption and L_t labor supply in period t. σ denotes the intertemporal substitution elasticity, and ψ parametrizes the disutility of working. I suppose the period utility function is the result of indivisible labor, see Hansen (1985) and Rogerson (1988).[12] The household owns all firms and receives aggregate profits denoted Π_t. The problem of the household is

$$\max_{C_t, L_t} U(C_t, L_t) \quad \text{s.t.} \quad C_t \leq w_t L_t + \Pi_t, \tag{2.4}$$

[12] These preferences are common in related general equilibrium models with non-convex capital adjustment frictions, see Khan and Thomas (2008) and Bloom et al. (2014) for example.

where w_t is the wage. Due to household ownership, firms discount future profits by

$$Q_{t,t+1} = \beta \frac{p_{t+1}}{p_t}, \tag{2.5}$$

with $p_t = C_t^{-\sigma}$ the marginal utility of consumption. The household's optimal labor supply requires $w_t = \psi / p_t$.

2.4.2 Engineering firms and capital suppliers

To invest in specific capital, producers of consumption goods need to hire an engineering firm that acts as an intermediary on the capital market. Engineering firms search for a capital good producer to supply the required goods. When search succeeds, the capital supplier produces all goods within one period.

Let me motivate the setup by the assembly line example in the introduction. Since assembly lines are complex, the investing firm needs to hire an assembly line producer (engineering firm). This producer, in turn requires a network of suppliers that provide the various inputs that compose an assembly line. On top, assembly lines are specific, and thus require different supplier networks across orders. While many individual business relationships are firmly established, the producer may need to search for some new suppliers given a new assembly line order. In my model, the capital supplier is a shortcut for a supply network.

In detail, I assume a continuum of capital submarkets indexed by cost parameter ξ, distributed by G. Consumption good producers randomly access a submarket ξ. The remainder of this subsection focuses on an arbitrary submarket ξ. There is a large mass of engineering firms (short: engineers) and capital suppliers. The mass of active engineers be E_t, and the mass of active capital suppliers S_t. Formally, the matching technology between engineers and capital suppliers is

$$M_t = m_t E_t^{\eta} S_t^{1-\eta}, \tag{2.6}$$

where m_t is stochastic matching efficiency that follows an AR(1) process in logs

$$\log(m_t) = (1-\rho^m)\log(\mu^m) + \rho^m \log(m_{t-1}) + \sigma^m \epsilon_t^m,\ \epsilon_t^m \overset{iid}{\sim} \mathcal{N}(0,1). \tag{2.7}$$

I define market tightness as $\theta_t = E_t / S_t$. The order filling probability for an engineer is $q_t = m_t \theta_t^{\eta-1}$, and the matching probability for a supplier is $\theta_t q_t$. Once matched, the probability of match separation is χ.

Suppliers and engineers need to hire ξ workers to participate in the market and workers are mobile across sectors so the wage is equal across sectors. When matched for any given investment order i_t, the capital supplier produces within the period and delivers the order to the engineer for unit price p_t^S. Capital suppliers have unit marginal costs to transform consumption goods into capital. Given the stochastic discount factor in (2.5), the value of an unmatched and matched capital supplier is

$$V_t^S = -\xi w_t + \theta_t q_t J_t^S + (1 - \theta_t q_t)\mathbb{E}_t[Q_{t,t+1} V_{t+1}^S], \tag{2.8}$$

$$J_t^S = p_t^S i_t - i_t + (1 - \chi)\mathbb{E}_t[Q_{t,t+1} J_{t+1}^S], \tag{2.9}$$

respectively. Engineers are hired on a spot market for investment orders, they can perfectly commit and are perfectly competitive. A consumption good producer can only hire one engineering firm. Thus, the number of engineers equals the number of orders. Conditional on delivery, engineers receive unit price p_t^E. To deliver, the engineer needs to find a matching capital supplier. The value of an unmatched and matched engineer is, respectively,

$$V_t^E = -\xi w_t + q_t J_t^E + (1 - q_t)\mathbb{E}_t[Q_{t,t+1} V_{t+1}^E], \tag{2.10}$$

$$J_t^E = p_t^E i_t - p_t^S i_t + (1 - \chi)\mathbb{E}_t[Q_{t,t+1} J_{t+1}^E]. \tag{2.11}$$

In equilibrium, engineers make zero profits on the spot market for investment orders, and I assume that capital suppliers satisfy a free entry condition.

$$V_t^E = V_t^S = 0. \tag{2.12}$$

When matched, engineer and capital supplier split the match surplus by Nash bargaining over the unit price p_t^S, where ϕ is the engineer's bargaining weight

$$\max_{p_t^S} (J_t^E - V_t^E)^{\phi} (J_t^S - V_t^S)^{1-\phi}. \tag{2.13}$$

The two equations in (2.12) together with the solution to (2.13) jointly define the equilibrium values of θ_t, p_t^S, p_t^E.

Assumption: Matches are formed for a single period, $\chi = 1$.

The assumption considerably simplifies the problem and appears less strong when reconsidering the capital supplier as shortcut for a supplier network. Under $\chi = 1$, the solution to (2.13) is $p_t^S = \phi + (1-\phi)p_t^E$ and the unit price engineers receive becomes $p_t^E = 1 + \frac{\xi w_t}{\phi q_t}\frac{1}{i_t}$. Thus, investment expenditure $p_t^E i_t = i_t + f_t$ consists of a size-dependent component with unit price of one, and a fixed cost component

$$f_t = \frac{\xi w_t}{\phi q_t}. \tag{2.14}$$

It further follows that equilibrium tightness is constant

$$\theta_t = \frac{\phi}{1-\phi}. \tag{2.15}$$

Hence, lower matching efficiency m_t unambiguously lowers delivery probability q_t.

2.4.3 Consumption good producers

The economy consists of a fixed unit mass of perfectly competitive consumption good firms, indexed by j, that produce a homogeneous consumption good

$$y_{jt} = z_t x_{jt} k_{jt}^{\alpha} \ell_{jt}^{\nu}, \tag{2.16}$$

using firm-specific capital, k_{jt}, labor, ℓ_{jt}, and subject to aggregate productivity, z_t, and idiosyncratic productivity, x_{jt}. The production function has decreasing returns to scale in the control variables, $0 < \alpha + \nu < 1$. Aggregate productivity has a deterministic trend but throughout the paper, the model is formulated along the balanced

growth path. Both idiosyncratic and aggregate productivity follow an AR(1) process

$$\log(z_t) = \rho^z \log(z_{t-1}) + \sigma^z \epsilon_t^z, \quad \epsilon_t^z \overset{iid}{\sim} \mathcal{N}(0,1), \tag{2.17}$$

$$\log(x_{jt}) = \rho^x \log(x_{jt-1}) + \sigma^x \epsilon_{jt}^x, \quad \epsilon_{jt}^x \overset{iid}{\sim} \mathcal{N}(0,1), \tag{2.18}$$

respectively. Labor adjustment is frictionless and I define gross cash flow as

$$cf_{jt} \equiv \max_{\ell_{jt} \in \mathbb{R}^+} \left\{ z_t x_{jt} k_{jt}^{\alpha} \ell_{jt}^{\nu} - w_t l_{jt} \right\}. \tag{2.19}$$

Capital adjustment is not frictionless. Firm-specific capital evolves over time according to $\gamma k_{jt+1} = (1-\delta)k_{jt} + i_{jt}$, where δ denotes the depreciation rate, i_{jt} is investment, and γ denotes constant, aggregate growth of labor productivity.

Let me detail the capital adjustment frictions. First, to invest, consumption good producers need to hire an engineering firm that searches for capital suppliers to supply the required capital goods. As a result of frictional capital markets, investment orders are not delivered instantaneously, but with probability q_t implying average time to build of $1/q_t$. Second, investment entails a fixed cost, f_t, see (2.14). The fixed cost depends on the capital submarket ξ. The submarket in which the consumption good producer can order capital is random and iid across firms and investment orders. Third, re-adjusting an outstanding order before delivery is prohibitively costly. Fourth, I assume resale losses of capital.[13]

In the dynamic firm problem, I distinguish between consumer good producers with and without outstanding orders. For firms without outstanding orders, the idiosyncratic state is described by $(k_{jt}, x_{jt}, \xi_{jt})$ with probability distribution μ^V defined for space $S^V = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$. For firms with outstanding order, the idiosyncratic state consists of $(k_{jt}, i_{jt}^o, x_{jt}, \xi_{jt})$, where i_{jt}^o denotes the outstanding investment order. The probability distribution is μ^W defined for space is $S^W = \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$. The cross-sectional distribution of all consumption good firms over their idiosyncratic states is $\mu_t = (\mu_t^V, \mu_t^W)$ defined for $S = S^V \times S^W$. The economy's aggregate state is denoted by $\mathbf{s}_t = (\mu_t, z_t, m_t)$. In the following, I drop time and firm indices and use $'$ notation to indicate subsequent periods. The value of a firm without outstanding

[13] I assume reselling is also subject to time to build: Disinvesting producers need to hire an engineer that searches for a capital supplier that transforms the capital into consumption goods.

order is given by

$$V(k,x,\xi,\mathbf{s}) \;=\; \max\Big\{V^A(k,x,\xi,\mathbf{s}), V^{NA}(k,x,\mathbf{s})\Big\}. \tag{2.20}$$

Conditional on not ordering investment (not adjusting), the firm value is

$$V^{NA}(k,x,\mathbf{s}) \;=\; cf(k,x,\mathbf{s}) + \mathbb{E}[Q(\mathbf{s},\mathbf{s}')V((1-\delta)k/\gamma, x', \xi', \mathbf{s}')|x,k,\mathbf{s}]. \tag{2.21}$$

Conditional on ordering investment (adjusting), the firm value is

$$V^A(k,x,\xi,\mathbf{s}) \;=\max_{i^o\in\mathbb{R}}\Big\{W(k,i^o,x,\xi,\mathbf{s})\Big\}, \tag{2.22}$$

The resale loss of divestment is expressed by the investment price function $p^i(i^o)$, which equals $0 \leq \bar{p}^i \leq 1$ if investment $i^o < 0$, and which equals one if investment is positive. Total investment expenditure is

$$ac(i^o,\xi,\mathbf{s}) \;=\; (1-p^i(i^o))i^o + f(\xi,\mathbf{s}) \tag{2.23}$$

The value of the consumption good firm with outstanding orders is

$$\begin{aligned} W(k,i^o,x,\xi,\mathbf{s}) \;=\; & cf(k,x,\mathbf{s}) \\ & + q(\mathbf{s})\Big[-ac(i^o,\xi,\mathbf{s}) + \mathbb{E}[Q(\mathbf{s},\mathbf{s}')V\left(((1-\delta)k+i^o)/\gamma, x', \xi', \mathbf{s}'\right)|x,\mathbf{s}]\Big] \\ & + (1-q(\mathbf{s}))\Big[\mathbb{E}[Q(\mathbf{s},\mathbf{s}')W\left((1-\delta)k/\gamma, i^o/\gamma, x', \xi, \mathbf{s}'\right)|k,i^o,x,\mathbf{s}]\Big]. \end{aligned} \tag{2.24}$$

The extensive margin of the capital adjustment decision is described by the threshold value $\hat{\xi}(k,x,\mathbf{s})$ that satisfies

$$V^A(k,x,\hat{\xi}(k,x,\mathbf{s}),\mathbf{s}) \;=\; V^{NA}(k,x,\mathbf{s}). \tag{2.25}$$

Adjustment is optimal whenever fixed adjustment costs $\xi < \hat{\xi}(k,x,\mathbf{s})$. Note that this formulation of the firm problem nests the conventional firm problem with one period time to build whenever $q(\mathbf{s}) = 1\ \forall\mathbf{s}$.

2.4.4 Recursive competitive equilibrium (RCE)

Before I define the equilibrium conditions, I define important macroeconomic aggregates. The aggregate production of the consumption good is

$$Y(\mathbf{s}) = \int_S z x k^{\alpha} \ell(k,x,\mathbf{s})^{\nu} \mu(d[k \times i^o \times x \times \xi]), \tag{2.26}$$

where $\ell(k,x,\mathbf{s})$ is the solution to (2.19). Aggregate investment expenditure is

$$\begin{aligned} I(\mathbf{s}) = &\int_{S^V} \mathbf{1}_{\{\xi < \hat{\xi}(k,x,\mathbf{s})\}} q(\mathbf{s}) ac(i^o(k,x,\mathbf{s}),\xi,\mathbf{s}) \mu^V(d[k \times x \times \xi]) \\ &+ \int_{S^W} q(\mathbf{s}) ac(i^o,\xi,\mathbf{s}) \mu^W(d[k \times i^o \times x \times \xi]). \end{aligned} \tag{2.27}$$

$\mathbf{1}_{\{\cdot\}}$ is an indicator function, that equals one if the argument is true and zero otherwise. I define aggregate order backlog as the total volume of investment orders at the beginning of the period, after new orders have been made

$$\begin{aligned} B(\mathbf{s}) = &\int_{S^V} \mathbf{1}_{\{\xi < \hat{\xi}(k,x,\mathbf{s})\}} ac(i^o(k,x,\mathbf{s}),\xi,\mathbf{s}) \mu^V(d[k \times x \times \xi]) \\ &+ \int_{S^W} ac(i^o,\xi,\mathbf{s}) \mu^W(d[k \times i^o \times x \times \xi]), \end{aligned} \tag{2.28}$$

A *RCE* is a list of functions $(w, f, q, \ell, i^o, \hat{\xi}, C, L, \Pi, Q, V, W, \mu')$ that satisfies:

1. *Consumption good producers*: Labor demand ℓ, intensive and extensive margin investment demand $(i^o, \hat{\xi})$, and value function (V, W) solve (2.19)–(2.25).

2. *Engineering firms and capital good producers*: Capital prices f and delivery probability q satisfy (2.14) and (2.15).

3. *Household*: Consumption demand C and labor supply L solve (2.4).

4. *Consistency*:
   (a) Π is consistent with profit maximization of consumption good firms.
   (b) Q is given by (2.5).

(c) μ', the law of motion of μ, is consistent with functions $(q, i^o, \hat{\xi})$ describing capital adjustment.

5. *Labor market clearing*: Labor supply L equals labor demand for consumption good production ℓ and labor demand for fixed costs of engineers and suppliers, described by $\hat{\xi}$ and G, the distribution of ξ.

6. *Goods market clearing*: $C = Y - I$, with Y and I given by (2.26) and (2.27).

2.4.5 Solution

The recursive competitive equilibrium is not computable, because the solution depends on the infinite-dimensional distribution μ. Instead, I solve for an approximate equilibrium building on the algorithms in Campbell (1998) and Reiter (2009). The general idea is to use global approximation methods with respect to the individual states, but local approximation methods with respect to the aggregate states. I solve the steady state of my model using projection methods and perturb the model locally around the steady state to solve for the model dynamics in response to aggregate shocks.

Compared to the Krusell-Smith algorithm, see Krusell et al. (1998), the perturbation approach does not require simulating the model with respect to aggregate shocks (in order to update the parameters of the forecasting rules). Further it can easily handle a large number of aggregate shocks. Terry (2015) compares the Krusell-Smith algorithm with the Campbell-Reiter algorithm for a Khan and Thomas (2008) economy. He finds that the Campbell-Reiter algorithm is more than 100 times faster. Ahn et al. (2016) combine the Campbell-Reiter algorithm to compute aggregate dynamics for a general class of heterogeneous agent economies in continuous time. More closely related to this paper, Winberry (2016b) uses (and extends) the Campbell-Reiter algorithm to solve a variation of the Khan and Thomas (2008) economy.

My adaptation of the Reiter method uses cubic B-splines with collocation to approximate the value functions. For the baseline calibration of the model, it takes one minute to solve the steady state, aggregate dynamics, and compute the impulse response functions. Appendix 2.A.2 contains the details of my solution method.

2.5 Calibration

This section discusses the model calibration, which broadly follows the literature on non-convex capital adjustment frictions in general equilibrium models, see Khan and Thomas (2008) for example. I calibrate the model at quarterly frequency. I set the discount factor $\beta = 0.99$ to match an annual risk-free rate of 4%. I assume log-utility in consumption, $\sigma = 1$. The parameter governing the household's disutility from work, ψ, is calibrated to match one third of time spent working.

Table 2.1. Quarterly model calibration

Description	Parameter	Value
Households		
Discount factor	β	0.990
Intertemporal elasticity	σ	1.000
Preference for leisure	ψ	2.400
Engineers and capital suppliers		
Bargaining power	ϕ	0.500
Mean matching efficiency	μ^m	0.542
Persistence of matching efficiency	ρ^m	0.959
Dispersion of matching efficiency	σ^m	0.041
Consumption good producers		
Capital share	α	0.250
Labor share	ν	0.580
Depreciation rate	δ	0.025
Aggregate growth	γ	1.004
Idiosyncratic persistence	ρ^x	0.970
Idiosyncratic dispersion	σ^x	0.065
Aggregate persistence	ρ^z	0.950
Aggregate dispersion	σ^z	0.007
Fixed adjustment cost (upper bound)	$\bar{\xi}$	0.010
Resale loss	$\bar{p}^i$	0.830

The parameters that describe the technology of consumption good producers are set to $\alpha = 0.25$ and $\nu = 0.58$. These values are well within the range of estimates in Cooper and Haltiwanger (2006) and Kehrig (2015), and similar to the values assumed in Khan and Thomas (2008) and Bachmann and Bayer (2013).[14] I assume

[14] Interpreting the production function as revenue production function derived in a model of monopolistic competition, the value for output elasticities would imply a markup of roughly 20%.

$\delta = 0.025$ consistent with an annual depreciation rate of 10%. Following Khan and Thomas (2008), I calibrate γ to an annualized aggregate labor productivity growth of 1.6%.

On capital markets, I assume symmetric Nash bargaining between engineers and suppliers, $\phi = 0.5$. This implies delivery probability $q_t = m_t$, which is independent of η. To calibrate mean, persistence, and variance of matching efficiency, I target the corresponding first and second moments of the empirical baseline measure of time to build, the backlog ratio. To this end, I use (2.28) to compute aggregate order backlog. Since the delivery probability is state-independent and since shipments equal investment in the model, the backlog ratio in the model is $B_t / S_t = q_t$. I set the mean matching efficiency to satisfy an average time to build of 5.5 months corresponding to the mean of the backlog ratio. Given that the backlog ratio has a weak, non-linear time trend, I detrend the quarterly time series using a low-frequency HP filter with $\lambda = 100,000$ and fit persistence and standard deviations to the residual. This yields ρ^m and σ^m for the quarterly matching efficiency process.

I assume that G, the distribution of ξ, is uniform with lower bound zero and upper bound $\bar{\xi}$. To calibrate $\bar{\xi}$ and resale loss $\bar{p}^i$, I target the share of spike investment rates in micro data. Since the idiosyncratic productivity process, described by ρ^x and σ^x, also determines the investment rate distribution, it is key to calibrate these four parameters jointly using the same dataset. I use manufacturing establishment-level data from the Longitudinal Research Database. In particular, I use the estimates in Cooper and Haltiwanger (2006) based on revenue function $\tilde{x}k^\theta$, which I take as the production technology after maximizing out labor with $\theta = \alpha/(1-\nu)$. Given ν, I translate their estimates of the profitability process at annual frequency into the parameters describing the quarterly process of x, where $x = \tilde{x}^{1-\nu}$. This yields $\rho^x = 0.97$ and $\sigma^x = 0.065$. To calibrate adjustment cost parameters $\bar{\xi}$ and $\bar{p}^s$ I target the share of positive and negative spike adjusters, documented in Cooper and Haltiwanger (2006). The two model parameters can exactly match the 18.6% share of positive spikes and the 1.5% share of negative spikes. The fixed cost is important to generate fat tails, while the resale loss is particularly important in generating the large difference between positive and negative spikes. Appendix 2.A.3 provides more details and robustness on the calibration.

2.6 Macroeconomic effects of matching technology shocks

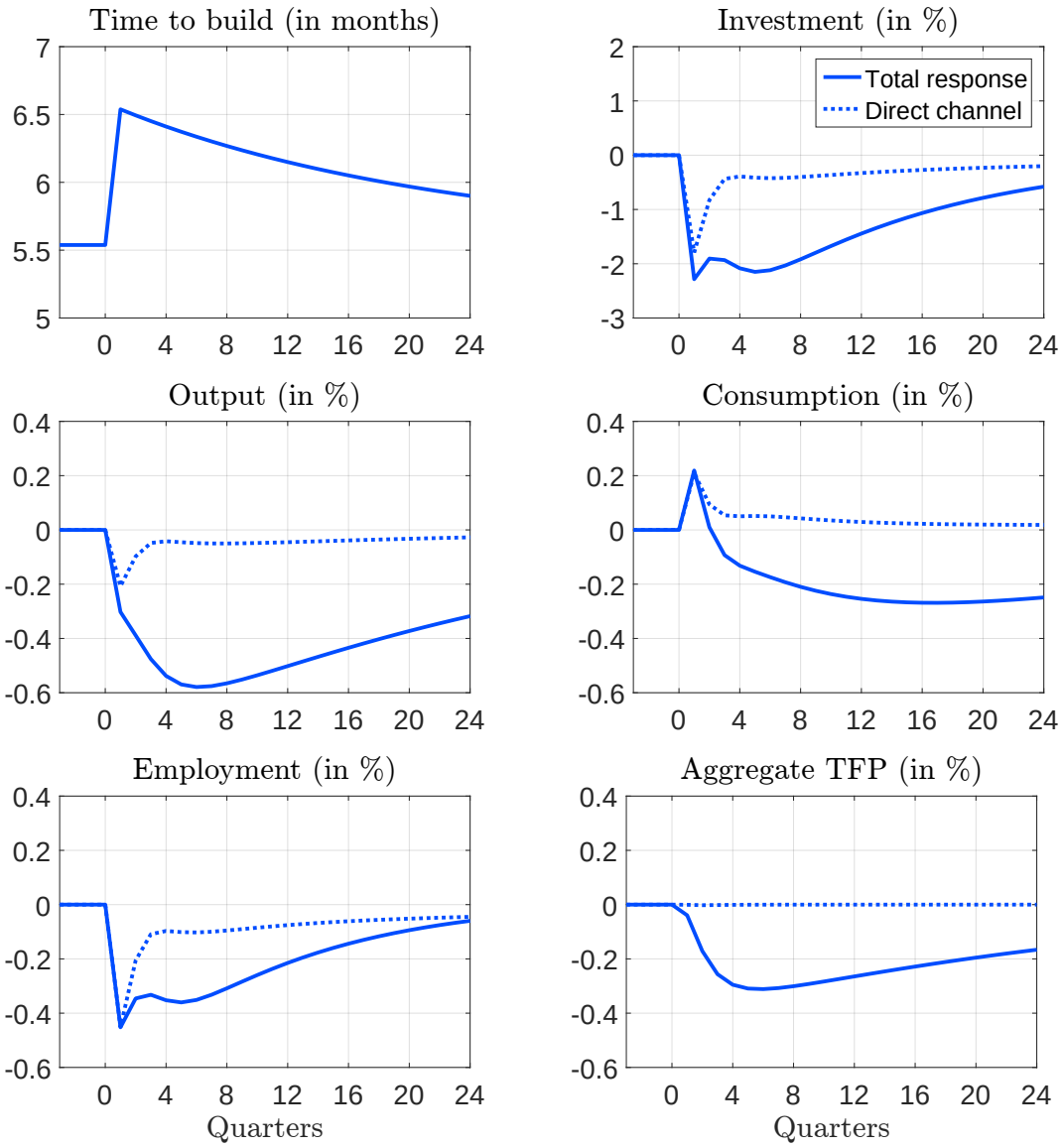
This section discusses the quantitative effects of shocks to the matching technology. In short, a shock to the matching technology that raises time to build by one month lowers investment by 2 percent and output by 0.5 percent. These shocks explain up to one third of the decline in output and investment during the early 1990s recession and the 2007-09 Great Recession.

In more detail, Figure 2.3 shows the responses to an adverse shock to the matching technology. The shock increases time to build by exactly one month, which is roughly an increase by one standard deviation of the filtered time to build series. The shock causes substantial fluctuations in output, investment, and consumption. Investment (first row, right) has the strongest impact response to the match efficiency shocks. It falls by 2 percent initially and remains strongly depressed during the first two years after the shock. Output (second row, right) falls by 0.3 percent on impact and reaches its trough of 0.55 percent five quarters later. Measured aggregate total factor productivity (last row, right) declines gradually and reaches its trough at 0.3 percent 5 quarters after the shock.

The aggregate effects of adverse shock to the matching technology are explained by a *direct* and an *indirect* channel. The direct channel captures that longer time to build delays delivery of outstanding investment orders and thus reduces investment and output. The indirect channel captures that longer time to build affects firm-level investment policies: firms invest less frequently and, if they invest, their investment reflect less their contemporaneous productivity. In turn, the alignment between firm-level capital and productivity weakens. Thus, longer time to build lowers measured aggregate total factor productivity. For a more detailed discussion of the indirect channel revisit Section 2.3.

To understand the relative quantitative importance of the two transmission channels, I suggest a simple exercise. While the indirect channel affects measured aggregate total factor productivity, the direct channel has no impact on measured productivity. To isolate the direct channel, I compute a series of exogenous shocks to aggregate productivity that exactly offset the effects the matching technology shock has

Figure 2.3. Responses to an adverse shock to the matching technology



Notes: The impulse response functions are for a shock to the matching technology that decreases time to build by one (unconditional) standard deviations starting from steady state and using the baseline calibration. 'Direct channel' denotes the impulse responses when aggregate TFP changes are eliminated through opposing aggregate productivity (z) shocks. Aggregate TFP is computed as $TFP = \log(Y_t) - \alpha \log(K_t) - \nu \log(L_t)$.

on measured aggregate productivity. Thus, measured aggregate productivity remains at its steady state level. The effects of the direct channel are the macroeconomic re-

sponses to the joint occurrence of the initial match efficiency shock and the series of productivity shocks.

The resulting 'direct channel' responses are shown as dotted lines in Figure 2.3. The direct channel is only central to understand the immediate responses, while the medium-term effect is by and large explained by the indirect channel operating through capital misallocation. The prominent role of aggregate productivity in my model corresponds to the finding in Chari et al. (2007) on the efficiency wedge. The direct channel is most important on impact of the shock because in subsequent periods firms adjust their investment policies. Firms prepone investment orders as delivery takes longer, see Figure 2.7 in Appendix 2.A.4. Abstracting from the on-impact effect, capital adjustment frequency falls, consistent with wait-and-see behavior.

Note that I evaluate the quantitative impact of shocks to the matching technology in general equilibrium. Accounting for general equilibrium effects is important because household consumption smoothing motives may substantially dampen the investment and output responses that would arise in partial equilibrium, see Khan and Thomas (2008). The initial increase in consumption (second row, right) reflects a general equilibrium mechanism. Since prices are flexible in my model, the intratemporal household optimality condition dictates that consumption has to increase initially in response to the initial decrease in investment. The reason is that the capital input in production is predetermined and labor demand falls.

As robustness check of the results, I consider a model driven by an exogenous process for the delivery probability q_t. This will turn off the equilibrium effects that matching technology shocks have on fixed capital adjustment costs f_t. Figure 2.9 in the Appendix shows that the responses are somewhat weaker, but the effects of time to build fluctuations remain quantitatively important.

The responses in Figure 2.3 show quantitatively important and persistent effects of match efficiency shocks. Next, I assess the importance of time to build for understanding past business cycles. To this end, I compute a matching technology shock series that fits the empirical time to build series. This confines my analysis to the period from 1968 through 2015. Using the model, I can compute the time series for output, investment, consumption, and employment. To be clear, in this exercise fluctuations in these series are only driven by shocks to the matching technology. To

make the quarterly series comparable to the data, I HP filter both the simulated series and their empirical counterparts using the same low-frequency filter ($\lambda = 100,000$) I used in the calibration. More details on the empirical time series are provided in Section 2.7.

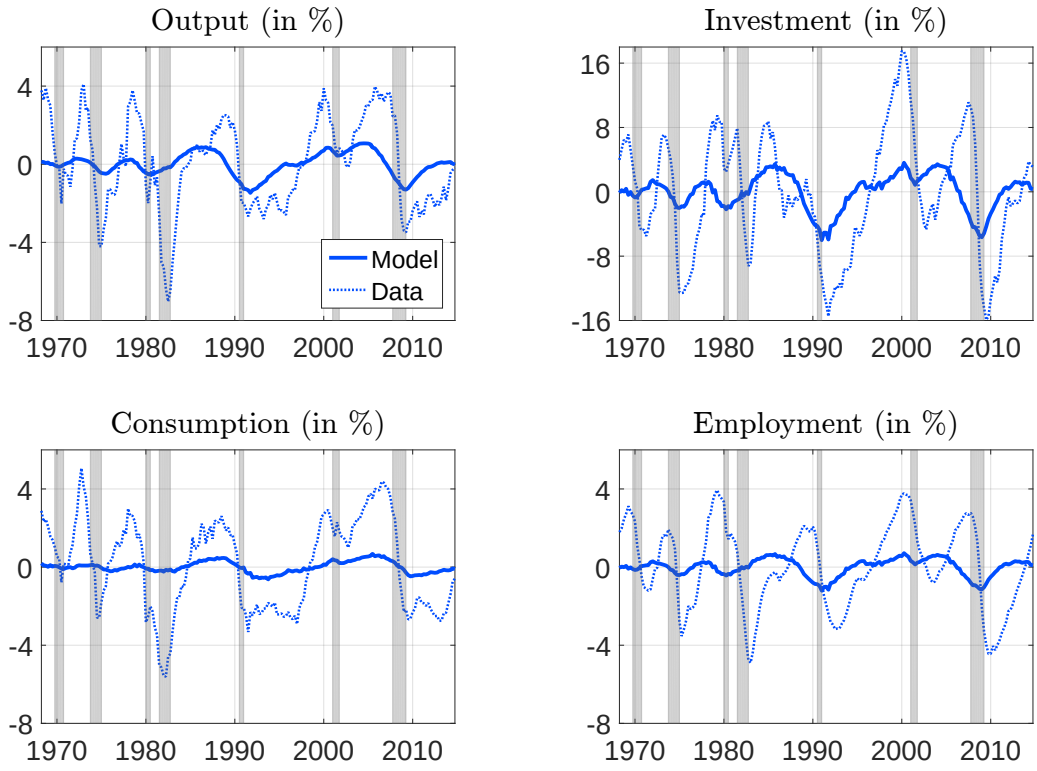
Figure 2.4 plots the model-implied time series against their empirical counterparts. Two observations stand out. First, the official recession periods (grey-shaded areas) are relatively well matched by periods where shocks to the matching technology induce below-trend output. The figure also reveals some phase shifts for the timing of expansions and contractions in aggregate production. This may not be surprising given that this paper does not claim that shocks to the matching technology are the sole driver of business cycles and other business cycle shocks may follow distinct time patterns. Second, shocks to the matching technology explain an important share of the observed business cycle variations. These shocks alone explain a drop in investments of more than 5% during the Great Recession and the early 1990s recession, compared to a drop of 16% in the data. For output, the model also explains more than a third of the empirically observed drop during these two recessions and for consumption it is almost a quarter. The exercise, thus, suggests that time to build fluctuations are important drivers of business cycles.

Finally, Table 2.5 in Appendix 2.A.4 reports business cycle moments for both the empirical data and based on simulations of the model. The model generates autocorrelation in the detrended series for output, consumption, investment, and employment close to the empirical estimates. Further, the volatility of investment relative to output in the model is very similar to the data. The magnitudes of fluctuations generated by the model are between five and ten times lower than in the data. This reflects the observation that time to build exhibits large fluctuations only in two of the seven recessions for which data is available. Conversely, while shocks to the matching technology account for an important share of the early 1990 recession and the Great Recession, these shocks are less important for other recessions.

2.7 Time series evidence

In this section, I assess the importance of structural time to build shocks using vector autoregressions. The identification requires few assumptions and I compare the iden-

Figure 2.4. Role of time to build in understanding past business cycles



Notes: These time series are computed matching the empirically observed (filtered) movements in time to build through shocks to the matching technology and otherwise using the baseline model calibration. Grey-shaded areas indicate NBER recession dates.

tified shocks to the shocks to matching technology in my general equilibrium model. The main finding is that the qualitative effects of time to build shocks are similar to the effects of matching technology shocks in the model above, while the quantitative effects are even larger. In addition, the identified shocks are largely uncorrelated with various external measures of business cycle variation, which supports the notion that time to build is driven by an independent source of variation.

2.7.1 Baseline model

I estimate a medium-scale, eight-variable vector autoregression (VAR) that allows for rich dynamic interactions between the baseline time to build measure, see

Section 2.2, and several macroeconomic series, prominent in both structural and empirical business cycle models. The vector of endogenous variables is:

$$Y = \begin{bmatrix} \text{Time to Build} \\ \text{Real GDP} \\ \text{Real Consumption} \\ \text{Real Investment} \\ \text{Consumer Prices} \\ \text{Real Wage} \\ \text{Federal Funds Rate} \\ \text{Labor Productivity} \end{bmatrix}$$

I use data at quarterly frequency that covers 1968Q1 through 2014Q4. All macroeconomic series except time to build are sourced from FRED.[15] All variables but the federal funds rate are transformed by the natural logarithm. Notice that I use non-durable consumption goods, because durable consumption goods include equipment goods that time to build shocks may directly affect. Similarly, my preferred investment time series is nonresidential investments because only firms invest in my model. The results are robust against using total consumption and total investment instead.

The baseline structural VAR model is in levels with linear time trend (D)

$$Y_t = A_0 + Dt + \sum_{j=1}^{4} A_j Y_{t-j} + Bu_t, \quad \text{Cov}(u_t) = \mathbb{I}_8, \quad \Sigma = \text{Cov}(Bu_t) = BB', \tag{2.29}$$

where Bu_t denotes reduced-form shocks and u_t structural shocks. The covariance matrix of u_t is the identity matrix of dimension eight, $\mathbb{I}_8$. I assume the match efficiency

[15] The FRED series names are GDPC96 (*Real GDP*), DNDGRA3Q086SBEA (*Real Personal Consumption Expenditures: Nondurable goods*), B008RA3Q086SBEA (*Real Private Fixed Investment: Nonresidential*), CPI, AHETPI/CPI (*Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private; deflated by CPI*), FEDFUNDS (*Effective Federal Funds Rate*), PAYEMS (*All Employees: Total Nonfarm Payrolls*). Labor productivity is real GDP over employment.

shock is the last element in vector u_t. The structural impulse responses of Y_t to the match efficiency shock are identified by the last vector in B, denoted B_8.

2.7.2 A conservative identification scheme

The baseline identification assumption is that time to build increases in response to a structural time to build shock while all other macroeconomic time series do not respond contemporaneously, i.e. $B_8 = [0, \dots, 0, b_{88}]'$. Combining this identification restriction with $BB' = \Sigma$, it follows that $b_{88} = \sqrt{e_8' \Sigma^{-1} e_8}$, where e_i is the i-th column of $\mathbb{I}_8$. B_8 is point-identified by the identification restriction.

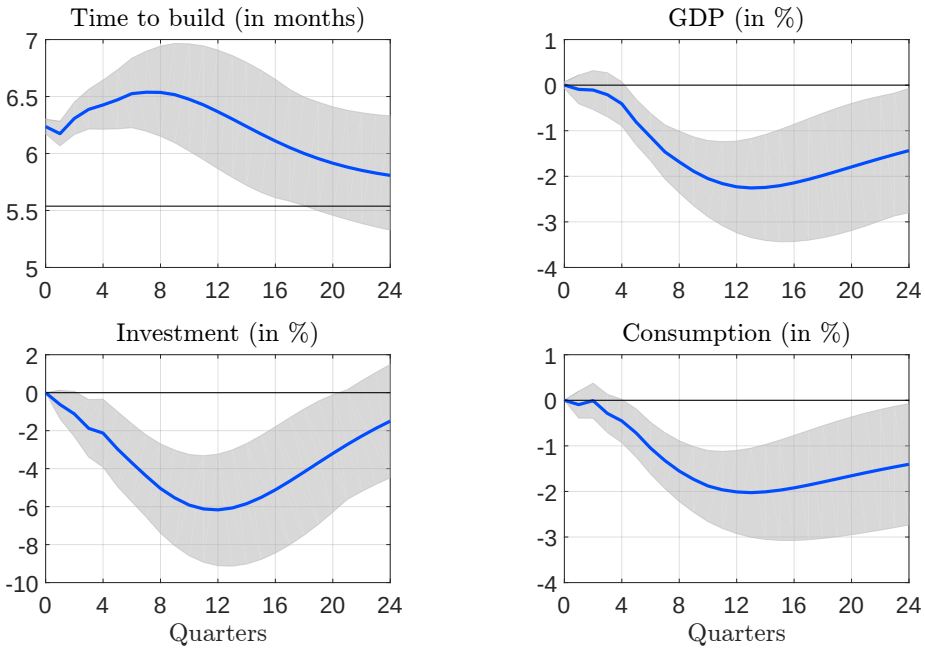
This identification scheme is conservative in the following sense. Except for the time to build shock, all structural shocks may affect time to build contemporaneously, while the time to build shock may affect all variables except time to build only through a one-quarter lag.[16] The identification is also conservative relative to the general equilibrium model where all variables are contemporaneously affected by shocks to the matching technology. Later, I reassess the importance of time to build shocks using a model-consistent identification scheme.

Figure 2.5 shows the impulse responses to a positive time to build shock that raises time to build by one month at peak. I have chosen the size of the shock to mimic the exercise in the general equilibrium model. The shock has a persistent, significant effect on time to build. More interestingly, GDP and its two main components, investment and consumption, significantly fall in response to the match efficiency shock. Not only are the responses statistically significant, but their magnitudes are also economically relevant: GDP and consumption fall by up to 2%, and investment by up to 6% within the first three years.

To get a sense of the role of time to build shocks to explain variation in macroeconomic variables, Table 2.2 shows the shares of forecast error variance explained by time to build shocks. Albeit conservatively identified, the time to build shock explains an important fraction of macroeconomic fluctuations: more than 20% of GDP and consumption, and 7% of investment. This provides further evidence in support of this paper's suggestion that time to build fluctuations are important for a better understanding of business cycle fluctuations. Importantly, at business cycle frequency

[16] The identification strategy resembles Christiano et al. (2005) for monetary policy shocks.

Figure 2.5. Impulse responses to a one month time to build shock



Notes: Solid, blue lines show (selected) responses to a match efficiency shock, under the baseline identification scheme. Shaded, gray areas illustrate the associated 90% confidence intervals.

Table 2.2. Forecast error variance decomposition

	1 year	*2 years*	*3 years*	*4 years*	*5 years*	∞
GDP	0.2	7.6	18.1	22.6	23.4	18.2
Investment	0.3	0.9	2.8	4.9	6.6	6.7
Consumption	0.8	9.8	22.2	26.9	28.2	24.6
Time to build	73.4	57.0	48.8	44.8	42.2	31.1

Note: The shares of forecast error variance explained by time to build shocks are expressed as percentages for different forecast horizons ranging from 1 year to infinity.

the time to build shock explains almost 50% of the forecast error variance of time to build itself. That is, other structural shocks explain only 50%. This supports the modeling choice of the general equilibrium model, in which time to build is directly driven by a shock to the matching technology, and not by a conventional business cycle shock, such as a shock in aggregate productivity.

The importance of time to build shocks could potentially reflect other structural shocks that are not well identified in my model. To address this concern, I correlate my identified time to build shock series with various business cycles shocks, constructed in a number of papers outside my empirical framework. These business cycles shocks include direct estimates of productivity shocks and numerous policy shocks. Table 2.3 provides the correlation of the time to build shock series with leads and lags of the external business cycle shock series. By and large, I find time to build shocks uncorrelated with external shocks. This finding further supports to the importance of exogenous shocks to time to build.

Table 2.3. Correlogram of time to build shocks with external business cycle shocks

	quarterly lags/leads								
	-4	*-3*	*-2*	*-1*	*0*	*+1*	*+2*	*+3*	*+4*
TFP	-0.07	-0.05	0.00	0.00	-0.03	-0.04	-0.07	-0.08	0.00
UA-TFP	-0.09	-0.13*	0.04	0.07	-0.05	-0.06	-0.04	0.00	0.07
UA-TFP-I	-0.03	-0.15**	0.02	0.11	0.03	-0.03	0.03	0.00	0.05
UA-TFP-C	-0.10	-0.08	0.04	0.03	-0.09	-0.07	-0.07	0.00	0.05
MP	0.02	0.08	0.04	0.02	-0.01	0.06	0.04	0.09	0.11
Oil	-0.01	0.00	-0.02	-0.02	0.01	0.00	0.01	0.09	-0.04
Defense	-0.12	-0.15**	-0.02	-0.03	-0.16**	-0.04	-0.08	-0.01	-0.10
Tax	0.02	-0.06	0.02	0.04	0.01	-0.13	0.09	0.04	-0.02

Note: The table shows the correlation of time to build shocks with various shock series at lags/forwards between -4 and +4 quarters. */**/*** denote 10%/5%/1% significance levels, respectively. Productivity shock series are from Fernald (2014): TFP, Utilization-Adjusted (UA) TFP, UA-TFP in equipment and durables, and UA-TFP in nondurables. Monetary policy shocks (MP) are based on Romer and Romer (2004) and Coibion (2012). Oil price shocks are based on Ramey and Vine (2010). Surprise defense expenditures as fiscal shocks are from Ramey and Shapiro (1998), and tax shocks from Mertens and Ravn (2011).

The identified time to build shock series appears not to reflect investment-specific productivity shocks along the lines of Justiniano et al. (2010) and Justiniano et al. (2010). Beyond the evidence in Table 2.3, this conclusion is supported by the finding that extending my VAR model by the relative price of investment goods only marginally affects the results presented here. By the same argument, identified time to build shocks appear not to reflect uncertainty shocks. The identified shocks further do not appear to reflect changes in aggregate financial conditions. This conclusion is

based on the following result. The VAR exercise in Gilchrist et al. (2014) finds that uncertainty shocks are crucially transmitted through credit spreads. When replacing uncertainty by time to build, I do not find evidence for the transmission of time to build shocks through credit spreads.

2.7.3 Robustness

Appendix 2.A.5 provides robustness for the empirical results. First, I evaluate the importance of the linear time trend assumption by estimating the VAR under the same identification restrictions but expressing all variables in first differences. The results are broadly robust. Within the first three years, GDP, investment, and consumption respond significantly to a time to build shock, and the magnitudes are similar to the baseline model in levels. Second, I compare the role of my identification scheme by suggesting an alternative identification scheme, in which time to build shocks can affect all variables contemporaneously, but no other structural shock can affect time to build contemporaneously. Importantly, this restriction is consistent with the restrictions imposed in the general equilibrium model. The responses to a time to build shock tend to be stronger under the alternative restriction, albeit the differences are small. Third, I suggest a new robustness for frequentist, point-identified structural VARs. Based on the findings in Gafarov, Meier, and Montiel Olea (2016), I replace zero restrictions by elasticity bounds. To provide robustness for time to build shocks, I replace the contemporaneous zero restrictions of the baseline restriction by constraining the elasticity of the contemporaneous response of variables other than time to build to be bounded by ±1%. I do find my baseline results to be robust against such relaxation of identification restrictions.

Beyond the robustness in Appendix 2.A.5, the results are also robust against estimating a monthly VAR, in which I replace GDP by IP and investment by new orders for non-defense capital goods. Further, the results are not solely driven by the Great Recession period. The VAR results are robust against cutting the sample from 2008.

2.8 Conclusion

This paper contributes to our understanding of business cycles by addressing a novel question: What are the business cycle implications of fluctuations in time to build? To address this question, I develop a dynamic stochastic general equilibrium model, in which capital good markets are characterized by search frictions. Fluctuations in time to build are driven by shocks to the matching technology. Calibrating the model to US data, I show that the empirically observed fluctuations in time to build are quantitatively of first-order importance for business cycles. Of particular quantitative importance is the interaction of time to build and firm investment policies leading to capital misallocation. To corroborate the model-implied results, I provide time series evidence on the importance of structural time to build shocks. I find that the effects of time to build shocks are even stronger than in the structural model.

An important follow-up question is to better understand the micro-foundations behind fluctuations in time to build. In particular, it may be useful to study capital good supply networks. Small changes at critical points in such networks, for example the exit of an important supplier, could have non-trivial aggregate implications for time to build. A complementary explanation may revolve around trade credit. While the empirical evidence rejects an important role for aggregate financial conditions, trade credit in capital good production networks might be important to understand the observed time to build fluctuations. For example, suppose capital suppliers produce subject to cash-in-advance constraints. During recessions short-run liquidity in the form of trade credit may become scarce. As a result, suppliers may need to slow down production despite long order books.

The long-run time series pattern of time to build shows that it has become more volatile since the mid-1980s. In fact, this coincides with the Great Moderation period from the mid-1980s until 2007. The Great Moderation is characterized by less volatile business cycles. One popular explanation for the decline in volatility is 'just-in-time' inventory practices, which mitigates inventory volatility. Possibly, the flip side of lower inventory volatility is larger volatility in order backlog, and thus time to build.

Appendix 2.A Appendix

2.A.1 Time to build fluctuations

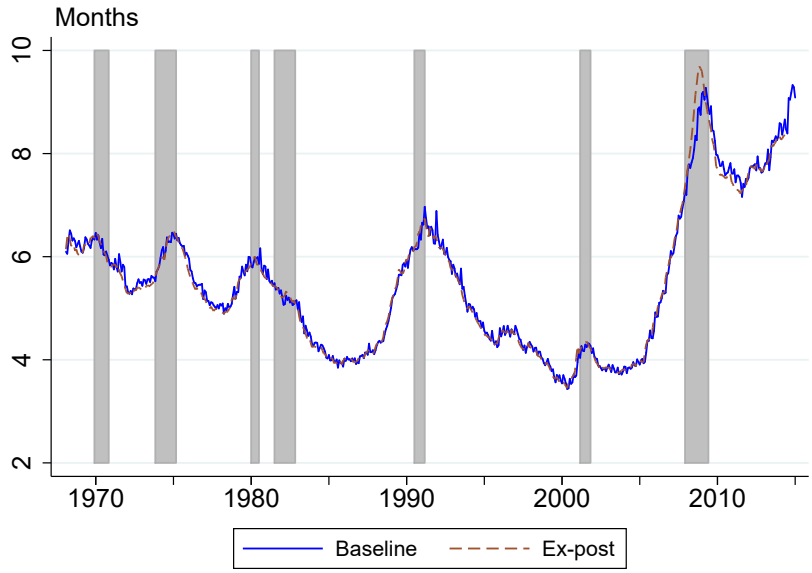
My *ex-post* measure of time to build captures the time which new orders remain in the capital good producers' order books using ex-post realizations of shipments (instead of current shipments). To be precise, I compute the lowest number of future periods required to deplete the given order backlog

$$\widetilde{TTB}_t \equiv \min_{\tau}\left(\sum_{j=1}^{\lfloor\tau\rfloor} S_{t+j} + (\tau - \lfloor\tau\rfloor)(S_{t+\lfloor\tau\rfloor+1} - S_{t+\lfloor\tau\rfloor}) - B_t\right)^2,$$

where $\lfloor\cdot\rfloor$ denotes the floor function. The second term in above formula captures a linear interpolation of shipments between two periods, by which the ex-post time to build measure becomes continuous.

Figure 2.6 compares my baseline measure with the ex-post measure of time to build. Differences between the two series are barely visible, which mainly reflects the high auto-correlation of monthly shipments.

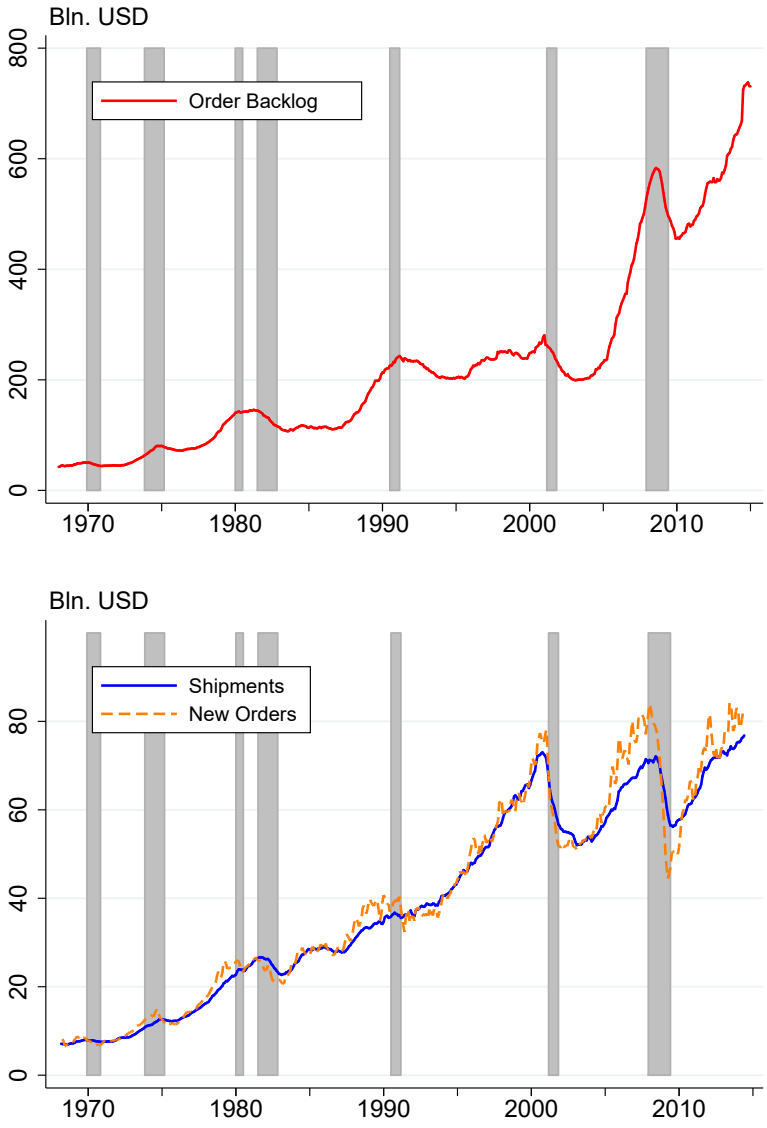
Figure 2.6. Alternative measurement of time to build



Notes: Time to build is measured as the ratio of order backlog to monthly shipments, for non-defense equipment goods. Shaded, gray areas indicate NBER recession dates.

The two panels of Figure 2.7 show the individual series defining the order stock-flow equation. The series are plotted in nominal values because the stock-flow equation is defined over nominal values.

Figure 2.7. Responses of investment orders to an adverse match efficiency shock



Notes: The time series for order backlog, shipments, and new orders refer to the non-defense equipment goods sector and are expressed in nominal values. Shaded, gray areas indicate NBER recession dates.

2.A.2 Solution algorithm

2.A.2.1 Simplified consumption good firm problem

To solve the model most efficiently, I rewrite the firm problem. First, I transform the firm problem. Instead of i^o, the investment order, I let firms choose k^o, the new capital stock upon delivery. Computationally, this transformation has the advantage that I can use the same grid for k^o as for k, and this grid can be tighter than the one for i^o. To leave the firm problem unchanged, k^o needs to evolve over time to guarantee the implicitly defined investment order satisfies $i^{o\prime} = \frac{i^o}{\gamma}$. Using the identity, $i^o = \gamma k^o + (1-\delta)k$, the evolution of k^o over time (conditional on no delivery) according to $k^{o\prime} = \frac{k^o}{\gamma} - \frac{\delta(1-\delta)k}{\gamma^2}$. Second, in slight abuse of notation, I drop the aggregate state $\mathbf{s}$ and instead use time subscripts for functions that depend on the aggregate state. I express the firm value functions in utils, see Khan and Thomas (2008), and redefine the value function such that the expectation with respect to idiosyncratic productivity does not have to be computed within the maximization problem. This raises computational efficiency and it tends to smooth the value functions. More precisely, I define $\tilde{V}_t(k,x,\xi) = p_t \mathbb{E}_x \mathbb{E}_\xi V(k,x',\xi')$, $\tilde{V}_t^A(k,x,\xi) = p_t V_t^A(k,x,\xi)$, $\tilde{V}_t^{NA}(k,x) = p_t V^{NA}(k,x)$, $\tilde{W}_t(k,x,\xi) = \mathbb{E}_x \bar{W}_t(k,x',\xi)$, $\bar{W}_t(k,x,\xi) = p_t W_t(k,x,\xi)$, where $\mathbb{E}_x$ ($\mathbb{E}_\xi$) denotes the expectation with respect to x' (ξ') conditional on x (ξ) and $p_t = C_t^{-\sigma}$ as before. Then equations (2.21), (2.20), (2.22), and (2.24) can be rewritten as:

$$
\begin{aligned}
\tilde{V}_t(k,x,\xi) &= \mathbb{E}_x \mathbb{E}_\xi \max\left\{\tilde{V}_t^A(k,x',\xi'), \tilde{V}_t^{NA}(k,x')\right\} \\
\tilde{V}_t^{NA}(k,x) &= p_t c f_t(k,x) + \beta \mathbb{E}_t[\tilde{V}_{t+1}((1-\delta)k/\gamma, x, \xi)] \\
\tilde{V}_t^A(k,x,\xi) &= \max_{k_t^o \in \mathbb{R}^+} \left\{\bar{W}_t(k,k_t^o,x,\xi)\right\} \\
\bar{W}_t(k,k^o,x,\xi) &= p_t c f_t(k,x) \\
&\quad + q_t\Big[- p_t[(1-p^i(k,k^o))(\gamma k^o - (1-\delta)k) + f_t^E(\xi)] + \beta \mathbb{E}_t[\tilde{V}_{t+1}(k^o,x,\xi)]\Big] \\
&\quad + (1-q_t)\Big[\beta \mathbb{E}_t[\tilde{W}_{t+1}\left((1-\delta)k/\gamma, k^o/\gamma - \delta(1-\delta)k/\gamma^2, x, \xi\right)]\Big] \\
\tilde{W}_t(k,k^o,x,\xi) &= \mathbb{E}_x \bar{W}_t(k,x',\xi)
\end{aligned}
$$

where $\mathbb{E}_t$ denotes the expectation with respect to aggregate state $\mathbf{s}_{t+1}$ conditional on $\mathbf{s}_t$. The net present value of the fixed adjustment cost can be expressed by $fac_t\xi$, where fac_t is defined recursively

$$fac_t = q_t p_t \frac{w_t}{\phi q_t} + (1-q_t)\beta \mathbb{E}_t fac_{t+1}.$$

In turn, this allows me to simplify the firm problem as

$$\begin{aligned}
\tilde{V}_t(k,x) &= \mathbb{E}_x \mathbb{E}_\xi \max\left\{\tilde{V}_t^A(k,x') - fac_t\xi', \tilde{V}_t^{NA}(k,x')\right\} \\
fac_t &= q_t p_t \frac{w_t}{\phi q_t} + (1-q_t)\beta \mathbb{E}_t fac_{t+1} \\
\tilde{V}_t^{NA}(k,x) &= p_t cf_t(k,x) + \beta \mathbb{E}_t[\tilde{V}_{t+1}((1-\delta)k/\gamma, x)] \\
\tilde{V}_t^{A}(k,x) &= \max_{k_t^o \in \mathbb{R}^+} \left\{\bar{W}_t(k,k_t^o,x)\right\} \\
\bar{W}_t(k,k^o,x) &= p_t cf_t(k,x) \\
&\quad + q_t\Big[-p_t(1-p^i(k,k^o))(\gamma k^o - (1-\delta)k) + \beta \mathbb{E}_t[\tilde{V}_{t+1}(k^o,x)]\Big] \\
&\quad + (1-q_t)\Big[\beta \mathbb{E}_t[\tilde{W}_{t+1}\big((1-\delta)k/\gamma, k^o/\gamma - \delta(1-\delta)k/\gamma^2, x\big)]\Big] \\
\tilde{W}_t(k,k^o,x) &= \mathbb{E}_x \bar{W}_t(k,x')
\end{aligned}$$

Importantly, this allows me to compute the extensive margin adjustment policy in closed form,

$$\hat{\xi}_t = \frac{\tilde{V}_t^A(k,x') - \tilde{V}_t^{NA}(k,x')}{fac_t}.$$

Next, I approximate firm values using collocation where Φ denotes basis functions in matrix representation and c denotes vectors of coefficients

$$\begin{aligned}
\tilde{V}_t(k,x) &\simeq \Phi^V(k,x)c_t^V \\
\tilde{W}_t(k,k^o,x) &\simeq \Phi^W(k,k^o,x)c_t^W
\end{aligned}$$

The approximations are exact at the n_k collocation nodes $k_1,...,k_{n_k}$ and $k_1^o,...,k_{n_k}^o$. In practice, I choose the same collocation nodes for k and k^o.

As baseline we use cubic B-splines to approximate the firm value functions. This does not only have the advantage of being computationally fast, but also conditional on the coefficients we know the Jacobian in closed form. In particular, I can write down the optimality condition for intensive margin capital adjustment (k_t^o) as

$$q_t p_t p^s(k, k_t^o)\gamma \; = q_t \beta \mathbb{E}_t \Phi_k^V(k_t^o, x) c_{t+1}^V + (1 - q_t)\beta \mathbb{E}_t \Phi_{k^o}^W((1-\delta)k_t/\gamma, k_t^o, x) c_{t+1}^W,$$

where $\Phi_k^V = (\partial \Phi^V)/(\partial k)$ and $\Phi_{k^o}^W = (\partial \Phi^W)/(\partial k^o)$.

I approximate the AR(1) process of idiosyncratic productivity using Tauchen's algorithm. I denote the discrete grid points of x by $x_1, ..., x_{n_x}$ consisting of n_x grid points and the transition probability from state x_j to state $x_{j'}$ one period later by $\pi_x(x_{j'}|x_j)$.

To render the infinite-dimensional distribution μ_t tractable, I approximate it with a discrete histogram. That is, μ_t measures the share of firms for each discrete combination of capital stock k_{i_1}, outstanding order $k_{i_2}^o$ (both correspond to the collocation nodes), and productivity x_j. A further distinction is useful: Let μ_t^V denote the cross-sectional distribution of firms without outstanding orders over idiosyncratic states (k_i, x_j) and μ_t^W the distribution of firms with outstanding orders over $(k_{i_1}, k_{i_2}^o, x_j)$. It holds that $\mu_t = (\mu_t^V, \mu_t^W)$.

2.A.2.2 Campbell-Reiter algorithm

Using the preceding approximation and simplification steps, the model equilibrium is described by the following non-linear equations:

$$\Phi^V(k,x)c_t^V = \mathbb{E}_x\mathbb{E}_\xi \max\left\{\tilde{V}_t^A(k,x') - fac_t\xi', \tilde{V}_t^{NA}(k,x')\right\} \tag{2.30}$$

$$\hat{\xi}_t(k,x) = (\tilde{V}_t^A(k,x) - \tilde{V}_t^{NA}(k,x))/fac_t$$

$$\tilde{V}_t^{NA}(k,x) = p_t cf_t(k,x) + \beta\mathbb{E}_t\Phi^V((1-\delta)k/\gamma, x)c_{t+1}^V$$

$$\tilde{V}_t^A(k,x) = \bar{W}_t(k,k_t^o,x)$$

$$\begin{aligned}\bar{W}_t(k,k^o,x) = {} & p_t cf_t(k,x) \\ & + q_t\Big[-p_t(1-p^i(k,k^o))(\gamma k^o - (1-\delta)k) + \beta\mathbb{E}_t\Phi^V(k^o,x)c_{t+1}^V\Big] \\ & + (1-q_t)\Big[\beta\mathbb{E}_t\Phi^W((1-\delta)k/\gamma, k^o/\gamma - \delta(1-\delta)k/\gamma^2, x)c_{t+1}^W\Big]\end{aligned}$$

$$cf_t(k_t,x_t) = (1-\nu)(\nu/w_t)^{\nu/(1-\nu)}(z_t x_t)^{1/(1-\nu)}k_t^{\alpha/(1-\nu)}$$

$$w_t = \psi/p_t$$

$$q_t = m_t(\phi/(1-\phi))^{\eta-1}$$

$$\Phi^W(k,k^o,x)c_t^W = \mathbb{E}_x\bar{W}_t(k,x') \tag{2.31}$$

$$fac_t = q_t p_t\frac{w_t}{\phi q_t} + (1-q_t)\beta\mathbb{E}_t fac_{t+1} \tag{2.32}$$

$$q_t p_t p^s(k,k_t^o)\gamma = q_t\beta\mathbb{E}_t\Phi_k^V(k_t^o,x)c_{t+1}^V + (1-q_t)\beta\mathbb{E}_t\Phi_{k^o}^W((1-\delta)k_t/\gamma, k_t^o, x)c_{t+1}^W \tag{2.33}$$

$$\frac{1}{p_t} = Y_t - I_t \tag{2.34}$$

$$Y_t = \sum_{i_1,i_2,j}\mu_t(k_{i_1},k_{i_2},x_j)(\nu/w_t)^{\nu/(1-\nu)}(z_t x_j)^{1/(1-\nu)}k_{i_1}^{\alpha/(1-\nu)}$$

$$\begin{aligned}I_t = {} & \sum_{i,j}\mu_t^V(k_i,x_j)G(\hat{\xi}_t(k_i,x_j))q_t p^s(k_i,k_t^o(x_j))\Big[\gamma k_t^o(x_j) - (1-\delta)k_i\Big] \\ & + \sum_{i_1,i_2,j}\mu_t^W(k_{i_1},k_{i_2}^o,x_j)q_t p^s(k_{i_1},k_{i_2}^o)\Big[\gamma k_{i_2}^o - (1-\delta)k_{i_1}\Big]\end{aligned}$$

$$\begin{aligned}\mu_{t+1}^V(k_{i'},x_{j'}) = {} & \sum_{i,j}\pi_x(x_{j'}|x_j)\mu_t^V(k_i,x_j)[\omega_t^{V,V,A}(i,i',j) + \omega_t^{V,V,NA}(i,i',j)] \\ & + \sum_{i_1,i_2,j}\pi_x(x_{j'}|x_j)q_t\mu_t^W(k_{i_1},k_{i_2}^o,x_j)\omega_t^{W,V}(i_1,i_2,i',j)\end{aligned} \tag{2.35}$$

$$\begin{aligned}\mu_{t+1}^W(k_{i_1'},k_{i_2'},x_{j'}) = {} & \sum_{i,j}\pi_x(x_{j'}|x_j)\mu_t^V(k_i,x_j)\omega_t^{V,W}(i,i_1',i_2',j) \\ & + \sum_{i_1,i_2,j}\pi_x(x_{j'}|x_j)\mu_t^W(k_{i_1},k_{i_2},x_j)\omega_t^{W,W}(i_1,i_2,i_1',i_2',j)\end{aligned} \tag{2.36}$$

$$\log(m_{t+1}) = (1-\rho^m)\log(\mu^m) + \rho^m\log(m_t) \tag{2.37}$$

$$\log(z_{t+1}) = \rho^z\log(z_t) \tag{2.38}$$

With the following auxiliary equations for the law of motion of the distribution:

$$\omega_t^{V,V,A}(i,i',j) = \begin{cases} G(\hat{\xi}_t(k_i,x_j))q_t \frac{k_{i'}-k_t^o(x_j)}{k_{i'}-k_{i'-1}} & \text{if} \quad k_t^o(x_j) \in [k_{i'-1},k_{i'}] \\ G(\hat{\xi}_t(k_i,x_j))q_t \frac{k_t^o(x_j)-k_{i'}}{k_{i'+1}-k_{i'}} & \text{if} \quad k_t^o(x_j) \in [k_{i'},k_{i'+1}] \\ 0 & \text{else} \end{cases}$$

$$\omega_t^{V,V,NA}(i,i',j) = \begin{cases} [1-G(\hat{\xi}_t(k_i,x_j))]\frac{k_{i'}-(1-\delta)k_i/\gamma}{k_{i'}-k_{i'-1}} & \text{if} \quad (1-\delta)k_i/\gamma \in [k_{i'-1},k_{i'}] \\ [1-G(\hat{\xi}_t(k_i,x_j))]\frac{(1-\delta)k_i/\gamma-k_{i'}}{k_{i'+1}-k_{i'}} & \text{if} \quad (1-\delta)k_i/\gamma \in [k_{i'},k_{i'+1}] \\ 0 & \text{else} \end{cases}$$

$$\omega_t^{V,W}(i,i_1',i_2',j) = \begin{cases} G(\hat{\xi}_t(k_i,x_j))(1-q_t)\frac{k_{i_1'}-(1-\delta)k_i/\gamma}{k_{i_1'}-k_{i_1'-1}}\frac{k_{i_2'}-k_t^o(x_j)}{k_{i_2'}-k_{i_2'-1}} \\ \quad \text{if} \quad k_t^o(x_j) \in [k_{i_2'-1},k_{i_2'}] \quad \text{and} \quad (1-\delta)k_i/\gamma \in [k_{i'-1},k_{i'}] \\ G(\hat{\xi}_t(k_i,x_j))(1-q_t)\frac{(1-\delta)k_i/\gamma-k_{i_1'}}{k_{i_1'+1}-k_{i_1'}}\frac{k_{i_2'}-k_t^o(x_j)}{k_{i_2'}-k_{i_2'-1}} \\ \quad \text{if} \quad k_t^o(x_j) \in [k_{i_2'-1},k_{i_2'}] \quad \text{and} \quad (1-\delta)k_i/\gamma \in [k_{i'},k_{i'+1}] \\ G(\hat{\xi}_t(k_i,x_j))(1-q_t)\frac{k_{i_1'}-(1-\delta)k_i/\gamma}{k_{i_1'}-k_{i_1'-1}}\frac{k_t^o(x_j)-k_{i_2'}}{k_{i_2'+1}-k_{i_2'}} \\ \quad \text{if} \quad k_t^o(x_j) \in [k_{i_2'},k_{i_2'+1}] \quad \text{and} \quad (1-\delta)k_i/\gamma \in [k_{i'-1},k_{i'}] \\ G(\hat{\xi}_t(k_i,x_j))(1-q_t)\frac{(1-\delta)k_i/\gamma-k_{i_1'}}{k_{i_1'+1}-k_{i_1'}}\frac{k_t^o(x_j)-k_{i_2'}}{k_{i_2'+1}-k_{i_2'}} \\ \quad \text{if} \quad k_t^o(x_j) \in [k_{i_2'},k_{i_2'+1}] \quad \text{and} \quad (1-\delta)k_i/\gamma \in [k_{i'},k_{i'+1}] \\ 0 \quad \text{else} \end{cases}$$

$$\omega_t^{W,V}(i_1,i_2,i',j) = \begin{cases} q_t \frac{k_{i'}-k_{i_2}}{k_{i'}-k_{i'-1}} & \text{if} \quad k_{i_2} \in [k_{i'-1},k_{i'}] \\ q_t \frac{k_{i_2}-k_{i'}}{k_{i'+1}-k_{i'}} & \text{if} \quad k_{i_2} \in [k_{i'},k_{i'+1}] \\ 0 & \text{else} \end{cases}$$

$$\omega_t^{W,W}(i_1,i_2,i_1',i_2',j) = \begin{cases} (1-q_t)\frac{k_{i_1'}-(1-\delta)k_{i_1}/\gamma}{k_{i'}-k_{i_1'-1}} & \text{if} \quad (1-\delta)k_{i_1}/\gamma \in [k_{i_1'-1},k_{i_1'}] \quad \text{and} \quad i_{2'} = i_2 \\ (1-q_t)\frac{(1-\delta)k_{i_1}/\gamma-k_{i_1'}}{k_{i_1'+1}-k_{i_1'}} & \text{if} \quad (1-\delta)k_{i_1}/\gamma \in [k_{i_1'},k_{i_1'+1}] \quad \text{and} \quad i_{2'} = i_2 \\ 0 & \text{else} \end{cases}$$

Labeled equations (2.30)–(2.38) are the main equations, and all other unlabeled equations are auxiliary in defining the model equilibrium. Given n_k collocation nodes and n_x discrete grid points of x, equations (2.30)–(2.38) are $n_f = 2n_k^2 n_x + 3n_k n_x + 4$. I organize these equations in

$$\mathbb{E}_t[f(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_t, \mathbf{y}_{t+1})] = 0, \tag{2.39}$$

where $\epsilon_t = (\epsilon_t^m, \epsilon_t^z) \in \mathbb{R}^2$ denotes the vector of aggregate shocks. $\mathbf{x}_t$ denotes predetermined state variables and $\mathbf{y}_t$ denotes non-predetermined state variables

$$\mathbf{x}_t = [\mu_t; \log(m_t); \log(z_t)] \in \mathbb{R}^{n_x \equiv n_k^2 n_x + n_k n_x + 2}, \tag{2.40}$$

$$\mathbf{y}_t = [c_t^V; c_t^W; \log(ac_t); \log(k_t^o); \log(p_t)] \in \mathbb{R}^{n_y \equiv n_k^2 n_x + 2n_k n_x + 2}. \tag{2.41}$$

The non-stochastic steady state is defined as $f(\bar{\mathbf{x}}, \bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{y}}) = 0$. In the general case, the model solution is given by

$$\mathbf{y}_t = g(\mathbf{x}_t, \zeta), \tag{2.42}$$

$$\mathbf{x}_{t+1} = h(\mathbf{x}_t, \zeta) + \zeta\tilde{\sigma}\epsilon_{t+1}, \tag{2.43}$$

where ζ is the perturbation parameter and $g : \mathbb{R}^{n_x} \times \mathbb{R}^+ \to \mathbb{R}^{n_y}$ and $f : \mathbb{R}^{n_x} \times \mathbb{R}^+ \to \mathbb{R}^{n_x}$. The exogenous shocks are collected in $\epsilon_{t+1} \in \mathbb{R}^{n_\epsilon}$, and $\tilde{\sigma} \in \mathbb{R}^{n_x \times n_\epsilon}$ attributes shocks to the right equations while also scaling them (by σ^m, σ^z). To solve the two policy functions, I use a first-order approximation. I follow the perturbation algorithm in Schmitt-Grohe and Uribe (2004). This requires to compute the Jacobians of function f (locally) at steady state. Importantly, the algorithm in Schmitt-Grohe and Uribe (2004) checks for existence and uniqueness of a model solution.

2.A.2.3 Krusell-Smith algorithm

This subsection suggests how the model can be solved using the Krusell-Smith algorithm. Following Krusell et al. (1998), and the adaption for heterogeneous firms by Khan and Thomas (2008), I assume agents in my model only observe a finite set of moments, informative about the entire distribution, instead of observing μ directly. The agents approximate equilibrium prices and the evolution of the observed moments by a log-linear rule.

I approximate the distribution μ by the aggregate capital stock,

$$K_t = \int_S k d\mu, \tag{2.44}$$

and the stock of investments outstanding from the preceding period

$$I_t^o = \int_{S^W} (\gamma k^o - (1-\delta)k) d\mu^W. \tag{2.45}$$

If time-to-build dropped to zero $q = 1$, I_t^o would constitute the investments activated in addition to new orders. I suggest the following log-linear forecast rules

$$\log K_{t+1} = \beta_k^0(z_t, m_t) + \beta_k^1(z_t, m_t) \log K_t + \beta_k^2(z_t, m_t) \log I_t^o, \tag{2.46}$$

$$\log I_{t+1}^o = \beta_i^0(z_t, m_t) + \beta_i^1(z_t, m_t) \log K_t + \beta_i^2(z_t, m_t) \log I_t^o, \tag{2.47}$$

and the log-linear pricing rule

$$\log p_t = \beta_p^0(z_t, m_t) + \beta_p^1(z_t, m_t) \log K_t + \beta_p^2(z_t, m_t) \log I_t^o. \tag{2.48}$$

The forecasting and pricing rules are described by coefficients that depend on the exogenous aggregate shock. For discretized processes of z and m, the equilibrium under bounded rationality with the above rules becomes computable. I use these rules to solve for the optimal policy functions and then simulate the economy and compute equilibrium prices p_t in every period t. The simulated economy allows price series are then used to update the coefficients of the log-linear rules. I stop the procedure when the coefficients have converged.

2.A.3 Additional information on the model calibration

Cooper and Haltiwanger (2006) targets the spike investment shares, but also persistence of investment rates and the correlation of investment rates with idiosyncratic productivity, when estimating a richer specification of capital adjustment costs including convex adjustment costs. I exclude the latter two moments because they may depend sensitively on the specific time to build setup. Nonetheless, the model matches these moments reasonably well with a persistence of 1.6% (empirically 5.8%), and a productivity correlation of 24% (empirically 14%).

An alternative strategy to calibrate adjustment costs is to target cross-sectional skewness and kurtosis of investment rates, see Bachmann and Bayer (2013). In fact, our calibrated model closely matches these moments in the data: skewness/kurtosis

in the model are 5.1/48.3, while in a balanced panel of Census data these are 6.5/67.4 for total investment and 5.5/47.9 for equipment investment, see Kehrig and Vincent (2016). Since skewness and kurtosis monotonically increase in the adjustment cost parameters, this indicates the calibrated adjustment costs may be too low.

Table 2.4. Calibration targets

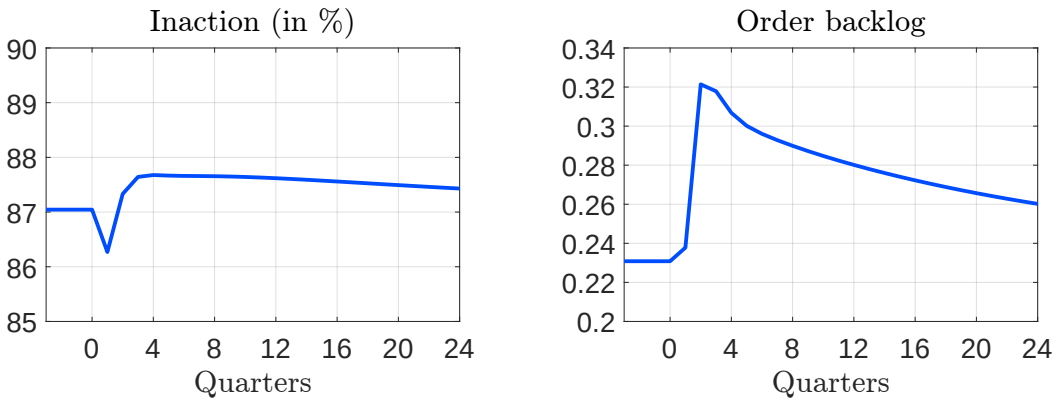
	Model	Data
Targeted (LRD)		
Positive spikes	18.6%	18.6%
Negative spikes	1.5%	1.5%
Non-targeted (LRD)		
Persistence	0.016	0.058
Productivity correlation	0.14	0.24
Non-targeted (Census)		
Skewness	5.1	6.5
Kurtosis	48.3	67.4

Notes: All moments relate to annual investment rates computed as I/K. Positive and negative spikes denote the share of investment rates larger than 20% and smaller than -20%, resp. LRD moments are from Cooper, Haltiwanger (2006), Census moments are from Kehrig, Vincent (2016).

Alternative data sources used to calibrate and estimate similar models are the IRS tax data, see, e.g., Winberry (2016a), and Compustat data, see, e.g., Bloom (2009). Both datasets are at the firm-level. The IRS does includes only positive investments, and Compustat is biased to large private firms. The main disadvantage of the LRD dataset is that it covers manufacturing only.

2.A.4 Additional results from the model simulation

Figure 2.8. Responses of investment orders to an adverse match efficiency shock



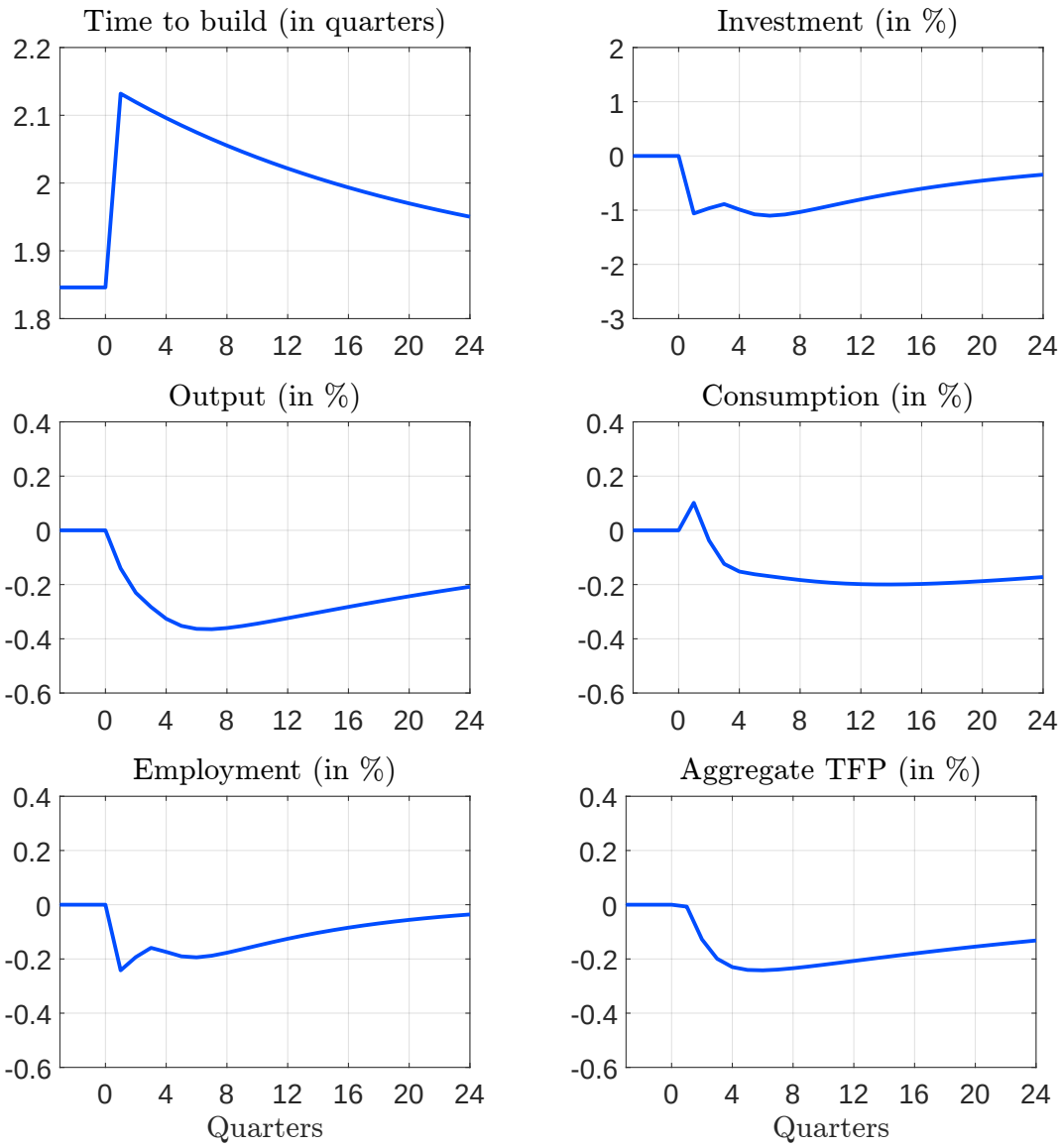
Notes: The impulse response functions are based on a *decrease* in match efficiency by one (unconditional) standard deviations starting from steady state and using the baseline calibration. Inaction measures the share of firms without outstanding orders that do not make a new order in a given period. The order backlog is the total of investments outstanding for delivery.

Table 2.5. Business cycle statistics

	Data	Model
Volatility of output (%)	2.37	0.31
Volatility of consumption (%)	2.08	0.16
Volatility of investment (%)	7.27	1.23
Volatility of employment (%)	2.11	0.24
Autocorrelation of output	0.94	0.96
Autocorrelation of consumption	0.94	0.87
Autocorrelation of investment	0.96	0.89
Autocorrelation of employment	0.97	0.89
Correlation of consumption with output	0.86	0.61
Correlation of investment with output	0.72	0.92
Correlation of employment with output	0.71	0.89

Note: All series, from data and model simulations, are expressed in logs and HP-filtered with a quarterly smoothing parameter of 100,000.

Figure 2.9. Responses under alternative fixed adjustment costs: $f(\xi, \mathbf{s}) = \frac{\xi w(\mathbf{s})}{\phi \bar{q}}$ with $q(\mathbf{s}) = \bar{q}$ in steady state



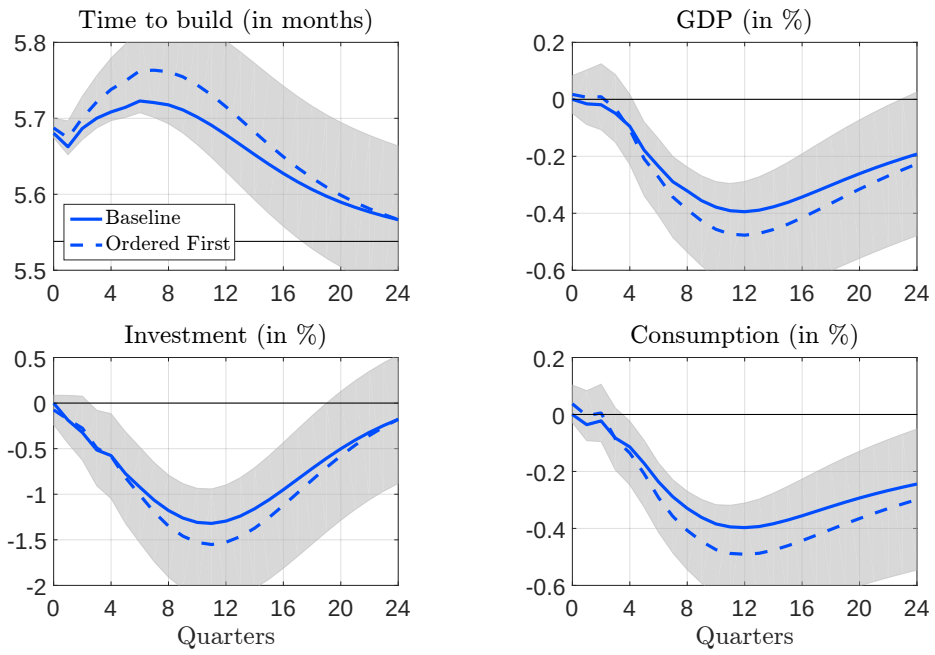
Notes: The impulse response functions are based on a *decrease* in match efficiency by one (unconditional) standard deviations starting from steady state and using the baseline calibration. 'Direct effect' are the impulse responses when aggregate TFP changes are eliminated through opposing aggregate productivity (z) shocks. Aggregate TFP is computed as $TFP = \log(Y_t) - \alpha \log(K_t) - \nu \log(L_t)$.

2.A.5 Robustness of the structural VAR results

2.A.5.1 Alternative identification scheme and first differences

First, I investigate the results under an alternative identification assumption. While the baseline identification scheme tends to be conservative, its restrictions are stronger than the restrictions of the general equilibrium model. As alternative identification, I suggest to have the time to build shock *'ordered first'*. This term refers to the ordering of variables in the VAR. It means that time to build shocks can contemporanously affect all other variables in the VAR, but no shock other than time to build shocks can affect time to build contemporaneously. Figure 2.10 shows that the baseline identification implies smaller macroeconomic respones to time to build shocks compared to the alternative identification, albeit the differences are not large. Impulse responses under the alternative identification remain significant.

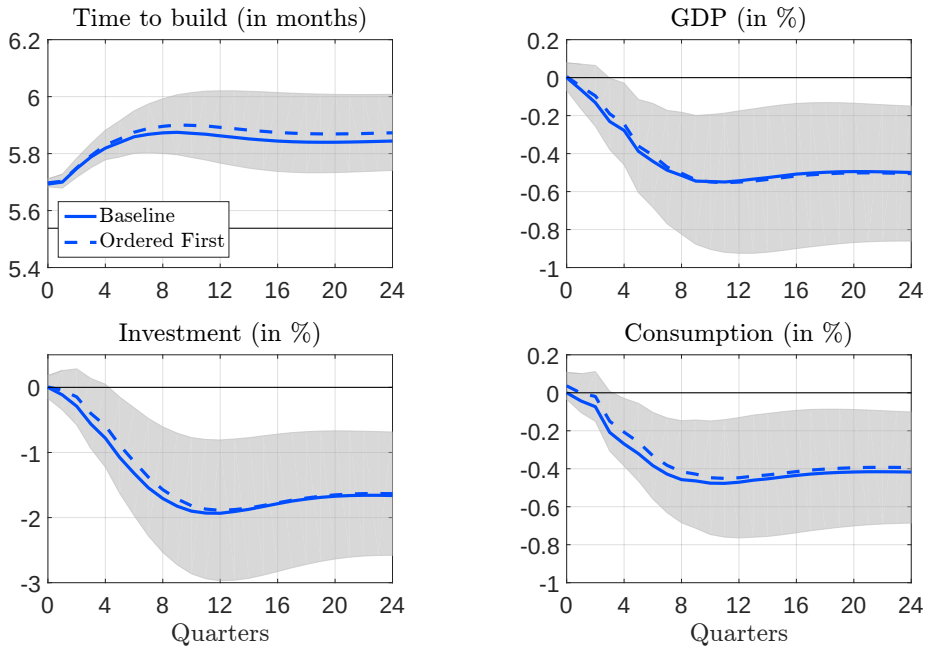
Figure 2.10. Impulse responses to a one standard deviation time to build shock (model in levels with linear time trend, alternative identification schemes)



Notes: Solid lines show (selected) impulse responses to a time to build shock under the baseline identification scheme. Dashed lines show the impulse responses under the alternative identification scheme, in which time to build is *'ordered first'*. Shaded, gray areas illustrate the 90% confidence intervals associated with the alternative identification scheme.

Figure 2.11 shows the cumulative impulse responses when estimating a VAR, in which all variables enter in first differences and the linear time trend is dropped. At the same time, the figure compares the two identification schemes. The differences of the impulse responses across identification schemes appears negligible. The important finding is that the impulse responses are similar to the ones in Figure 2.10. While I assumed a linear time trend for the latter, the findings on time to build shocks appear robust to non-linear time trends.

Figure 2.11. Cumulative impulse responses to a one standard deviation time to build shock (model in first differences, two alternative identification schemes)



Notes: Solid lines show (selected) cumulative impulse responses to a time to build shock under the baseline identification scheme. Dashed lines show the impulse responses under the alternative identification scheme, in which time to build is *'ordered first'*. Shaded, gray areas illustrate the 90% confidence intervals associated with the alternative identification scheme.

2.A.5.2 Elasticity bounds

In this subsection, I propose a new approach to provide robustness for point-identified structural VAR models in a frequentist setup. Structural VAR models, such as Gali (1999), Christiano et al. (2005), and Bloom (2009), impose various zero restrictions on contemporaneous and long-run responses to obtain point identification.

As robustness, I propose to replace some or all of the zero restrictions by bounds on the elasticity with respect to the shock of interest.[17] For example, instead of assuming an uncertainty shock does not contemporaneously affect GDP, as robustness I would restrict the elasticity of GDP with respect to a change in uncertainty due to an uncertainty shock to be bounded between $\pm c\%$. This nests the point-identified model in the limit case when all bounds are zero ($c = 0$). The structural VAR model is no longer point-identified when replacing a zero restriction with strictly positive bounds on the elasticities ($c > 0$).

I implement this robustness exercise using the results in Gafarov, Meier, and Montiel Olea (2016), which provide inference for set-identified structural VAR models. Formally, to apply their results, I need to assume that for a given IRF either the lower and upper elasticity bound may not hold jointly. Notice that confidence sets are estimated based on Delta method inference. In fact, bootstrap inference is not necessarily valid here because the endpoints of the identified sets are not fully differentiable.

The suggested robustness is similar to Conley et al. (2012) which proposes as robustness to relax the exclusion restriction when using IV methods. I suggest the following robustness for the conservative baseline identification. Instead of zero restrictions on contemporaneous responses, I constrain the elasticity of all variables (except for the backlog ratio) with respect to the match efficiency shock to be between -1% and +1%, see Table 2.6. For an increase in the backlog ratio of 2.5%, the contemporaneous responses are bound to be between -0.025% and +0.025%.

Figure 2.12 shows the resulting impulse responses under the *robustness* identification scheme. Instead of a single impulse response, there is an interval with admissible impulse responses (dotted lines). The confidence set is adjusted accordingly based on Gafarov, Meier, and Montiel Olea (2016). Notice that the main findings of the baseline model in Figure 2.12 are 'robust' in the sense that the declines in GDP, investment, and consumption remain significant.

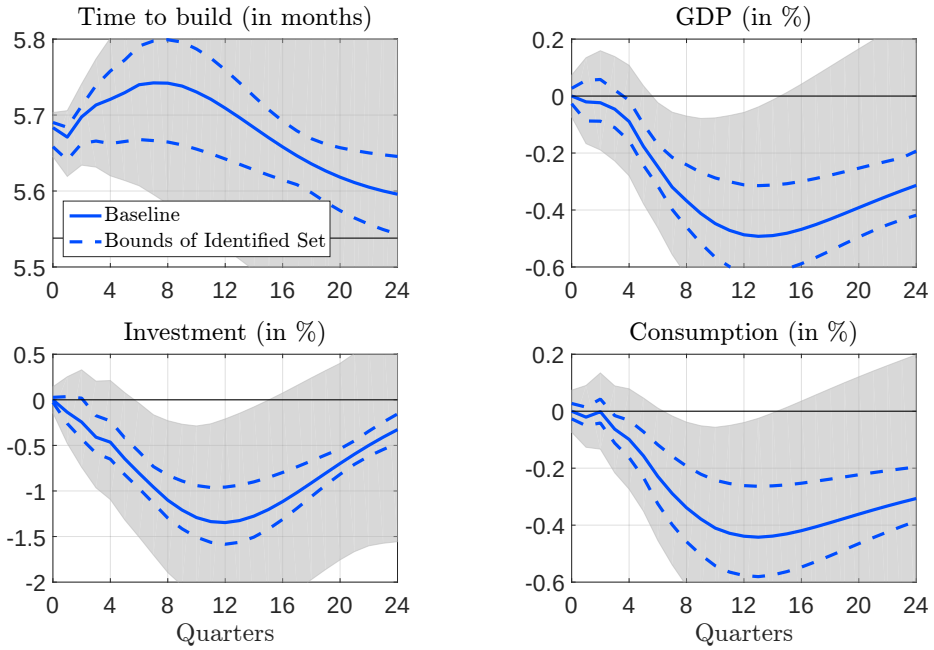
[17] Elasiticity bounds have recently gained popularity in the Bayesian structural VAR literature, see, e.g., Kilian and Murphy (2012a) and Baumeister and Hamilton (2015b).

Table 2.6. Identification schemes: constraints on contemporaneous elasticities

	TTB	GDP	Inv	Con	CPI	Wag	FFR	LaP
Baseline	+	0	0	0	0	0	0	0
Robustness	+	±1%	±1%	±1%	±1%	±1%	±1%	±1%

Notes: +/0/±1% indicate that the elasticity is constrained to be positive/exactly zero/between -1% and +1%, respectively. The contemporaneous elasticity of variable i and time to build in response to time to build shocks is given by $(e_i' B_1)/(e_1' B_1)$, where e_i is the i-th column of the identity matrix $\mathbb{I}_8$. *TTB*: Time to build, *GDP*: Real GDP, *Con*: Real Consumption, *Inv*: Real Investment, *CPI*: Consumer Prices, *Wag*: Real Wage, *FFR*: Federal Funds Rate, *LaP*: Labor Productivity.

Figure 2.12. Impulse responses to a one standard deviation time to build shock (model in levels with linear time trend, two alternative identification schemes)



Notes: Solid lines show (selected) responses to a time to build shock under the baseline identification scheme. Dashed lines show the bounds of the identified set under elasticity constraints, see Table 2.6. Shaded, gray areas illustrate the 90% confidence intervals for the identified sets.

3

Do Plants Freeze Upon Uncertainty Shocks?

Joint with Ariel Mecikovsky

3.1 Introduction

One of the most active areas of business cycle research since the Great Recession is the literature studying fluctuations in uncertainty. The following briefly revises this literature: First, there is ample evidence that uncertainty shocks are contractionary.[1] Second, a number of transmission channels have been put forward that rationalize the broad empirical finding.[2] Third, we know relatively little about the empirical relevance of these various transmission channels.

However, understanding the transmission of uncertainty shocks is of central importance when designing counter-cyclical policy interventions. For example, in models with (non-convex) factor adjustment frictions, in which the transmission is characterized by plants adopting wait-and-see behavior (freezing) in response to higher

[1] Empirical evidence includes Bloom (2009), Bachmann et al. (2013), Caggiano et al. (2014), Jurado et al. (2015), Baker et al. (2016), and many, many more.

[2] Transmission channels suggested in the literature build on capital adjustment frictions, e.g. Bloom (2009), Bachmann and Bayer (2013), Bloom et al. (2014); labor adjustment (search) frictions, e.g. Schaal (2012), Leduc and Liu (2014), Riegler (2014); price rigidities, e.g. Bundick and Basu (2014), Born and Pfeifer (2016), Fernandez-Villaverde et al. (2015), Vavra (2014); and financial frictions, e.g. Alfaro et al. (2016),Christiano et al. (2010), Arellano et al. (2012), Gilchrist et al. (2014), Dyrda (2015)

uncertainty, an effective policy may be an investment or hiring subsidy that targets those plants close to their adjustment threshold, e.g. small plants as in Winberry (2016a). Instead, if financial frictions are key, an effective policy intervention may target the financing conditions of firms close to defaulting. Finally, if price rigidities are key, adequate monetary policy rules are important, see Basu and Bundick (2015).

This paper contributes to the uncertainty literature by providing evidence on the channels through which uncertainty shocks affect employment. For this purpose, we first analyze models with various frictions – labor adjustment frictions, capital adjustment frictions, price rigidities, and financial frictions, respectively – and show how the response of job flows to uncertainty shocks depends on the strength of these frictions.[3] This provides testable implications for our empirical analysis, which uses highly disaggregated industry-level data on job creation and job destruction. We identify the response of these job industry-level flows to uncertainty shocks and document substantial variation. We then relate the estimated responses to measures that capture the strength of various frictions. Using our model-based findings, we argue that the data strongly suggests that financial frictions are important for the transmission of uncertainty shocks, while we find no evidence in support of factor adjustment frictions or price rigidities.

In models with non-convex labor adjustment frictions, plants postpone employment changes under higher uncertainty. Such freezing lowers both job flows. We show that this decline is magnified by larger adjustment costs. If instead capital adjustment is frictional, plants postpone investment in response to higher uncertainty. This unambiguously lowers job creation because non-investing plants shrink due to capital depreciation. Quantitatively, this decline is magnified by larger adjustment costs. In models with nominal rigidities, adjusting plants increase prices and thus markups when uncertainty increases. That is because profits can turn negative if prices are too low, but remain non-negative for too high prices. Assuming that prices are set one period in advance, we find that the job flow responses are mitigated if prices are more rigid. Finally, with financial frictions, higher uncertainty raises the

[3] We study the responses of (gross) job creation and (gross) job destruction. Following Davis and Haltiwanger (1992), we define job creation as the total employment change of plants with net employment gains and job destruction as the total employment change of plants with net employment losses.

probability of (costly) default and thus raises borrowing costs. In return, job creation falls and job destruction increases. Quantitatively, the effects are stronger the more severe financial frictions.

Guided by our model-based insights, we empirically examine the interaction of job flow responses to uncertainty shocks and the severity of various frictions. In this paper, we create a new panel dataset of four-digit manufacuring industry job creation and job destruction series from 1972 to 2013. We estimate job flow responses to uncertainty shocks using a sectoral vector autoregressive (VAR) model. The model includes aggregate variables - stock market level, uncertainty, aggregate job flows - and industry-level job flows. The VAR model is restricted such that the identified uncertainty shock series is common across industries while industry-level responses may differ.

At the aggregate level, job creation decreases while job destruction increases in response to uncertainty shocks. While we find substantial variation in the industry-level responses, 80% of industries exhibit a joint decrease in job destruction and an increase in job creation. This finding by itself is hard to reconcile with (non-convex) labor adjustment frictions playing a key role for the shock transmission since that would imply all employment adjustment, both downward and upward, shrinks in response to uncertainty shocks.

To study the interaction of job flow responses with frictions, we construct industry-level indices that capture the severity of frictions. For example, we consider the within-industry kurtosis of gross investment rates and employment growth as indicators of capital and labor adjustment costs, respectively. The justification is that larger costs lead to more lumpy adjustment, which in turn raises the kurtosis, see, for example, Caballero et al. (1997) and Bachmann and Bayer (2013). Similarly, we construct various industry-level variables that capture the degree of nominal rigidities and financial frictions.

When relating our friction indices to the estimated job flow responses, we find strong evidence in support of financial frictions as transmission channel of uncertainty shocks. In particular, we find that the job flow responses are significantly larger in industries with stronger financial frictions. The relations between the other friction indices and job flow responses are either statistically insignificant, or significant but

of the opposite sign when compared to our model-based predictions. For example, in industries with stronger measured labor adjustment frictions, job destruction increases by more.

This paper relates to recent empirical work examining the effects of uncertainty shocks on labor markets. The evidence concentrates on the effects at unemployment rate (Leduc and Liu, 2014), job finding rate (Guglielminetti, 2013), and separation rate in connection with job finding rate (Riegler, 2014). Our empirical findings complement their results.

The structure of the paper is as follows. Section 3.2 examines the model-based relation of frictions and job flow responses to uncertainty shocks. Section 3.4 describes the data used and created. Our estimation strategy is outlined in Section 3.3. Section 3.5 provides the main empirical results, and Section 3.6 concludes. An Appendix follows.

3.2 Theoretical background: frictions and uncertainty

This section examines the role of various plant-level frictions for the transmission of uncertainty shocks on labor markets. We separately consider plants facing labor adjustment frictions, capital adjustment frictions, price rigidities, and financial frictions. For each friction, we study the response of job flows to a shock that increases uncertainty.

3.2.1 Model primitives

We consider an economy with a unit mass of plants. Each plant i produces output y_{it} using a neoclassical production function that combines capital k_{it} and labor l_{it},

$$y_{it} = l_{it}^{\alpha} k_{it}^{1-\alpha}.$$

Plants are monopolistically competitive and face an isoelastic demand curve,

$$p_{it} = z_{it} \left(\frac{y_{it}}{Y_t} \right)^{-1/\xi} P_t,$$

where z_{it} is a stochastic demand shifter for the product of plant i, Y_t is the aggregate demand, P_t is the aggregate price, and ξ is the elasticity of demand. We assume that idiosyncratic demand z_{it} follows a log-normal AR(1) process

$$\log(z_{it+1}) \;=\; \rho \log(z_{it}) + \epsilon_{it+1}, \quad \epsilon_{it+1} \overset{iid}{\sim} \mathcal{N}(-\sigma_t^2/2, \sigma_t^2).$$

The volatility of demand shocks, σ_t, also follows a first-order Markov process. We adopt the timing convention that changes in uncertainty are observed one period before these affect the distribution of new demand shock realizations, cf. Bloom (2009).

3.2.2 Labor adjustment frictions

There is ample evidence suggesting that non-convex labor adjustment frictions at the plant level are important. For example, Caballero et al. (1997) shows that the distribution of net employment growth at the establishment-level exhibits excess kurtosis, which suggests lumpiness in employment adjustment. Further, indirect inference as in Cooper and Willis (2009) or Bloom (2009) estimate significant non-convex labor adjustment costs.

Such non-convex labor adjustment frictions bear important implications for the impact of uncertainty shocks on labor markets. When uncertainty, broadly about plant profitability, is high, the option value of postponing labor adjustment increases. The intuition is that higher volatility makes reversals in profitability more likely, which might necessitate reversals in employment size. However, such employment reversals need to be evaluated against partially sunk employment adjustment costs, leading plants to avoid such reversals and consequently adjust labor less frequently. In other words, some plants *freeze* their employment adjustment plans.

To understand the relation between uncertainty shocks and the degree of labor adjustment frictions, we consider a dynamic problem of the plant. We assume capital is adjusted every period and abstract from nominal rigidities and financial frictions. We abstract from general equilibrium effects setting $Y_t = P_t = 1$. Let us drop firm and time indices and define revenues net of capital expenditures as

$$CF^l(z,l) \;=\; \max_k \left\{ z l^{\alpha\mu} k^{(1-\alpha)\mu} - (r+\delta)k \right\}, \quad \mu \;=\; (\xi-1)/\xi.$$

We assume that adjustment costs apply for net employment changes as in Cooper and Willis (2009). In the hypothetical presence of exogenous quits, labor adjustment costs were zero if plants adjust employment to offset quits.[4] Employment adjustment costs are given by

$$AC^l(l,l') \;=\; ac_f^l \mathbf{1}\{l' \neq l\} + ac_p^l |l' - l|,$$

where $\mathbf{1}\{\cdot\}$ is an indicator function, ac_p^l denotes partial irreversibility costs, and ac_f^l fixed costs. The wage is w and the problem of the plant is given by

$$V(z,l,\sigma) \;=\; \max_{l'} \left\{ CF^l(z,l') - wl' - AC^l(l,l') + \beta E[V(z',l',\sigma')] \right\}.$$

A shock that increases uncertainty unambiguously lowers both job creation and job destruction on impact. That is due to the timing convention for demand shocks, which exclusively give rise to a real option effect on impact, cf. Bloom (2009). What is less clear is how the job flow response to uncertainty shocks changes in the level of labor adjustment friction. We will come back to this question in Section 3.2.6.

3.2.3 Capital adjustment frictions

Similar to the employment change distribution, the gross investment rates exhibit excess kurtosis and negative skewness, see, for example, Bachmann and Bayer (2013) and Kehrig and Vincent (2016). This suggests lumpy investment as explained by non-convex capital adjustment costs or investment irreversibilities. Given irreversibilities, a shock raising uncertainty raises the option value of waiting and leads plants to freeze investment plans. The capital stock of inactive plants decreases because of physical depreciation. This indirectly depresses labor demand, which lowers job creation. The effect on job destruction is ambiguous. The freezing of plants that had disinvested absent an increase in uncertainty contributes to less job destruction. However, the freezing of plants that had invested absent an increase in uncertainty contributes to more job destruction.

[4] The literature alternatively considers costs in adjusting gross flows in the presence of exogenous quits. However, the combination of exogenous labor attrition and sufficiently high adjustment costs for gross employment changes implies a negative median net employment growth in the cross-section of plants. Empirically, however, this median is non-negative, see Davis and Haltiwanger (1992).

We aim to understand the relation between uncertainty shocks and the degree of capital adjustment frictions. We assume labor is adjusted every period and abstract from nominal rigidities and financial frictions. Analogous to the model above, we define revenues net of labor expenditures as

$$CF^k(z,k) \;=\; \max_l \left\{ z l^{\alpha\mu} k^{(1-\alpha)\mu} - wl \right\}, \quad \mu \;=\; (\xi-1)/\xi.$$

Capital adjustment costs are given by

$$AC^k(k,k') \;=\; k' - (1-\delta)k + ac_f^k \mathbf{1}\{k' \;\neq\; (1-\delta)k\} + ac_p^k \mathbf{1}\{k' \;<\; (1-\delta)k\},$$

where ac_p^k are resale losses, and ac_f^k fixed costs. The problem of the plant is given by

$$V(z,k,\sigma) \;=\; \max_{k'} \left\{ CF^k(z,k) - AC^k(k,k') + \beta E[V(z',k',\sigma')] \right\}.$$

In Section 3.2.6 we show how the job flow response to uncertainty shocks changes in the level of capital adjustment frictions.

3.2.4 Price rigidities

In an economy with monopolistic competition and staggered prices, plants respond to an uncertainty shock by setting a higher price. Important for the upward-pricing result is the asymmetry of the profit function in the price: it is costlier setting a too low relative price than setting a too high relative price. Consequently, less jobs are created and more jobs destroyed. Upward-pricing emerges under price setting à la Rotemberg or Calvo and when prices are set before shocks realize, see Fernandez-Villaverde et al. (2015) and Born and Pfeifer (2016).[5]

To understand the relation between uncertainty shocks and the degree of nominal price rigidity, we study a plant problem subject to Calvo (1983) rigidities. Uncertainty of idiosyncratic demand does not affect plants' price policy. As a shortcut to generate precautionary price setting, we let the aggregate price level be stochastic

[5] On the contrary, in a model with menu costs and time-varying uncertainty as in Vavra (2014), an uncertainty shock increases the frequency and volatility of price changes, resulting in more jobs created and destroyed.

with stochastic volatility. In particular, we assume it follows

$$\log(P_{t+1}) \;=\; \rho \log(P_t) + \epsilon^P_{t+1}, \quad \epsilon^P_{t+1} \overset{iid}{\sim} \mathcal{N}(-\sigma_t^2/2, \sigma_t^2).$$

We abstract from factor adjustment frictions and financial frictions and set $Y_t = 1$. Prices are set before observing realizations of the stochastic states. The dynamic problem of the plant is given by

$$V(z,p,P,\sigma) \;=\; E\left[\theta \max_{p'} \tilde{V}(z',p',P',\sigma') + (1-\theta)\tilde{V}(z',p,P',\sigma')\right]$$

where θ denotes the price adjustment probability, and

$$\tilde{V}(z,p,P,\sigma) \;=\; \left[\frac{p}{P} - \left(\frac{w}{\alpha}\right)^{\alpha}\left(\frac{r+\delta}{1-\alpha}\right)^{1-\alpha}\right]\left(\frac{p}{zP}\right)^{-\xi} + \beta V(z,p,P,\sigma).$$

In Section 3.2.6 we study the role of the degree of price stickiness for the job flow response to uncertainty shocks.

3.2.5 Financial frictions

Fluctuations in uncertainty may affect the economy through financial frictions. Uncertainty may raise default probabilities and thereby increase plant borrowing costs, which in turn depresses economic activity. In addition, liquidity may become more valuable in periods of high uncertainty and plants may optimally downscale to preserve liquidity. In the following, we consider a plant model which highlights the role of financial frictions and liquidity, in the spirit of the model by Arellano et al. (2012).

The key financial friction in our model is that plants cannot borrow against expected future profits. Whenever period profits, dividends, are negative, the plant defaults, irrespective of their continuation value. For simplicity, we abstract from capital and assume technology $y = l$. Plants need to choose employment before observing their demand shock, which implies that profits can turn negative. Plants thus default for a sufficiently low demand shock. The plant may finance its expenditures by issuing a defaultable one-period bond. The debt contract pays b' units when not defaulting, and provides qb' in return. In addition, plants face fixed operating costs

f such that plant dividends are

$$d \;=\; zl^{\mu} - wl - b + q(l', b', z_t, \sigma_t)b' - f \;\geq\; 0, \quad \mu \;=\; (\xi - 1)/\xi.$$

The dynamic problem a continuing plant is given by

$$V(z, l, b, \sigma) \;=\; \max_{b', l', d} d \;+\; \nu\beta E[V(z', l', b', \sigma')],$$

where plants exogenously exits the market with probability $1 - \nu$ every period.[6]

Plants sign one-period loan contracts with a perfectly competitive financial intermediary. The plant saves ($b < 0$) at the risk free rate, with the price of the bond $q(l', b'|z, \sigma) = \beta$. If the plant accumulates debt, default may occur, denoted by indicator $\psi(z, l, b, \sigma) = 0$, while one is no default. In the case of default, the lender recovers part of the outstanding debt by taking possession of the plant with debt reset at zero and at cost η.[7] The parameter η can be thought of as a cost for processing bankruptcy, but also determines how much from plants' value can be collateralized. The price of the bond when borrowing is thus given by

$$\begin{aligned} &q(l', b'|z, \sigma)b' \\ &\quad= \beta E\Big[\psi(z', l', b', \sigma')b' + \big(1 - \psi(z', l', b', \sigma')\big)\min\big\{b', \bar{V}(z', l', 0, \sigma')\big\}\Big], \end{aligned}$$

where $\bar{V}(z', l', 0, \sigma') = \max\big\{V(z', l', 0, \sigma') - \eta, 0\big\}$ is the recovery value. In the next section we study the job flows response to an uncertainty shock for different levels of default costs, η.

3.2.6 Quantitative results

We calibrate the various models at quarterly frequency, in line with the frequency of our empirical analysis. We use standard assumptions for parameters α, ξ, δ, and

---

[6] The assumption of exogenous exit further motivates the use of debt by plants, see, for example, Bernanke et al. (1999) and Gertler and Kiyotaki (2010). We assume defaulting or exiting plant are replaced by an identical plant with zero debt next period.

[7] We express η proportional to steady state revenues. As in Gilchrist et al. (2013), the exit rate does not affect the loan price. We implicitly assume plants make payment decisions before the exit shock realizes.

r. Based on Bloom et al. (2014), we calibrate the stochastic demand process and assume stochastic uncertainty, σ_t, follows a two-state Markov chain. We denote low and high uncertainty by σ_L and σ_H, respectively. The transition probability, e.g. from low to high uncertainty, is denoted $\pi^{\sigma}_{L,H}$. As in Cooper and Willis (2009), we calibrate the wage rate to match the average plant employment of 600 workers observed in the Longitudinal Research Database. Finally, following Gilchrist et al. (2013), we calibrate the exogenous exit rate based on establishment entry and exit tabulations from the Business Employment Dynamics, and the fixed operation costs based on the ratio of general expenses to sales from Compustat data.[8]

Table 3.1. Model parameters

Parameter	Value	Explanation
α	0.65	Labor share in the economy
ξ	4	Markup of 33%
β	0.99	Discount factor at quarterly frequency
δ	2.6%	Annual capital depreciation of 10%
r	1.01%	Annual risk free interest rate of 4%
ρ	0.95	Serial correlation process
σ_L	0.051	Baseline uncertainty level
σ^H	4 x σ	High uncertainty is four times the baseline level
$\pi^{\sigma}_{L,H}$	0.03	Probability from low to high uncertainty
$\pi^{\sigma}_{H,H}$	0.92	Probability of remaining at high uncertainty state
All models except financial frictions		
w	0.113	Average plant employs 600 workers
Model with financial frictions		
w	0.07	Average plant employs 600 workers
ν	0.95	Survival probability
f	8	Fixed costs (11% of revenues)

[8] Based on firm-level data from Compustat, the median ratio of sales, general, and administrative expenses to sales is 22%. As Gilchrist et al. (2013), we assume that 50% of those expenditures represent fixed costs of operations.

After solving the dynamic problems, we independently simulate 5,000 economies with 1,000 plants, for 80 quarters, respectively. Each economy is hit with an uncertainty shock in the same quarter. In order to calculate the average response of aggregate job flows to an uncertainty shock, we average across simulated economies.

Figure 3.1 shows the immediate response of job flows to an uncertainty shock under different degrees of frictions. Stronger labor adjustment frictions increase the magnitude of the job creation and job destruction response to an uncertainty shock. While it is well known that adjustment frequency falls in adjustment costs, and thus job creation and destruction fall, we find that adjustment costs positively interact with uncertainty shocks and magnify the responses. Critically important for this result is the change in the employment adjustment triggers in response to higher uncertainty. If the relative increase of these triggers is at least as large under high capital adjustment costs as under low capital adjustment costs, then on impact job creation and job destruction must fall by more under high costs. That is because under high costs, plants' employment policies are such that adjustment frequency is lower even in normal times. Thus the probability mass of plants close to adjustment triggers is lower. If the triggers move by at least as much under high costs as under low costs, then the share of plants that adjust drops by more under high costs.[9]

As explained in Section 3.2.3, with capital adjustment frictions the immediate response of job creation to higher uncertainty is an unambiguous drop, while the response of job destruction is less clear. We find that job creation falls by more the large capital adjustment costs are. The explanation is analogous to our discussion on labor adjustment frictions. As for job destruction, however, we indeed find an ambiguous effect of capital adjustment costs on the immediate response to uncertainty.

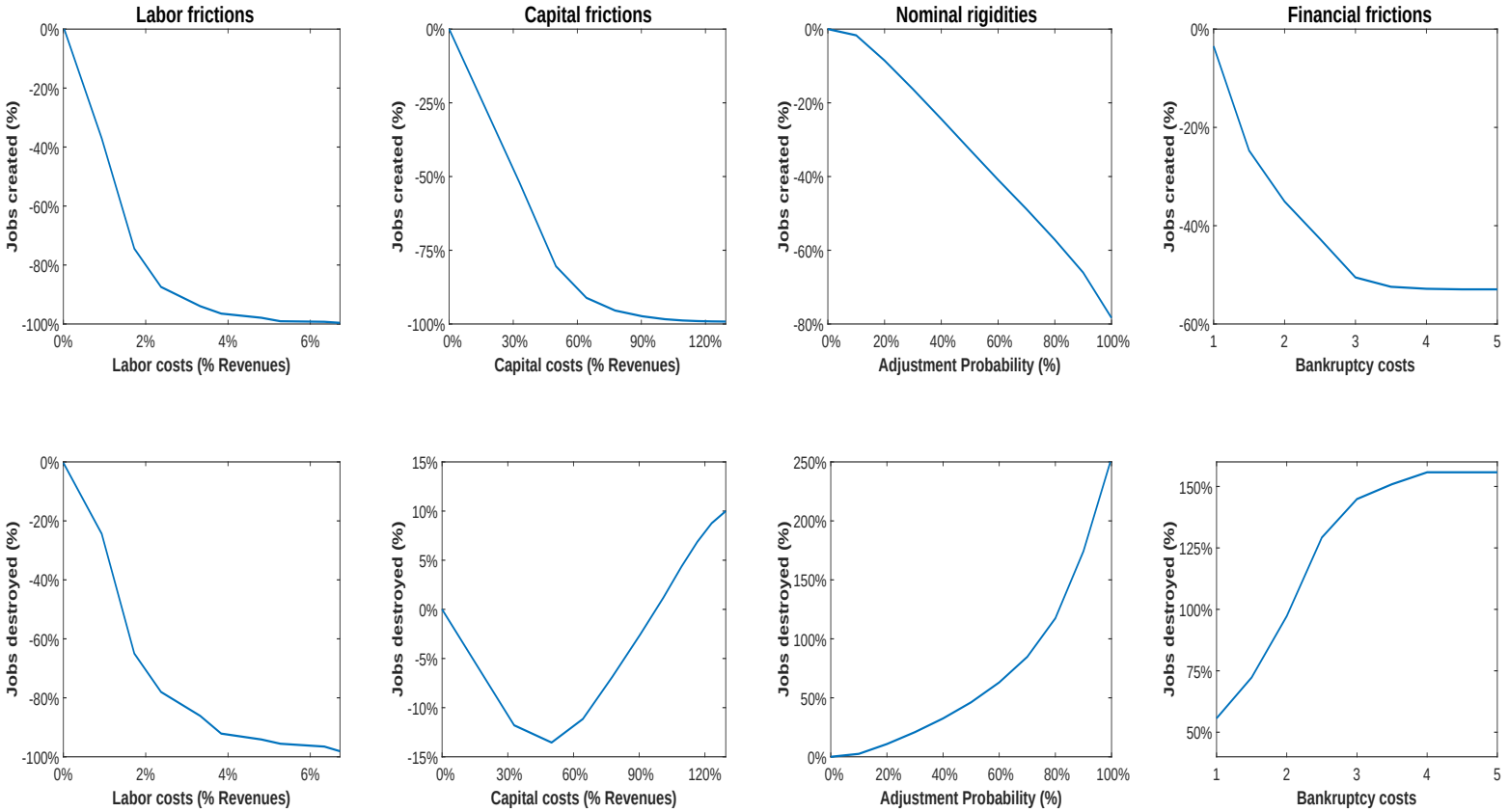
Under nominal rigidity, upward-price setting leads to lower job creation and higher job destruction in response to higher uncertainty. We find that these response are stronger the less rigid prices are. Notice that in our setup even when prices can be adjusted every period ($\theta = 0$), prices are set before observing the period profitability. Thus, upward-price setting exists even for $\theta = 0$. The longer-lived prices are, i.e. the

[9] The analytical results in Abel and Eberly (1996) lend support to these quantitative results. Applying their results to our model with labor adjustment frictions by setting $\delta = 0$, we can show that the shift of adjustment triggers in response to higher uncertainty is magnified through the level of adjustment costs.

larger θ, the less are prices set only for a period of high uncertainty, but also for subsequent periods during which plants expect lower uncertainty. This weakens the incentive for upward-price setting. In addition, if less plants adjust prices, less plants can raise them. This explains the mitigated responses of job flows to uncertainty shocks when prices rigidity increases.

In an economy with financial frictions, higher uncertainty increases the probability of default. Therefore, the bond price declines, which leads to more jobs destroyed and less jobs created. Importantly, the costlier default is, the more the bond price declines in response to a shock that raises uncertainty. As a result, the adverse responses of job flows to an uncertainty shock are magnified by the costs of default. Alternatively, we could consider an economy where borrowing is subject to collateral constraints. Higher uncertainty makes default more likely, which induces lenders to charge a higher risk premium and raise collateral requirements. Consequently, plants scale down production.

Figure 3.1. First-quarter response of job flows to an uncertainty shock for varying degrees of frictions



Notes: The figures show the percentage change of aggregate jobs flows in the first quarter after an uncertainty shock hits when varying the degree of labor adjustment frictions, capital adjustment frictions, price rigidity, and financial frictions. Labor/Capital costs (% Revenues) denote the average costs of adjusting plants relative to revenues. Bankruptcy costs are proportional to optimal revenues in the steady state. The model with labor frictions is solved for $ac_f^l \in \{0, 0.0005, ..., 0.05\}$ and $ac_p^l \in \{0, 0.1, .., 1\}$, with capital frictions for $ac_f^k \in \{0, 0.0005, ..., 0.05\}$ and $ac_p^k \in \{0, 0.1, .., 1\}$, with price rigidity for $\theta \in \{0, 0.1, .., 1\}$, and with financial frictions for $\eta \in \{1, 0.5, ..., 5\}$.

3.3 Empirical estimation strategy

This paper aims to provide empirical evidence on the transmission of uncertainty shocks with a focus on labor markets. Our empirical approach builds on the structural relation between various frictions and the transmission of uncertainty shocks established in the previous section. In particular, our estimation strategy is to exploit variation in the job flow responses to uncertainty shocks of narrowly defined industries and link variation in the estimated responses to industry-level observables that are informative about the severity of the frictions we consider.

To identify industry-level responses of job flows to uncertainty shocks, we consider a structural vector autoregressive (VAR) model

$$Y_t = A_0 + Dt + \sum_{j=1}^{p} A_j Y_{t-j} + Bu_t \tag{3.1}$$

where Y_t denotes a vector of endogenous variables, A_0 is a vector of constants, D is a vector capturing time trends, $A_1, \ldots, A_p$ are lag matrices of slope coefficients up to maximum lag length p, u_t are structural shocks with the identity matrix as covariance matrix, and B is the matrix of structural coefficients such that Bu_t are reduced-form residuals.

The model is six-variate including aggregate and industry-specific variables, specifically the log of the S&P 500 stock market index, uncertainty, the log of aggregate manufacturing job creation and destruction and the log of industry-specific job creation and destruction. We include the aggregate variables to make sure we identify responses in job flows that are industry-specific. To avoid identifying different shocks from the same uncertainty series, we do not allow industry-specific job flows to feed back into aggregate time series. To be precise, we restrict the slope coefficients that capture such feedbacks to be zero. Denoting by $A_j^{k,l}$ the slope coefficent that captures the effect of variable k on variable l at lag j, then the restrictions we impose are

$$A_j^{ijc,agg} = A_j^{ijd,agg} = 0, \quad \forall\, j = 1, \ldots, p, \quad \text{and } \forall agg = s, u, jc, jd,$$

where ijc and ijd denote industry-specific job creation and destruction, respectively, whereas aggregates s, u, jc, jd denote the stock market, uncertainty and aggregate job creation and destruction, respectively. This approach mimics Davis and Haltiwanger (2001), who identify the effect of oil price shocks on industry-level job flows.

To identify uncertainty shocks, we impose restrictions on B. We assume that stock market level is not contemporaneously affected by an uncertainty shock, while shocks that contemporaneously affect the stock market may also contemporaneously affect uncertainty. All other variables in the VAR model may be contemporaneously affected by uncertainty shocks. Following convention of the VAR literature, we let the uncertainty shock be the second structural shock. The impulse response to the uncertainty shock are then identified by the second column of B, here denoted B_2. The recursive identification scheme exactly identifies B_2.

Using this identification scheme, we directly estimate the impulse responses to uncertainty shocks using local projections, see Jorda (2005). Compared to the alternative of estimating the VAR model and inverting it to obtain impulse responses, the local projections is more robust to model misspecification and importantly more flexible to handle nonlinearities or the extra zero restrictions we impose on the reduced-form model. We can use local projection to estimate reduced-form impulse responses and transform these into structural impulse responses. To obtain reduced-form impulse responses we project the vector of endogenous variables at different forecast horizons $h = 0, 1, \ldots, H$ on its own lags

$$Y_{t+h} \;=\; \mu^h + \delta^h t + \sum_{j=1}^{p} M_j^{h+1} Y_{t-j} + v_{t+h}, \tag{3.2}$$

where v_{t+h} is the forecast error. For consistency with the VAR model, we restrict $M_j^{h+1,ijc,agg} = M_j^{h+1,ijd,agg} = 0$, $\forall\ j = 1, \ldots, p$, $\forall\ h = 0, 1, \ldots, H$, and $\forall\ agg = s, u, jc, jd$. Matrices M_1^h contain the reduced-form impulse responses, while the structural impulse response to the uncertainty shocks at horizon h are given by vector

$$M_1^h B_2. \tag{3.3}$$

After collecting the impulse responses of job creation and job destruction for all industries, we construct indices that capture the strength of the frictions we consider. We then investigate whether cross-industry variation in these indices predicts the estimated cross-industry variation in job flow responses to uncertainty shocks. We propose to test for the importance of various transmission channels by whether the responses of both job creation and job destruction significantly varies in the strength of frictions with the sign as suggested by theory in the preceding section.

3.4 Data

This section describes the data used in the empirical application. In addition to aggregate data used in the VAR model, we use industry-level data on job flows, and data to capture cross-industry variation in the strength of labor adjustment frictions, capital adjustment frictions, price rigidities, and financial frictions.

3.4.1 Uncertainty

Uncertainty is not directly observable. Instead we consider the series of conditional foracast uncertainty over macroeconomic and financial variables as estimated in Jurado et al. (2015) and Ludvigson et al. (2015). Importantly, these uncertainty measures condition on predicted variability. Macroeconomic uncertainty is based on real economic activity, prices, bond and stock market indexes, among others. Financial uncertainty is based on financial variables, such as credit spreads, valuation ratios, risk factors. Our baseline analysis uses the quarterly averaged macroeconomic uncertainty.[10] Appendix 3.A.3 shows the robustness of our results when using financial uncertainty.

3.4.2 Job flows

A secondary contribution of this paper is to construct a panel of quarterly industry-level job flows from 1972 until 2013. We do so by combining data from two sources, the Longitudinal Research Database (LRD) and the Quarterly Workforce Indicator

[10] We use the 3-month ahead conditional forecast uncertainty, but our findings are robust against alternative forecast horizons.

(QWI). LRD-based industry-level data on job flows are made publicly available by Davis et al. (1998). The panel covers the years 1972-1998, and provides industry disaggregation at the 4-digit SIC level, totaling 456 manufacturing industries.[11] We complement this panel with the QWI, publicly available through the United States Bureau of the Census (2015). The QWI measures worker and job flows disaggregated at the 4-digit NAICS level.[12] The underlying data is provided by states. State participation in the QWI has not been complete initially, and has increased over time. We consider only those states who started providing information before 2000Q2. The selected sample constitutes 90% manufacturing employment in United States. We use X-13 ARIMA to remove the seasonal component from the series. To create a common industry classification, we use a correspondence table, provided by the National Bureau of Economic Research, which maps NAICS into SIC codes. Notice that our final panel has a gap from 1999 to 2001.

Figure 3.2 shows the aggregate time series of manufacturing employment based on our new panel compared to the corresponding series published by the Bureau of Labor Statistics (BLS). As expected, employment is about 90% lower in the QWI-based series because of some missing states. Yet, the two series display strong comovement with the correlation being 98%. We refer the reader to Appendix 3.A.2 for additional details about this data.

3.4.3 Indices of industry-level frictions

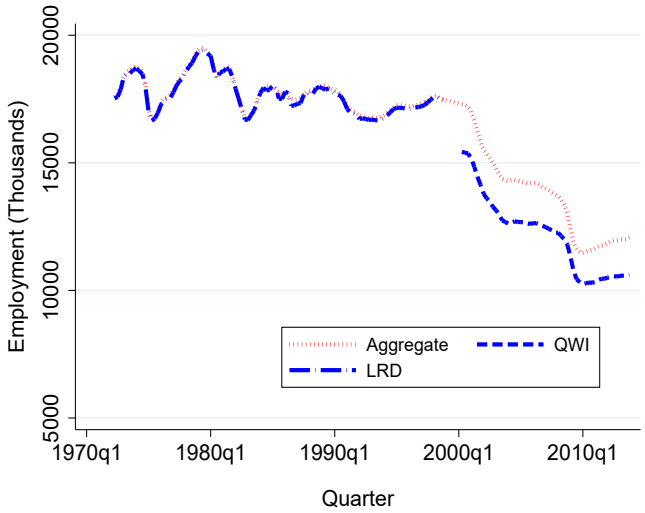
To provide empirical evidence on the transmission mechanisms of uncertainty shocks, we propose cross-sectional indices intended to capture the strength of factor adjustment frictions, price rigidities, and financial frictions at the industry-level. These indices will be constructed such that higher index values imply higher degree of frictions in a given industry.

---

[11] The LRD collects employment data from all US manufacturing plants with at least five employees and accounts for more than 99% of total manufacturing employment.

[12] The QWI is based on the Longitudinal Employer-Household Dynamics (LEHD). The LEHD consists of linked employer-employee data covering over 95% of US private sector jobs. It considers employer's state-specific UI account number as the business identifier.

Figure 3.2. Aggregate employment in US manufacturing



Notes: Manufacturing employment based on the connected industry sample from *LRD* and *QWI*. *Aggregate* manufacturing employment based on tabulations from the Bureau of Labor Statistics.

3.4.3.1 Labor adjustment frictions

Labor market regulation tends to be common to all industries within the same country. Yet, the cost of raising hours worked, hiring or firing workers may differ across industries. Following Botero et al. (2004), we use the industry-level share of workers in full-time positions as measure of the flexibility of employment contracts and the cost of firing workers. Further, we include the share of workers affiliated to labor unions to capture union power at the industry.[13] Finally, we consider the industry-level kurtosis of the cross-sectional net employment growth distribution based on Compustat data. As argued in Section 3.2.2, under the presence of non-convex employment adjustment costs, labor adjustments are infrequent and lumpy, which implies excess kurtosis.[14]

[13] The share of full-time workers and union density are based on the March Supplements from the Current Population Survey, see Table 3.2. In order to map the industry classification from CPS into 1987 SIC we use a concordance table provided by David Dorn at http://www.dorn.net/data.htm.

[14] We compute the kurtosis of net employment growth at the industry if we have at least ten observations at the industry. Given that we lack of sufficient information at the 4-digit SIC level, we compute it at the 3-digit SIC level. Furthermore, for 10% of the industries, the information at the 3-digit SIC level is not available, so we impute it by the mean value at the 2-digit SIC level.

3.4.3.2 Capital adjustment frictions

Capital adjustment costs can be estimated through indirect inference using the distribution of cross-sectional gross investment rates, see, for example, Cooper and Haltiwanger (2006) and Bachmann and Bayer (2014). A striking feature of the cross-sectional distribution is substantial positive skewness and excess kurtosis. Non-convex capital adjustment costs can account well for these empirical observations, because they lead to lumpy investment, generating excess kurtosis, while in between adjustments the capital stock depreciates leading to positive skewness. Importantly, the large skewness and excess kurtosis, the larger adjustment costs must be to match these moments. Using Compustat data, we consider the within-industry skewness and kurtosis of the gross investment rate distribution to capture the degree of capital adjustment frictions.[15] In addition, we consider the ratio of structures over equipment at the industry. Since structures are more costly to adjust than equipment capital, see, for example, Caballero and Engel (1999), a large structure share implies larger capital adjustment costs for a given total stock of capital.

3.4.3.3 Price rigidity

We measure industry-level price rigidity using the estimates in Petrella and Santoro (2012). The authors consider sector-specific New Keynesian Philips Curve (NKPC) to back-out the degree of price rigidity. They evaluate alternative measures of marginal costs and different specifications of the sectoral NKPC. We consider the average sectoral price adjustment probabilities based on the intermediate input share as a proxy for marginal costs, which fits best their model predictions.

3.4.3.4 Financial frictions

Industries differ in their liquidity and borrowing needs, which affect their vulnerability to financial frictions. Following Raddatz (2006), we construct measures of liquidity needs as the industry-level median ratio of inventories to sales, and labor

[15] We compute these moments if we have at least ten observations for a industry. Given that we lack of sufficient information at the 4-digit SIC level, we compute the indicators at the 3-digit SIC level. Furthermore, for 10% of the industries, the information at the 3-digit SIC level is not available, so we impute it by the mean value at the 2-digit SIC level.

costs to sales.[16] These variables capture the share of inventory investments or labor costs which can be commonly financed by revenues. The larger these ratio, the higher the need for external finance. In principle, the constructed ratios may not be entirely technological. For example, businesses may opt to accumulate liquid assets to avoid financial dependence. To circumvent this problem, we follow the literature and construct the measures using information from publicly traded U.S. companies. The underlying assumption is that observed industry differences at these large publicly traded companies are not driven by the supply of credit, which is assumed to be perfectly elastic.

We complement this information with the employment share of young firms (below 5 years old) per industry. There is ample evidence which associates larger borrowing costs to young businesses, as they have higher degree of idiosyncratic risk, lower amount of collateral, and shorter credit records.[17]

3.4.3.5 Industry-level indices

Table 3.2 summarizes the variables we use to capture cross-industry variation in the severity of various frictions. We aggregate the information available by creating industry-level indices for the severity of each friction. Our baseline approach is to average the variables after standardizing them. For example, the baseline labor friction index for a given industry is the mean over the standardized values of the share of full-time workers, the unionization rate, and the employment growth kurtosis.[18] Our final panel includes 443 manufacturing industries. Table 3.3 presents the correlations between the indexes. On the one hand, industries with larger capital adjustment costs tend to have larger labor adjustment costs. On the other hand,

---

[16] We consider only those industries with at least 5 firms at the industry. Given that we lack of sufficiently information at the 4-digit SIC level, we compute the indicators at the 3-digit SIC level. Furthermore, for 10% of the industries, the information at the 3-digit SIC level is not available, so we impute it by the mean value at the 2-digit SIC level.

[17] The literature used size of the business as alternative indicator for access to credit and tightness of the borrowing constraint. However, there is recent evidence which suggests that financial frictions do not lead to different business dynamics across firm size, once controlling by the age of the firm. See for example Hurst and Pugsley (2011), Dyrda (2015), and Fort et al. (2013).

[18] As alternative to such index, we estimate the first principal component over the set of included variables. Our main findings are robust to this specification, see Appendix 3.A.3.2.

industries more vulnerable to financial conditions tend to exhibit lower factor adjustment frictions.

Table 3.2. Variables used to measure the strength of frictions at the industry-level

Variable	Source
Labor adjustment frictions	
Share of (35 hours or more) full-time workers	March CPS: 1970-2011
Unionization rate of workers	March CPS: 1990-2011
(Within-industry) net employment growth kurtosis	Compustat: 1968-2006
Capital adjustment frictions	
(Within-industry) gross investment rate skewness	Compustat: 1968-2006
(Within-industry) gross investment rate kurtosis	Compustat: 1968-2006
Share of structure per equipment capital	NBER-CES: 1958-2011
Price rigidities	
Price adjustment probability (model-based estimates)	Petrella and Santoro (2012)
Financial frictions	
Share of inventory value over sales	Compustat: 1968-2006
Share of labor cost over sales	Compustat: 1968-2006
Employment share at young firms (below 5 years)	QWI: 2000-2013

Table 3.3. Correlation between indexes

	Labor index	Capital index	Price index	Financial index
Labor index	1			
Capital index	.268***	1		
Price index	-.022	−.03	1	
Financial index	−.311***	−.083*	0.087*	1

Notes: This table presents pairwise correlations between our indexes. See Table 3.2 for for a detail description of the industry indexes. Significance: 1% (***), 5% (**), 10% (*).

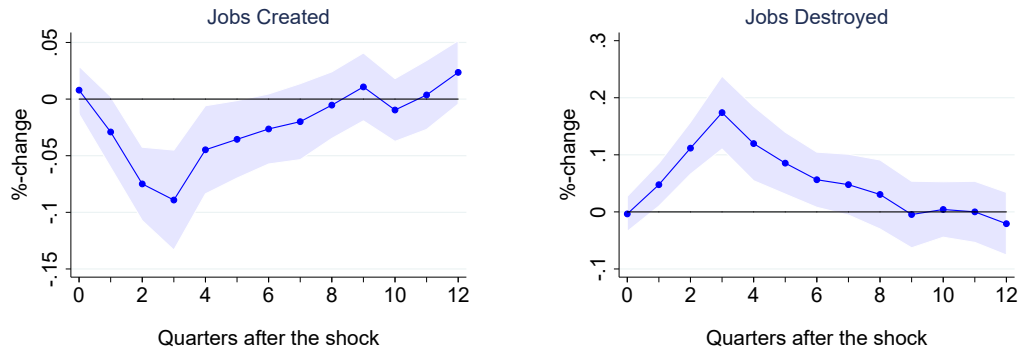
3.5 Empirical evidence on the transmission of uncertainty shocks

3.5.1 Job flow responses to uncertainty shocks

We fit our VAR model in (3.1) using quarterly data from 1972Q2 to 2013Q4 and assuming four lags. To account for the missing states in the QWI data used, we add a step dummy to (3.1), which takes the value of one from 2000Q2 onwards, and zero otherwise. Further, we allow for different time trends in the first part of the panel (1972-1998) and the second part (2000-2013).

Figure 3.3 shows the effects of an uncertainty shock on aggregate job creation and job destruction. Note that job creation significantly falls while job destruction significantly increases. This finding is of interest by itself. In particular, in a simple model where net employment change is subject to non-convex costs, we should expect job destructions to fall in response to higher uncertainty, because plants freeze, see Section 3.2.2. Therefore, the response of the aggregate series suggests that such employment frictions are likely not key for the transmission of uncertainty shocks, which further motivates the subsequent industry-level analysis.

Figure 3.3. Response of aggregate job flows to uncertainty shock

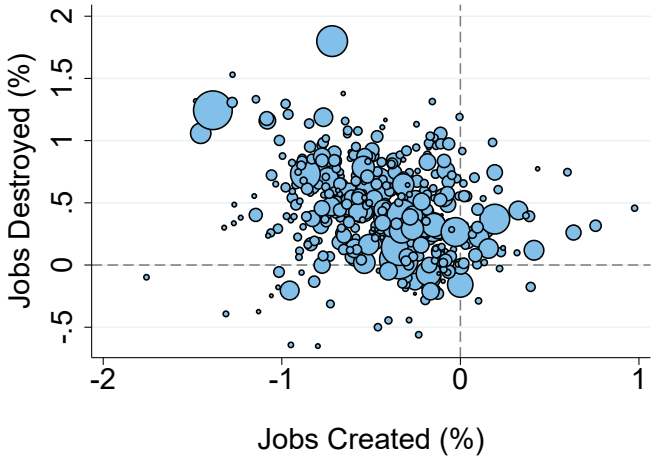


Notes: The blue lines show the responses of aggregate manufacturing job flows to a positive, three-standard deviation uncertainty shock. The shaded area is the 90% confidence interval using block bootstraps as in Kilian and Kim (2009).

To compress the information contained in the industry-level impulse response functions, we focus on the average response within the first year after the uncertainty shock hits. We restrict attention to the short-term responses because in models with factor adjustment frictions the real options effect, i.e. plant freezing, has a predominantly short-term effect on the

response of job flows to uncertainty shocks. Figure 3.4 shows the cross-sectoral variation in job flow responses to uncertainty shock. Reconfirming the result in Figure 3.3, we find for 80% of the industries a joint decline in job creation and an increase in job destruction.

Figure 3.4. Cross-industry variation in job flow responses to uncertainty shock



Notes: Response of job flows averaged over the first year horizon to a three-standard deviation uncertainty shock. Marker size is proportional to employment of an industry.

3.5.2 Transmission channels

Given the estimated aggregate and industry-level responses, the central question we ask is through which channel uncertainty shocks affect labor markets. To tackle this question, we relate the first-year average industry-level impulse responses to the labor frictions index, capital frictions index, price rigidity index, financial friction index as constructed in Section 3.4.3.

In Table 3.4, we rank industries according to the indices and provide mean and standard error of the job flow response of those industries in the first (bottom) quintile and those industries in the fifth (top) quintile. Only for financial frictions, the responses across quintile groups are both significantly different and the differences correspond to theoretical predictions: The stronger financial frictions, the larger the drop in job creation and the larger the rise in job destruction. For labor frictions, the quintile-group difference for job destruction is significant but of the wrong sign: Empirically, industries in which our index suggest stronger labor adjustment frictions destroy more jobs. If labor adjustment frictions are the key frictions for the transmission of uncertainty shocks on labor markets, we would expect the opposite

sign. We find a significant negative effect of capital frictions for the job destruction response. Yet, our theoretical prediction were ambiguous. To summarize, while we do find evidence in support of financial frictions being key for the transmission of uncertainty shocks, we do not find such supportive evidence for any of the other three frictions.

Table 3.4. Job flow responses and quintiles of friction indices

	Job creation		Job destruction	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
Labor frictions index	-0.64	-0.75	0.49	0.85
	(0.04)	(0.04)	(0.05)	(0.04)
Capital frictions index	-0.55	-0.60	0.68	0.54
	(0.04)	(0.05)	(0.04)	(0.05)
Price rigidity index	-0.70	-0.70	0.66	0.62
	(0.05)	(0.04)	(0.05)	(0.04)
Financial frictions index	-0.45	-0.61	0.26	0.64
	(0.06)	(0.04)	(0.06)	(0.04)

Notes: Bottom (Top) quintile: First-year average job flow response of industries in the first (fifth) quintile of the cross-industry distribution of a given index. Standard errors are in parenthesis.

Complementing the quintile-based friction-by-friction analysis in Table 3.4, we estimate a regression of the job flow response on cubic polynomials of all four friction indices jointly.[19] Figure 3.5 shows the fitted relationships between the job flow responses and each individual friction index, while evaluating the remaining indices at their median. The results broadly reconfirm and strengthen the findings in Table 3.4. We do not find supporting evidence of capital frictions and price rigidities as channels through which uncertainty affects job flows. While the share of jobs destroyed declines with the size of capital frictions at the industry, there is no significant relation between the decline in job creation and the size of capital adjustment frictions. Similarly, the relation between price rigidity and industry-level job flow responses is almost flat. On the contrary, financial frictions seems to play an important role for explaining the effects of uncertainty shocks. The more severe financial frictions in an industry, the stronger the response of job flows from an uncertainty shock. The role of financial frictions is quantitatively important: In response to a three-standard deviation uncertainty shock, job creation barely falls for industries with weak frictions, while it falls by up to 0.7% for strong

[19] Davis and Haltiwanger (2001) inspired us to conduct this exercise.

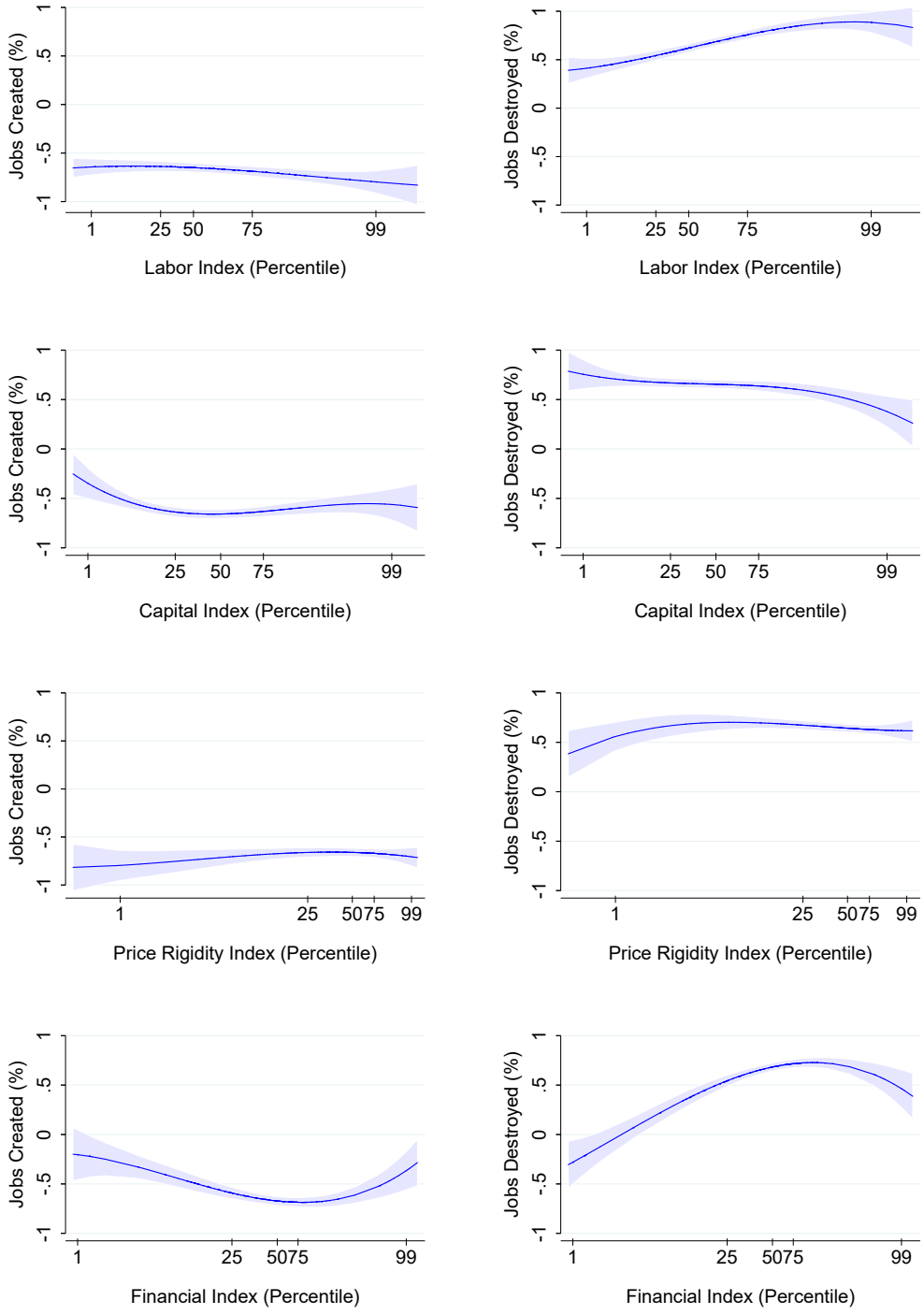
frictions. Similarly, Job destruction does not increase for weak-friction industries, while it increases by up to 0.7% for strong-friction industries.

3.5.3 Robustness of empirical findings

Our main empirical findings, in particular the empirical support of financial frictions as transmission mechanism for uncertainty shocks, is robust along various dimensions. In Appendix 3.A.3, we show that the results are robust against substituting macroeconomic uncertainty based on Jurado et al. (2015) by a financial uncertainty series based on Ludvigson et al. (2015). Moreover, instead of constructing indices to capture the strength of various frictions across industries, we extract the first component form the set of variables considered for the various frictions. In the appendix, we show that this leaves our conclusions broadly unaffected. To address concerns about the construction of our job flow panel, we separately consider the LRD-based panel from 1972 to 1998 and the QWI-based panel from 2000 to 2013. The former panel reconfirms our baseline findings. This also addresses concerns whether our results are driven by the extraordinary uncertainty spike during the Great Recession. For the second sample, we receive the same qualitative results, with the only exception that the job creation response varies insignificantly across quintiles of different degrees of financial frictions. Obviously, using a 14-year panel to study business cycles is challenging in terms of statistical power.

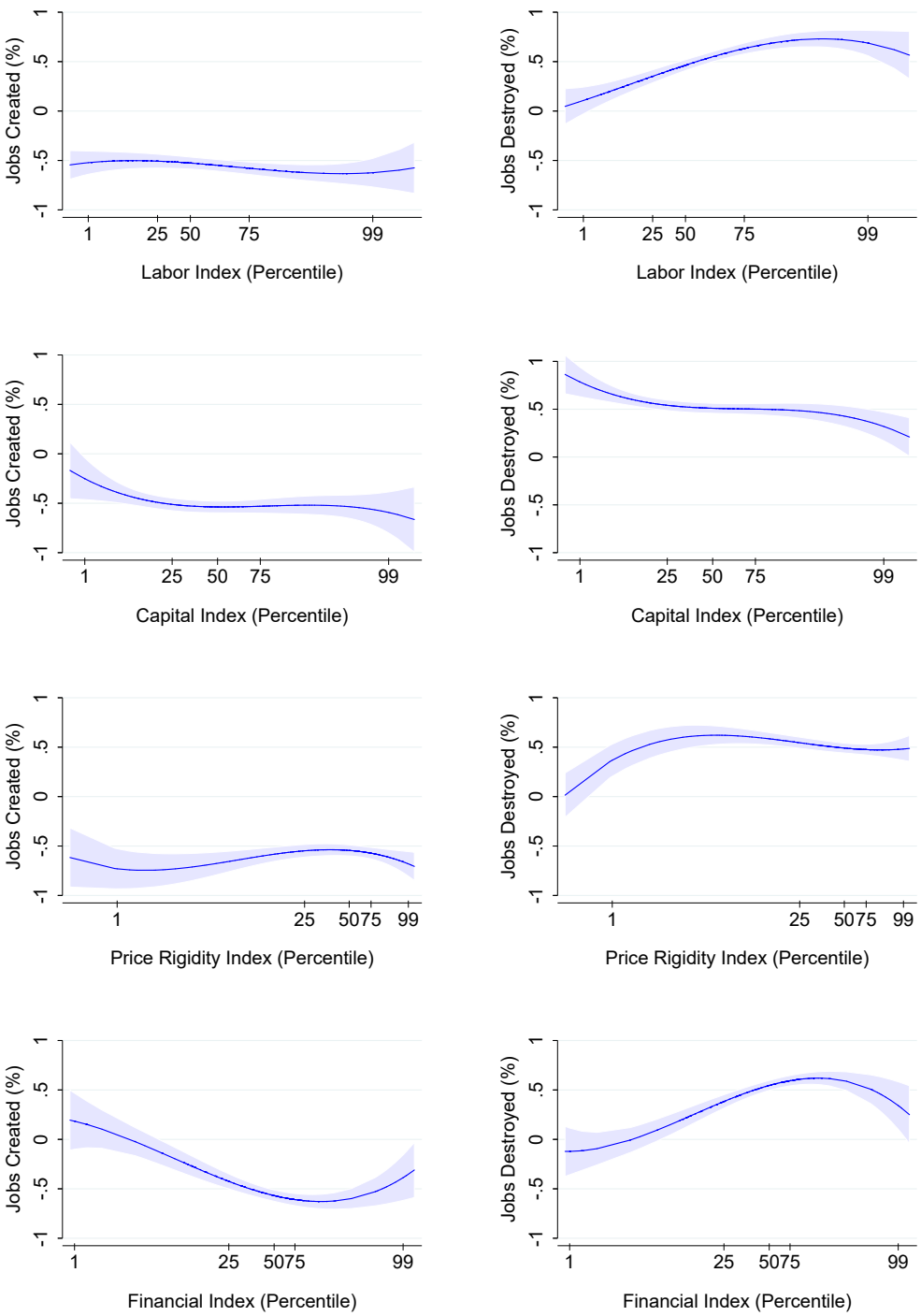
In addition, we assess whether our findings are robust in a richer VAR system which explicitly controls for monetary and fiscal shocks. We augment our baseline specification with monetary and fiscal shocks identified through narrative approaches. For the former, we include the shocks series by Coibion et al. (2012), for the latter the series in Mertens and Ravn (2014). Data availability limits this analysis until 2006Q4. We place the tax and monetary shocks first in the recursive ordering. We find that in more than 60% of the industries, an uncertainty shock leads to a joint increase in job destruction and a decrease in job creation. Figure 3.6 shows that the relation between job flow responses and financial vulnerability remains significant and of similar quantitative magnitude when compared to our baseline results.

Figure 3.5. Nonlinear relation between job flow responses and friction indices



Notes: The figures are based on regressions of industry-level job flow responses on cubic polynomials of all friction indices. Blue lines show the estimated relation between job flow responses and one friction index when keeping the other friction indices at their medians, respectively. Shaded areas denote 90% confidence interval. We weight industry-level responses by the estimated absolute effect relative to its standard error.

Figure 3.6. Nonlinear relation between job flow responses and friction indices when explicitly controlling for monetary and fiscal shocks



Notes: The figures are based on regressions of industry-level job flow responses on cubic polynomials of all friction indices. Blue lines show the estimated relation between job flow responses and one friction index when keeping the other friction indices at their medians, respectively. Shaded areas denote 90% confidence interval. We weight industry-level responses by the estimated absolute effect relative to its standard error.

3.6 Conclusion

This paper revises a number of transmission channels for uncertainty shocks studied in the literature, in particular labor adjustment frictions, capital adjustment frictions, price rigidities, and financial frictions. Focusing on labor markets, we provide empirical evidence on the aggregate and industry-level response of job flows to uncertainty shocks. The key contribution of this paper is to exploit the cross-industry variation in job flow responses to provide evidence for the transmission mechansims of uncertainty shocks.

We create industry-level data on job flows for 1972-2013 in the US and find that a positive uncertainty shock jointly raises job destruction and lowers job creation in 80% of the industries. The (absolute) magnitude of these responses strongly and significantly responds to how vulnerable industries are to financial conditions, which supports financial frictions as transmission channel of uncertainty shocks. On the contrary, we do not find evidence in support of factor adjustment frictions or price rigidities as transmission channels of uncertainty shocks.

Appendix 3.A Appendix

3.A.1 Computation

We solve the models by discretizing the state space. For the exogenous demand process z, we apply a Tauchen discretization with 19 grid points while taking into account time-varying volatility, see Bloom et al. (2014). For the model with labor adjustment frictions, we use 1,000 log-linear grids points for labor. For the model with capital adjustment frictions, we use 1,000 grid points for capital. In the model with price rigidities, we further discretize the exogenous aggregate price process P with 19 grid points, applying again a Tauchen algorithm, and consider 1,000 grid points for the price of the plant. Finally, for the financial frictions model, which is relatively more complex than the other dynamic problems, we discretize the demand process with 16 grid points, and consider 40 log-spaced grid points for labor, and 48 equi-distant grid points for debt. We solve the models using value function iteration.

To numerically solve the model with financial frictions in Section 3.2.5 we consider the following steps: First, guess the price for the bond the value of the plant. Second, given the bond price, calculate the value function using value function iteration. Third, using the updated value function, construct the default function and update the bond price. Fourth, based on the updated bond price, return to point 1 until the price of the bond convergences.

3.A.2 Data description

We use a concordance table provided by the National Bureau of Economic Research (NBER) to connect the LRD-based information at the 4-digit 1987 SIC level with the data series from QWI, disaggregated at the 4-digit 2007 NAICS level, see http://www.nber.org/nberprod/. We create consistent a consistent industry classification using this concordance table together with weights that reflect the share of employment at the SIC level which corresponds to an industry in NAICS. Before proceeding with this concordance, we need to conduct some adjustments. First, the available concordance between SIC and NAICS is based on the 1997 NAICS. Therefore, we translate 6-digit 2007 NAICS into 6-digit 1997 NAICS using the table given by US Census Bureau at http://www.census.gov/eos/www/naics/concordances/concordances.html. Second, we adjust the concordance table from the 6-digit NAICS level to the 4-digit NAICS level, and re-compute the weights from SIC into NAICS based on the share of employment of the 6-digit NAICS industry at the 4-digit NAICS level. At the end, we are

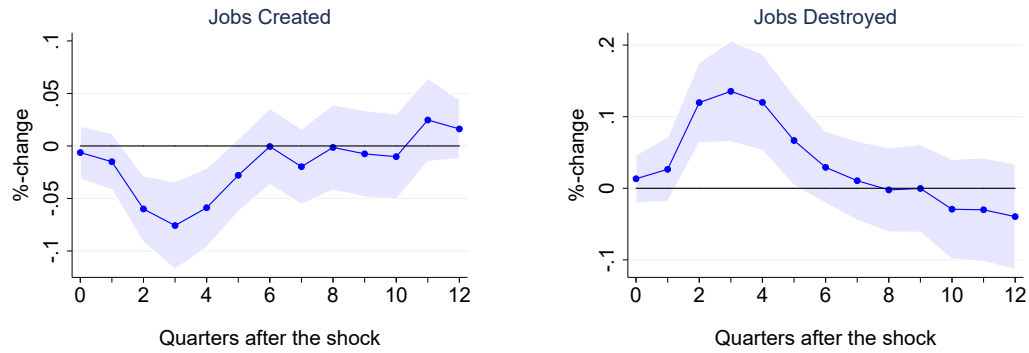
able to map all industry-level job flows from NAICS with the LRD-based data from Davis et al. (1998).

3.A.3 Robustness

3.A.3.1 Job flow responses from a financial uncertainty shock

Figures 3.7 and Table 3.5 show the effect of uncertainty shocks on job flows when using financial uncertainty as a proxy for uncertainty. Ludvigson et al. (2015) distinguish between macroeconomic and financial uncertainty where the latter seems to have larger negative impact on real economic activity. Yet, the quantitative effect on job flows and its relation with the industry indexes are remarkably similar to our baseline results, when using as a proxy macroeconomic uncertainty.

Figure 3.7. Response of aggregate job flows to financial (instead of macroeconomic) uncertainty shock



Notes: The blue lines show the responses of aggregate manufacturing job flows to a positive, three-standard deviation uncertainty shock. The shaded area is the 90% confidence interval using block bootstraps as in Kilian and Kim (2009).

3.A.3.2 Results based on the first principal component

As alternative to the baseline indices, mean of standardized variables, to capture the strength of various frictions, we consider the first principal component of of the variables used to capture frictions. Our results are remarkably robust against these alternative indices.

Table 3.5. Job flow responses and quintiles of friction indices when using financial (instead of macroeconomic) uncertainty

	Job creation		Job destruction	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
Labor frictions index	-0.63	-0.65	0.50	0.73
	(0.04)	(0.05)	(0.05)	(0.04)
Capital frictions index	-0.59	-0.58	0.64	0.52
	(0.04)	(0.05)	(0.04)	(0.05)
Price rigidity index	-0.61	-0.63	0.60	0.57
	(0.05)	(0.04)	(0.06)	(0.04)
Financial frictions index	-0.52	-0.64	0.22	0.57
	(0.05)	(0.04)	(0.06)	(0.04)

Notes: Bottom (Top) quintile: First-year average job flow response of industries in the first (fifth) quintile of the cross-industry distribution of a given index. Standard errors are in parenthesis.

Table 3.6. Job flow responses and quintiles of friction indices when using first principal components to construct friction indices

	Job creation		Job destruction	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
Labor frictions index	-0.70	-0.77	0.52	0.86
	(0.04)	(0.04)	(0.05)	(0.04)
Capital frictions index	-0.52	-0.60	0.65	0.58
	(0.04)	(0.05)	(0.04)	(0.05)
Price rigidity index	-0.71	-0.72	0.66	0.62
	(0.05)	(0.04)	(0.05)	(0.04)
Financial frictions index	-0.48	-0.62	0.25	0.66
	(0.06)	(0.04)	(0.06)	(0.04)

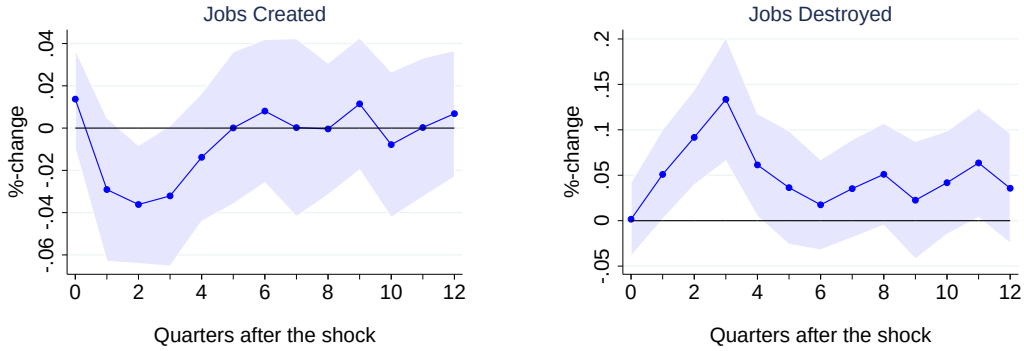
Notes: Bottom (Top) quintile: First-year average job flow response of industries in the first (fifth) quintile of the cross-industry distribution of a given index. Standard errors are in parenthesis.

3.A.3.3 Effects from uncertainty shocks in the first and second sample of analysis

We explore whether our findings hold if we focus on the first sample of analysis (1972-1998), see Figures 3.8 and Table 3.7, and second sample of analysis (2000-2013), see Figures 3.9 and 3.7. While the samples differ in regards to the quantitative effect of uncertainty shocks,

as higher the vulnerability to financial frictions, stronger is the effect of uncertainty shocks on job flows at both series.

Figure 3.8. Response of aggregate job flows to uncertainty shock when using the sample 1972-1998



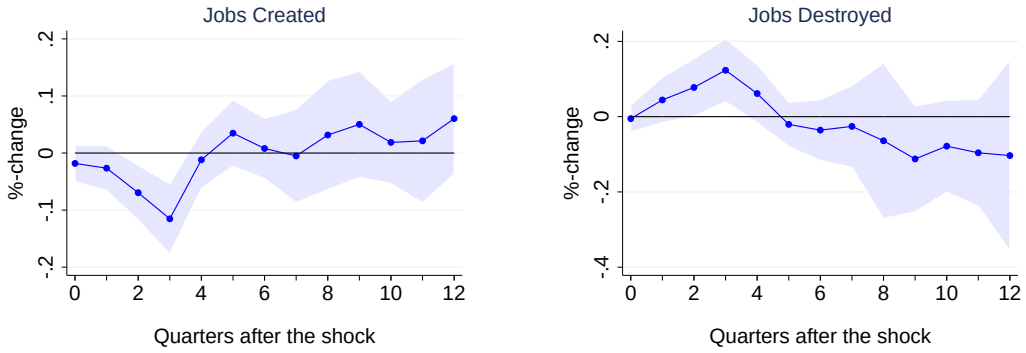
Notes: The blue lines show the responses of aggregate manufacturing job flows to a positive, three-standard deviation uncertainty shock. The shaded area is the 90% confidence interval using block bootstraps as in Kilian and Kim (2009).

Table 3.7. Job flow responses and quintiles of friction indices when using the sample 1972-1998

	Job creation		Job destruction	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
Labor frictions index	-0.58	-0.70	0.33	0.86
	(0.06)	(0.07)	(0.07)	(0.06)
Capital frictions index	-0.55	-0.67	0.68	0.51
	(0.07)	(0.07)	(0.06)	(0.07)
Price rigidity index	-0.71	-0.70	0.57	0.51
	(0.08)	(0.06)	(0.07)	(0.06)
Financial frictions index	-0.31	-0.56	0.12	0.57
	(0.09)	(0.07)	(0.08)	(0.06)

Notes: Bottom (Top) quintile: First-year average job flow response of industries in the first (fifth) quintile of the cross-industry distribution of a given index. Standard errors are in parenthesis.

Figure 3.9. Response of aggregate job flows to uncertainty shock when using the sample 2000-2013



Notes: The blue lines show the responses of aggregate manufacturing job flows to a positive, three-standard deviation uncertainty shock. The shaded area is the 90% confidence interval using block bootstraps as in Kilian and Kim (2009).

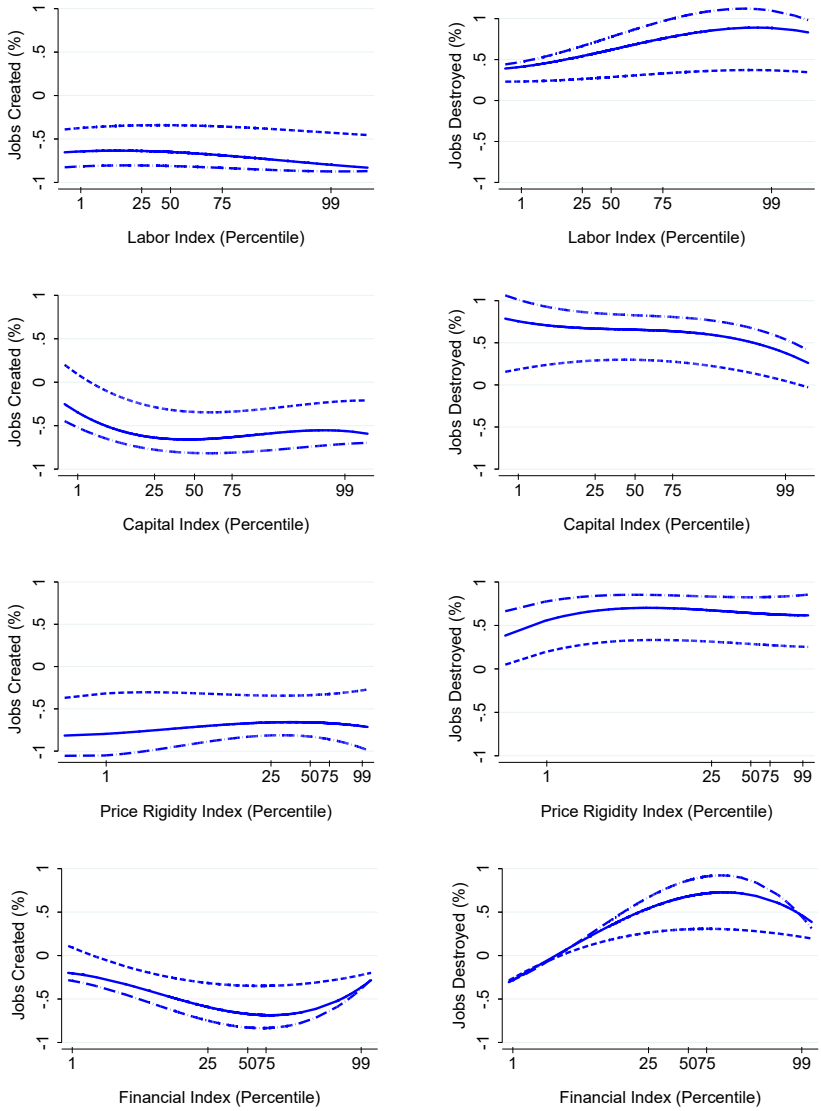
Table 3.8. Job flow responses and quintiles of friction indices when using the sample 2000-2013

	Job creation		Job destruction	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile
Labor frictions index	-0.31	-0.42	0.40	0.50
	(0.02)	(0.03)	(0.02)	(0.02)
Capital frictions index	-0.42	-0.33	0.47	0.37
	(0.03)	(0.03)	(0.02)	(0.02)
Price rigidity index	-0.36	-0.29	0.42	0.40
	(0.03)	(0.02)	(0.02)	(0.02)
Financial frictions index	-0.28	-0.31	0.20	0.41
	(0.03)	(0.03)	(0.04)	(0.02)

Notes: Bottom (Top) quintile: First-year average job flow response of industries in the first (fifth) quintile of the cross-industry distribution of a given index. Standard errors are in parenthesis.

3.A.3.4 Alternative time horizons for average job flow responses

Figure 3.10. Nonlinear relation between job flow responses and friction indices when using different horizons of job flow responses



Notes: The figures are based on regressions of industry-level job flow responses on cubic polynomials of all friction indices. Dashed/solid/dash-dotted lines show the estimated relation between first-two-quarters/first-year/first-six-quarters average job flow responses and one friction index when keeping the other friction indices at their medians, respectively. Shaded areas denote 90% confidence interval. We weight industry-level responses by the estimated absolute effect relative to its standard error.

4

Productivity Dispersions: Could it Simply be Technology Choice?

Joint with Christian Bayer and Ariel Mecikovsky

4.1 Introduction

The allocation of factors to their most productive use is often seen as one of the key determinants of economic prosperity (Foster et al., 2008). While first-best efficiency requires that factors produce the same marginal revenue across all production units, many studies show this condition to be violated in micro-data: factor productivities differ substantially within industries.[1]

We ask whether these micro-level differences can be understood as a result of frictions in technology choice. We suggest a setup, where firms may in principle choose from a broad set of technologies, but it is costly to search for them, to install them, and to acquire the know-how necessary to use them. This leads firms to operate one single technology which they adjust only occasionally. In between adjustments, the capital-labor ratio, the capital intensity, remains fixed: firms operate a Leontief production technology. As the economic environment changes and firms asynchronously adapt their technology in response, cross-sectional differences in factor productivities and capital intensity emerge.

This cross-sectional dispersion, however, is not the only empirical implication of frictional technology choice. Across all firms, differences in factor productivities and capital intensity

[1] See Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Peters (2013), Asker et al. (2014), Gopinath et al. (2015), and Restuccia and Santaeulalia-Llopis (2015) to name a few.

should be predominantly long-lived. Moreover, there must be a trade-off involved. Firms with persistently high productivity in one factor should have a persistently low productivity in another factor. Further, as long as the capital intensity is fixed, i.e. in the short run, labor and capital productivity can only move in the same direction. Finally, the extent of competition limits the scope of technologies used in the economy. The more competitive the environment, the larger is the pressure to abandon cost-inefficient technologies.

To explore whether these implications are borne out empirically, we compute micro-level labor and capital productivity controlling for industry and time effects, and decompose them into their persistent and transitory components. To have a broad empirical base, we exploit micro data from Germany (firm-level), Chile, Colombia, and Indonesia (plant-level). We find that between 61% and 94% of the cross-sectional variance in labor and capital productivity is explained by their persistent components. The result is even stronger for capital intensity where the persistent component explains more than 77% for all countries. Furthermore, the persistent components of labor and capital productivity are negatively correlated, while their transitory components are positively correlated. In addition, persistent differences in capital intensity are less dispersed in more competitive environments, i.e. where markups are persistently lower. Firms/plants in the most competitive quintile exhibit a 30-50% lower variance of capital intensity than those in the least competitive quintile. In summary, the data qualitatively supports the idea of a friction in technology choice driving productivity dispersions.

We use this framework to quantify the effects of a frictional technology choice in aggregate productivity. Despite the large cross-sectional productivity dispersion, our estimated efficiency losses from misallocation are on average 5%, which is small relative to the estimates from the literature. Important for this is our focus on productive efficiency, i.e. deviations from optimal capital intensity. In contrast, studies like Hsieh and Klenow (2009) have taken a broader focus including allocative efficiency, i.e. deviations from optimal scale. We disregard those deviations, showing up as dispersions in markups, for our efficiency calculations for two reasons. First, these dispersions might reflect efficient differentiation within industry. For example, they might stem from alternative strategies on product quality or range (e.g. Bar-Isaac et al., 2012), think of generics vs. patented pharmaceuticals. Second, there is already a broad set of theories predicting markup dispersions to which we have little to add. Think of models with price setting frictions á la Calvo (1983), with building a customer base (Gourio and Rudanko, 2014), or with entry dynamics and innovation as in Peters (2013). All of these provide explanations of productivity dispersions through heterogeneous markups as endogenous objects. At the same time, our data suggests that markup dispersions themselves explain only a minority of all productivity dispersion.

Our results are linked to the traditional putty-clay assumption (Johansen, 1959), which has been advocated to address a broad array of other empirical phenomena (Gilchrist and Williams, 2000, 2005; Gourio, 2011). Particularly closely related is Kaboski's (2005) model of putty-clay technology choice under factor price uncertainty. An important insight from this paper that carries over to our setup is that firms underreact to current prices in setting their technology, such that the regression techniques usually used to identify the long-run elasticity of substitution (see e.g. Raval (2014) or Oberfield and Raval (2014) for recent contributions or Chirinko (2008) for an overview) are subject to a downwards bias. In fact, we provide evidence that this downwards bias is likely substantial. This high elasticity not only has important implications for income-shares (see e.g. Solow, 1956; Piketty, 2011; Piketty, 2014; Karabarbounis and Neiman, 2013) but is also key to compute the efficiency losses from a friction in technology choice.

The remainder of this paper is organized as follows: Section 4.2 describes our technology choice model. This guides our empirical analysis in Section 4.3. In Section 4.4 we discuss the potential gains from eliminating this friction. Section 4.5 concludes and an Appendix follows.

4.2 Technology choice model

This section suggests a simple model of technology choice. We then study the implications of technology choice for productivity dispersions and for aggregate productive efficiency.

4.2.1 Model setup

We consider a two-period model. We assume a mass of firms of measure one. Each firm, i, is endowed with one plant that has an exogenously given capital intensity $k_i = \frac{K_i}{N_i}$, where K_i is the physical amount of capital and N_i is labor. Furthermore, wages, W, and user costs of capital, R, are exogenously given but stochastic.

Each firm has a constant returns to scale production technology and faces monopolistic competition for its product, where the elasticity, ξ_i, of demand for the product, y_i, of firm i is firm-specific and constant, such that prices are given by

$$p_i = \frac{1}{1-\xi_i} z_i^{\xi_i} y_i^{-\xi_i},$$

where z_i is the stochastic market size for firm i's product. Unit costs of production depend on the plant's capital intensity and factor prices, $c_i = c(k_i, W, R)$. The firm maximizes profits,

and we assume that the firm needs to decide about output before knowing actual factor prices and demand. The optimal policy will choose output in order to stabilize the expected markup at its optimal level. The expected gross markup is constant, $\frac{1}{1-\xi_i} > 1$. Denoting the expectations operator as $\mathbb{E}$, it is straightforward to show that the profit maximizing output, y_i^*, and expected profits under the optimal policy, π^*, are given by

$$y_i^* = \left[\frac{\mathbb{E}z_i^{\xi_i}}{\mathbb{E}c(k_i,R,W)}\right]^{1/\xi_i}; \qquad \pi_i^* = \frac{\xi_i}{1-\xi_i} y_i^* \mathbb{E}c(k_i,R,W). \tag{4.1}$$

4.2.2 Revenue productivities

Profit maximization implies that firms facing higher demand elasticities, ξ_i, have larger markups on average. In addition, deviations from expected costs, $\mathbb{E}c_i/c_i$, and deviations from expected demand, $z_i^{\xi_i}/\mathbb{E}z_i^{\xi_i}$, lead to fluctuations in realized markups, given by:

$$\frac{p_i y_i^*}{WN_i + Rk_iN_i} = \frac{1}{1-\xi_i}\frac{z_i^{\xi_i}}{\mathbb{E}z_i^{\xi_i}}\frac{\mathbb{E}c_i}{c_i}. \tag{4.2}$$

Splitting up realized markups in two components, the same terms affect revenue factor productivities, the capital and labor expenses per value added:

$$\frac{p_i y_i^*}{WN_i} = \frac{1}{1-\xi_i}\frac{z_i^{\xi_i}}{\mathbb{E}z_i^{\xi_i}}\frac{\mathbb{E}(W+Rk_i)}{W} \tag{4.3}$$

$$\frac{p_i y_i^*}{Rk_iN_i} = \frac{1}{1-\xi_i}\frac{z_i^{\xi_i}}{\mathbb{E}z_i^{\xi_i}}\frac{\mathbb{E}(W+Rk_i)}{Rk_i} \tag{4.4}$$

(4.3) and (4.4) show that firms with higher (target) markups, $\frac{1}{1-\xi_i}$ exhibit both higher labor and capital productivities. Similarly, positive and unforeseen demand shocks, $z_i^{\xi_i}/\mathbb{E}z_i^{\xi_i}$, increase both factor productivities. Notice that in a more general multi-period setup, these deviations from expectations can only be transitory. Importantly, firms with higher capital intensity have a lower capital revenue productivity and higher labor revenue productivity, even when these capital intensity differences are expected.

To summarize, productivities differ across firms either because of differences in size relative to demand (the first two terms) or due to differences in capital intensity and factor prices (the last term) in (4.3) and (4.4).[2]

4.2.3 Choice of technology

We assume that in the period preceding production, the firm can opt to replace its existing plant, setting up a new one with different capital intensity k. In doing so, the firm compares expected profits with and without technology adjustment to decide the period preceding production whether to produce with its initially given capital intensity or to invest in changing the technology. We assume adjustment is costly as it disrupts production. This disruption summarizes all costs of searching for a technology, installing it and learning to operate it. Upon adjustment the firm forgoes a fraction ϕ_i of next period's profits, where ϕ_i stochastic and drawn from a distribution Φ. The firm draws ϕ_i before it decides about adjustment. If the firm adjusts, it chooses $\hat{k}$, the capital intensity that minimizes expected unit costs. Adjustment is optimal whenever $(1-\phi_i)E\pi(\hat{k}) > E\pi(k_i)$. Using (4.1), this simplifies to

$$(1-\phi_i) \; > \; \left(\frac{\mathbb{E}c(k_i,R,W)}{\mathbb{E}c(\hat{k},R,W)}\right)^{\frac{\xi_i-1}{\xi_i}}. \tag{4.5}$$

Since $\mathbb{E}c(k_i,R,W) \geq \mathbb{E}c(\hat{k},R,W)$, firms with higher elasticity of demand, ξ_i, are less likely to adjust for a given ex ante capital intensity k_i. The reason is that firms with high market power can offload their higher unit costs to consumers and hence have less incentive to invest in efficient capital intensities. This is reminiscent of Leibenstein's (1966) X-inefficiency of monopolies or Bester and Petrakis's (1993) results for oligopolies.[3]

As a result, ex-post capital-intensity will be less dispersed within the group of firms with low markups than among high-markup firms if the ex-ante distribution of capital intensities is centered around the cost minimizing level $\hat{k}$.

---

[2] As evident from equation 4.2, in this environment, adding an additional shock to unit costs (a TFP shock) has the same implications as a demand shock.

[3] There is, however, one interesting side result of our setup. One can easily show that under the specific assumption of an isoelastic demand curve and monopolistic competition, producer profits and consumer rents are equal and therefore, total social surplus of adjustment as well as the social costs of adjustment need to be scaled by factor two such that the individual optimal adjustment choice is socially optimal.

4.2.4 Productive efficiency

The friction in technology choice is costly. We study the aggregate costs of this friction focusing on the losses in productive efficiency, i.e. deviations from optimal capital intensity. We compute these losses by characterizing the difference of unit costs under the friction compared against a frictionless benchmark.

To specify more concretely the relation between capital intensity and unit costs, we assume that the long-run technology is given by a constant elasticity of substitution (CES) production function with σ the substitution elasticity between labor and capital. The output of a plant with capital intensity k_i is given by

$$y_i = \left[\alpha k_i^{\frac{\sigma-1}{\sigma}} + (1-\alpha)A^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} N_i, \tag{4.6}$$

where A captures (Harrod neutral) labor-augmenting technological change, and α is the distribution parameter. This implies that *realized* unit costs, $c_i = \frac{Rk_iN_i+WN_i}{y_i}$ are minimal at capital intensity k^*, given by

$$k^* = \left[\frac{\alpha}{1-\alpha}\frac{W}{R}\right]^{\sigma} A^{1-\sigma}. \tag{4.7}$$

Now, to obtain an expression that allows us to relate the cross-sectional average unit costs to the first two moments of the capital intensity distribution, we use a log second-order approximation around that minimum:

$$\mathbb{E}^x\left[\log\frac{c(k_i,R,W)}{c(k^*,R,W)}\right] \approx \frac{1}{2\sigma}s^*(1-s^*)\left\{\left[\mathbb{E}^x\left(\log\frac{k_i}{k^*}\right)\right]^2 + \mathbb{V}^x(\log k_i)\right\}, \tag{4.8}$$

where s^* is the capital expenditure share in the cost-minimizing optimum

$$s^* = Rk^*/(W+Rk^*),$$

and $\mathbb{E}^x$ denotes the cross-sectional average and $\mathbb{V}^x$ the cross-sectional variance.[4] In words, the efficiency loss is composed of the average relative difference of capital intensity from its optimum, $\mathbb{E}^x \log(k_i/k^*)$, and the cross-sectional dispersion of capital intensity across plants, $\mathbb{V}^x(\log k_i)$. Importantly, the higher the elasticity of substitution between labor and capital, σ, the lower the efficiency loss from not re-setting capital intensities to their optimum.

[4] See Appendix 4.A.2 for details.

4.3 Empirical analysis

4.3.1 Data description

We document factor revenue productivity and capital intensity dispersion in firm-level data from Germany, and plant-level data from Chile, Colombia and Indonesia. For Germany, we use the balance sheet data base of the Bundesbank, USTAN, which is a private sector, annual firm-level data available for 26 years (1973-1998).[5] For Chile, Colombia and Indonesia, we have plant level data from the ENIA survey for 1995-2007, the EAM census for 1977-1991 and the IBS dataset for 1988-2010, respectively. These datasets are focused on the manufacturing sector, with the exception of Germany, which provides information for the entire private non-financial business sector.[6]

When preparing the data for our analysis, we make sure to treat the various data sets in the most comparable way. From each survey, we use a firm's/plant's four-digit industry code, wage bill, value-added and book or current value of capital stock. To obtain economically consistent capital series for each firm/plant, we re-calculate capital stocks using the perpetual inventory method whenever the data reports capital stocks in book values. In particular, we exploit information of capital disaggregated into structures and equipment, which allows us to control for heterogeneity in capital composition across firms/plants.

Our capital productivity measure requires information on depreciation and the real interest rate. We do not rely on depreciation as reported by firms/plants, as it is potentially biased for tax purposes. Instead we use economic depreciation rates by type of capital good obtained from National Statistics. We then take the different capital good mixes across firms/plants into account.[7] We set the real rate to 5% for all economies. This implies user costs of capital $R_{it} = 5\% + \delta_{it}$. In generating cross-sectional statistics, time variations in user costs are controlled for by taking out four-digit industry-year fixed effects. The data treatment and sample selection is described in detail in Appendix 4.A.1.2.

---

[5] See Bachmann and Bayer (2014) for a detailed description.

[6] In particular, private non-financial business sector includes Agriculture, Energy and Mining, Manufacturing, Construction, and Trade.

[7] The economic depreciation rates of equipment and structures for Germany is obtained from *Volkswirtschaftliche Gesamtrechnung* (VGR) while for Chile we obtain time series from Henriquez (2008). As for Colombia and Indonesia, we consider the average depreciation in Chile for the available period given the absence of national data sources. The depreciation rate values are 15.1% (equipment) and 3.3% (structures) in Germany, and on average 10.5% (equipment) and 4.4% (structures) for the other three countries.

4.3.2 Productivities and their transitory and persistent component

We compute average factor productivities for capital and labor per firm and year using the reported value added per firm/plant at current prices, $p_{it}y_{it}$, labor expenses, W_tN_{it} as reported in the profit and loss statements, and imputed capital expenses, $R_{it}K_{it}$. Taking logs, we define revenue productivities of labor and capital:

$$\alpha_{it}^N := \log(p_{it}y_{it}) - \log(W_tN_{it}); \qquad \alpha_{it}^K := \log(p_{it}y_{it}) - \log(R_{it}K_{it}). \tag{4.9}$$

Using expenditures and value added implicitly controls for quality differences in both inputs and outputs (c.f. Hsieh and Klenow, 2009). In addition, we construct markups as value added relative to total expenditures on labor and capital

$$mc_{it} := \log(p_{it}y_{it}) - \log(R_{it}K_{it} + W_tN_{it}). \tag{4.10}$$

Finally, we calculate the factor price-weighted capital intensity,

$$\kappa_{it} := \log(R_{it}K_{it}) - \log(W_tN_{it}). \tag{4.11}$$

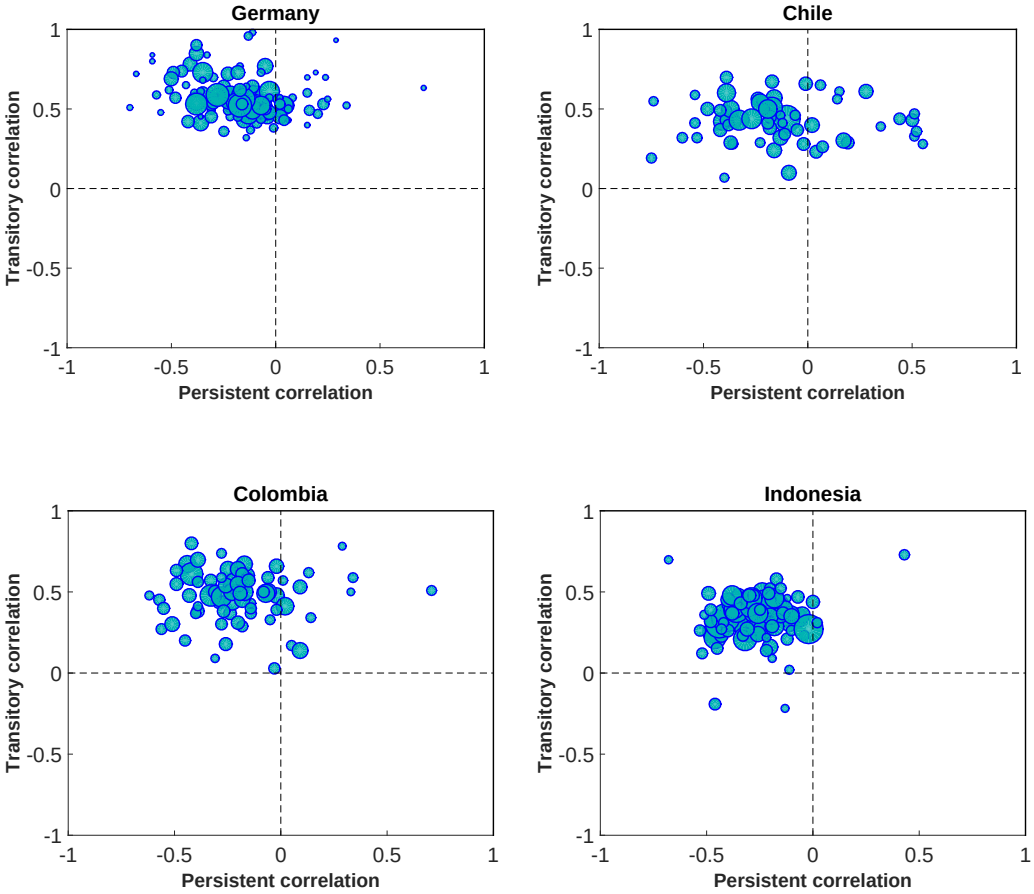
For any of these variables, say x_{it}, we calculate 5-year moving averages, denoted $\bar{x}_{it} := \frac{1}{5}\sum_{s=-2}^{2} x_{it+s}$, to identify the persistent component. Deviations thereof, $\hat{x}_{it} := x_{it} - \bar{x}_{it}$, identify the transitory component.

We then take out four-digit industry-year fixed effects and calculate dispersions and correlations between the factor productivities for the transitory and persistent component.

4.3.3 Empirical findings

Table 4.1 reports standard deviations and correlation for labor and capital productivity and for all four countries. Three observations stand out: First, capital and labor productivity are positively correlated in the transitory component ($\rho \approx 40\%$) while they are negatively correlated in the persistent component ($\rho \approx -20\%$). Using the expressions for factor productivities in Section 4.2, see (4.3) and (4.4), deviations from optimal size are more important in the short run, while deviations from optimal capital intensity are more important in explaining long-run productivity differences. Second, the persistent components in productivity explain the vast majority of cross-sectional productivity differences (between 60% and 92% for labor and between 79% and 94% for capital). Third, the developing economies show larger productivity dispersions.

Figure 4.1. Correlations of factor productivities by four-digit industry



Notes: *Transitory (Persistent) Correlation*: Correlation between the transitory (persistent) component of labor and capital productivity at the firm/plant level, controlling for time-fixed effects. Each circle represents a four digit industry, where the size of a circle reflects aggregate employment in that industry. For this figure, we restrict industries to include at least 20 firms/plants. The number of industries inside the upper-left quadrant is 99 (out of 125) in Germany, 45 (out of 61) in Chile, 62 (out of 73) in Colombia, and 85 (out of 90) in Indonesia.

Table 4.1. Transitory and persistent components of factor productivities

	$\text{std}(\hat{\alpha}_{it}^L)$	$\text{std}(\hat{\alpha}_{it}^K)$	$\rho(\hat{\alpha}_{it}^L, \hat{\alpha}_{it}^K)$	$\text{std}(\bar{\alpha}_{it}^L)$	$\text{std}(\bar{\alpha}_{it}^K)$	$\rho(\bar{\alpha}_{it}^L, \bar{\alpha}_{it}^K)$
	Transitory Component			Persistent Component		
DE	0.066	0.119	0.352	0.229	0.456	-0.207
	(0.000)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
CL	0.184	0.281	0.449	0.232	0.577	-0.190
	(0.006)	(0.008)	(0.017)	(0.009)	(0.028)	(0.021)
CO	0.144	0.172	0.517	0.257	0.568	-0.234
	(0.003)	(0.004)	(0.012)	(0.008)	(0.023)	(0.018)
ID	0.211	0.369	0.343	0.255	0.669	-0.269
	(0.003)	(0.005)	(0.007)	(0.004)	(0.013)	(0.009)

Notes: Cross-sectional standard-deviations (std) and correlation (ρ) of transitory and persistent components of labor- and capital productivity, α_{it}^L and α_{it}^K as in (4.9). DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia. Transitory and persistent components are obtained by applying a five year moving average filter. Factor productivities are demeaned by 4-digit industry and year, and expressed in logs. In parentheses: Clustered standard errors at the firm/plant level.

As the positive/negative correlation pattern between labor and capital productivity is a particularly important prediction of technology choice, we investigate whether this pattern holds for individual four-digit industries. Figure 4.1 shows that this is the case for the vast majority of industries.

In light of our results in Section 4.2, it is useful to look at markup and capital intensity differences, see Table 4.2. In particular, (4.8) allows us to relate capital intensity differences across firms/plants directly to increases in unit costs, and thus to losses in productive efficiency. For all countries, differences in capital intensity are very persistent. The transitory component makes up only between 4% (Germany) and 17% (Indonesia) of the total variance. At the same time, persistent differences in capital intensity are substantially more dispersed in Chile, Colombia, and Indonesia than they are in Germany with variances being twice as high in Indonesia than in Germany.

On the contrary, the dispersion of persistent cross-sectional markup differences is strikingly similar across countries, and transitory differences in markups are an important com-

Table 4.2. Transitory and persistent components of markup and capital intensity

	$\text{std}(\hat{mc}_{it})$	$\text{std}(\hat{\kappa}_{it})$	$\rho(\hat{mc}_{it}, \hat{\kappa}_{it})$	$\text{std}(\bar{mc}_{it})$	$\text{std}(\bar{\kappa}_{it})$	$\rho(\bar{mc}^L_{it}, \bar{\kappa}_{it})$
	Transitory Component			Persistent Component		
DE	0.064	0.114	-0.155	0.172	0.551	0.062
	(0.000)	(0.001)	(0.002)	(0.001)	(0.004)	(0.004)
CL	0.177	0.258	-0.090	0.184	0.661	-0.085
	(0.005)	(0.009)	(0.017)	(0.005)	(0.029)	(0.022)
CO	0.134	0.157	-0.016	0.206	0.676	-0.232
	(0.003)	(0.004)	(0.012)	(0.005)	(0.025)	(0.018)
ID	0.203	0.357	-0.120	0.195	0.778	-0.021
	(0.002)	(0.005)	(0.007)	(0.003)	(0.014)	(0.010)

Notes: Capital intensities, κ_{it}, and markups, mc_{it}, as defined in (4.10) and (4.11). See notes of Table 4.1 for further explanation.

ponent of the total cross-sectional variance of markups – at least in the developing economies (30% in Colombia, 50% in Chile and Indonesia) but less so in Germany (12%).[8]

Through the lense of the technlogy choice model, in particular (4.3) and (4.4), these results suggest that a major component in the persistent differences in productivity is the choice of capital intensities. Deviations in optimal scale are important but minor.

To understand to what extent firms actively take these unit cost increases into account, we split the sample according to firm/plant characteristics – age, size, and importantly a firm's average markup – and compute again the dispersions of the persistent component of capital intensity, see Table 4.3. While there are some differences in these dispersions according to age and size, these are neither large nor systematic. What stands out is splitting the sample according to the average markup. The highest markup quintile exhibits between 30% and 60% higher capital intensity dispersions (in terms of variances) than the lowest markup quintile. This is in line with the qualitative predictions of our model.

[8] This might relate to the fact that demand is less stable in the developing economies. In fact, the cross-sectional standard deviation of value-added growth is two to four times larger in these economies than in Germany.

Table 4.3. Persistent component of capital intensity by firm/plant characteristics

	$\text{std}(\bar{\kappa}_{it})$					
	Markups		Size		Age	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Young	Old
DE	0.545	0.622	0.610	0.509	n.a.	n.a.
	(0.010)	(0.010)	(0.009)	(0.011)		
CL	0.568	0.713	0.749	0.622	n.a.	n.a.
	(0.042)	(0.075)	(0.068)	(0.058)		
CO	0.547	0.694	0.763	0.669	0.697	0.699
	(0.035)	(0.061)	(0.051)	(0.061)	(0.100)	(0.048)
ID	0.716	0.834	0.830	0.816	0.770	0.801
	(0.028)	(0.035)	(0.034)	(0.035)	(0.058)	(0.038)

Notes: Bottom (top) markup quintile: firm/plant average markup below the 20th percentile (above the 80th percentile). Old (young): Plant age below 4 years (above 15 years). Bottom (top) size quintile: firm/plant average employment below the 20th percentile (above 80th percentile). The micro data from Germany and Chile does not include age. See notes of Table 4.1 and 4.2 for further explanation.

4.3.4 Robustness

We conduct some robustness checks. First, we show that our empirical findings are robust to alternative way of decomposing factor productivities into transitory and persistent components (Table 4.4), and to weighting of the moments (Table 4.5). We also show that persistent capital intensity differences are more dispersed for high-markup firms/plants even when controlling for size and age (Table 4.6). In particular, we first remove cross-sectional differences in log capital intensity that can be explained by markups, size and age in logs. Second, the square unexplained component from the first stage is regressed on standarized markups, size and age in logs. For all countries, except Colombia, markups are as important as size for explaining persistent differences in capital intensity.

Table 4.4. Robustness: Transitory and persistent components (HP-filtered) of factor productivities, markups, and capital intensity

	$\text{std}(\hat{a}^L_{it})$	$\text{std}(\hat{a}^K_{it})$	$\rho(\hat{a}^L_{it}, \hat{a}^K_{it})$	$\text{std}(\bar{a}^L_{it})$	$\text{std}(\bar{a}^K_{it})$	$\rho(\bar{a}^L_{it}, \bar{a}^K_{it})$
	Transitory Component (HP)			Persistent Component (HP)		
DE	0.062	0.113	0.352	0.236	0.471	-0.223
CL	0.169	0.260	0.447	0.231	0.578	-0.191
CO	0.134	0.159	0.516	0.257	0.569	-0.234
ID	0.196	0.343	0.344	0.256	0.670	-0.270
	$\text{std}(\hat{mc}_{it})$	$\text{std}(\hat{\kappa}_{it})$	$\rho(\hat{mc}_{it}, \hat{\kappa}_{it})$	$\text{std}(\bar{mc}_{it})$	$\text{std}(\bar{\kappa}_{it})$	$\rho(\bar{mc}_{it}, \bar{\kappa}_{it})$
	Transitory Component (HP)			Persistent Component (HP)		
DE	0.073	0.134	-0.184	0.157	0.490	0.089
CL	0.183	0.295	-0.123	0.152	0.552	-0.097
CO	0.145	0.186	-0.066	0.178	0.594	-0.230
ID	0.207	0.412	-0.130	0.160	0.672	-0.027

Notes: Labor productivity, a^L_{it}, and capital productivity, a^K_{it}, as defined in (4.9). Markups, mc_{it}, and capital intensity, κ_{it}, as defined in (4.10) and (4.11). HP: results based on the decomposing between transitory and persistent using a HP-filter ($\lambda = 6.25$). Factor productivities are demeaned by 4-digit industry and year and expressed in logs. Standard errors are clustered standard errors at the firm/plant level. ρ denotes correlation. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

4.4 Efficiency losses from a friction in technology choice

In this section, we quantify the efficiency losses from frictional technology choice. Based on the results in Section 4.2, we can compute these losses by combining the dispersion in capital intensity with the capital cost share and the elasticity of substitution between labor and capital.

We estimate the capital cost share as the ratio of capital expenditures to total expenditures directly from the micro-data. We obtain a capital share of 21% (Germany), 40% (Colombia), 32% (Chile), and 23% (Indonesia). The elasticity of substitution between labor and capital can be recovered from time-series information of the aggregate capital intensity and the relative factor price. In a frictionless economic environment, the elasticity is determined

Table 4.5. Robustness: Weighted second moments of factor productivities, markups, and capital intensity at different frequencies

	$\text{std}(\hat{\alpha}^L_{it})$	$\text{std}(\hat{\alpha}^K_{it})$	$\rho(\hat{\alpha}^L_{it}, \hat{\alpha}^K_{it})$	$\text{std}(\bar{\alpha}^L_{it})$	$\text{std}(\bar{\alpha}^K_{it})$	$\rho(\bar{\alpha}^L_{it}, \bar{\alpha}^K_{it})$
	Transitory Component (5Y MA)			Persistent Component (5Y MA)		
DE	0.050	0.101	0.316	0.196	0.457	-0.176
CL	0.187	0.281	0.457	0.239	0.551	-0.205
CO	0.143	0.170	0.520	0.260	0.562	-0.239
ID	0.216	0.370	0.349	0.263	0.672	-0.275
	$\text{std}(\hat{mc}_{it})$	$\text{std}(\hat{\kappa}_{it})$	$\rho(\hat{mc}_{it}, \hat{\kappa}_{it})$	$\text{std}(\bar{mc}_{it})$	$\text{std}(\bar{\kappa}_{it})$	$\rho(\bar{mc}_{it}, \bar{\kappa}_{it})$
	Transitory Component (5Y MA)			Persistent Component (5Y MA)		
DE	0.052	0.090	-0.161	0.172	0.503	0.067
CL	0.179	0.259	-0.090	0.183	0.645	-0.087
CO	0.133	0.155	-0.016	0.209	0.670	-0.237
ID	0.207	0.356	-0.123	0.198	0.787	-0.021

Notes: abor productivity, a^L_{it}, and capital productivity, a^K_{it}, as defined in (4.9). Markups, mc_{it}, and capital intensity, κ_{it}, as defined in (4.10) and (4.11). Cross-sectional standard-deviations (std) and correlation (ρ) of transitory and persistent components. Transitory and persistent components are obtained by applying a five year moving average filter (5Y MA). Moments are weighted based on the value-added of the plant/firm. Variables under interest are demeaned by 4-digit industry and year and expressed in logs. Standard errors in parentheses are clustered standard errors at the firm/plant level. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

by the contemporaneous correlation between these variables. However, the identification is problematic in the presence of frictions which prevents immediate adjustment of production factors: The contemporaneous response of capital intensity to price movements (short-run elasticity) then differs from the long-run elasticity.

To uncover the long-run elasticity of substitution, we instrument observed relative factor prices with the top marginal income tax rate on domestic corporations at the country level.[9] As our instrumental variable is highly persistent, we capture movements in factor pri-

[9] Given that we do not have information on real interest rate from all countries, we approximate the risk-free interest rate using the Federal Funds rate (yearly average). We consider country panel

Table 4.6. Robustness: Dispersion of capital intensity and markups

	DE	CL	CO	ID
	$\epsilon^2_{\tilde{\kappa}_{it}}$			
Log-Markup	0.024	0.069	0.036	0.057
	(0.003)	(0.017)	(0.019)	(0.011)
Log-Size	-0.026	-0.068	-0.057	0.017
	(0.003)	(0.017)	(0.024)	(0.015)
Log-Age		-	0.044	0.009
		-	(0.018)	(0.011)

Notes: The results are obtained based on a two step procedure. First, we remove cross sectional differences in log capital intensity (κ_{it}) that can be explained by the log of markups, size and age. Second, the squared estimated residual based on the first stage ($\epsilon^2_{\tilde{\kappa}_{it}}$), is regressed on the standarized log of markups, size and age. Standard errors in parentheses are clustered standard errors at the firm/plant level. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

ces that are long-lived, and thus, we obtain a better approximation of the long-run elasticity of substitution.[10]

Table 4.7 provides the results of this exercise. Once we instrument the relative factor price with corporate taxes, we obtain an estimated elasticity of 1.28. In contrast, the simple contemporaneous regression implies a 50% lower estimated elasticity.

Based on these elasticity estimates, we compute the efficiency losses from a friction in technology choice. On average, unit costs increase by 5% compared to their minimum obtained by always setting capital intensity to the optimal level, the values range goes from 2.5% in Germany to 6.3% in Indonesia.

---

data on labor, capital, and hourly wage from Feenstra et al. (2015). We impute hours worked at those countries with missing information by the average hours worked at each year based on those countries with available data. Finally, we construct tax series using the World tax Database available at http://www.bus.umich.edu/otpr/otpr/default.asp.

[10] Alternatively, the literature aims to estimate the long-run elasticity of substitution using cointegration properties, cross country variation in the trends of factor prices, or low-pass filters. See Chirinko (2008) for more details.

Table 4.7. Estimation of long-run elasticity of substitution

	Dependent variable: $\log\left(\frac{K}{N}\right)$		
$\log\left(\frac{W}{R}\right)$	0.68	0.43	1.28
	(0.01)	(0.01)	(0.35)
Constant	39.41	19.40	135.97
	(1.82)	(1.40)	(48.24)
Trend	Yes	Yes	Yes
Country fixed effects	No	Yes	Yes
Instrument	No	No	Yes
R^2	0.76	0.75	0.71
Countries	99	99	99
Obs	2609	2609	2609

Notes: Regressions based on country panel data for the period 1956-2002. Period length differs by country due to data availability. We instrument relative factor price using the top marginal income tax rate on domestic corporations at the country level. Standard errors in parenthesis.

Notice that these efficience loss estimates do not consider the time-series component $\left(E_t^x \log k_{it} - \log k^*\right)^2$. To do so, we require a dynamic version of the model described in Section 4.2. Therefore, our estimates constitute a lower bound of the potential efficiency losses from a friction in technology choice.

4.5 Conclusion

This paper asks whether productivity dispersions should be understood as a result of frictions in technology choice. We have derived qualitative implications of such friction and show that these are borne out empirically.

In line with the existing literature, we find large productivity differences across firms/ plants even within narrowly defined industries. We show that most of the differences are long-lived and related to highly persistent differences in capital intensity. Despite the strong relative differences across countries, our estimated efficiency losses from frictional technology choice are modest, on average 5%.

For future work it would be important to explore whether a dynamic model of technology choice is able to explain our empirical results not only qualitatively, but also quantitatively.

Appendix 4.A Appendix

4.A.1 Empirics

4.A.1.1 Description of the data

German Firm Data: USTAN (Unternehmensbilanzstatistiken)

USTAN is itself a byproduct of the Bundesbank's rediscounting and lending activity. The Bundesbank had to assess the creditworthiness of all parties backing promissory notes or bills of exchange put up for rediscounting (i.e. as collateral for overnight lending). It implemented this regulation by requiring balance sheet data of all parties involved, which were then archived and collected, see Bachmann and Bayer (2013) for details. Our initial sample consists of 1,846,473 firm-year observations. We remove observations from East German firms to avoid a break of the series in 1990. Finally, we drop the following sectors: hospitality (hotels and restaurants), financial and insurance institutions, public health and education sectors. The resulting sample covers roughly 70% of the West-German real gross value added in the private non-financial business sector. In particular, it includes Agriculture, Energy and Mining, Manufacturing, Construction, and Trade.

Chilean Plant Data: ENIA (Encuesta Nacional Industrial Anual)

ENIA is collected by the National Institute of Statistics (*Instituto Nacional de EstadÃŋsticas*, INE) and provides plant-level data from 1995 to 2007. ENIA contains information for all manufacturing plants with total employment of at least ten. For the period under analysis, we have a sample of 70,217 plant-year observations. According to INE, this sample covers about 50% of total manufacturing employment.

Colombian Plant Data: EAM (Encuesta Anual Manufacturera)

EAM is a plant-level survey collected by National Institute of Statistics (*Departamento Administrativo Nacional de Estaditicas*, DANE) for the period 1977 to 1991. The survey covers information for all manufacturing plants during 1977-1982, while it only contains data on plants above 10 employees for 1983-1984, and from 1985, small plants are included in small proportion. This results in 103,011 plant-year observations.

Indonesian Plant Data: IBS (Survei Tahunan Perusahaan Industri Pengolahan)

IBS is the Indonesian Manufacturing Survey of Large and Medium Establishments, provided by the National Institute of Statistics (*Badan Pusat Statistik*, BPS). The survey covers all plants with 20 or more employees in the manufacturing sector. Given that the capital stock is reported since 1988 onwards, we exclude earlier years and focus on the period 1988-2010, with 485,052 plant-year observations.

4.A.1.2 Sample selection

Starting from the raw data set, we concentrate on describing the general cleaning steps common to all countries, and we provide more information about country-specific cleaning steps at Table 4.8.

To begin with, we remove observations where firms or plants report extraordinarily large depreciation rates (e.g. due to fire or accident). The reason is that our dynamic model does not capture such cases, and the perpetual inventory method (PIM) will inaccurately measure the actual capital stock after such incidents occur.[11] Next, for those countries where current values of capital stock is not provided (Germany and Colombia), we recompute capital stocks using the PIM. In conducting the PIM, we drop a small amount of outliers, as explained in Section 4.A.1.4. Further, we do not consider observations where value-added, capital stock, or employment is non-positive or missing.

Moreover, we do not consider observations where firms/plants have missing values in the changes of employment (N), real capital (K) and real value-added (VA).[12] To construct capital productivity, we use the lagged value of capital stock, so we effectively discard the first year of each micro unit. We remove outliers in the levels and in the relative changes of employment, capital, value-added, and factor shares based on 3 standard deviations from the industry-year mean. In addition, we drop firm/plant-year observations whenever the total factor expenditures share is either below 1/3 or above 3/2, and whenever the firm/plant

[11] At some cases in the ENIA, EAM, and IBS surveys, plants do not report depreciation conditional on positive capital stock. In order to not loose this observations, we impute the depreciation by capital type and two-digit industry, estimating a random effect model, using as explanatory variable the log-capital stock. To discard rare depreciation events, we drop observations whenever the reported depreciation rate in structures (equipment) is above 40% (60%) yearly. Additionally, we do not consider those cases where the reported depreciation is below 0.1% (1%) in structures (equipment), yearly.

[12] To construct measures of real capital stock we consider an index price by each capital type (when available) using the information of gross fixed capital formation at current and constant prices from National Accounts, while for for value added we use the GDP price deflator.

average total factor expenditure share is above 1. These two cleaning steps should exclude units from our analysis which report continuously unreasonably large markups or losses.

Finally, as our empirical results rely on a 5-year moving average filter, we do not consider firm/plant-year observations that have less than 5 consecutive years.

Table 4.8. Sample selection

Criterion/Country	Germany	Chile	Colombia	Indonesia
Initial sample	1,846,473	70,217	103,006	485,052
East Germany	-115,201	–	–	–
Additional cleaning steps	–	–	–	-32,618
Imputation capital stock	–	–	–	+37,341
Rare depreciation events	-54,280	-8,197	-6,176	-8,775
Outliers in PIM	-73,784	–	-4,280	–
Missing values	-422,739	-19,589	-29,804	-235,280
Outliers in factor variables	-176,232	-12,375	-24,651	-86,070
Less than 5 consecutive years	-312,452	-15,479	-14,264	-84,885
Final sample	689,665	14,307	23,831	74,765

Notes: Missing values denote the sum of missing values at log value added, log capital, factor shares and log changes in employment, capital and value added. Outliers in factor variables is the sum of all identified outliers at log changes in employment, real capital and real value added, and factor shares. For more information with respect to *Additional cleaning steps* and *Imputation of capital stock* in Indonesia, see Section 4.A.1.3.

4.A.1.3 Specific cleaning and imputation steps for IBS

Before proceeding with the general cleaning steps applied to all datasets, we need to implement some specific corrections at the Indonesian micro-data. In doing so, we closely follow Blalock and Gertler (2009). First, we correct for mistakes due to data keypunching. If the sum of the capital categories is a multiple of 10^n (with n being an integer) of the total reported capital, we replace the latter with the sum of the categories. Second, we drop duplicate observations within the year (i.e. observations which have the same values for all variables in the survey but differ in their plant identification number). Third, we re-compute value added whenever their values are not consistent with the formula provided by BPS. Finally, the survey changed their industry classification from ISIC Rev. 2 in 1998 to ISIC Rev. 3 in 1999 and to ISIC Rev. 4 in 2010. We use United Nations concordance tables to construct a consistent time series of four digit industry classification.

Further, the surveys from 1996 and 2006 provides only information on the aggregate capital stock, yet, not disaggregated by capital type (structure and equipment). To construct an economically reasonable estimate of these variables for these years, we use the average reported investment share and capital share of capital type in the preceding and subsequent year, and impute it, multiplying the aggregate capital stock and investment with the respective share.

Finally, we impute capital stock for plants, whenever the survey presents missing values for this variable in plants which reported information in previous and/or subsequent years. Following Vial (2006), we impute capital by type (machinery, vehicles, land and buildings), using the following regression by two-digit sectoral level:

$$\log K_{it} \;=\; \beta_0 + \beta_1 \log K_{it-1} + \theta \ln X_{it-1} + \mu_i + \epsilon_{it}$$

where K_{it} is the capital stock of type i, μ_i plant fixed effects and X_{it-1} a set of explanatory variables (total output, input, employees, wages, fuel costs and expenditures on materials, leasing, industrial services and taxes).[13]

4.A.1.4 Perpetual inventory method

Whenever the dataset does not directly provide information on a firm's/plant's capital stock at current values (USTAN and EAM), we re-calculate capital stocks using the perpetual inventory method (PIM), in order to obtain economically meaningful capital series. In doing so, we follow Bachmann and Bayer (2014). To begin with, we compute nominal investment series using the accumulation identity for capital stocks:

$$p^I_t I_{i,k,t} \;=\; K^r_{i,k,t+1} - K^r_{i,k,t} + D^r_{i,k,t},$$

where $K^r_{i,k,t}$ and $D^r_{i,k,t}$ are firm/plant i's reported capital stock and depreciation for capital type k at time t, respectively. Given that capital is reported at historical prices and does not reflect the productive (real) level of capital stock, we apply the PIM to construct economic real capital stock at each type of capital:

$$K_{i,k,1} \;=\; \frac{p^I_1}{p^I_{base}} K^a_{i,k,1}; \quad K_{i,k,t+1} \;=\; K_{i,k,t}(1-\delta_{i,k,t}) + \frac{p^I_t}{p^I_{base}} I_{i,k,t}, \;\; \forall t \;\in\; [0,T]$$

[13] We evaluate the robustness of the imputation procedure, using linear interpolation as an alternative approach. Our empirical findings are robust to this alternative specification.

where $K^a_{i,k,1}$ is the accounting value of the capital stock of type k for the first period we observe the unit, $\frac{p_t}{p_{base}} I_{i,k,t}$ is the real investments in capital k of firm/plant i at time t and $\delta_{i,k,t}$ is the reported depreciation rate of capital k by firm/plant i at time t.[14]

Even though the aforementioned procedure makes sure that values follows a economically meaningful real capital stock series from second period onwards, it is not clear whether the starting (accounting) input of capital at the unit, $K^a_{i,k,t}$, reflects the productive real value. To account and adjust the first period value of capital we use an iterative approach. In specific, we construct a time average factor ϕ_k for each type of capital. It the first iteration step, the adjustment factor takes value of 1 while capital is equal to its balanced sheet value. That is, $K^n_{i,k,t} = \frac{p^I_t}{p^I_{base}} K^a_{i,k,1}$ for $n = 1$. For the subsequent iterations, capital is computed using PIM:

$$K^n_{i,k,t+1} \;=\; K^n_{i,k,t}(1 - \delta_{i,k,t}) + \frac{p_t}{p_{base}} I_{i,k,t},$$

while the ajdustment factor is constructed using the ratio between the capital of consecutive iterations

$$\phi^n_k \;=\; \frac{1}{NT} \sum_{i,t} \frac{K^n_{i,k,t}}{K^{n-1}_{i,k,t}}.$$

Finally, the capital stock at the first period we observe the unit is adjusted by the factor ϕ^n_k. We apply the procedure iteratively until ϕ_k converges[15]

$$K^n_{i,k,1} \;=\; \phi^{n-1}_k K^{n-1}_{i,k,1}.$$

[14] The reported depreciation rate is adjusted such that, on average, it coincides with the economic depreciation rate given by National Accounts. To deflate investment series, we compute an investment good price deflator from each country using the information of gross fixed capital formation at current and constant prices from National Accounts.

[15] We stop whenenever the value of ϕ_k is below 1.1. At each iteration step we drop 0.1% from the bottom and the top of the capital distribution. This cleaning step makes sure to not consider episodes of extraordinary depreciation at the plant, which implies that using reported depreciation rate (adjusted to have the same average value from National Accounts) do not reflect the capital stock given by the PIM.

4.A.2 Second order approximation of unit costs

For convenience, let us define the relative factor price by $\tilde{R_t} := \frac{R_t}{W_t}$ and (physical) output per worker by

$$f(k_{it}) := \frac{Y_{it}}{N_{it}} = \left[\alpha k_{it}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}.$$

Subsequently, marginal costs may be expressed as

$$c_{it} = W_t \frac{1+\tilde{R_t}k_{it}}{f(k_{it})}$$

and the first derivative of (log) marginal costs with respect to (log) capital intensity,

$$\begin{aligned}\frac{\partial \log(c_{it})}{\partial \log(k_{it})} &= \frac{\tilde{R_t}k_{it}}{1+\tilde{R_t}k_{it}} - \frac{k_{it}f'(k_{it})}{f(k_{it})} \\ &= \frac{(1-\alpha)\tilde{R_t}k_{it} - \alpha k_{it}^{\frac{\sigma-1}{\sigma}}}{(1+\tilde{R_t}k_{it})(\alpha k_{it}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)A_t^{\frac{\sigma-1}{\sigma}})}\end{aligned}$$

Let us denote above denominator by $D \equiv (1+\tilde{R_t}k_{it})(\alpha k_{it}^{\frac{\sigma-1}{\sigma}} + (1-\alpha)A_t^{\frac{\sigma-1}{\sigma}})$, and obtain the second derivative as

$$\frac{\partial^2 \log(c_{it})}{\partial \log(k_{it})^2} = \frac{\left[(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t} - \frac{\sigma-1}{\sigma}\alpha k_{it}^{-\frac{1}{\sigma}}\right]k_{it}D - \left[(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t}k_{it} - \alpha k_{it}^{\frac{\sigma-1}{\sigma}}\right]D'k_{it}}{D^2}.$$

The cost-minimizing capital intensity k^* implies $\frac{\partial \log(c_{it})}{\partial \log(k_{it})}\Big|_{k_{it}=k^*} = 0$, and the second derivative evaluated at $k_{it} = k^*$, where $(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t}k_{it}^* = \alpha k_{it}^{*\frac{\sigma-1}{\sigma}}$, is

$$\begin{aligned}\frac{\partial^2 \log(c_{it})}{\partial \log(k_{it})^2}\Big|_{k_{it}=k^*} &= \frac{(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t}k_{it}^* - \frac{\sigma-1}{\sigma}\alpha k_{it}^{*\frac{\sigma-1}{\sigma}}}{D} \\ &= \frac{(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\frac{1}{\sigma}\tilde{R_t}k_{it}^*}{(1+\tilde{R_t}k^*)((1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t}k^* + (1-\alpha)A_t^{\frac{\sigma-1}{\sigma}})} = \frac{1}{\sigma}\frac{\tilde{R_t}k^*}{(1+\tilde{R_t}k^*)^2},\end{aligned}$$

where the second equation results again from $(1-\alpha)A_t^{\frac{\sigma-1}{\sigma}}\tilde{R_t}k^* = \alpha {k^*}^{\frac{\sigma-1}{\sigma}}$. The 2nd order Taylor expansion directly follows as

$$\log(c_{it}) - \log(c^*) \;\approx\; \sigma^{-1}\frac{\tilde{R_t}k^*}{(1+\tilde{R_t}k^*)^2}\frac{1}{2}(\log(k_{it}) - \log(k^*))^2.$$

5

Delta-Method Inference for a Class of Set-Identified SVARs

Joint with Bulat Gafarov and José Luis Montiel Olea

5.1 Introduction

An increasingly popular practice in empirical macroeconomics is to set-identify the parameters of a Structural Vector Autoregression [SVAR]. This approach was pioneered by Faust (1998), Canova and Nicoló (2002) and Uhlig (2005). Most of the follow-up studies have relied on Bayesian methods to construct posterior credible sets for the structural coefficients of the impulse-response function.

There has been recent interest in studying non-Bayesian approaches to summarize uncertainty in set-identified SVARs. Moon et al. (2013) [MSG13] propose Projection/Bonferroni frequentist inference based on a moment-inequality-minimum-distance framework. Giacomini and Kitagawa (2015) [GK14] propose robust-Bayesian inference using multiple priors for rotation matrices. Gafarov et al. (2016) [GMM16] propose frequentist inference based on the projection of a Wald ellipsoid for the SVAR reduced-form parameters. None of these approaches requires the specification of prior beliefs by the researcher.

This paper contributes to the non-Bayesian analysis of set-identified SVARs by proposing a novel delta-method confidence interval for the coefficients of the impulse-response function [IRF]. Broadly speaking, our approach is based on a closed-form characterization of the endpoints of the identified set (given a vector of reduced-form parameters and a collection of binding inequality constraints). Our delta-method confidence interval takes the form of a

plug-in estimator for the identified set plus/minus standard errors. In terms of theoretical results, we establish the pointwise consistency in level of our confidence interval. In terms of practical considerations, we argue that the computational cost of our procedure compares very favorably with other non-Bayesian procedures and also with the standard Bayesian algorithm described in Uhlig (2005).

The main limitation of our approach is that the delta-method confidence interval is only defined for SVAR models that impose equality and inequality restrictions on a single structural shock (e.g., a monetary shock). Admittedly, this is problematic, as some popular applications of set-identified SVARs feature restrictions on multiple structural innovations.[1] In spite of this observation, single-shock set-identified models have been applied in several empirical studies: the effects of monetary policy on output [Uhlig (2005)], the impact of monetary policy on the housing market [Vargas-Silva (2008)], the effects of labor market shocks on worker flows [Fujita (2011)], the effects of exchange rates on aggregate prices [An and Wang (2012)], and the effect of optimism shocks on business cycles fluctuations [Beaudry et al. (2014)]. Thus, we think there is room for our results to have an impact on empirical work.

Empirical Application—Unconventional Monetary Policy Shocks: To illustrate the usefulness of our main results, we estimate a monetary Structural Vector Autoregression using monthly U.S. data from July 1979 to December 2007 (a sample that deliberately ends one semester before the financial crisis begins). The goal of our exercise is to use pre-crisis data to learn about the responses of macroeconomic variables to shocks that have effects similar to the 'unconventional' monetary policy interventions implemented after the crisis.

In 'conventional' descriptions of monetary policy, the short-term nominal interest rate is assumed to be the central bank's policy instrument. Following any adjustment by the monetary authority, the market participants—households and firms, both domestic and foreign—use available information to form expectations about the future level of longer-term real interest rates relevant for their consumption and investment decisions.

The recent *Great Recession* has forced the Federal Reserve to consider alternative mechanisms to affect market beliefs about the future of real interest rates. Two examples of such unconventional policies are the Federal Open Market Committee's *forward guidance* an-

[1] SVAR applications for the oil market set-identify both demand and supply shocks using sign restrictions and elasticity bounds [Kilian and Murphy (2012b)]. The same is true for recent labor market applications [Baumeister and Hamilton (2015a)]. Mountford and Uhlig (2009)—one of the most cited applications of set-identified SVARs—use sign restrictions to identify a government revenue shock as well as a government spending shock, while controlling for a generic business cycle shock and a monetary policy shock.

nouncements and the Federal Reserve's *large-scale asset purchases program*. Broadly speaking, through forward guidance "the Federal Open Market Committee provides an indication to households, businesses, and investors about the stance of monetary policy expected to prevail in the future".[2] In a similar fashion, the asset purchase program of the Federal Reserve intends to "put downward pressure on yields of a wide range of longer-term securities, support mortgage markets, and promote a stronger economic recovery".[3]

With this motivation in mind, we set-identify an *unconventional* monetary policy [UMP] shock as an innovation that decreases the two-year government bond rate upon impact, but has no effect over the nominal federal funds rate.[4] We consider two additional sign restrictions on the contemporaneous responses of inflation and output. Namely, we assume that—upon impact—neither inflation nor output can respond negatively to a UMP shock. Since the model is only set-identified, our analysis effectively captures the effects of any historical economic shock that affected the economy in the same way as an UMP shock.

We apply our delta-method approach to construct a confidence interval for the dynamic responses of Industrial Production, inflation, the two-year government bond rate, and the nominal federal funds rate. We use our confidence bands to assess the effects of the announcement of the second part of the so-called Quantitative Easing program (QE2) in August 2010. Pre-crisis data turns out to be extremely useful to learn about the post-crisis response of macroeconomic aggregates to unconventional monetary policy.

The remainder of the paper is organized as follows. Section 5.2 presents an overview of the main methodological results in this paper. Section 5.3 introduces our empirical application, which is used as a running example throughout the paper. Section 5.4 presents our algorithm to evaluate the endpoints of the identified set. Section 5.5 establishes the differentiability properties of the endpoints. Section 5.6 presents our delta-method approach and establishes its asymptotic validity. Section 5.7 presents the confidence intervals for the dynamic responses to the QE2 program. Section 5.8 concludes. All the proofs are collected in the Appendix.

GENERIC NOTATION: If A is a matrix, A_{ij} denotes the ij-th element of A, vec(A) denotes the vectorization of A, and vech(A) denotes half-vectorization (applicable only if A is symmetric). The Kronecker product between matrices A and B is denoted by $A \otimes B$. The vector $e_i^m \in \mathbb{R}^m$

[2] Link: http://www.federalreserve.gov/faqs/money_19277.htm

[3] Link:http://www.federalreserve.gov/faqs/what-are-the-federal-reserves-large-scale-asset-purchases.htm

[4] The paper focuses on the two-year rate as this variable changed considerably after the announcement of the second round of the Quantitative Easing program. See Krishnamurthy and Vissing-Jorgensen (2011)

denotes the i-th column of the identity matrix of dimension m. If B is a matrix of dimension $n \times n$, $B_i \equiv Be_i^n$ denotes its i-th column. If the dimension of e_i^n is obvious, we ignore the superscript n.

5.2 Model and overview of the main theoretical results

This section introduces notation, the class of SVAR models we consider, and presents a brief overview of the main methodological results in the paper. It is our hope that this brief summary (which contains references to the main propositions and lemmas in the paper) contributes to the understanding of the theoretical basis behind our delta-method confidence interval.

Notation: This paper studies the n-dimensional Structural Vector Autoregression (SVAR) with p lags; i.i.d. structural innovations distributed according to F; and unknown $n \times n$ structural matrix B:

$$Y_t \;=\; A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \;\; \mathbb{E}_F[\varepsilon_t] \;=\; 0_{n\times 1}, \;\; \mathbb{E}_F[\varepsilon_t \varepsilon_t'] \;\equiv\; \mathbb{I}_n, \tag{5.1}$$

see Lütkepohl (2007), p. 362.

The object of interest is the k-th period ahead structural impulse response function of variable i to a particular shock j (e.g., a monetary shock). In the SVAR model this parameter is given by the (k,i,j)-*coefficient of the structural impulse-response function*:

$$\lambda_{k,i,j}(A,B) \;\equiv\; e_i' C_k(A) B_j, \tag{5.2}$$

where $B_j \equiv Be_j$ and e_i and e_j denote the i-th and j-th column of the identity matrix $\mathbb{I}_n$.[5]

An auxiliary object in the estimation of the structural parameters is the vector of *reduced-form parameters* in the SVAR model:

$$\mu \;\equiv\; (\mathrm{vec}(A)', \mathrm{vec}(\Sigma)')' \;\in\; \mathscr{M} \;\subseteq\; \mathbb{R}^d, \quad A \;\equiv\; (A_1, A_2, \ldots, A_p), \quad \Sigma \;\equiv\; BB'. \tag{5.3}$$

---

[5] The transformation $C_k(A)$ that appears in equation (5.2) is defined recursively by the formula $C_0 \equiv \mathbb{I}_n$:

$$C_k(A) \equiv \sum_{m=1}^{k} C_{k-m}(A)\, A_m, \quad k \in \mathbb{N},$$

$A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116.

The parameter A denotes the autoregressive coefficients in the VAR model, while Σ denotes the covariance matrix of residuals. These parameters can be estimated directly from the data by Ordinary Least-Squares. The (reduced-form) parameter space is $\mathscr{M}$.

Set-Identifying Restrictions: Let $\mathscr{R}(\mu) \subseteq \mathbb{R}^n$ be a set of inequality and equality restrictions imposed on B_j.[6] A common practice in empirical macroeconomics is to use the restrictions in $\mathscr{R}(\mu)$ to *set-identify* the structural parameter in (5.2) as a function of the reduced-form parameters in (5.3). In our paper, the set $\mathscr{R}(\mu)$ takes the form:

$$\mathscr{R}(\mu) \equiv \left\{ B_j \in \mathbb{R}^n \;\middle|\; Z(\mu)'B_j = \mathbf{0}_{m_z \times 1} \text{ and } S(\mu)'B_j \geq \mathbf{0}_{m_s \times 1} \right\}, \tag{5.4}$$

where $Z(\mu)$ is a matrix of dimension $n \times m_z$ matrix and $S(\mu)$ is a matrix of dimension $n \times m_s$. The matrix $Z(\mu)$ collects the equality restrictions specified by the researcher (we assume there are m_z of them). The matrix $S(\mu)$ collects the inequality restrictions (we assume there are m_s of them). We assume that both $Z(\mu)$ and $S(\mu)$ are differentiable functions of μ.

Scope: The simple formulation in (5.4) allows the researcher to incorporate any restriction of the form $R(\mu)'B_j \geq 0$, where $R(\mu)$ is differentiable. Thus, our analysis allows for the following identifying restrictions:

1. Restrictions on the responses of variable i at horizon k to an impulse on the j-th shock:

$$e_i' C_k(A) B_j \geq \text{ or } = 0,$$

   as in Uhlig (2005).

2. Long-run restrictions on the response of variable i to an impulse on the j-th shock:

$$e_i' (\mathbb{I}_n - A_1 - \ldots - A_p)^{-1} B_j \geq \text{ or } = 0,$$

   as in Blanchard and Quah (1989).

3. Restrictions on the j-th column of $(H')^{-1}$:

$$e_i' (H')^{-1} e_j = e_i' \Sigma^{-1} B_j \geq \text{ or } = 0,$$

   as in Rubio-Ramirez et al. (2015).

[6] For example, a contractionary monetary policy shock increases interest rates and does not affect prices upon impact.

4. Elasticity bounds as in Kilian and Murphy (2012b); for example, for some $b \in \mathbb{R}$:

$$e_i' B_j / e_{i'}' B_j \geq b \iff (e_i - b e_{i'})' B_j \geq 0,$$

provided $e_{i'}' B_j > 0$.

Endpoints of the identified Set: The main results in this paper concern the endpoints of the identified set for the structural parameters, given μ. The endpoints of the identified set (which we sometimes refer to as the *maximum and minimum* response) are defined as follows:

Definition (Endpoints of the identified set): Given a vector of reduced-form parameters μ we define the endpoints of the identified set for $\lambda_{k,i,j}$ as the functions:

$$\overline{v}_{k,i,j}(\mu) \equiv \sup_{B \in \mathbb{R}^{n \times n}} e_i' C_k(A) B e_j, \text{ s.t. } BB' = \Sigma \text{ and } B e_j \in \mathscr{R}(\mu), \tag{5.5}$$

and

$$\underline{v}_{k,i,j}(\mu) \equiv \inf_{B \in \mathbb{R}^{n \times n}} e_i' C_k(A) B e_j, \text{ s.t. } BB' = \Sigma \text{ and } B e_j \in \mathscr{R}(\mu). \tag{5.6}$$

The function $\overline{v}_{k,i,j}(\mu)$ corresponds to the largest value of the structural parameter, $\lambda_{k,i,j}$ subject to the restriction that $B_j \in \mathscr{R}(\mu)$ and also that B_j is the j-th column of a square root of Σ. The lower bound is defined analogously.

Overview of the main results: Our delta-method confidence interval is supported by the three theoretical results described in the abstract. Our results can be summarized as follows:

• Summary of Lemma 1 *(Characterization of the maximum and minimum response given a fixed set of active constraints)*: We show that $\overline{v}_{k,i,j}(\mu)$ and $\underline{v}_{k,i,j}(\mu)$ are the *value functions* of a mathematical program whose Karush-Kuhn-Tucker points can be described analytically—up to a set of 'active' inequality constraints.[7] More concretely, we show that the maximum

[7] The term 'active constraints' or 'active set of is constraints' is the common terminology used in numerical optimization; see p. 308 in Nocedal and Wright (2006).

response for $\lambda_{k,i,j}$ is equal to either plus or minus the function:

$$v_{k,i,j}(\mu; r) \equiv \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r} \Sigma^{1/2} C_k(A)' e_i \right)^{1/2},$$

where

$$M_{\Sigma^{1/2} r} \equiv \mathbb{I}_n - \Sigma^{1/2} r (r' \Sigma r)^{-1} r' \Sigma^{1/2},$$

and r' is a matrix collecting the gradient vectors of the constraints in $\mathscr{R}(\mu)$ that are active at a maximum. The minimum response is obtained analogously.

• Summary of Proposition 1 *(Algorithm to evaluate the maximum and minimum response)*: We use the closed-form expressions of Lemma 1 to present an algorithm that allows a researcher to evaluate the endpoints of the identified set given a vector of reduced-form parameters. The algorithm evaluates different collections of active constraints (different matrices r) and selects the constraints that generate the largest (or smallest) value function—after checking that the inequality constraints not included in r are satisfied.

• Summary of Lemma 2: *(Differentiability of the maximum and minimum response for a fixed set of active constraints)* We establish the differentiability of the function $v_{k,i,j}$, which depends on a fixed set of active constraints. We allow the constraints in the $n \times l$ matrix r (with $l \leq n-1$) to depend on the reduced-form parameters. We show that the derivative of $v_{k,i,j}$ w.r.t. μ is given by:

$$\dot{v}_{k,i,j}(\mu) \equiv \begin{bmatrix} \frac{\partial v_{k,i,j}(\mu;r)}{\partial \text{vec}(A)} \\ \\ \frac{\partial v_{k,i,j}(\mu;r)}{\partial \text{vec}(\Sigma)} \end{bmatrix} = \begin{bmatrix} \frac{\partial \text{vec}(C_k(A))}{\partial \text{vec}(A)} (x^* \otimes e_i^n) - \sum_{k=1}^{l} w_k^* \frac{\partial r_k}{\text{vec}(A)} x^* \\ \\ \lambda^* (\Sigma^{-1} x^* \otimes \Sigma^{-1} x^*) - \sum_{k=1}^{l} w_k^* \frac{\partial r_k}{\text{vec}(\Sigma)} x^* \end{bmatrix},$$

where r_k is the k-th column of r, w_k^* is the k-th component of w^*, and

$$x^* \equiv \Sigma^{1/2} \Big(M_{\Sigma^{1/2} r} \Big) \Sigma^{1/2} C_k(A)' e_i \Big/ v_{k,i,j}(\mu; r),$$

$$\lambda^* \equiv \frac{1}{2} v_{k,i,j}(\mu; r), \quad w^* \equiv [r' \Sigma r]^{-1} r' \Sigma C_k(A)' e_i.$$

We argue that λ^* and w^* can be interpreted as the Lagrange multipliers associated to the constraints $BB' = \Sigma$ and to the active constraints in r.

• Summary of Proposition 2 *(Directional Differentiability of the endpoints)*: We use the formula in Lemma 2 to show that the functions $\overline{v}_{k,i,j}(\cdot)$ and $\underline{v}_{k,i,j}(\cdot)$ are *directionally* differentiable, in a sense we make precise. We relate the expression of the directional derivative with the generalized versions of the envelope theorems in the work of Fiacco and Ishizuka

(1990) and Bonnans and Shapiro (2000). We argue that directional differentiability of the value functions (as opposed to full differentiability) arises due to the possibility that different structural models lead to the maximum (or minimum) response. In particular, let $R^*(\mu)$ denote the sets of active restrictions that yield the same maximum response and assume, for simplicity, that $\overline{v}_{k,i,j}(\mu) > 0$. We show that:

$$\sqrt{n}\Big(\overline{v}_{k,i,j}(\mu + h_n/\sqrt{n}) - \overline{v}_{k,i,j}(\mu)\Big) \to \max_{r \in R^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r)'h\Big].$$

• Summary of Proposition 3 *(Delta-Method Confidence Interval)*: We establish the pointwise consistency in level of a delta-method confidence interval, which takes the form:

$$\text{CS}_T(1-\alpha; \lambda_{k,i,j}) \equiv \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big],$$

where $\widehat{\mu}_T$ is the typical OLS estimator for the VAR reduced-form parameters, $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of a standard normal, and $\widehat{\sigma}_{(k,i,j),T}$ is our formula for the standard errors based on the directional derivatives.

Outline: In the remaining part of the paper, we formalize these propositions and apply them to conduct inference about the responses to an *unconventional monetary shock*.

5.3 Running example: unconventional monetary policy shocks

This section introduces our empirical application, which will be used as a running example to illustrate our assumptions and results.

Monetary Svar: We consider a simple 4-variable model that includes the Consumer Price Index (CPI_t), the Industrial Production Index (IP_t), the 2-year Treasury Bond rate ($2yTB_t$), and the Federal Funds rate (FF_t).[8] We take a logarithmic transformation of CPI_t, IP_t and then work with first differences for all variables. Thus, our vector of macro variables is:

$$Y_t \equiv \Big(\ln CPI_t - \ln CPI_{t-1},\quad \ln IP_t - \ln IP_{t-1},\quad 2yTB_t - 2yTB_{t-1},\quad FF_t - FF_{t-1}\Big)'.$$

[8] All these variables are sourced from the data set of Gertler and Karadi (2015). We thank Peter Karadi for making their data set available to us.

We set the number of lags equal to 11 using the Bayesian Information Criterion ($p = 11$). The time span of the monthly series is July 1979 to August 2008 ($T = 342$). To keep our exposition as simple as possible, we ignore potential co-integration issues between short-term and long-term interest rates. Without loss of generality, we assume that column of B corresponding to an UMP shock is the first column; $B_1 \equiv Be_1$. Our equality/inequality restrictions are summarized in the following Table:

Table 5.1. Set-identification of an unconventional monetary policy shock: restrictions

Series	Acronym	UMP	Notation
Consumer Price Index	CPI	$+$	$e_1' B_1 \geq 0$
Industrial Production	IP	$+$	$e_2' B_1 \geq 0$
2-year Treasury Bond rate	2yTB	$-$	$e_3' B_1 \leq 0$
Fed Funds Rate	FF	0	$e_4' B_1 = 0$

DESCRIPTION: Restrictions on contemporaneous responses to a UMP shock. '0' stands for a zero restriction, '—' stands for a negative sign restriction and '+' for positive sign restriction. These sign restrictions can be justified by the DSGE model calibrated in the work of Bhattarai et al. (2014).

The suggested set-identification strategy in this paper is not new. Baumeister and Benati (2013) study an analogous 'spread' monetary shock that leaves the short-term nominal rate unchanged, but affects the spread between the ten-year Treasury-bond yield and the policy rate. They consider a Bayesian SVAR with time varying parameters and stochastic volatility combined with demand and supply structural shocks that satisfy zero/sign restrictions as in Rubio-Ramirez et al. (2010). Their main result is that the long-term yield spread exerts a powerful effect on both output growth and inflation. All their inference is Bayesian, while ours is frequentist. In addition, our SVAR model does not consider time-varying parameters, stochastic volatility, and restrictions on other nonmonetary shocks.

5.4 The endpoints of the identified set

In this section we formalize Lemma 1 and Proposition 1. We consider the problem of finding the maximum response to an impulse in the j-th structural shock subject to m_z equality ('zero') restrictions and m_s inequality ('sign') restrictions. The focus on the maximum and the minimum is an intermediate step to conduct frequentist inference about the coefficients of the impulse-response function.

This section makes two assumptions on the sign and zero restrictions allowed in the model. First, we require the number of zero restrictions to be less than $n-1$. Second, we assume that every collection of $n-1-m_z$ inequality restrictions and the m_z equality restrictions are linearly independent everywhere in the parameter space.

Just as before, the set $\mathscr{R}(\mu)$ is given by

$$\mathscr{R}(\mu) \equiv \left\{ B_j \in \mathbb{R}^n \;\middle|\; Z(\mu)' B_j = \mathbf{0}_{m_z \times 1} \text{ and } S(\mu)' B_j \geq \mathbf{0}_{m_s \times 1} \right\},$$

where $Z(\mu)$ is a matrix of dimension $n \times m_z$ matrix and $S(\mu)$ is a matrix of dimension $n \times m_s$.

Running Example—$\mathscr{R}(\mu)$: In the UMP example, the set of restrictions $\mathscr{R}(\mu)$ corrresponds to (see Table I):

$$\left\{ B_1 \in \mathbb{R}^4 \;\middle|\; e_4' B_1 = 0, (e_1, e_2, -e_3)' B_1 \geq \mathbf{0}_{3\times 1} \right\}.$$

Consequently:

$$Z(\mu) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ and } S(\mu) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

We note that the equality and inequality restrictions in our example do not depend on the reduced-form parameters (neither Z nor S depend on μ).

Main Assumptions: The first assumption in this section requires the number of zero restrictions to be strictly smaller than $n-1$. The rationale behind Assumption 1 is as follows: if $m_z > n-1$, then $\mathscr{R}(\mu) = \{\mathbf{0}_{n\times 1}\}$ for every μ for which there are n linearly independent equality restrictions. This is problematic, as the latter implies there is no $B_j \in \mathbb{R}^n$ such that $B_j \in \mathbb{R}^n$ and $B_j = Be_j$ for some $BB' = \Sigma$ (provided B is invertible).[9]

Assumption 1. $m_z \leq n-1$.

Our second assumption imposes a 'linear independence' condition on the equality and inequality restrictions on the model (given a particular value of the reduced-form parameter μ). Let $e_1^{m_s}, e_2^{m_s}, \ldots e_{m_s}^{m_s}$ denote the m_s different columns of the identity matrix $\mathbb{I}_{m_s}$. Let $e(k)$

[9] If B is invertible, then Σ is invertible and $B'\Sigma^{-1}B = \mathbb{I}_n$, which implies $B_j'\Sigma^{-1}B_j = 1$. Therefore, if $B_j = Be_j$ for some square root of Σ then B_j must be different from $\mathbf{0}_{n\times 1}$.

denote an $m_s \times k$ matrix formed by collecting any of the $k \leq n-1-m_z$ columns of $\mathbb{I}_{m_s}$. Note that for any matrix S, the matrix $Se(k)$ selects k columns of S.

Definition. We say that $Z(\mu)$ and $S(\mu)$ are linearly independent at μ if for any $k \in \mathbb{N}, k \leq n-1-m_z$ and any $e(k)$ the matrix

$$R(\mu; e(k)) \equiv [Z(\mu), S(\mu)e(k)] \in \mathbb{R}^{n \times (m_z+k)},$$

is assumed to have full column rank (rank $m_z + k$).

We use this definition to state our following assumption:

Assumption 2. The parameter space $\mathscr{M}$ is such that $Z(\mu)$ and $S(\mu)$ are linearly independent at every $\mu \in \mathscr{M}$.

This 'linear independence' property plays an important role in the characterization of the maximum and minimum response in terms of *Karush-Kuhn-Tucker* conditions.

Running Example—Assumption 1 and 2: In our application $m_z = 1$, so we only need to verify Assumption 2. As we mentioned before, the matrix $e(k)$ is a selector matrix. For example, let $e(2)$ be given by the first and third column of $\mathbb{I}_3$; that is

$$e(2) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

This implies that

$$S(\mu)e(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix},$$

and moreover:

$$R(\mu; e(2)) \equiv [Z(\mu), S(\mu)e(2)] = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{4 \times (2+1)}.$$

Thus, the matrix $R(\mu; e(2))$ is formed by collecting the gradient of the unique equality restriction and the first and third inequality restrictions in S. Note that regardless of the

number of k columns selected from S—and regardless of what $\mathscr{M}$ is—the resulting matrix $R(\mu, e(k))$ will always have full column rank.

Verifying Assumption 2 with more general restrictions requires additional work. For example, suppose that the researcher is interested in including the restriction:

$$e_2' C_1(A) B_1 \geq 0.$$

This restriction says that the UMP shock cannot decrease the growth rate in Industrial Production even one-period after the shock. Since $C_1(A) = A_1$, the vector $e_2' C_1(A)$ is equal to the second row of A_1, which we can denote as $(A_{1,(2,1)}, A_{1,(2,2)}, A_{1,(2,3)}, A_{1,(2,4)})$. The matrix $S(\mu)$ is now given by:

$$S(\mu) = \begin{pmatrix} 1 & 0 & 0 & A_{1,(2,1)} \\ 0 & 1 & 0 & A_{1,(2,2)} \\ 0 & 0 & -1 & A_{1,(2,3)} \\ 0 & 0 & 0 & A_{1,(2,4)} \end{pmatrix}.$$

Hence, we conclude that Assumption 2 will be satisfied as long as $\mathscr{M}$ is such that $A_{1,(2,j)} \neq 0$ for all $j = 1, \ldots 4$, which means that each of the entries in the first lag of Y_{t-1} has predictive power on Y_t after controlling for the rest of the lags.

The third assumption is the following:

Assumption 3. The matrices $Z(\mu)$ and $S(\mu)$ are differentiable functions of the reduced-form parameter μ.

We are not aware of equality/inequality restrictions in the SVAR literature that does not satisfy this property. In particular, all the examples given in p. 5 of this paper satisfy Assumption 3.

5.4.1 Lemma 1: Closed-form solution for the maximum response given an active set of constraints

In this section we show that that given a collection $r \in \mathbb{R}^{m \times n}$ of 'active' constraints ($m \leq n-1$) the maximum response is determined in closed-form (and up to sign) by the Karush-Kuhn-Tucker conditions of the program (5.5) and (5.6).

Lemma 1. *Let r be a matrix of dimension $n \times m$, $m \leq (n-1)$ collecting the gradients of the 'active' (binding) constraints at a solution of the mathematical program (5.5). Suppose that Assumption 1 holds and suppose that $Z(\mu)$ and $S(\mu)$ are linearly independent at μ. Then, if $\overline{\nu}_{k,i,j}(\mu) \neq 0$:*

a) $\overline{\nu}_{k,i,j}(\mu)$ is given by either plus or minus the norm of the residual of the projection of $\Sigma^{1/2} C_k(A)' e_i$ into the space spanned by the columns of $\Sigma^{1/2} r$; that is:

$$\overline{\nu}_{k,i,j}(\mu) \;=\; \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2}, \tag{5.7}$$

or

$$\overline{\nu}_{k,i,j}(\mu) \;=\; -\left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2}, \tag{5.8}$$

where

$$M_{\Sigma^{1/2} r} \equiv \mathbb{I}_n - \Sigma^{1/2} r (r' \Sigma r)^{-1} r' \Sigma^{1/2}.$$

b) In addition, there is a unique maximizer $x^*(\mu; r)$ such that $r' x^*(\mu; r) = \boldsymbol{0}_{m \times 1}$ and is given by:

$$x^*(\mu; r) = \Sigma^{1/2} \left(M_{\Sigma^{1/2} r}\right) \Sigma^{1/2} C_k(A)' e_i \Big/ \overline{\nu}_{k,i,j}(A, \Sigma).$$

Consequently, the sign of $\overline{\nu}_{k,i,j}(\mu)$ depends on which of the two solutions $x^*(\mu; r)$ (the one with (5.7) in the denominator or the one with (5.8)) satisfies the sign restrictions that are not in r.

Proof: See Appendix 5.A.1 for the proof, which uses the necessary Karush-Kuhn-Tucker conditions of the optimization problem to characterize the maximizers given a set r of active constraints.[10] Figure 5.1 presents a graphical representation of the mathematical program of interest. Figure 5.2 presents an intuitive description of the solution.

One way to think about the solution to the problem of interest is explained in Figure 5.2. Suppose there are only equality constraints. Note that $Z' B_j = \mathbf{0}_{m \times 1}$ implies that re-parameterized choice variable $\widetilde{x} \equiv \Sigma^{-1/2} B_j$ must lie on the orthogonal space of $\Sigma^{1/2} Z$. That is, the selected value of $\widetilde{x}$ should be of the form:

$$\widetilde{x} = M_{\Sigma^{1/2} Z} y, \quad M_{\Sigma^{1/2} Z} \equiv \left(\mathbb{I}_n - \Sigma^{1/2} Z (Z' \Sigma Z)^{-1} Z' \Sigma^{1/2}\right), \; y \in \mathbb{R}^n.$$

[10] To guarantee the existence of Karush-Kuhn-Tucker multipliers we use the fact that $Z(\mu)$ and $S(\mu)$ are linear independence at μ. Our assumption implies that the mathematical program defining the endpoints of the identified set satisfies a linear independence constraint qualification (see Fiacco and Ishizuka (1990), p. 224).

Figure 5.1. The mathematical program defining $\overline{v}_{k,i,j}(A, \Sigma)$ $(n = 3)$ with one zero restriction.

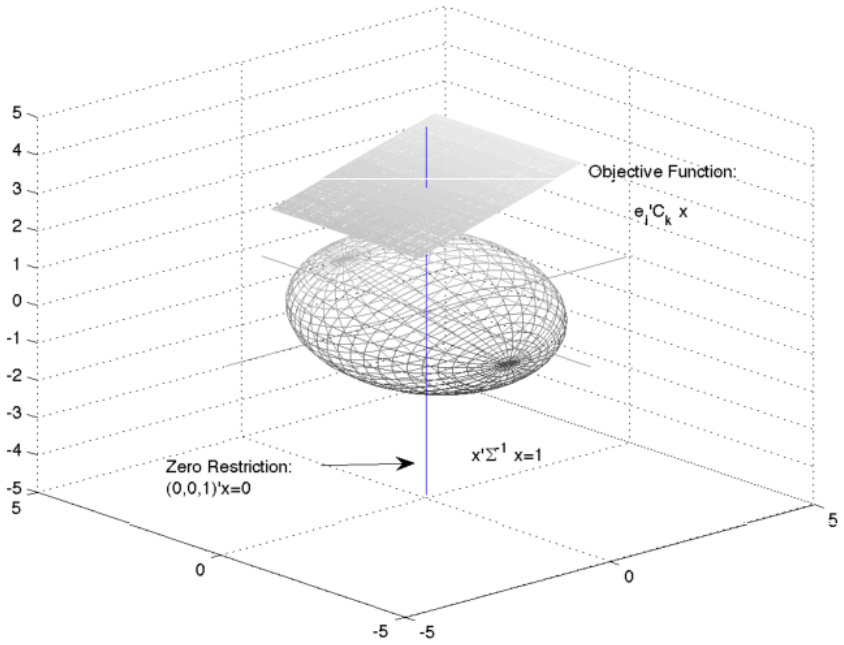


Figure 5.1 provides a graphical representation of the mathematical program (5.5), where $BB' = \Sigma$ has been replaced by the 'ellipsoid' constraint $x'\Sigma^{-1}x = 1$, $x \equiv B_j \in \mathbb{R}^3$ (this equivalence will not hold, in general, if there are restrictions on multiple shocks). The objective function corresponds to the hyperplane with normal vector $C_k(A)'e_i \in \mathbb{R}^3$. In this example, there is only one equality restriction with normal vector given by the (BLUE, SOLID) line. This restriction requires the contemporaneous impact of the j-th shock on the third variable to be zero. Note that without the equality restriction the maximizer and minimizer will be given by the point at which the hyperplane is tangent to the ellipsoid.

The quadratic equality constraint also restricts the choice variable $\widetilde{x}$ to satisfy $\widetilde{x}'\widetilde{x} = 1$. Consequently, the problem can be re-written as

$$\max_{y \in \mathbb{R}^n} e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z} y \quad \text{s.t.} \quad y' M_{\Sigma^{1/2}Z} y = 1.$$

An application of the Cauchy-Schwartz inequality shows that the positive value in (5.7) gives the maximum response in (5.5).[11]

[11] Using the fact that $M_{\Sigma^{1/2}Z}$ is idempotent and using the assumption that

$$\left(e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z} \Sigma^{1/2} C_k' e_i\right)^{1/2} \neq 0,$$

Figure 5.2. Solving for $\overline{v}_{k,i,j}(A, \Sigma)$ ($n = 3$, $\Sigma = \mathbb{I}_3$) with one equality restriction.

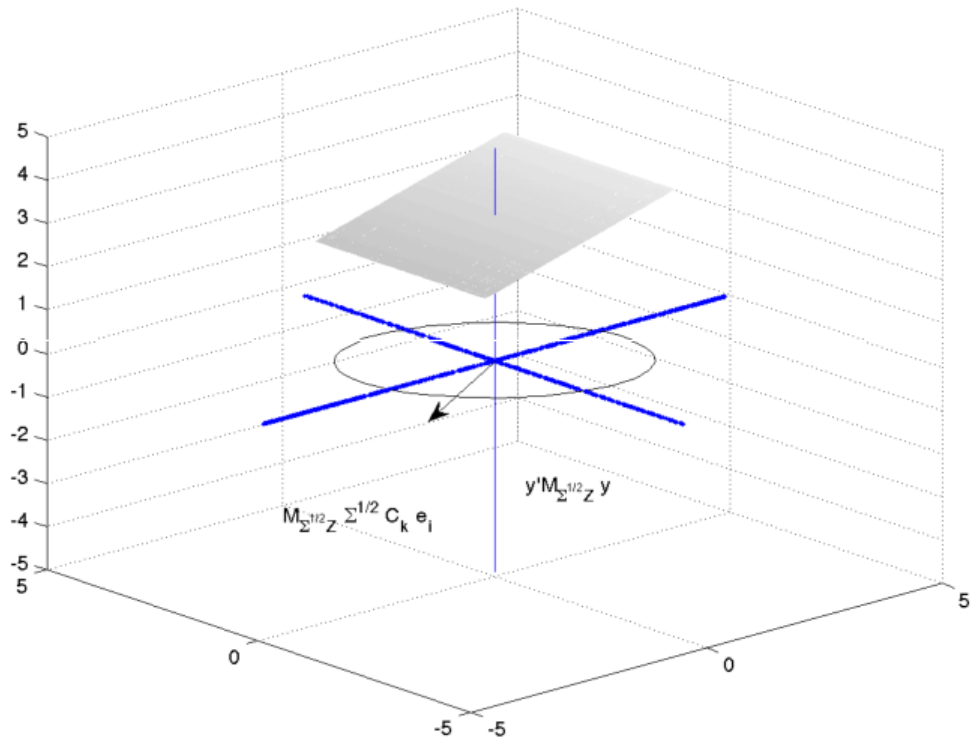


Figure 5.2 provides a graphical representation of the solution to the mathematical program (5.5) when $\Sigma = \mathbb{I}_3$ and there is only one zero restriction. The solution to the program must lie in the orthogonal complement of Z (BLUE, THIN, SOLID). In this picture the orthogonal complement corresponds to the space spanned by the BLUE, THICK, SOLID lines. This implies that the rotated solution, denoted $\widetilde{x} \equiv \Sigma^{-1/2}x$, must be of the form $M_{\Sigma^{1/2}Z}y$ for some $y \in \mathbb{R}^3$. Hence, the only relevant part of $x'\Sigma^{-1}x = 1$ becomes the projected version of it: $y'M_{\Sigma^{1/2}}y = 1$, represented by the BLACK, SOLID ellipsoid. One can find the value of this problem by projecting the gradient of the objective function on the orthogonal complement of $\Sigma^{1/2}z$ (arrow) and selecting a direction in the ellipsoid proportional to it. The value function $\overline{v}_{k,ij}(A, \Sigma)$ will be given by the norm of the arrow.

---

the problem of interest becomes:

$$\max_{y \in \mathbb{R}^n} \left(e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z} \Sigma^{1/2} C_k' e_i \right)^{1/2} \left[\frac{e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z}}{\left(e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z} \Sigma^{1/2} C_k' e_i \right)^{1/2}} \right] M_{\Sigma^{1/2}Z} y,$$

s.t. $y' M_{\Sigma^{1/2}Z} y = 1$. By the Cauchy-Schwartz inequality this program is bounded above by $(e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}Z} \Sigma^{1/2} C_k' e_i)^{1/2}$. This value can be achieved by $x^*(A, \Sigma; Z)$ in Lemma 1.

5.4.2 Proposition 1: Algorithm to evaluate the maximum and minimum response

We have provided a closed-form expression (up to a sign) for the maximum response $\overline{\nu}_{k,i,j}(\mu)$, given a collection r of active restrictions. We now answer the following question: how does one compute the maximum response $\overline{\nu}_{k,i,j}(\mu)$ for a given value of μ?

We use the result in Lemma 1 to state the solution of the mathematical program (5.5) that includes both equality and inequality restrictions. The main result in this section is that such problem can be solved by 'activating' different combinations of inequality constraints. In other words, the problem in (5.5) can be solved by finding the largest value among the Karush-Kuhn-Tucker points that satisfy a feasibility constraint.

Additional Notation Illustrated with our Example—1) Collection of Active Constraints: Fix (A, Σ) and, in a slight abuse of notation let Z and S denote $Z(\mu)$ and $S(\mu)$. Define first:

$$R_0 \equiv Z,$$

as the $\mathbb{R}^n$ matrix that collects all of the m_z zero restrictions. Hence, in our empirical application:

$$R_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Define also:

$$R_0 \equiv \left\{ R \in \mathbb{R}^{n \times (m_z+1)} \,\middle|\, R = [R_0, Se_i^{m_s}], \quad i \in \{1, \dots m_s\} \right\},$$

as the collection of all matrices that *activate* one of the m_s inequality restrictions; analgously, R_1 corresponds to the collection of matrices that impose one of the inequality restrictions as

an equality restriction. In our example:

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore,

$$R_1 = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \right\}.$$

More generally, for $l \leq n - m_z - 1$, consider the collection:

$$R_l \equiv \left\{ R \in \mathbb{R}^{n \times (m_z + l)} \,\middle|\, R = [R_0, Se_{m_1}^{m_s}, \ldots Se_{m_l}^{m_s}],\ \{m_i\}_{i=1}^{l} \text{is a subsequence of } \{1, \ldots m_s\} \right\}.$$

The matrix $r_l \in R_l$ activates l of the m_s sign-restrictions in the SVAR model. Note that the collection R_l has $m_s!/(l!(m_s - l)!)$ elements and R_{m_s} has a unique element in which all the sign restrictions of the model are active (provided $m_s \leq n - m_z - 1$). In our example, $n - m_z - 1 = 2$. There are 3 different subsequences of two elements from the sequence $\{1, 2, 3\}$: $\{1, 2\}, \{1, 3\}$, and $\{2, 3\}$. Therefore,

$$R_2 = \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

Thus, R_1 and R_2 denote the different collection of active constraints formed by choosing one and two of the elements of S, respectively.

Additional Notation Illustrated with our Example—2) Feasibility: We define the feasibility of a vector $x \in \mathbb{R}^n$ (with respect to the sign restrictions) as the indicator function

$$\mathbf{1}_{m_s}(x) \;=\; \mathbf{1}\left\{ S'x \;\geq\; \mathbf{0}_{m_s \times 1} \right\} \;\in\; \mathbb{R}, \tag{5.9}$$

where, following convention, $\geq$ is taken component-wise whenever the binary relation is applied to vectors. Hence, $x \in \mathbb{R}^n$ is a feasible point for the mathematical program (5.5) if and only if $\mathbf{1}_{m_s}(x) = 1$ and x satisfies the equality restrictions in Z.[12] In the context of our example:

$$x = \begin{pmatrix} 4 \\ 3 \\ -2 \\ 0 \end{pmatrix} \Longrightarrow \mathbf{1}_{m_s}(x) = 1, \text{ but } y = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 0 \end{pmatrix} \Longrightarrow \mathbf{1}_{m_s}(y) = 0.$$

The point x is a feasible point for the mathematical program in (5.5) as it satisfies both the equality and inequality restrictions.

Proposition 1 (**Algorithm to evaluate the maximum and minimum response**). *For $\mu \equiv (vec(A)', vec(\Sigma)')'$ consider the mathematical programs*

$$\overline{\nu}_{k,i,j}(\mu) \equiv \max_{B_j} e_i' C_k(A) B_j, \quad \underline{\nu}_{k,i,j}(\mu) \equiv \min_{B_j} e_i' C_k(A) B_j,$$

with each program subject to

$$BB' = \Sigma,$$

with equality and inequality restrictions:

$$Z'B_j = \mathbf{0}_{m_z \times 1}, \quad S'B_j \geq \mathbf{0}_{m_s \times 1}.$$

Let

$$R \equiv \bigcup_{l=0}^{\min\{n-1-m_z, m_s\}} R_l,$$

denote all possible combinations of up to $n-1$ active constraints and for $r \in R$ define $\nu_{k,i,j}(\mu; r)$ as the function:

$$\nu_{k,i,j}(\mu; r) = \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r} \Sigma^{1/2} C_k(A)' e_i \right)^{1/2},$$

and let c be a positive and large constant (penalty term) such that $-c < \underline{\nu}_{k,i,j}(A, \Sigma) \leq \overline{\nu}_{k,i,j}(A, \Sigma) < c$ and such that $-c < -\nu_{k,i,j}(A, \Sigma, r) < \nu_{k,i,j}(A, \Sigma, r) < c$ for all $r \in R$.

---

[12] We use the term primal feasibility in contrast with *dual* feasibility, which obtains whenever the Lagrange multipliers for the sign restrictions that are not active are all positive (or negative).

Consider the candidate value functions:

$$v^{+}_{k,i,j}(\mu;r) = v_{k,i,j}(A,\Sigma;r), \quad v^{-}_{k,i,j}(\mu;r) = -v_{k,i,j}(A,\Sigma;r).$$

If $v_{k,i,j}(\mu;r) \neq 0$*, set:*

$$\begin{aligned}
f^{+}_{max}(\mu;r) &\equiv v^{+}_{k,i,j}(\mu;r) - 2(1 - \mathbf{1}_{m_s}(x^{*}_{+}(\mu;r)))c,\\
f^{-}_{max}(\mu;r) &\equiv v^{-}_{k,i,j}(\mu;r) - 2(1 - \mathbf{1}_{m_s}(x^{*}_{-}(\mu;r)))c,\\
f^{+}_{min}(\mu;r) &\equiv v^{+}_{k,i,j}(\mu;r) + 2(1 - \mathbf{1}_{m_s}(x^{*}_{+}(\mu;r)))c,\\
f^{-}_{min}(\mu;r) &\equiv v^{-}_{k,i,j}(\mu;r) + 2(1 - \mathbf{1}_{m_s}(x^{*}_{-}(\mu;r)))c,
\end{aligned}$$

where

$$x^{*}_{+}(\mu;r) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu),$$
$$x^{*}_{-}(\mu;r) = -\Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu).$$

If $v_{k,i,j}(\mu;r) = 0$ *and there is a point* $x^{*} \neq 0$ *satisfying the equality restrictions in* r *and also the inequality restrictions that are not included in* r*, set:*

$$f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = f^{+}_{min}(\mu;r) = f^{-}_{min}(\mu;r) = 0.$$

If $v_{k,i,j}(\mu;r) = 0$ *and there is no point* $x^{*} \neq 0$ *satisfying the equality restrictions in* r *and the inequality restrictions that are not in* r*, set:*

$$f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = -c \quad \text{and } f^{+}_{min}(\mu;r) = f^{-}_{min}(\mu;r) = c.$$

Then:

$$\overline{v}_{k,i,j}(\mu) = \max_{r\in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big),$$

and

$$\underline{v}_{k,i,j}(\mu) = \min_{r\in R}\Big(\min\{f^{+}_{min}(\mu;r), f^{-}_{min}(\mu;r)\}\Big).$$

That is, the value function $\overline{v}_{k,i,j}(\mu)$ *is obtained by computing the Karush-Kuhn-Tucker points in Lemma 1 for each* r*, penalizing the value* $\overline{v}_{k,i,j}(\mu;r)$ *if unfeasible, and maximizing over all the possible values of* r*. The minimum value is obtained analogously.*

PROOF: The intuition behind the proof is as follows. Note that any combination of active sign restrictions r for which $x_+^*(A,\Sigma;r)$ (or $x_-^*(A,\Sigma;r)$) is well-defined and feasible must be, by definition, no larger than $\overline{\nu}_{k,i,j}(A,\Sigma)$. Thus, we only have to show that $\max_{r\in R}\Big(\max\{f_{max}^+(A,\Sigma;r), f_{max}^-(A,\Sigma;r)\}\Big) \geq \overline{\nu}_{k,i,j}(A,\Sigma)$. Since Lemma 1 showed that the value of the program (5.5) should be of the form $f_{\max}^+(A,\Sigma,r)$ or $f_{\max}^-(A,\Sigma,r)$ for some $r \in R$, the result must follow. The proof is formalized in Appendix 5.A.2.

ALGORITHM TO EVALUATE THE ENDPOINTS OF THE IDENTIFIED-SET: The proposition above shows that in order to solve the mathematical problem in (5.5) it is sufficient to apply the following algorithm:

1. *Activate* different combinations of the m_s sign restrictions. Collect the original m_z equality restrictions and the inequality restrictions that were activated in the matrix r. The matrix should have no more than $n-1$ columns. Note that the total number of matrices r will be given by:
$$\sum_{k=0}^{\min\{m_s,n-1-m_z\}} \binom{m_s}{k}.$$

2. Compute the candidate value functions $\pm\nu_{k,i,j}(A,\Sigma;r)$ for each of the elements $r \in R$.

3. If $\nu_{k,i,j}(A,\Sigma;r) \neq 0$, verify if $x_+^*(A,\Sigma;r)$ satisfies the sign restrictions that were not included in r. That is, verify the *feasibility* of the solution $x_+^*(A,\Sigma;r)$. If the primal feasibility condition is satisfied set
$$f_{max}^+(A,\Sigma;r) = f_{min}^+(A,\Sigma;r) = \nu_{k,i,j}(A,\Sigma;r).$$
If the primal feasibility condition is violated penalize $\nu_{k,i,j}(A,\Sigma;r)$ to guarantee that it is never a solution by setting:
$$f_{max}^+(A,\Sigma;r) \equiv \nu_{k,i,j}(A,\Sigma;r) - 2c,$$
and
$$f_{min}^+(A,\Sigma;r) \equiv \nu_{k,i,j}(A,\Sigma;r) + 2c.$$
Check the feasibility of $x_-^*(A,\Sigma,r)$ and proceed in the same way. If $\nu_{k,i,j}(A,\Sigma;r) = 0$ do the adjustment described in Proposition 1.

4. Select the maximum value of $\max\{f^{+}_{max}(A,\Sigma;r), f^{-}_{max}(A,\Sigma;r)\}$ over $r \in R$; that is, consider the different combinations of active restrictions and select the maximum value $\pm\nu_{k,i,j}(A,\Sigma,r)$ over them. This gives $\overline{\nu}_{k,i,j}(A,\Sigma)$. Taking the minimizer over $f^{+}_{min}(A,\Sigma;r), f^{-}_{min}(A,\Sigma;r)$ gives the smallest value.

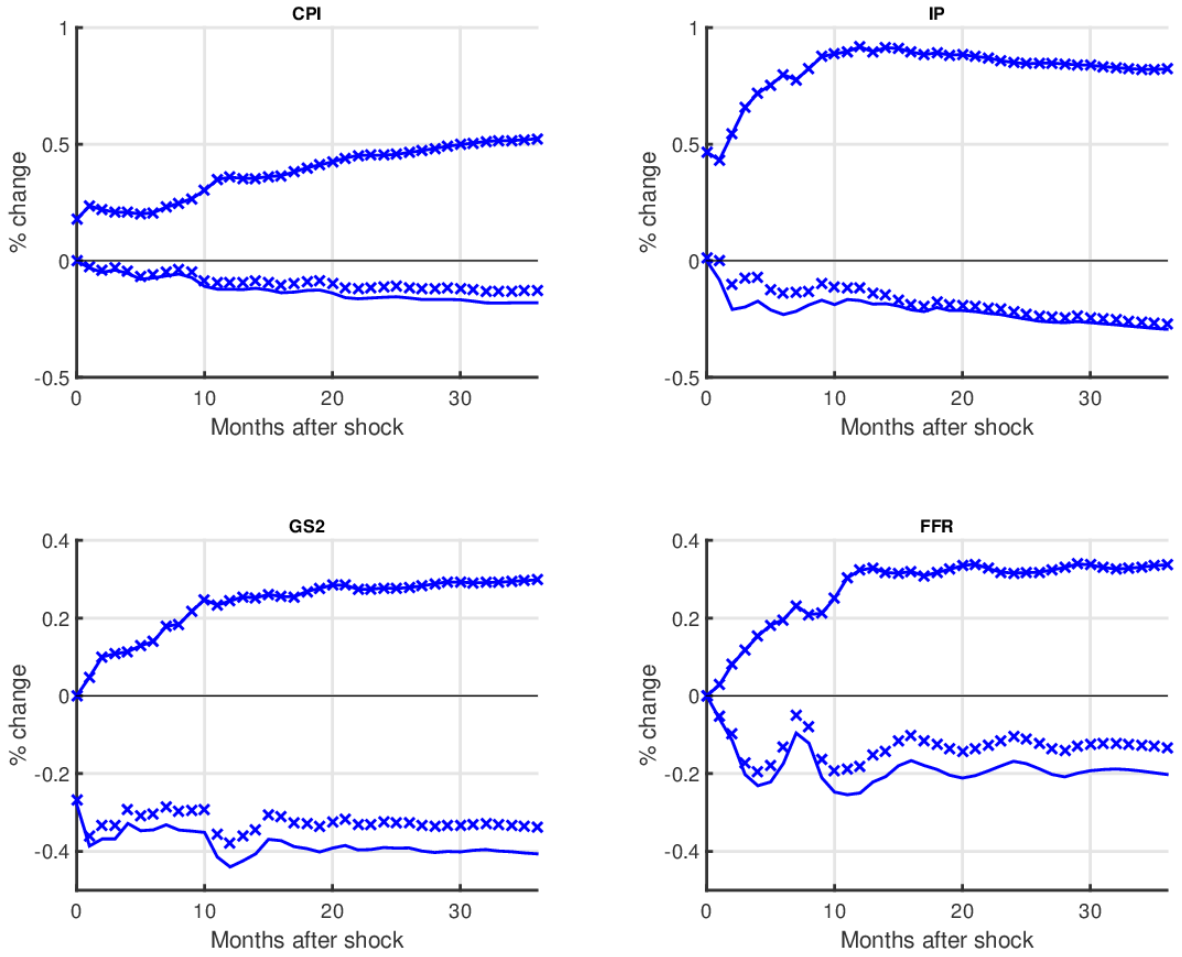
Using the algorithm in the UMP example: We use the algorithm to evaluate the identified set in the running example. We fix μ at its estimated OLS values, denoted $\widehat{\mu}_T$, and we report $\overline{\nu}_{k,i,j}(\widehat{\mu}_T)$ and $\underline{\nu}_{k,i,j}(\widehat{\mu}_T)$ for the cumulative IRFs.[13] The scale in Figure 5.3 corresponds to a one standard deviation structural UMP shock.

We consider first the equality/inequality restrictions in Table I. We note that evaluating the endpoints of the identified for the 4 variables in the VAR, over 40 horizons, takes around .1 seconds. We then include an additional inequality restriction on the response of output to an expansionary UMP shock. Namely, we assume that even one period after the shock, the cumulative effect on IP cannot be negative ($e_2'(C_0 + C_1(A))B_1 \geq 0$). A comparison between the two collections of restrictions suggests that, in this example, the noncontemporaneous constraint has almost no additional identification power.

Of course, one could use the Bayesian algorithm in Uhlig (2005) to approximate the value of the endpoints. Given D draws of the reduced-form parameters (A,Σ) and a unit vector $q \in \mathbb{R}^n$, one could report the maximum and minimum value for $\{\lambda^{d}_{k,i,j}(A,\Sigma,q)\}_{d=1}^{D}$ over the different draws. This algorithm is a random grid search approach to solve the programs (5.5) and (5.6). Figure 5.6 in the appendix presents a comparison between the different approaches. The *grid search* takes around 300 seconds to run and underestimates the identified set.

[13] The formula for the maximum (minimum) k-th period ahead *cumulative* IRF replaces $C_k(\widehat{A}_T)$ by $C_0(\widehat{A}_T) + C_1(\widehat{A}_T) + \ldots, + C_k(\widehat{A}_T)$.

Figure 5.3. Identified set for the cumulative impulse response functions to a one standard deviation UMP shock (given $\widehat{\mu}_T$)



(Solid, Blue Line) Endpoints of the identified set for the cumulative responses given $\widehat{\mu}_T$ and the equality/inequality restrictions in Table I. (Blue, Crosses) Endpoints of the identified set with the additional restriction $e_2'(C_0 + C_1(A))B_1 \geq 0$.

5.5 Directional differentiability of the endpoints

Once again, let

$$R(\mu) \equiv \bigcup_{l=0}^{\min\{n-1-m_z, m_s\}} R_l(\mu),$$

denote the set of all possible combinations up to $n-1$ active constraints.[14] Let us denote the typical element in $R(\mu)$ as $r(\mu)$, which we take to be an $n \times l$ matrix with $l \leq n-1$. We will continue working with the auxiliary function:

$$v_{k,i,j}(\mu; r(\mu)) = \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r(\mu)} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2},$$

where we now explicitly acknowledge the possible dependence of r on μ. In Lemma 1 we have shown that if $r(\mu)$ is the active set of constraints at a solution of the program (5.5), then:

$$\overline{v}_{k,i,j}(\mu) = v_{k,i,j}(\mu; r(\mu)) \quad \text{or} \quad \overline{v}_{k,i,j}(\mu) = -v_{k,i,j}(\mu; r(\mu));$$

as long as $\overline{v}_{k,i,j}(\mu) \neq 0$. In order to establish the differentiability of $\overline{v}_{k,i,j}(\mu)$ we prove the following intermediate result.

Lemma 2 (**Differentiability results for a given active set of constraints**). *If $r(\mu)$ is differentiable with respect to μ and $v_{k,i,j}(\mu; r(\mu)) \neq 0$, then $v_{k,i,j}(\mu; r(\mu))$ is differentiable with respect to μ with derivative $\dot{v}_{k,i,j}(\mu; r(\mu))$ given by:*

$$\begin{bmatrix} \frac{\partial v_{k,i,j}(\mu; r(\mu))}{\partial vec(A)} \\ \\ \frac{\partial v_{k,i,j}(\mu; r(\mu))}{\partial vec(\Sigma)} \end{bmatrix} = \begin{bmatrix} \frac{\partial vec(C_k(A))}{\partial vec(A)} (x^*(\mu; r(\mu)) \otimes e_i) - \sum_{k=1}^{l} w_k^* \frac{\partial vec(r_k(\mu))}{vec(A)} x^*(\mu, r(\mu)) \\ \\ \lambda^* (\Sigma^{-1} x^*(\mu, r(\mu)) \otimes \Sigma^{-1} x^*(\mu, r(\mu))) - \sum_{k=1}^{l} w_k^* \frac{\partial vec(r_k(\mu))}{vec(\Sigma)} x^*(\mu, r(\mu)) \end{bmatrix},$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$,

$$x^*(A, \Sigma; r(\mu)) = \Sigma^{1/2} \left(M_{\Sigma^{1/2} r(\mu)}\right) \Sigma^{1/2} C_k(A)' e_i \Big/ v_{k,i,j}(A, \Sigma; r(\mu)),$$

$$\lambda^* \equiv \frac{1}{2} v_{k,i,j}(A, \Sigma; r(\mu)), \quad w^* \equiv [r(\mu)' \Sigma r(\mu)]^{-1} r(\mu)' \Sigma C_k(A) e_i,$$

[14] The dependence of the set R on the parameter μ was omitted in the previous section for notational simplicity.

and w_k^* is the k-th component of the vector w^*.

Proof: See Appendix 5.A.3.

The *envelope theorem* sheds light on the derivative formula in Lemma 2. Note first that

$$v_{k,i,j}(\mu, r(\mu)) = \max_{x \in \mathbb{R}^n} e_i' C_k(A) x \quad \text{s.t.} \quad x' \Sigma^{-1} x = 1 \text{ and } r'(\mu) x = \mathbf{0}_{l \times 1}.$$

The auxiliary *Lagrangian function* of this problem is given by:

$$\mathscr{L}(x; \mu, r(\mu)) = (x' \otimes e_i') \text{vec}(C_k(A)) - \lambda \Big((x' \otimes x') \text{vec}(\Sigma^{-1}) - 1 \Big) - w'(r(\mu)' x),$$

where λ is the Lagrange multiplier corresponding to the quadratic equality restriction and $w \in \mathbb{R}^l$ is the vector of Lagrange multipliers corresponding to the l equality restrictions. The envelope theorem suggests that $\dot{v}_{k,i,j}(\mu; r(\mu))$ is given by the formula in Lemma 2. We confirm this intutition in the prove of Lemma 2; provided $v_{k,i,j}(\mu; r(\mu)) \neq 0$.

We now establish the differentiability of $\overline{v}_{k,i,j}(\mu)$. Without loss of generality, assume that $\overline{v}_{k,i,j}(\mu) > 0$ (and also that $\underline{v}_{k,i,j}(\mu) < 0$). For a fixed vector of reduced-form parameters define the sets:

$$R^*(\mu) \equiv \Big\{ r(\mu) \in R(\mu) \mid \overline{v}_{k,i,j}(\mu) = v_{k,i,j}(\mu, r(\mu)) \Big\},$$

and

$$R_*(\mu) \equiv \Big\{ r(\mu) \in R(\mu) \mid \underline{v}_{k,i,j}(\mu) = -v_{k,i,j}(\mu, r(\mu)) \Big\}.$$

The set $R^*(\mu)$ collects the different active constraints that could lead to the maximum value. The set $R^*(\mu)$ is a singleton if and only if the program (5.5) has a unique solution. The set $R_*(\mu)$ is defined analogously.

Proposition 2 (**Directional differentiability of the endpoints of the identified set**). *Suppose, w.l.o.g., that $\overline{v}_{k,i,j}(\mu) > 0$ and $\underline{v}_{k,i,j}(\mu) < 0$. Then, for any sequence $h_n \in \mathbb{R}^d$ such that $h_n \to h$ and any sequence $t_n \to \infty$:*

$$t_n \Big(\overline{v}_{k,i,j}(\mu + h_n / t_n) - \overline{v}_{k,i,j}(\mu) \Big) \to \max_{r(\mu) \in R^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu))' h \Big],$$

and

$$t_n \Big(\underline{v}_{k,i,j}(\mu + h_n / t_n) - \underline{v}_{k,i,j}(\mu) \Big) \to \min_{r(\mu) \in R_*(\mu)} \Big[-\dot{v}_{k,i,j}(\mu; r(\mu))' h \Big].$$

Thus, $\overline{v}_{k,i,j}$ and $\underline{v}_{k,i,j}$ are directionally differentiable functions of the reduced-form parameters with directional derivative:

$$\max_{r(\mu)\in R^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu))' h \Big],$$

and

$$\min_{r(\mu)\in R_*(\mu)} \Big[-\dot{v}_{k,i,j}(\mu; r(\mu))' h \Big],$$

respectively.

Whenever $R^*(\mu)$ is a singleton—i.e., $R^*(\mu) = \{r(\mu)\}$—the value function $\overline{v}_{k,i,j}(\mu)$ is fully differentiable with derivative $\dot{v}_{k,i,j}(\mu, r(\mu))$. Likewise if $R_*(\mu)$ is a singleton, the value function $\underline{v}_{k,i,j}(\mu)$ is differentiable with derivative $-\dot{v}_{k,i,j}(\mu, r(\mu))$.

Proof: See Appendix 5.A.4.

Theorem 4.2, p. 223 in Fiacco and Ishizuka (1990) and Theorem 4.24, p. 280 in the book of Bonnans and Shapiro (2000) present a generalized version of the envelope theorem. They show that—under suitable constraint qualifications—the directional derivative (in direction h and evaluated at parameter μ) of the largest and smallest value in a mathematical program with equality and inequality constraints is given by:

$$\sup_{r\in R^*(\mu)} \Big[\nabla_\mu \mathcal{L}(x^*(\mu; r); \mu) h \Big],$$

and

$$\inf_{r\in R_*(\mu)} \Big[\nabla_\mu \mathcal{L}(x^*(\mu; r); \mu) h \Big],$$

provided there is a unique set of Lagrange Multipliers supporting the optimal solutions $x^*(\mu; r)$. Proposition 2 uses the results in Lemma 1 and Lemma 2 to verify this formula.

Delta-Method vs. Bootstrap: We also note that directionally differentiable functions have been a topic of recent research. Fang and Santos (2015) show that the standard bootstrap is not consistent when applied to parameters of the form $v(\mu)$, where v is a directionally differentiable function. Kitagawa et al. (2016) show that Bayesian credible sets based on the quantiles of the posterior distribution of $v(\mu)$ will be asymptotically equivalent to the frequentist bootstrap (which is not consistent in this case).

These results imply that typical frequentist and Bayesian inference for directionally differentiable functions is problematic. The next section shows that the special form of the di-

rectional derivative in the class of SVARs models studied in this paper allows the researcher to conduct delta-method inference, with a slight adjustment on the standard errors.

5.6 Delta-method inference

This section proposes a delta-method confidence interval of the form

$$\text{CS}_T(1-\alpha) \equiv \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big],$$

where

$$\widehat{\mu}_T \equiv (\text{vec}(\widehat{A}_T)', \text{vec}(\widehat{\Sigma}_T)'),$$

is the OLS estimator for μ defined as:

$$\widehat{A}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^{T} Y_t X_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\Big)^{-1}, \quad \widehat{\Sigma}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\widehat{\eta}_t\widehat{\eta}_t',$$

with

$$X_t \equiv (Y_{t-1}', \ldots, Y_{t-p}')', \quad \widehat{\eta}_t \equiv Y_t - \widehat{A}_T X_t.$$

We work under the assumption that $\sqrt{T}(\widehat{\mu}_T - \mu)$ is asymptotically normal with some covariance matrix Ω. A common formula to estimate the asymptotic variance of $\widehat{\mu}_T$ is:

$$\widehat{\Omega}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^{T}\text{vec}\Big([\widehat{\eta}_t X_t', \widehat{\eta}_t\widehat{\eta}_t' - \widehat{\Sigma}_T]\Big)\text{vec}\Big([\widehat{\eta}_t X_t', \widehat{\eta}_t\widehat{\eta}_t' - \widehat{\Sigma}_T]\Big)'.$$

We use the results in Proposition 2 and the asymptotically normality of $\widehat{\mu}_T$ to suggest the following formula for $\widehat{\sigma}_{(k,i,j),T}$:

$$\widehat{\sigma}_{(k,i,j),T} \equiv \max_{r\in R(\widehat{\mu}_T)}\Big[\dot{v}_{k,i,j}(\widehat{\mu}_T; r)'\widehat{\Omega}_T\dot{v}_{k,i,j}(\widehat{\mu}_T; r)\Big], \tag{5.10}$$

where $R(\widehat{\mu}_T)$ is the set of all possible active constraints evaluated at $\widehat{\mu}_T$.

Main Result in this section: Let P denote the data generating process and let $\mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))$ denote the identified set for the structural parameter $\lambda_{k,i,j}$ given the equality/inequality restrictions in $\mathscr{R}(\mu)$. This section shows that under our proposed specification of $\widehat{\sigma}_{(k,i,j),T}$:

$$\liminf_{T\to\infty} \inf_{\lambda\in\mathscr{I}^{R}(\mu(P))} P\Big(\lambda \in \mathrm{CS}_T(1-\alpha)\Big) \geq 1-\alpha.$$

Consequently, the delta-method confidence interval presented in this paper is *pointwise consistent in level*. We now describe the main large-sample assumptions concerning P.

Data Generating Process: The SVAR parameters $(A_1,\ldots,A_p,B,F)$ define a probability measure, denoted P, over the data observed by the econometrician. Our main assumption concerning P is as follows:

Assumption 4 (**Asymptotic Normality of $\widehat{\mu}_T$**)**.** The data generating process P is such that for $\mu(P)\in\mathbb{R}^d$:

$$\sqrt{T}(\widehat{\mu}_T - \mu(P)) \xrightarrow{d} \zeta(P) \sim \mathscr{N}_d\Big(\mathbf{0}\,,\,\Omega(P)\Big),$$

and

$$\widehat{\Omega}_T \xrightarrow{p} \Omega(P).$$

Thus, our only restriction on $(A_1,A_2,\ldots,A_p,B,F)$ is that, whatever these parameters are, the OLS estimator $\widehat{\mu}_T$ is asymptotically normal with a covariance matrix that can be estimated consistently.

Delta-method for Directionally Differentiable functions: Dümbgen (1993), Shapiro (1991), and Fang and Santos (2015) have shown if v is a directionally differentiable function with directional derivative $\dot{v}_\mu(h)$ (in direction h evaluated at μ) then:

$$\sqrt{T}(v(\widehat{\mu}) - v(\mu)) \xrightarrow{d} \dot{v}_\mu(\zeta),$$

whenever Assumption 4 holds.[15] Proposition 2 in the previous section established that the directional derivative of $\overline{v}_{k,i,j}$—in direction h evaluated at μ—is given by:

$$\max_{r \in R^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r))' h \Big].$$

where $R^*(\mu)$ collects the active constraints that generate $\overline{v}_{k,i,j}(\mu)$. Thus, Proposition 2 and Assumption 4 imply that:

$$\sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}) - \overline{v}_{k,i,j}(\mu)) \xrightarrow{d} \max_{r \in R^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu, r)' \zeta \Big],$$

where

$$\dot{v}_{k,i,j}(\mu, r(\mu))' \zeta \sim \mathcal{N}_1\Big(0, \dot{v}_{k,i,j}(\mu, r(\mu))' \Omega \dot{v}_{k,i,j}(\mu, r(\mu))\Big).$$

Inference for Directionally Differentiable Functions: How can we use the delta-method result above to construct a confidence set for $\lambda_{k,i,j}$? Our suggestion—which exploits the specific form of the directional derivative in the SVAR context—is to consider:

$$\widehat{\sigma}_{(k,i,j),T} \equiv \max_{r \in R(\widehat{\mu}_T)} \Big[\dot{v}_{k,i,j}(\widehat{\mu}_T; r)' \widehat{\Omega}_T \dot{v}_{k,i,j}(\widehat{\mu}_T; r) \Big],$$

where $R(\widehat{\mu}_T)$ is the set of all the different collections of active constraints evaluated at $\widehat{\mu}_T$. The resulting standard error will then be used to enlarge the plug-in estimator of the endpoints of the identified set. The suggested confidence interval is shown to be pointwise consistent in level.[16] This is formalized in the following proposition.

Proposition 3 (**Pointwise Consistency in Level of the Delta-Method Confidence Interval**)**.** *Let $\widehat{\sigma}_{(k,i,j),T}$ be defined as in (5.10). Suppose there are at most $n-1$ equality restrictions and*

[15] The map $v : \mathbb{R}^r \to \mathbb{R}$ is said to be Hadamard directionally differentiable at $\mu \in \mu \subseteq \mathbb{R}^r$, tangentially to $\mathbb{R}^r$, if there is a continuous (not necessarily linear) map $\dot{v}(\cdot, \mu) : \mathbb{R}^r \to \mathbb{R}$ such that:

$$\lim_{T \to \infty} \left| \frac{v(\mu + t_T h_T) - v(\mu)}{t_T} - \dot{v}(h; \mu) \right| = 0,$$

for all sequences $\{h_T\} \subseteq \mathbb{R}^r$ and $\{t_T\} \subseteq \mathbb{R}_+$ such that $t_T \to 0^+$, $h_T \to h \in \mathbb{R}^r$ and $\mu + t_T h_T \in \mu$ for all T. The function $v(\cdot)$ is Fully Differentiable at μ if and only if the mapping $\dot{v}(\cdot; \mu)$ is linear. See Fang and Santos (2015) for a recent elegant exposition on directionally differentiable functions. See also Shapiro (1990).

[16] The question of how to build a *uniformly consistent in level*, delta-method confidence set for a set-identified parameter is beyond the scope of this paper. For the readers interested in uniform inference for set-identified parameters in SVARs our suggestion is to apply the projection approach developed in Gafarov et al. (2016). Compared to GMM16, the delta-method approach described in this paper is faster to implement.

that data generating process P is such that $Z(\mu(P))$ and $S(\mu(P))$ are linearly independent at $\mu(P)$ as defined in Section 4, and also differentiable. Suppose in addition that:

$$\min_{r\in R(\mu)}\Big[\dot{v}_{k,i,j}(\mu(P),r)'\Omega(P)\dot{v}_{k,i,j}(\mu(P),r)\Big]>0.$$

Then, under Assumption 4:

$$\liminf_{T\to\infty}\inf_{\lambda\in\mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))}P\Big(\lambda\in\Big[\underline{v}_{k,i,j}(\widehat{\mu}_T)-z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T)+z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big]\Big)\ \geq\ 1-\alpha$$

Proof: See Appendix 5.A.5.

The intuition behind our proof is as follows. Note first that if λ belongs to the identified set, (i.e., $\lambda\in\mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))$, then such parameter must lie between the maximum and the minimum response; that is $\lambda\in[\underline{v}_{k,i,j}(\mu(P)),\overline{v}_{k,i,j}(\mu(P))]$. Consequently, one can show that:

$$P\Big(\lambda\in\Big[\underline{v}_{k,i,j}(\widehat{\mu}_T)-z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \ \overline{v}_{k,i,j}(\widehat{\mu}_T)+z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big]\Big),$$

is larger than or equal to

$$P\Big(\underline{v}_{k,i,j}(\mu(P)),\overline{v}_{k,i,j}(\mu(P))\ \in\ \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T)-z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T)+z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big]\Big).$$

Thus, a sufficient condition for the validity of the delta-method confidence interval is that it covers the identified set with probability at least $1-\alpha$. Note that the probability of covering the identified set can be written as one minus the sum of the following two terms:

$$P\Big(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T)-\underline{v}_{k,i,j}(\mu(P)))>z_{1-\alpha/2}\widehat{\sigma}_{(k,i,j),T}\Big),$$

and

$$P\Big(\sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T)-\overline{v}_{k,i,j}(\mu(P))<-z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}>)\Big).$$

Using our large sample assumptions and the delta-method for directional differentiable functions, these probabilities are approximately equal to:

$$P\Big(\min_{r\in R_*(\mu(P))}[-\dot{v}_{k,i,j}(\mu(P),r)'Z_d]>z_{1-\alpha/2}\sigma_{(k,i,j)}\Big),$$

and

$$P\Big(\max_{r\in R^*(\mu(P))}[\dot{v}_{k,i,j}(\mu(P),r)'Z_d] < -z_{1-\alpha/2}\sigma_{(k,i,j)}\Big),$$

where:

$$\sigma_{k,i,j} \equiv \max_{r\in R(\mu)}\Big[\dot{v}_{k,i,j}(\mu(P),r)'\Omega(P)\dot{v}_{k,i,j}(\mu(P),r)\Big], \quad Z_d \sim \mathcal{N}(0,\Omega).$$

Take any $r_* \in R_*(\mu(P))$ for which $\sigma_{(k,i,j)}(r_*) \equiv \dot{v}_{k,i,j}(\mu(P),r_*)'\Omega\dot{v}_{k,i,j}(\mu(P),r_*) > 0$. (we have assumed that such r_* exists). It follows that:

$$P\Big(\min_{r\in R_*(\mu(P))}[-\dot{v}_{k,i,j}(\mu(P),r)'Z] > z_{1-\frac{\alpha}{2}}\sigma_{(k,i,j)}\Big) \le P\Big(\sigma_{k,i,j}(r_*)N(0,1) > z_{1-\frac{\alpha}{2}}\sigma_{(k,i,j)}\Big),$$

and the last term is bounded above by $\alpha/2$. Analogously, we can select an $r^* \in R^*(\mu(P))$ for which $\sigma_{(k,i,j)}(r^*) \equiv \dot{v}_{k,i,j}(\mu(P),r^*)'\Omega\dot{v}_{k,i,j}(\mu(P),r^*) > 0$ and we can show that:

$$P\Big(\max_{r\in R^*(\mu(P))}[\dot{v}_{k,i,j}(\mu(P),r)'Z] < -z_{1-\frac{\alpha}{2}}\sigma_{(k,i,j)}\Big) \le P\Big(\sigma_{k,i,j}(r^*)N(0,1) < -z_{1-\frac{\alpha}{2}}\sigma_{(k,i,j)}\Big),$$

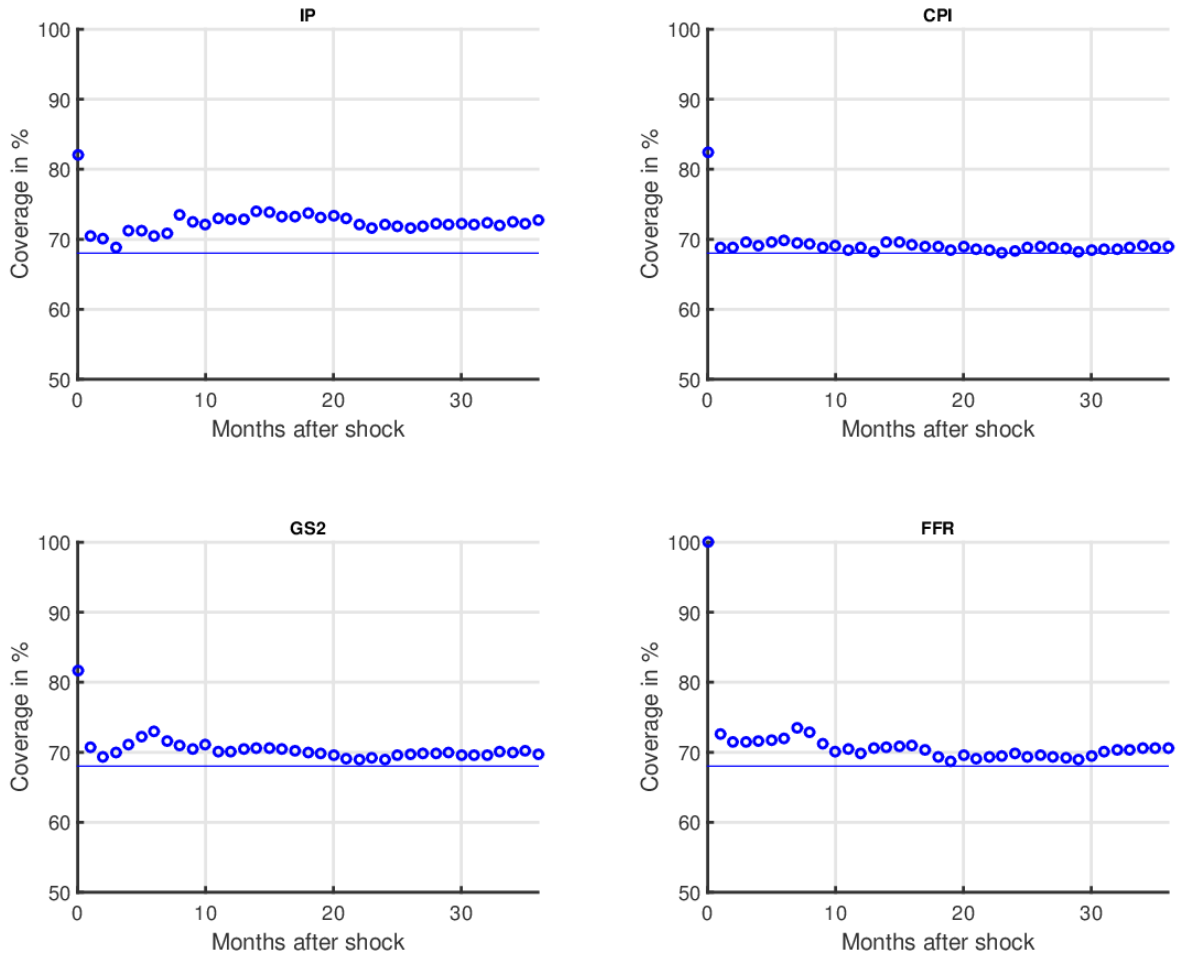
which is also bounded above by $\alpha/2$. These inequalities suffice to establish the pointwise validity of our delta-method approach.

Monte-Carlo Evidence: We conduct a simple Monte-Carlo exercise to study the coverage probability of our delta-method confidence set. We set $(1-\alpha) = .68$, which implies that $z_{1-\alpha/2} = .9945$. We generate $10,000$ draws from the multivariate normal model $N_d(\widehat{\mu}_T, \widehat{\Omega}_T/T)$ and for each draw (denoted μ^*) we compute the confidence interval:

$$\Big[\underline{v}_{k,i,j}(\mu^*) - .9945\,\sigma^*_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\mu^*) + .9945\,\sigma^*_{(k,i,j),T}/\sqrt{T}\Big].$$

We check whether $[\underline{v}_{k,i,j}(\widehat{\mu}_T), \overline{v}_{k,i,j}(\widehat{\mu}_T)]$ is contained in the confidence interval or not. The estimated probability provides a lower bound on the coverage of the identified parameter. The results are reported in Figure 5.4.

Figure 5.4. Monte-Carlo coverage probability based on the model $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T / T)$, $T = 342$.



(Blue, Circles) Monte-Carlo estimate of the probability $P\Big([\underline{v}_{k,i,j}(\widehat{\mu}_T), \overline{v}_{k,i,j}(\widehat{\mu}_T)] \in [\underline{v}_{k,i,j}(\mu^*) - .9945\,\sigma^*_{(k,i,j),T} / \sqrt{T},\ \overline{v}_{k,i,j}(\mu^*) + .9945\,\sigma^*_{(k,i,j),T} / \sqrt{T}]\Big)$ for the model $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T)$, with $T = 342$. The values $\widehat{\mu}_T$ and $\widehat{\Omega}_T$ correspond, respectively, to the estimators of the reduced-form parameter and its asymptotic covariance matrix in the UMP application. (Blue, Solid Line) Nominal confidence level for the delta-method confidence interval (68%).

5.7 Unconventional monetary policy shocks

As we mentioned before, the identification strategy in this paper was motivated by two mechanisms used by the Federal Reserve to affect market beliefs during the Great Recession: forward guidance announcements and the large-scale asset purchase program. We will focus on one particular episode of the Great Recession illustrating the role of forward guidance. In August 2010 the Federal Open Market Committee announced: "*The Committee will keep constant the Federal Reserve's holdings of securities at their current level by reinvesting principal payments from agency debt and agency mortgage-backed securities in longer-term Treasury securities.*" This announcement was an important prelude for the second part of the Quantitative Easing program (QE2) (see p. 244 in Krishnamurthy and Vissing-Jorgensen (2011) for a detailed discussion). In addition, this announcement generated a drop in the intraday yield for two- and ten- year treasury bond. In fact, from the end of July 2010 to the end of August 2010 the 2 year Treasury bond rate fell by 10 basis points.

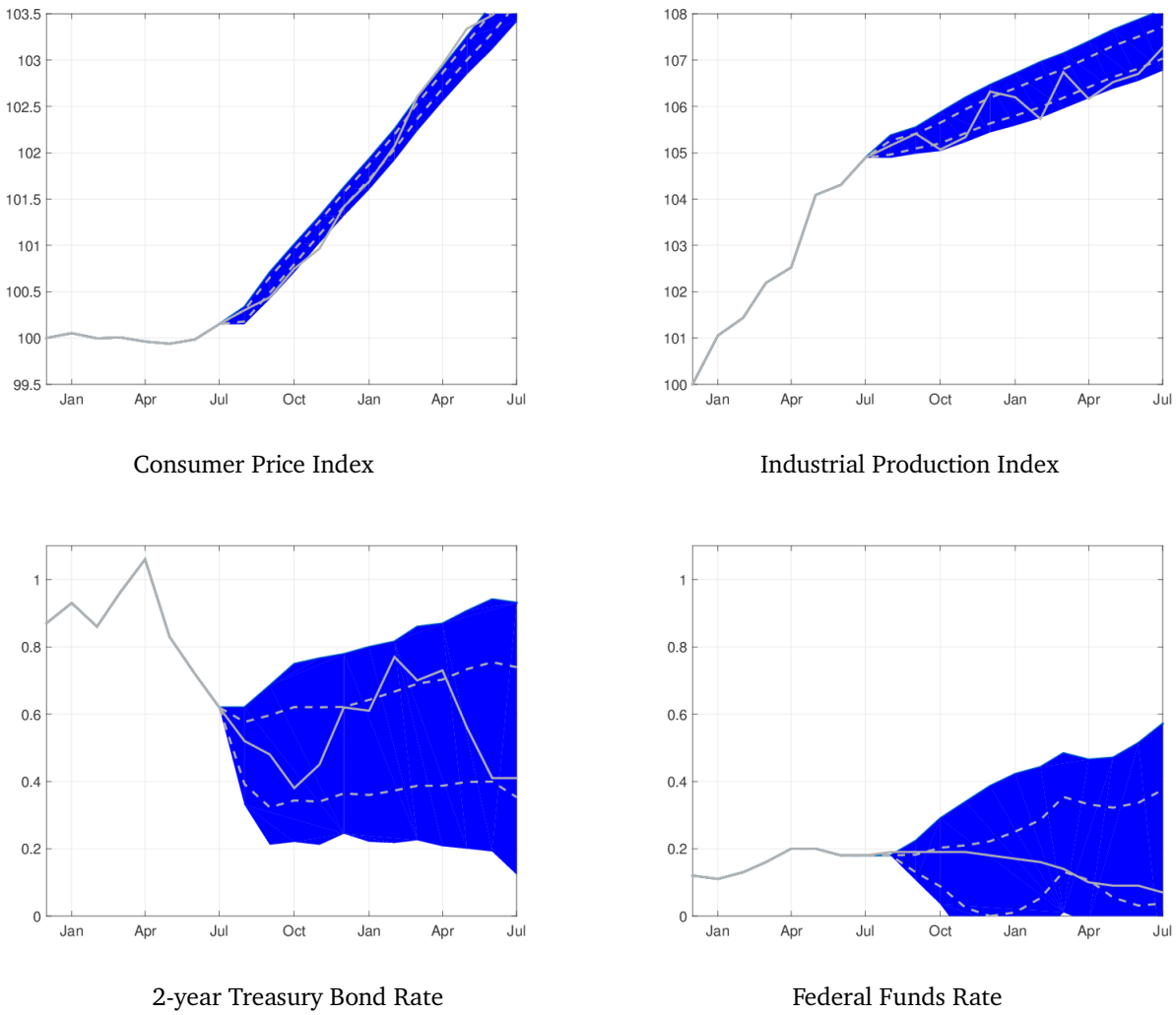
Figure 5.5 uses our delta-method approach to construct confidence bands for the evolution of the levels of the four variables in the monetary SVAR. We fix all the variables at their level on July 2010 and we trace their evolution (over a 12-month window) according to the confidence set for their cumulative responses. The motivation for this exercise is as follows. Suppose that—back in August 2010—an econometrician is asked to provide confidence bands for the evolution of IP, CPI, 2YTB, and FF after the August 2010 announcement of the Federal Open Market Committee (FOMC). The econometrician observes the realization of the macroeconomic variables from July 1979 until August 2010, but decides to deliberately ignore the two years of data after the crisis (to avoid introducing structural changes, stochastic volatility, or any other feature that will complicate the estimation of the VAR).

The econometrician uses the data until December 2007—one semester before the financial crisis—to conduct delta-method inference on the cumulative responses to a one standard deviation unconventional monetary policy shock. The econometrician then uses these cumulative responses to get a rough idea of the evolution of the variables (in levels) following the announcement of the Federal Reserve in August 2010. The econometrician assumes there is a linear trend for CPI/IP, and ignores sampling uncertainty coming from the trend estimation in reporting the bands.

An ex-post evaluation of this exercise (over a window of 12 months) is reported in Figure 5.5.[17] We note that the observed dynamics for CPI, IP, GS2, and FFR from August 2010 to July 2011 fall within the bounds motivated by our delta-method confidence interval. We also note that our delta-method confidence interval misses the observed value at most three out of 12 months, which means that our 68% confidence-set covers each of these variables at least 75% of the time. We also report the 68% Bayesian credible sets.

[17] The reason to focus in a 12-month window is to cover the period between the QE2 announcement and the announcement of the so-called "Operation Twist" in September 2011. See http://www.federalreserve.gov/newsevents/press/monetary/20110921a.htm.

Figure 5.5. Delta-method confidence interval for CPI, IP, 2yTB, FF after the August 2010 announcement



Consumer Price Index

Industrial Production Index

2-year Treasury Bond Rate

Federal Funds Rate

(Shaded, Blue Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Gray, Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Gray, Dashed Line) Evolution of the Levels CPI, IP, 2yTB, and FF based on the 68% credible set constructed using the priors in Uhlig (2005).

Computational Cost: We close this section with some comments regarding the computational cost of our delta-method procedure. Most of the work to compute the endpoints of the identified set and its derivatives is analytical. Consequently, practitioners can expect the computational burden of our procedure to be low. We note that the implementation of our delta-method confidence interval in the running example takes only around .15 seconds (using a standard Laptop @2.4GHz IntelCore i7). With the same equipment, the standard Bayesian implementation required around 327 seconds for 10,000 draws (which means that we could have constructed 2,000 delta-method confidence intervals while we generated the Bayesian credible set).

Comparison with the Projection Approach: Figure 5.7 presents a comparison between the delta-method approach and the *projection* confidence interval recently proposed by Gafarov et al. (2016) [GMM16]. The projection confidence interval has three theoretical properties that we were not able to verify for the delta-method approach. First, the projection confidence interval is *uniformly* consistent in level. Second, the projection confidence interval yields valid inference for the whole impulse-response function and not only its scalar coefficients. Third, the projection confidence interval has—in large samples—Bayesian credibility of at least the nominal level (for a large class of priors).

In order to exploit our formulas for the endpoints of the identified set, we followed a different algorithm to the one suggested in GMM16. We used a random grid over the Wald ellipsoid for the reduced-form parameters and reported the range of IRFs over this grid. The implementation of the projection confidence set (based on a random grid of 10,000 points) took around 1300 seconds (4 times slower than the Bayesian credible set and 8,000 times slower than the delta-method). We note that the projection confidence interval (which is wider than the delta-method bands) contains the realized value of IP, CPI, 2YTB, and FF for every horizon under consideration.

Comparison with Calibrated Projection: The projection confidence interval covers the structural parameters more often than necessary. GMM16 show that when the endpoints of the identified set are differentiable, one can project a Wald ellipsoid with a radius given by $z^2_{1-\alpha}$ to eliminate projection bias. The calibrated confidence set will be approximately given by:

$$\left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha}\,\widehat{\underline{\sigma}}_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha}\,\widehat{\overline{\sigma}}_{(k,i,j),T}/\sqrt{T}\right],$$

where $\widehat{\underline{\sigma}}_{(k,i,j),T}, \widehat{\overline{\sigma}}_{(k,i,j),T}$ are estimators of the asymptotic variance of the plug-in estimators for the endpoints of the identified set. Figure 5.8 reports the calibrated projection.

Under the differentiability assumption, both $\widehat{\underline{\sigma}}_{(k,i,j),T}, \widehat{\overline{\sigma}}_{(k,i,j),T}$ are smaller (for every data realization) than $\widehat{\sigma}_{(k,i,j),T}$. This suggests that—whenever the endpoints of the identified set are fully differentiable—the calibrated projection should deliver smaller confidence intervals than our delta-method approach, as our formula for the standard error takes into account the possibility that the endpoints are only directionally differentiable. In terms of computation time, the calibrated projection takes approximately 109,365 seconds (solving the nonlinear program described in GMM16).

Comparison with the Robust Approach: Finally, Figure 5.6 in the Appendix reports the robust-Bayesian credible set in Giacomini and Kitagawa (2015). The implementation of the robust-Bayes credible set (based on 10,000 posterior draws and using our algorithm to evaluate the endpoints) took around 9,106 seconds.[18]

5.8 Conclusion

This paper focused on set-identified SVARs that impose equality and inequality restrictions to set-identify only one structural shock. For this class of models, the endpoints of the identified set have special properties that allow an intuitive and computationally simple approach to conduct frequentist inference. Specifically, the paper made three contributions:

(i) We presented an algorithm to compute—for each horizon, each variable, a fixed vector of reduced-form parameters, and a given collection of equality and/or inequality restrictions—the largest and smallest value of the coefficients of the structural IRF (see Proposition 1). Our algorithm does not require random sampling from the space of rotation matrices or unit vectors. Instead, we treated the bounds of the identified set as the *maximum and minimum value* of a mathematical program whose solutions we can characterize analytically.

(ii) We provided sufficient conditions under which the largest and smallest value of the structural parameters are directionally differentiable functions of the reduced-form parameters (see Proposition 2). This result seems to be of interest in its own right, and for example, could be used to explore the frequentist properties of the robust-Bayesian procedure in Giacomini and Kitagawa (2015).

(iii) We proposed a computationally convenient delta-method confidence interval for the set-identified coefficients of the structural IRF. We presented sufficient conditions to guaran-

[18] Out of which 1,266 seconds were used just to compute the identified set for each posterior draw, and the remaining time to translate the posterior bounds into the GK robust bounds

tee the pointwise consistency in level of the suggested inference approach. The delta-method in this paper exploited the structure of the directional derivative.

Appendix 5.A Appendix

5.A.1 Lemma 1

Let $S(\mu)$ denote the $n \times m_s$ matrix of m_s 'sign' restrictions and let $Z(\mu)$ denote the $m_z \times n$ matrix of 'zero' restrictions. For notational simplicity, we deliberately ignore the dependence of the equality/inequality restrictions on μ. The problem in equation (5.5) is equivalent to:

$$\begin{aligned} \overline{v}_{k,i,j}(A,\Sigma) \equiv & \max_{x\in\mathbb{R}^n} e_i' C_k(A) x \\ \text{s.t. } & x'\Sigma^{-1}x = 1, \quad S'x = \mathbf{0}_{m_s\times 1}, \quad Z'x = \mathbf{0}_{m_z\times 1}. \end{aligned} \tag{5.11}$$

The auxiliary Lagrangian function is given by:

$$\mathscr{L}(x, w_1, w_2, w_3; A, \Sigma) = e_i' C_k x - w_1(x'\Sigma^{-1}x - 1) - w_2'(S'x) - w_3'(Z'x).$$

Since $Z(\mu)$ and $S(\mu)$ are linearly independent at μ we can characterize the maximum response using the Karush-Kuhn-Tucker conditions for the mathematical program in (5.5). The Karush-Kuhn-Tucker necessary conditions for this problem are as follows:

$$\begin{aligned} \text{Stationarity} \quad &: \quad C_k'(A)e_i - 2w_1\Sigma^{-1}x - Sw_2 - Zw_3 = \mathbf{0}_{n\times 1}, \\ \text{Primal Feasibility} \quad &: \quad x'\Sigma^{-1}x = 1, \\ &\quad\quad S'x \geq 0_{m_s\times 1,} \\ &\quad\quad Z'x = \mathbf{0}_{m_z\times 1,} \\ \text{Complementary Slackness} \quad &: \quad \omega_{2i}(e_i'Sx) = 0 \quad \forall \quad i = 1\ldots m_s, \end{aligned}$$

plus the additional dual feasibility constraint requiring the Lagrange multipliers, ω_{2i}, to be smaller than or equal to zero.

Let $x^*(A,\Sigma,r)$ be one (out of possibly many) maximizer of the program of interest and suppose that the $m \times n$ matrix ($m \leq n-1$) r collects all the restrictions that are active (binding). Because of Assumption 1 and the fact that Z and S are linearly independent at μ, the matrix r is of full rank m and m must be smaller than or equal $n-1$. Using stationarity, primal

feasibility, and complementary slackness at x^* we get:

$$\begin{aligned}
0 &= x^{*\prime}[C_k(A)'e_i - 2w_1^*\Sigma^{-1}x^* - Sw_2 - Zw_3] \\
&= x^{*\prime}C_k(A)'e_i - 2w_1^*x^{*\prime}\Sigma^{-1}x^* - x^{*\prime}Sw_2 - x^{*\prime}Zw_3 \\
&= x^{*\prime}C_k(A)'e_i - 2w_1^* - x^{*\prime}Sw_2 - x^{*\prime}Zw_3 \\
&\quad \text{(where we have used } x^{*\prime}\Sigma^{-1}x^* = 1\text{)} \\
&= x^{*\prime}C_k(A)'e_i - 2w_1^* \\
&\quad \text{(where we have used } x^{*\prime}Z = \mathbf{0}_{m_z\times 1} \text{ and complementary slackness)} \\
&= \overline{v}_{k,i,j}(A,\Sigma) - 2w_1^*.
\end{aligned}$$

where $\overline{v}_{k,i,j}(A,\Sigma)$ denotes the value of the maximum response when the constraints in r are active. Thus, the Lagrange multiplier w_1^* is unique and given by:

$$w_1^* = \frac{1}{2}\overline{v}_{k,i,j}(A,\Sigma).$$

Note also that $w_1^* \neq 0$ if and only if $\overline{v}_{k,ij}(A,\Sigma) \neq 0$. We now show that if $\overline{v}_{k,i,j}(A,\Sigma) \neq 0$ there are unique w_2^* and w_3^* that satisfy the Karush-Kuhn Tucker conditions. Note that left multiplying the stationarity condition by Σ we have:

$$2w_1^*x^{*\prime} = \Big(C_k(A)'e_i - rw\Big)'\Sigma.$$

where w collects the nonzero components of ω_2 and all the components of ω_3.

$$\begin{aligned}
\Big(C_k(A)'e_i - rw\Big)'\Sigma\Big(C_k(A)'e_i - rw\Big) &= 4w_1^2x'\Sigma^{-1}x \qquad (5.12)\\
&= 4w_1^2 \\
&\quad \text{(where we have used } x^{*\prime}\Sigma^{-1}x^* = 1\text{)} \\
&= 4\Big(\frac{1}{2}\overline{v}_{k,i,j}(A,\Sigma)\Big)^2 \\
&= \overline{v}_{k,i,j}(A,\Sigma)^2.
\end{aligned}$$

Consequently the value function given active constraints z is given by either:

$$\overline{v}_{k,i,j}(A,\Sigma) = \Big[\Big(C_k(A)'e_i - rw\Big)'\Sigma\Big(C_k(A)'e_i - rw\Big)\Big]^{1/2},$$

or

$$\overline{v}_{k,i,j}(A,\Sigma) = -\Big[\Big(C_k(A)'e_i - rw\Big)'\Sigma\Big(C_k(A)'e_i - rw\Big)\Big]^{1/2}.$$

We will use the first order conditions to find the vector of Lagrange multipliers w and show that they are unique given $\overline{v}_{k,i,j}(A,\Sigma) \neq 0$. Note that

$$\begin{aligned} 0 \;=\; 2w_1^* r' x^* \;&=\; \overline{v}_{k,i,j}(A,\Sigma)\Big[r'\Sigma(C_k(A)'e_i - rw)\Big] \\ &=\; \overline{v}_{k,i,j}(A,\Sigma)\Big[r'\Sigma C_k(A)'e_i - r'\Sigma r w\Big]. \end{aligned}$$

Under the assumptions of the Lemma, r is of rank m. If $\overline{v}_{k,i,j}(A,\Sigma) \neq 0$, the equation above holds if and only if

$$w^* = (r'\Sigma r)^{-1} r'\Sigma C_k(A)'e_i.$$

Consequently, the Lagrange multipliers for the active restrictions are unique. To conclude the proof, we get an explicit expression of the value function in terms of (A,Σ). To do so, note that:

$$\begin{aligned} \Sigma^{1/2}\Big(C_k(A)'e_i - rw\Big) \;&=\; \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}rw \\ &=\; \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma C_k(A)'e_i \\ &=\; \Big(\mathbb{I}_n - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma^{1/2}\Big)\Sigma^{1/2}C_k(A)'e_i \\ &=\; \Big(\mathbb{I}_n - P_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \\ &=\; M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i. \end{aligned}$$

Therefore, the equation above and (5.12) imply that if $\overline{v}_{k,i,j}(A,\Sigma) \neq 0$ then either:

$$\overline{v}_{k,i,j}(A,\Sigma) = \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}$$

or

$$\overline{v}_{k,i,j}(A,\Sigma) = -\Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

Furthermore, since any solution for which z is the set of binding constraints satisfies $2w_1^* x^{*\prime} = (C_k(A)'e_i - rw)'\Sigma$, then the solution vectors for which the constraints in r are binding are:

$$x^* = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2},$$

or

$$x^* = -\Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

In any case the lagrange multipliers for the active constraints are given (as shown above) by,

$$w^* = (r'\Sigma r)^{-1} r'\Sigma C_k(A)' e_i.$$

Remark to the proof of Lemma 1: If $\overline{v}_{k,i,j}(A,\Sigma) = 0$, then neither the maximizer x^* nor the Lagrange multipliers w^* are unique. Note that any $x \in \mathbb{R}^n$ orthogonal to $C_k(A)' e_i$ satisfying the ellipsoid constraint $x'\Sigma^{-1}x = 1$ and the sign/zero restrictions is a solution. In addition, any vector of Lagrange multipliers satisfying the equation $C_k' e_i - rw = \mathbf{0}_{n\times 1}$, satisfies the F.O.C.

5.A.2 Proof of Proposition 1

Let $x^* \in \mathbb{R}^n$ be a solution to the program (5.5) and let r^* be the set of constraints that are active at x^*.

Step 1: We show first that

$$\overline{v}_{k,i,j}(A,\Sigma) \geq \max_{r\in R}\Big(\max\{f^+_{max}(A,\Sigma;r), f^-_{max}(A,\Sigma;r)\}\Big).$$

We do so by considering two different cases. *Case 1.1:* Take any $r \in R$, and assume first that $v_{k,i,j}(A,\Sigma,r) \neq 0$. If $\mathbf{1}_{m_s}(x^*_+(A,\Sigma,r)) = 0$, then

$$f^+_{max}(A,\Sigma,r) = v_{k,i,j}(A,\Sigma,r) - 2c \leq c - 2c = -c < \overline{v}_{k,i,j}(A,\Sigma).$$

If, however, $r \in \mathbb{R}$ is such that $\mathbf{1}_{m_s}(x^*_+(A,\Sigma,r)) = 1$, then $x^*_+(A,\Sigma,r)$ satisfies all the equality and inequality restrictions in (5.5) and, by construction, also satisfies the ellipsoid constraint

$$x^*_+(A,\Sigma,r)'\Sigma^{-1}x^*_+(A,\Sigma,r) = 1.$$

Consequently, $\overline{v}_{k,i,j}(A,\Sigma) \geq f^+_{max}(A,\Sigma,r)$ for all $r \in R$. An analogou argument shows that $\overline{v}_{k,i,j}(A,\Sigma) \geq f^-_{max}(A,\Sigma,r)$. This implies that:

$$\overline{v}_{k,i,j}(A,\Sigma) \geq \max\{f^+_{max}(A,\Sigma,r), f^-_{max}(A,\Sigma,r)\},$$

for all $r \in \mathbb{R}$ such that $v_{k,i,j}(A,\Sigma,r) \neq 0$.

Case 1.2: Consider now any r such that $v_{k,i,j}(A,\Sigma,r) = 0$. If there is no feasible point x^* that gives such value, then $f^+_{max}(A,\Sigma,r) = f^-_{max}(A,\Sigma,r) = -c \leq \overline{v}_{k,i,j}(A,\Sigma)$. If there is such a feasible point $x^* \neq 0$ then $f^+_{max}(A,\Sigma,r) = f^-_{max}(A,\Sigma,r) = 0$ Since x^* is in the choice set

of the program 5.5, then $f^{+}_{max}(A,\Sigma,r) = f^{-}_{max}(A,\Sigma,r) = 0 \leq \overline{v}_{k,i,j}(A,\Sigma)$. Therefore, Case 1.1 and 1.2 imply that:

$$\overline{v}_{k,i,j}(A,\Sigma) \geq \max\{f^{+}_{max}(A,\Sigma,r), f^{-}_{max}(A,\Sigma,r)\} \quad \text{for all } r \in R.$$

Step 2: We now show that

$$\overline{v}_{k,i,j}(A,\Sigma) \leq \max_{r\in R} \max\{f^{+}_{max}(A,\Sigma,r), f^{-}_{max}(A,\Sigma,r)\}.$$

Again, we consider two cases.
Case 2.1: Assume first that $\overline{v}_{k,i,j}(A,\Sigma) \neq 0$. Without loss of generality, let $\overline{v}_{k,i,j}(A,\Sigma) > 0$. Let $r^* \in R$ denote the set of active sign restrictions at the solution x^*. By Lemma 1, we know that

$$\overline{v}_{k,i,j}(A,\Sigma) = \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r^*} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2},$$

and

$$x^*(A,\Sigma;r^*) = \Sigma^{1/2}\left(M_{\Sigma^{1/2} r^*}\right)\Sigma^{1/2} C_k(A)' e_i \Big/ \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r^*} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2}.$$

Since this point satisfies the sign restrictions not in r^*, then

$$\left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r^*} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2} = f^{+}_{max}(A,\Sigma,r^*).$$

Consequently,

$$\overline{v}_{k,i,j}(A,\Sigma) \leq \max_{r\in R} \max\{f^{+}_{max}(A,\Sigma,r), f^{-}_{max}(A,\Sigma,r)\}.$$

Case 2.2: If $\overline{v}_{k,i,j}(A,\Sigma) = 0$, there is an $x^* \neq 0$ in the choice set. Hence, the Karush-Kuhn-Tucker conditions implies that $C_k(A)' e_i$ is a linear combination of the active constraints that generate the value of zero (which means there is an r^* such that $f^{+}_{max}(A,\Sigma;r^*) = f^{-}_{max}(A,\Sigma;r^*) = 0$). Therefore, $\overline{v}_{k,i,j}(A,\Sigma) = f(A,\Sigma,r^*) \leq \max_{r\in R} \max\{f^{+}_{max}(A,\Sigma,r), f^{-}_{max}(A,\Sigma,r)\}$. Step 1 and Step 2 shows that the value function $\overline{v}_{k,i,j}(A,\Sigma)$ is obtained by computing the Karush-Kuhn-Tucker points in Lemma 1 for each r, penalizing the value $\overline{v}_{k,i,j}(A,\Sigma;r)$ if unfeasible, and maximizing over all the possible values of r. The proof for the lower bound is analogous.

5.A.3 Lemma 2

Note that if $r_l(\mu)$ is differentiable and $v_{k,i,j}(\mu, r_l(\mu)) \neq 0$, then the function:

$$v_{k,i,j}(\mu; r_l(\mu)) = \left(e_i' C_k(A) \Sigma^{1/2} M_{\Sigma^{1/2} r_l(\mu)} \Sigma^{1/2} C_k(A)' e_i\right)^{1/2},$$

is differentiable as well. Moreover, the function

$$x^*(\mu; r_l(\mu)) \equiv \Sigma^{1/2}\left(M_{\Sigma^{1/2} r_l(\mu)}\right)\Sigma^{1/2} C_k(A)' e_i \Big/ v_{k,i,j}(\mu; r_l(\mu))$$

is also differentiable. Therefore,

$$\begin{aligned}
\frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \mu} &= \frac{\partial [e_i' C_k(A) x^*(\mu, r_l(\mu))]}{\partial \mu} \\
&\quad \text{(since } v_{k,i,j}(\mu, r_l(\mu)) = e_i' C_k(A) x^*(\mu, r_l(\mu))\text{)} \\
&= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \mu} C_k'(A) e_i + \frac{\partial (x^*(\mu; r_l(\mu))' \otimes e_i') \mathrm{vec}(C_k(A))}{\partial \mu}, \\
&\quad \text{(where we have re-written } e_i' C_k(A) x^* \text{ as } (x^{*\prime} \otimes e_i') \mathrm{vec}(C_k(A))\text{)} \\
&= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \mu} C_k'(A) e_i + \frac{\partial \mathrm{vec}(C_k(A))}{\partial \mu} (x^*(\mu; r_l(\mu)) \otimes e_i) \\
&\quad \text{(where we have applied the chain rule for matrix derivatives).}
\end{aligned}$$

We now use the envelope theorem to compute this derivative. Note that we have shown the existence of unique multipliers $\lambda^* \in \mathbb{R}$ and $w^* \in \mathbb{R}^l$ such that:

$$C_k(A)' e_i = \lambda^* 2 \Sigma^{-1} x^*(\mu; r_l(\mu)) + r_l(\mu) w^*.$$

Therefore:

$$\begin{aligned}
\frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \mathrm{vec}(A)} &= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \mathrm{vec}(A)} \left[\lambda^* 2 \Sigma^{-1} x^*(\mu; r_l(\mu)) + r_l(\mu) w^*\right] \\
&\quad + \frac{\partial \mathrm{vec}(C_k(A))}{\partial \mathrm{vec}(A)} (x^*(\mu; r_l(\mu)) \otimes e_i).
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \mathrm{vec}(\Sigma)} &= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \mathrm{vec}(\Sigma)} \left[\lambda^* 2 \Sigma^{-1} x^*(\mu; r_l(\mu)) + r_l(\mu) w^*\right] \\
&\quad + \frac{\partial \mathrm{vec}(C_k(A))}{\partial \mathrm{vec}(\Sigma)} (x^*(\mu; r_l(\mu)) \otimes e_i).
\end{aligned}$$

Note also that:

$$0 = \frac{\partial x^*(\mu, r_l(\mu))' \Sigma^{-1} x^*(\mu, r_l(\mu))}{\partial \text{vec}(A)} = 2\frac{\partial x^*(\mu, r_l(\mu))}{\partial \text{vec}(A)} \Sigma^{-1} x^*(\mu, r_l(\mu))$$

and

$$\begin{aligned} 0 &= \frac{\partial r_l(\mu)' x^*(\mu, r_l(\mu))}{\partial \text{vec}(A)} \\ &= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \text{vec}(A)} r_l(\mu) \\ &\quad + \Big(\tfrac{\partial r_{1,l}(\mu)}{\text{vec}(A)} x^*(\mu, r_l(\mu)), \quad \cdots \quad , \tfrac{\partial r_{l,l}(\mu)}{\text{vec}(A)} x^*(\mu, r_l(\mu)) \Big), \end{aligned}$$

where $r_{k,l}(\mu)$ denotes the k-th column of $r_l(\mu)$. Consequently:

$$\begin{aligned} \frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \text{vec}(A)} &= \frac{\partial \text{vec}(C_k(A))}{\partial \text{vec}(A)} (x^*(\mu; r_l(\mu)) \otimes e_i) \\ &\quad - \sum_{k=1}^{l} w_k^* \frac{\partial \text{vec}(r_{k,l}(\mu))}{\text{vec}(A)} x^*(\mu, r_l(\mu)), \end{aligned}$$

where w_k^* is the k-th entry of the vector of lagrange multipliers w^*. This gives the partial derivative of $v_{k,i,j}(\mu, r_l(\mu))$ with respect to vec(A). We note that this derivative can also be written as:

$$\begin{aligned} \frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \text{vec}(A)} &= \frac{\partial \text{vec}(C_k(A))}{\partial \text{vec}(A)} (x^*(\mu; r_l(\mu)) \otimes e_i) \\ &\quad - \frac{\partial \text{vec}(r_l(\mu)')}{\text{vec}(A)} (x^*(\mu, r_l(\mu)) \otimes \mathbb{I}_l) w^*, \end{aligned}$$

which is the expression given in the overview. Finally, to get the derivative with respect to vec(Σ) we note that:

$$\begin{aligned} 0 &= \frac{\partial x^*(\mu, r_l(\mu))' \Sigma^{-1} x^*(\mu, r_l(\mu))}{\partial \text{vec}(\Sigma)} = 2\frac{\partial x^*(\mu, r_l(\mu))}{\partial \text{vec}(\Sigma)} \Sigma^{-1} x^*(\mu, r_l(\mu)) \\ &\quad - (\Sigma^{-1} x^*(\mu, r_l(\mu)) \otimes \Sigma^{-1} x^*(\mu, r_l(\mu))), \end{aligned}$$

and

$$\begin{aligned} 0 &= \frac{\partial r_l(\mu)' x^*(\mu, r_l(\mu))}{\partial \text{vec}(\Sigma)} \\ &= \frac{\partial x^*(\mu, r_l(\mu))}{\partial \text{vec}(\Sigma)} r_l(\mu) + \Big(\tfrac{\partial r_{1,l}(\mu)}{\text{vec}(\Sigma)} x^*(\mu, r_l(\mu)), \quad \cdots \quad , \tfrac{\partial r_{l,l}(\mu)}{\text{vec}(\Sigma)} x^*(\mu, r_l(\mu)) \Big). \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\partial v_{k,i,j}(\mu, r_l(\mu))}{\partial \text{vec}(\Sigma)} &= \lambda^* (\Sigma^{-1} x^*(\mu, r_l(\mu)) \otimes \Sigma^{-1} x^*(\mu, r_l(\mu))) \\ &\quad - \sum_{k=1}^{l} w_k^* \frac{\partial \text{vec}(r_{k,l}(\mu))}{\text{vec}(\Sigma)} x^*(\mu, r_l(\mu)). \end{aligned}$$

5.A.4 Proof of Proposition 2

Let $r_1(\mu), r_2(\mu), \ldots r_M(\mu)$ denote the elements of $R(\mu)$. Each of these elements activate a different collection of sign restrictions. Without loss of generality, assume that $R^*(\mu)$ contains the first L elements $r_1(\mu), \ldots, r_L(\mu)$ of $R(\mu)$. Consider any sequence (A_T, Σ_T) such that

$$(\text{vec}A_T', \text{vech}\Sigma_T')' = (\text{vec}A', \text{vech}\Sigma')' + h_T / r_T,$$

where $h_T \to h$, $r_T \to \infty$ and such that for large enough T, $(\text{vec}A_T, \text{vech}\Sigma_T)$ belongs to the parameter space μ. Suppose that $\overline{v}_{k,i,j}(\mu) > 0$. Note that by Proposition 1:

$$\overline{v}_{k,i,j}(A_T, \Sigma_T) =$$

$$\begin{cases} f(A_T, \Sigma_T, r_1(\mu_T)) & \text{if} \quad f(A_T, \Sigma_T, r_1(\mu_T)) = \max_{r \in R(\mu_T)} f(A_T, \Sigma_T, r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ f(A_T, \Sigma_T, r_L(\mu_T)) & \text{if} \quad f(A_T, \Sigma_T, r_L(\mu_T)) = \max_{r \in R(\mu_T)} f(A_T, \Sigma_T, r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ f(A_T, \Sigma_T, r_M(\mu_T)) & \text{if} \quad f(A_T, \Sigma_T, r_M(\mu_T)) = \max_{r \in R(\mu_T)} f(A_T, \Sigma_T, r). \end{cases}$$

Since each $r_l(\mu)$ is assumed to be continuous and only the first L elements of $R(\mu)$ belong to $R^*(\mu)$ then (for T large enough):

$$\overline{\nu}_{k,i,j}(A_T,\Sigma_T) \;=\; \begin{cases} \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_1(\mu_T)) & \text{if} \quad \nu(A_T,\Sigma_T,r_1(\mu_T)) \;=\; \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_L(\mu_T)) & \text{if} \quad \nu(A_T,\Sigma_T,r_L(\mu)) \;=\; \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r), \\ \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_{L+1}(\mu_T)) & \text{if} \quad \nu(A_T,\Sigma_T,r_{L+1}(\mu)) \;=\; \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r), \\ -2(1-\mathbf{1}_{m_s}(x_+^*(\mu,r_{L+1}(\mu_T)))) & \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_M(\mu_T)) & \text{if} \quad f(A_T,\Sigma_T,r_M(\mu)) \;=\; \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r). \\ -2(1-\mathbf{1}_{m_s}(x_+^*(\mu,r_M(\mu)))) & \end{cases}$$

Note that $\overline{\nu}_{k,i,j}(\mu) = \overline{\nu}_{k,i,j}(\mu; r_l(\mu))$ for all $r_l(\mu) \in R^*(\mu)$ and $\overline{\nu}_{k,i,j}(\mu) > \overline{\nu}_{k,i,j}(\mu; r_l(\mu))$ for all $r_l(\mu) \notin R^*(\mu)$. Therefore, the equation above implies that $r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T) - \overline{\nu}_{k,i,j}(A,\Sigma))$ equals

$$\begin{cases} r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_1(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_1(\mu_T))) & \text{if} \quad \nu(A_T,\Sigma_T,r_1(\mu_T)) = \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_L(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_L(\mu))) & \text{if} \quad \nu(A_T,\Sigma_T,r_L(\mu)) = \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_M(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_M(\mu))) & \text{if} \quad f(A_T,\Sigma_T,r_M(\mu)) = \max_{r\in R(\mu_T)} f(A_T,\Sigma_T,r). \\ \quad + r_T(\overline{\nu}_{k,i,j}(\mu,r_M(\mu)) - \overline{\nu}_{k,i,j}(\mu)) & \\ \quad - r_T 2(1-\mathbf{1}_{m_s}(x^*(\mu,r_M(\mu)))) & \end{cases}$$

For T large enough, this implies that $r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T) - \overline{\nu}_{k,i,j}(A,\Sigma))$ equals

$$\begin{cases} r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_1(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_1(\mu))) & \text{if} \quad \nu(A_T,\Sigma_T,r_1(\mu_T)) = \max_{r\in R^*(\mu_T)} \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r), \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_L(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_L(\mu))) & \text{if} \quad \nu(A_T,\Sigma_T,r_L(\mu)) = \max_{r\in R^*(\mu_T)} \overline{\nu}_{k,i,j}(A_T,\Sigma_T,r). \end{cases}$$

Since each of $r_T(\overline{\nu}_{k,i,j}(A_T,\Sigma_T,r_1(\mu_T)) - \overline{\nu}_{k,i,j}(\mu,r_l(\mu)))$ in the previous expression is, by Lemma 2, approximately equal to

$$\dot{\nu}_{k,i,j}(\mu; r_l(\mu))' h.$$

Then

$$r_T\Big(\overline{v}_{k,i,j}(A_T,\Sigma_T)-\overline{v}_{k,i,j}(A,\Sigma)\Big)\to\max_{r\in R^*(\mu)}\Big(\dot{v}_{k,i,j}(\mu;r)'h\Big).$$

5.A.5 Proof of Proposition 3

Let P denote the data generating process. For notational simplicity we write μ instead of $\mu(P)$ and Ω instead of $\Omega(P)$ whenever convenient. Note first that

$$P\Big(\lambda\in\Big[\underline{v}_{k,i,j}(\widehat{\mu}_T)-z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T},\ \ \overline{v}_{k,i,j}(\widehat{\mu}_T)+z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T}\Big]\Big) \tag{5.13}$$

equals

$$P\Big(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T)-\underline{v}_{k,i,j}(\mu))\le z_{1-\alpha/2}\,\widehat{\sigma}_T+\sqrt{T}(\lambda-\underline{v}_{k,i,j}(\mu))\text{ and}$$
$$\sqrt{T}(\lambda-\overline{v}_{k,i,j}(\mu))-z_{1-\alpha/2}\widehat{\sigma}_T\le\sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T)-\overline{v}_{k,i,j}(\mu))\Big).$$

Since

$$\sqrt{T}(\lambda-\underline{v}_{k,i,j}(\mu))\ge 0\text{ and }\sqrt{T}(\lambda-\overline{v}_{k,i,j}(\mu))\le 0,$$

(5.13) is bounded from below by

$$P\Big(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T)-\underline{v}_{k,i,j}(\mu))\le z_{1-\frac{\alpha}{2}}\,\widehat{\sigma}_T\text{ and }\ -z_{1-\frac{\alpha}{2}}\widehat{\sigma}_T\le\sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T)-\overline{v}_{k,i,j}(\mu))\Big),$$

which is itself bounded from below by:

$$\begin{aligned}P\Big(&\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T)-\underline{v}_{k,i,j}(\mu))\ \le\ z_{1-\frac{\alpha}{2}}\,\widehat{\sigma}_T,\\ &\text{and}\quad -z_{1-\frac{\alpha}{2}}\widehat{\sigma}_T\ \le\ \sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T)-\overline{v}_{k,i,j}(\mu)),\\ &\text{and }||\sqrt{T}(\widehat{\mu}_T-\mu)||\ \le\ M_\epsilon\Big),\end{aligned}$$

where M_ϵ is such that

$$P\Big(||\zeta(P)||>M_\epsilon\Big)\le\epsilon.$$

Since, by Proposition 2, both $\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\mu)$ are directionally differentiable with directional derivatives:

$$\min_{r\in R_*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'h\Big],$$

and

$$\max_{r\in R^*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'h\Big].$$

The directional differentiability implies that for any compact set K there is T large enough such that for any $h \in K$:

$$-\epsilon \leq \sqrt{T}(\underline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \underline{v}_{k,i,j}(\mu)) - \min_{r \in R_*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'h\right] \leq \epsilon,$$

and

$$-\epsilon \leq \sqrt{T}(\overline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \overline{v}_{k,i,j}(\mu)) - \max_{r \in R^*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'h\right] \leq \epsilon.$$

Therefore, for T large enough:

$$\inf_{\lambda \in \mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T},\ \ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T}\right]\Big)$$

is bounded from below by:

$$P\Big(\min_{r \in R_*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\sqrt{T}(\widehat{\mu}_T - \mu)\right] \leq z_{1-\alpha/2}\,\widehat{\sigma}_T \text{ and}$$

$$-z_{1-\alpha/2}\widehat{\sigma}_T \leq \max_{r \in R^*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\sqrt{T}(\widehat{\mu} - \mu)\right], \text{ and } ||\sqrt{T}(\widehat{\mu}_T - \mu)|| \leq M_\epsilon\Big),$$

which, by Assumption 4, converges in distribution to:

$$P\Big(\min_{r \in R_*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\zeta(P)\right] \leq z_{1-\alpha/2}\,\sigma \text{ and}$$

$$-z_{1-\alpha/2}\sigma \leq \max_{r \in R_*(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\zeta(P)\right], \text{ and } ||\zeta(P)|| \leq M_\epsilon\Big),$$

where σ is the probability limit of $\widehat{\sigma}_T$:

$$\sigma \equiv \max_{r \in R(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\Omega\dot{v}_{k,i,j}(\mu, r)\right],$$

and where we have used the fact that $\sigma > 0$.[19] Consequently,

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\frac{\alpha}{2}}\,\widehat{\sigma}_T/\sqrt{T},\right.$$
$$\left.\overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\frac{\alpha}{2}}\,\widehat{\sigma}_T/\sqrt{T}\right]\Big)$$

[19] This follows from the fact that:

$$\sigma \equiv \max_{r \in R(\mu)} \left[\dot{v}_{k,i,j}(\mu, r)'\Omega\dot{v}_{k,i,j}(\mu, r)\right] \geq \min_{r \in R(\mu)} \left[\dot{v}_{k,i,j}(\mu(P), r)'\Omega(P)\dot{v}_{k,i,j}(\mu(P), r)\right] > 0,$$

which we have assumed to be strictly bigger than zero.

is larger than or equal:

$$1 - P\Big(\min_{r\in R_*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'\zeta(P)\Big] > z_{1-\frac{\alpha}{2}}\,\sigma\Big) - P\Big(\max_{r\in R_*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'\zeta(P)\Big] < -z_{1-\frac{\alpha}{2}}\,\sigma\Big)$$

$$-P\Big(||\zeta(P)|| > M_\epsilon\Big).$$

Take some $r \in R_*(\mu)$ for which $\sigma(r) \equiv \dot{v}_{k,i,j}(\mu,r)'\Omega\dot{v}_{k,i,j}(\mu,r) > 0$ (one such r must exist for we have assumed that $\min_{r\in R(\mu)}[\dot{v}_{k,i,j}(\mu,r)'\Omega\dot{v}_{k,i,j}(\mu,r)] > 0$). Therefore:

$$\begin{aligned} P\Big(\min_{r\in R_*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'\zeta(P)\Big] \;>\; z_{1-\alpha/2}\,\sigma\Big) \;&\le\; P\Big(\dot{\bar{v}}_{k,i,j}(\mu,r)'\zeta(P) \;>\; z_{1-\alpha/2}\sigma\Big) \\ &\le\; \mathbb{P}\Big(N(0,1) \;>\; z_{1-\alpha/2}\frac{\sigma}{\sigma(r)}\Big), \end{aligned}$$

which is at most $\alpha/2$ since $\sigma \ge \sigma(r)$.

Now, take some $r \in R^*(\mu)$ for which $\sigma(r) \equiv \dot{v}_{k,i,j}(\mu,r)'\Omega\dot{v}_{k,i,j}(\mu,r) > 0$. Note that

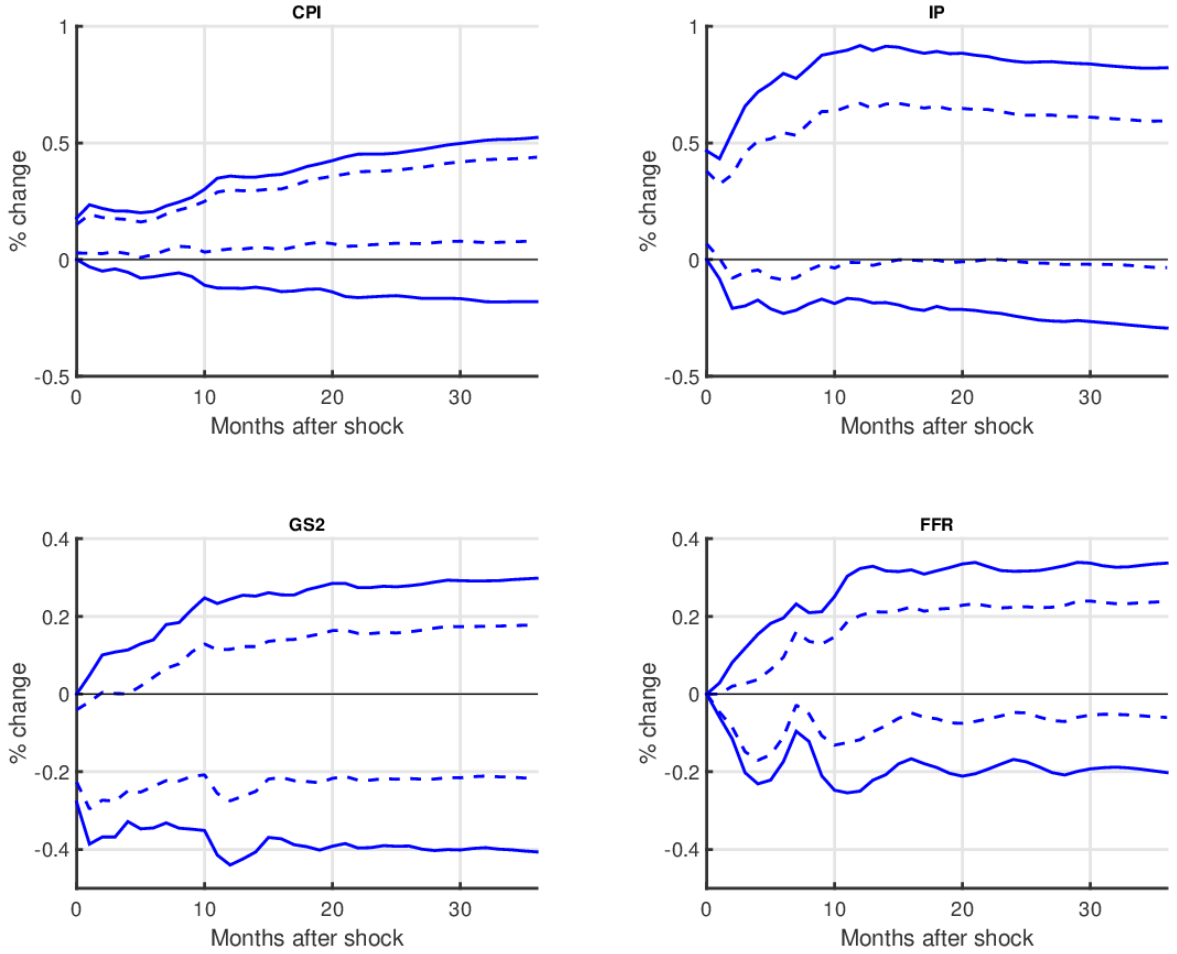
$$\begin{aligned} P\Big(\max_{r\in R^*(\mu)}\Big[\dot{v}_{k,i,j}(\mu,r)'\zeta(P)\Big] \;<\; -z_{1-\alpha/2}\,\sigma\Big) \;&\le\; P\Big(\dot{v}_{k,i,j}(\mu,r)'\zeta(P) \;<\; -z_{1-\alpha/2}\sigma\Big) \\ &\le\; \mathbb{P}\Big(N(0,1) \;<\; -z_{1-\alpha/2}\frac{\sigma}{\sigma(r)}\Big), \end{aligned}$$

which is less than or equal to $\le \alpha/2$ as $\sigma > \sigma(r)$. We conclude that

$$\begin{aligned} \liminf_{T\to\infty}\inf_{\lambda\in\mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \;\in\; \Big[&\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T}, \\ &\overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_T/\sqrt{T}\Big]\Big) \;\ge\; 1-\alpha-\epsilon. \end{aligned}$$

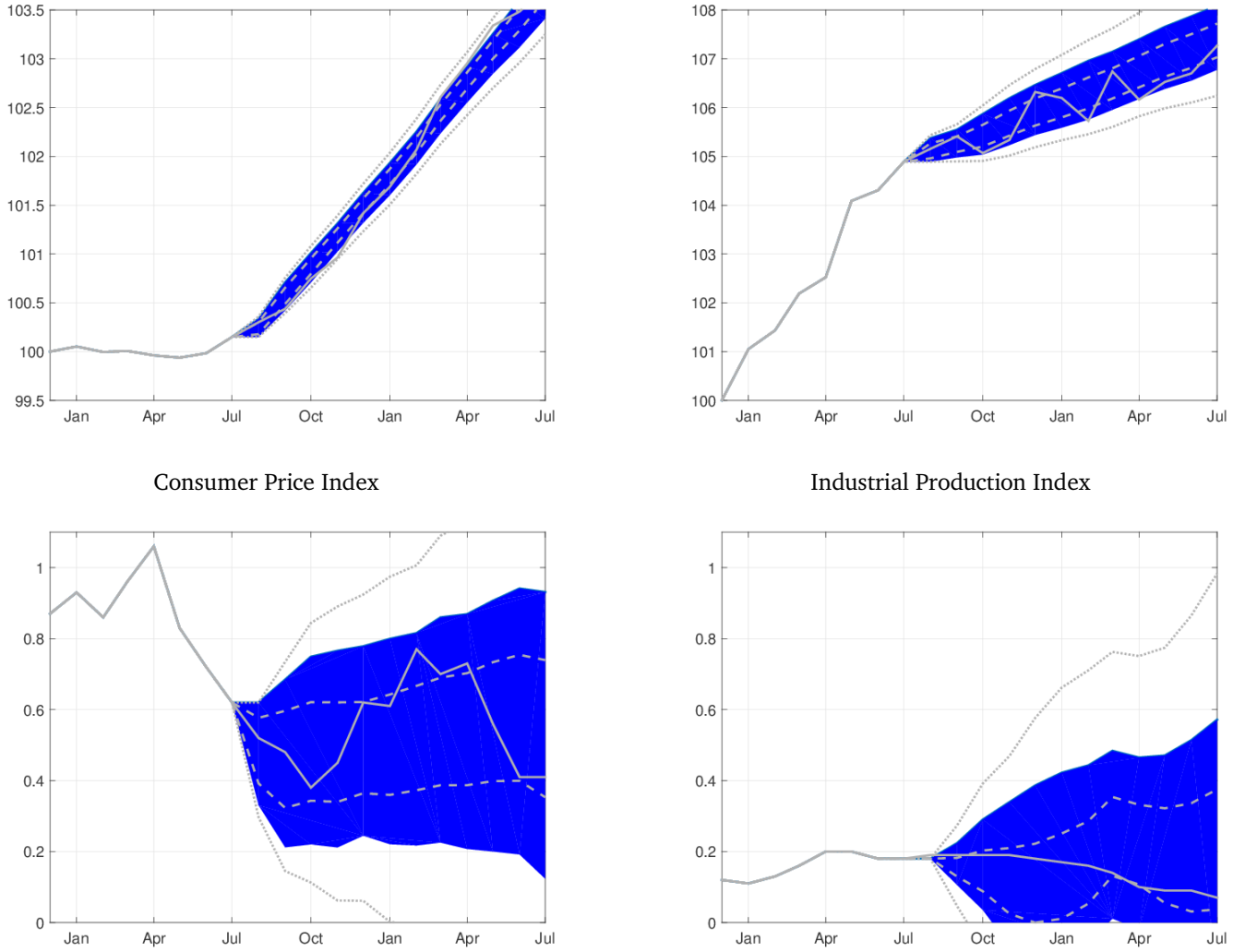
Since $\epsilon > 0$ is arbitrary, the desired result follows.

Figure 5.6. Identified set for the cumulative impulse response functions to an UMP shock (given $\widehat{\mu}_T$)



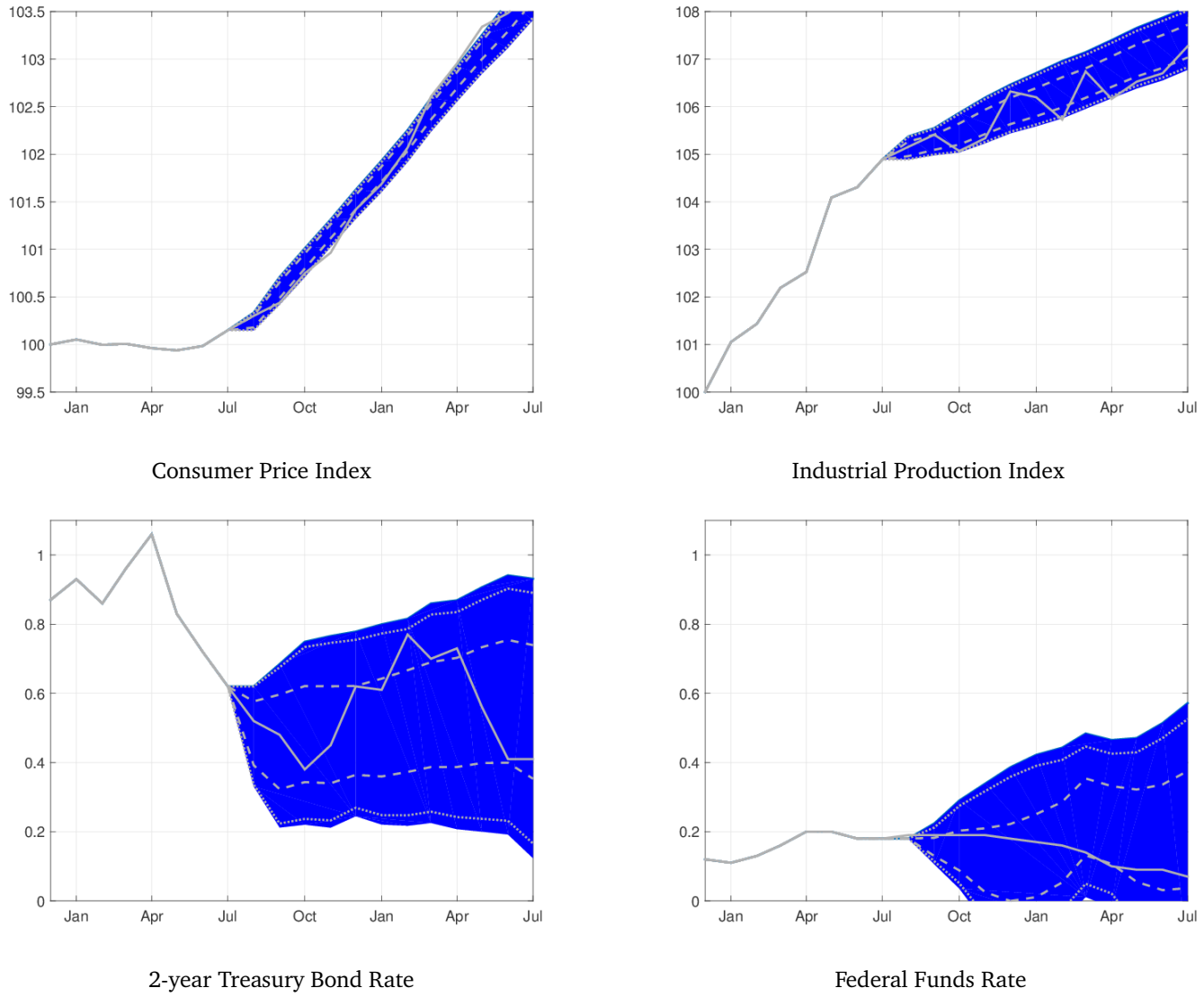
(Solid, Blue Line) Endpoints of the identified set for the cumulative responses given $\widehat{\mu}_T$ and the equality/inequality restrictions in Table I. (Blue, Dashed) Grid search approach to evaluate the identified set (10,000 uniform draws from a unit vector $q \in \mathbb{R}^4$).

Figure 5.7. Projection confidence interval for CPI, IP, 2yTB, FF after the August 2010 announcement



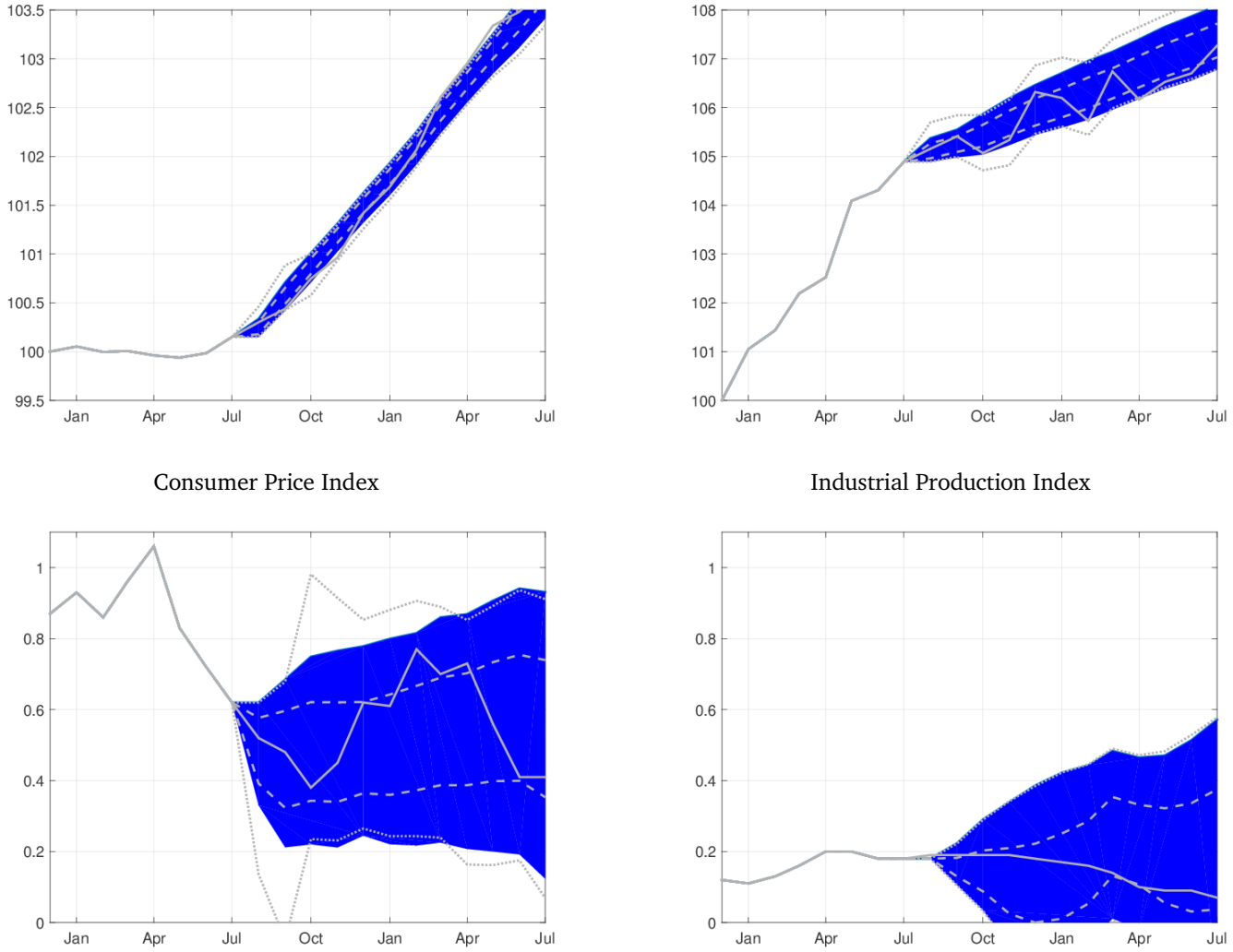
(Shaded, Blue Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Gray, Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Gray, Dashed Line) 68% credible set constructed using the priors in Uhlig (2005). (Gray, Dotted Line) Gafarov et al. (2016)'s 68% confidence interval based on the projection approach.

Figure 5.8. Calibrated projection confidence interval for CPI, IP, 2yTB, FF after the August 2010 announcement



(Shaded, Blue Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Gray, Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Gray, Dashed Line) 68% credible set constructed using the priors in Uhlig (2005). (Gray, Dotted Line) Gafarov et al. (2016)'s 68% *differentiable* projection.

Figure 5.9. Robust credible set for CPI, IP, 2yTB, FF after the August 2010 announcement



(Shaded, Blue Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Gray, Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Gray, Dashed Line) 68% credible set constructed using the priors in Uhlig (2005). (Gray, Dotted Line) Giacomini and Kitagawa (2015)'s 68% robust-Bayes credible set constructed using the priors for the reduced-form parameters in Uhlig (2005).

6

Projection Inference for Set-Identified SVARs

Joint with Bulat Gafarov and José Luis Montiel Olea

6.1 Introduction

A Structural Vector Autoregression (SVAR) (Sims (1980, 1986)) is a time series model that brings theoretical restrictions into a linear, multivariate autoregression. The theoretical restrictions are used to transform *reduced-form* parameters (regression coefficients and the covariance matrix of residuals) into *structural* parameters that are more amenable to policy interpretation. Depending on the restrictions imposed, the map between reduced-form and structural parameters can be one-to-one (a point-identified SVAR) or one-to-many (a set-identified SVAR).

It is now customary for empirical macroeconomic studies to impose sign and/or exclusion restrictions on structural dynamic responses in SVARs in order to set-identify the model, as in the pioneering work of Faust (1998) and Uhlig (2005). The vast majority of these studies use numerical Bayesian methods to construct posterior credible sets for the coefficients of the structural impulse-response function.

Despite the popularity of the Bayesian approach, a practical concern is the fact that posterior inference for the structural parameters continues to be influenced by prior beliefs even if the sample size is infinite. This point has been documented—in detail and generality—in the work of Poirier (1998), Gustafson (2009), and Moon and Schorfheide (2012). More recently,

Baumeister and Hamilton (2015a) provided an explicit characterization of the influence of prior beliefs on posterior distributions for structural parameters in set-identified SVARs.

This paper studies the properties of the *projection method*—which does not rely on the specification of prior beliefs for set-identified parameters—to conduct *simultaneous* inference about the coefficients of the structural impulse-response function (and their identified set). The proposal is to 'project' a typical Wald ellipsoid for the reduced-form parameters of a VAR. The suggested nominal $1-\alpha$ projection region consists of all the structural parameters of interest compatible with the reduced-form parameters in a nominal $1-\alpha$ Wald ellipsoid.

The attractive features of the projection approach—explained in more detail in the next section—are its general applicability, its computational feasibility, and the fact that a nominal $1-\alpha$ projection region has—asymptotically and under mild assumptions—both frequentist coverage and *robust* Bayesian credibility of at least $1-\alpha$.[1] Moreover, building on Kaido et al. (2016), we show that our baseline projection can be 'calibrated' to eliminate excessive robust Bayesian credibility.

The remainder of the paper is organized as follows. Section 6.2 presents an overview of the projection approach. Section 6.3 presents the SVAR model and establishes the frequentist coverage of projection. Section 6.4 establishes the asymptotic robust Bayesian credibility of the projection region. Section 6.5 presents the 'calibration' algorithm designed to eliminate the excess of robust Bayesian credibility and shows that, under regularity conditions, our calibration also removes the excess of frequentist coverage. Section 6.6 discusses the implementation of projection in the context of the demand/supply SVAR for the U.S. labor market. Section 6.7 concludes.

6.2 Overview and related literature

6.2.1 Overview

Let μ denote the parameters of a reduced-form vector autoregression; i.e., the slope coefficients in the regression model and the covariance matrix of residuals. Let λ denote the structural parameter of interest; i.e., the response of some variable i to a structural shock j at horizon k (or a vector of responses). In set-identified SVARs there is a known map between μ and the lower and upper bound for λ; see Uhlig (2005). Consequently, the smallest

[1] The robustness is relative to the choice of prior for the so-called 'rotation' matrix as in the recent work of Giacomini and Kitagawa (2015).

and largest value of a particular structural coefficient of interest can be written, simply and succinctly, as $\underline{v}(\mu)$ and $\overline{v}(\mu)$.

Our projection region for λ (and for its identified set) is based on a straightforward application of the classical idea of *projection* inference; see Scheffé (1953), Dufour (1990), and Dufour and Taamouti (2005, 2007). Let $\widehat{\mu}_T$ denote the sample least squares estimator for μ and let $\text{CS}_T(1-\alpha;\mu)$ denote its nominal $1-\alpha$ Wald confidence ellipsoid. If, asymptotically, $\text{CS}_T(1-\alpha;\mu)$ covers the parameter μ with probability $1-\alpha$, then, asymptotically, the interval

$$\text{CS}_T(1-\alpha;\lambda) \equiv \Big[\inf_{\mu\in\text{CS}_T(1-\alpha,\mu)} \underline{v}(\mu)\,,\ \sup_{\mu\in\text{CS}_T(1-\alpha,\mu)} \overline{v}(\mu)\Big] \tag{6.1}$$

covers the set-identified parameter λ (and its identified set) with probability at least $1-\alpha$ (uniformly over a large class of data generating processes).[2,3]

In many applications there is interest in conducting *simultaneous* inference on h structural parameters; for example, if one wants to analyze the response of variable i to a structural shock j for all horizons ranging from period 1 to h as in Jordà (2009), Inoue and Kilian (2013, 2016), and Lütkepohl et al. (2015). In this case, the projection region given by:

$$\text{CS}_T(1-\alpha;(\lambda_1,\ldots,\lambda_h)) \equiv \text{CS}_T(1-\alpha;\lambda_1)\times\ldots\times\text{CS}_T(1-\alpha;\lambda_h), \tag{6.2}$$

covers the structural coefficients $(\lambda_1,\ldots\lambda_h)$ and their identified set with probability at least $1-\alpha$ as the sample size grows large. The only assumption required to guarantee the frequentist coverage of our projection region is the asymptotic validity of the confidence set for the reduced-form parameters, μ.

General Applicability: The validity of our projection method requires no regularity assumptions (like continuity or differentiability) on the bounds of the identified set $\underline{v}(\cdot)$ and $\overline{v}(\cdot)$. This means we can handle the typical application of set-identified SVARs in the empirical macroeconomics literature (exclusion restrictions on contemporaneous coefficients, long-run restrictions, elasticity bounds, and of course sign/zero restrictions on the responses of different variables at different horizons for different shocks).

---

[2] Formally, we show that the confidence interval described in (6.1) is *uniformly consistent of level* $1-\alpha$ for the structural parameter λ (and its identified set) over some class of data generating processes.

[3] The application of projection inference to SVARs was first suggested by Moon and Schorfheide (2012) (p. 11, NBER working paper 14882). The projection approach is also briefly mentioned in the work Kline and Tamer (2015) (Remark 8) in the context of set-identified models. None of these papers established the properties for projection inference discussed in our work.

Computational Feasibility: The implementation of our projection approach requires neither numerical inversion of hypothesis tests nor sampling from the space of rotation matrices. Instead, we use state-of-the-art optimization algorithms to solve for the maximum and minimum value of a mathematical program to compute the two end points of the confidence interval in (6.1).

Robust Bayesian Credibility: In the spirit of making our results appealing to Bayesian decision makers, we show that our suggested nominal $1-\alpha$ projection region will have—as the sample size grows large—robust Bayesian credibility of at least $1-\alpha$. This means that the asymptotic posterior probability that the vector of structural parameters of interest belongs to the projection region will be at least $1-\alpha$; for a fixed prior on the reduced-form parameters, μ, and for *any prior* on the set-identified parameters. A sufficient condition to establish the robust Bayesian credibility of projection is that the prior for μ used to compute credibility satisfies the *Bernstein-von Mises* theorem.

'Calibrated' Projection: Despite the features highlighted above, projection inference is *conservative* both for a frequentist and a robust Bayesian. That is, both the asymptotic confidence level and the asymptotic robust credibility of projection can be strictly above $1-\alpha$. Kaido et al. (2016) [henceforth, KMS] refer to the excess of frequentist coverage as *projection conservatism* and develop an innovative *calibration* approach to eliminate it.[4]

The calibration exercise in KMS requires, in the SVAR context, the computation of Monte-Carlo coverage probabilities for the projection region over an exhaustive grid of values for the reduced-form parameters, μ. In several SVAR applications, the dimension of μ compromises the construction of an exhaustive grid.

Instead of insisting on removing excessive frequentist coverage, we suggest practitioners to calibrate projection to attain a robust Bayesian credibility of exactly $1-\alpha$. The calibration of robust credibility is computationally feasible even if μ is of large dimension, as no exhaustive grid for μ is needed. We provide a detailed description of our calibration procedure in Section 6.5. Broadly speaking, the calibration consists of drawing μ from its posterior distribution (or a suitable large-sample Gaussian approximation); evaluating the functions $\underline{\nu}(\mu), \overline{\nu}(\mu)$ for each draw of μ; and decreasing the radius defining the projection region until it contains exactly $(1-\alpha)\%$ of the values of $\underline{\nu}(\mu), \overline{\nu}(\mu)$ (for different horizons and different

[4] Another recent paper proposing a procedure to eliminate the frequentist excess coverage in moment-inequality models is Bugni et al. (2014). Adapting their *profiling* idea to our set-up could be of theoretical interest and of practical relevance. We leave this question open for future research.

shocks if desired).[5] We show that if $\underline{v}(\mu), \overline{v}(\mu)$ are differentiable, our suggested calibration also removes the excessive frequentist coverage *of the identified set*.

Illustrative Example: The illustrative example in this paper is a simple demand and supply model of the U.S. labor market. We estimate standard Bayesian credible sets for the dynamic responses of wages and employment using the Normal-Wishart-Haar prior specification in Uhlig (2005) and also the alternative prior specification recently proposed by Baumeister and Hamilton (2015a). The main set-identifying assumptions are sign restrictions on contemporaneous responses: an expansionary structural demand shock increases wages and employment upon impact; an expansionary structural supply shock decreases wages but increases employment, also upon impact.[6]

The Bayesian credible sets for this application illustrate the attractiveness of set-identified SVARs. The data, combined with prior beliefs, and with the (set)-identifying assumptions imply that the initial responses to demand and supply shocks persist in the medium-run, which was not restricted ex-ante.

The Bayesian credible sets for this application also illustrate how the quantitative results in set-identified SVARs could be affected by the prior specification. For example, under the prior in Baumeister and Hamilton (2015a) the 5-year ahead response of employment to a demand shock could be as large as 4%; whereas under the priors in Uhlig (2005) the same effect is at most 2%.

Our baseline projection approach (which takes around 15 minutes) allows us to get a prior-free assessment about the magnitude (and direction) of the structural responses of interest. For example, the largest value in our projection region for the 5-year response of employment to a structural demand shock is around 2.5%. This effect is larger than the one implied by the prior in Uhlig (2005), but smaller than the one implied by the priors in Baumeister and Hamilton (2015a).

Our baseline projection approach—though informative about the effects of demand shocks—is not conclusive about the medium-run effects of structural supply shocks on wages and employment (the projection region allows for both positive and negative responses). This could be a consequence of either the robustness of projection or its conservativeness. To disentangle these effects, we calibrate projection to guarantee that it has exact robust Bayesian credibility. The calibrated projection shows that an expansionary supply shock will decrease

[5] In Section 6.6 we provide more detials on the computation time of our calibrated projection (which is around 5 hours in our illustrative example).

[6] Following Baumeister and Hamilton (2015a) we also consider elasticity bounds on the wage elasticity of both labor demand and labor supply, and also bounds on the long-run impact of a demand shock over employment.

wages in each quarter over a 5 year horizon. However, the qualitative effects of supply shocks on employment remain undetermined. The simple SVAR for the labor market illustrates the usefulness of both the baseline and the calibrated projection to analyze the robustness of quantitative and qualitative results in SVARs to prior beliefs.

6.2.2 Related literature

There has been recent interest in departing from the standard Bayesian analysis of set-identified SVARs in an attempt to provide robustness to the choice of priors. Below we provide a short description of the similarities and differences between our projection approach and three alternative methods available in the literature. It is worth mentioning that our baseline projection approach is the only procedure (among the three alternative methods discussed) that has both a frequentist and a robust Bayes interpretation. In addition, none of the other approaches allow for simultaneous inference on a vector of impulse-response coefficients (a feature that has been deemed desirable in point-identified SVARs). Our baseline projection achieves all these properties while retaining computational tractability (we solve two mathematical programs per coefficient of interest).

a) In a pioneering paper, Moon et al. (2013) [MSG] proposed both projection and Bonferroni frequentist inference using a moment-inequality, minimum distance framework based on Andrews and Soares (2010). In terms of applicability, their procedures are designed for set-identified SVARs that impose restrictions on the dynamic responses of only one structural shock. It is possible to extend their approach to the same class of modes that we consider; there is, however, a serious issue regarding computational feasibility. Specifically, both the projection and Bonferroni approaches require the researcher to compute—by simulation—a critical value for each single orthogonal matrix of dimension $n \times n$, where n is the dimension of the SVAR. Our baseline implementation of the projection method does not require any type of grid over the space of orthogonal matrices and does not require the simulation of any critical value.

b) Giacomini and Kitagawa (2015) [GK] develop a novel and generally applicable robust Bayesian approach to conduct inference about a specific coefficient of the impulse-response function in a set-identified SVAR. In terms of our notation, their procedure can be described as follows. One takes posterior draws from μ and evalutes, at each posterior draw, the functions $\underline{v}(\mu), \overline{v}(\mu)$ by solving a nonlinear program. Their credible set is a numerical (grid-search) approximation to the *smallest* interval that covers $100(1-\alpha)\%$ of the posterior realizations of the identified set.

GK and Baseline projection: In terms of properties, our baseline projection is shown to admit both a frequentist and a robust Bayes interpretation (the GK approach has only been shown to admit the latter). In terms of implementation, GK solve as many nonlinear programs as posterior draws for μ. This means that our baseline procedure will be typically faster to implement than the GK robust procedure (since our baseline projection only needs to solve two nonlinear programs). The price to pay for the reduced computational cost is the excess of robust Bayesian credibility.

GK and Calibrated projection: Our calibrated projection requires a similar amount of work as the GK robust method (both procedures evaluate the bounds of the identified set for each posterior draw). There are two differences remaining between the two approaches. First, our calibrated projection allows for *simultaneous* credibility statements over different horizons, different variables, and different shocks. Second, our calibrated projection is guaranteed to have correct frequentist coverage whenever the bounds of the identified set are differentiable in μ.

c) Gafarov et al. (2015) [GMM1] establish the differentiability of the bounds of $\underline{\nu}(\mu), \overline{\nu}(\mu)$ for a class of SVAR models that impose restrictions only on the responses to one structural shock. Based on the differentiability results, they propose a 'delta-method' confidence interval given by the plug-in estimators of the bounds plus/minus r times standard errors. In Appendix 6.A.3 we show that, in large samples, the 'delta-method' procedure in GMM1 is equivalent to a projection region based on a Wald ellipsoid for μ with radius r^2.

6.3 Basic model, main assumptions, and frequentist results

6.3.1 Model

This paper studies the n-dimensional Structural Vector Autoregression with p lags; i.i.d. structural innovations—denoted ε_t—distributed according to F; and unknown $n \times n$ structural matrix B:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \;\; \mathbb{E}_F[\varepsilon_t] = 0_{n\times 1}, \;\; \mathbb{E}_F[\varepsilon_t \varepsilon_t'] \equiv \mathbb{I}_n. \tag{6.3}$$

see Lütkepohl (2007), p. 362.

The *reduced-form parameters* of the SVAR model are defined as the vectorized autoregressive coefficients and the half vectorized covariance matrix of reduced-form residuals:

$$\mu \equiv (\text{vec}(A)', \text{vech}(\Sigma)')' \in \mathbb{R}^d, \quad \text{where} \quad A \equiv (A_1, A_2, \ldots, A_p), \quad \Sigma \equiv BB'.$$

In applied work, these reduced-form parameters are estimated directly from the data using least squares. That is:

$$\widehat{\mu}_T \equiv (\text{vec}(\widehat{A}_T)', \text{vech}(\widehat{\Sigma}_T)')',$$

where

$$\widehat{A}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^{T} Y_t X_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\Big)^{-1}, \quad \widehat{\Sigma}_T \equiv \frac{1}{T}\sum_{t=1}^{T} \widehat{\eta}_t \widehat{\eta}_t',$$

and

$$X_t \equiv (Y_{t-1}', \ldots, Y_{t-p}')', \quad \widehat{\eta}_t \equiv Y_t - \widehat{A}_T X_t.$$

A common formula for the asymptotic variance of $\widehat{\mu}_T$ in stationary models is:

$$\widehat{\Omega}_T \equiv V_T\Big(\frac{1}{T}\sum_{t=1}^{T} \text{vec}\Big([\widehat{\eta}_t X_t', \widehat{\eta}_t \widehat{\eta}_t' - \widehat{\Sigma}_T]\Big)\text{vec}\Big([\widehat{\eta}_t X_t', \widehat{\eta}_t \widehat{\eta}_t' - \widehat{\Sigma}_T]\Big)'V_T'$$

where

$$V_T \equiv \begin{pmatrix} \mathbb{I}_n \otimes \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} & \mathbf{0} \\ \mathbf{0} & L_n \end{pmatrix},$$

and L_n is the matrix of dimension $n(n+1)/2 \times n^2$ such that $\text{vech}(\Sigma) = L_n \text{vec}(\Sigma)$, see Lütkepohl (2007), p. 662 equation A.12.1.

6.3.2 Assumptions for frequentist inference

The SVAR parameters $(A_1, \ldots, A_p, B, F)$ define a probability measure, denoted P, over the data observed by the econometrician. The measure P is assumed to belong to some class $\mathscr{P}$ which we describe in this section.

We state a simple high-level assumption concerning the asymptotic behavior of the $1-\alpha$ Wald confidence ellipsoid for μ, which is defined as:

$$CS_T(1-\alpha;\mu) \equiv \left\{\mu \in \mathbb{R}^d \quad | \quad T(\widehat{\mu}_T-\mu)'\widehat{\Omega}_T^{-1}(\widehat{\mu}_T-\mu) \leq \chi^2_{d,1-\alpha}\right\}.^{7} \tag{6.4}$$

The first assumption requires the *uniform consistency in level* (over the class $\mathscr{P}$) of the Wald confidence set for the reduced-form parameters. This is:

Assumption 1. $\liminf_{T\to\infty} \inf_{P\in\mathscr{P}} P\Big(\mu(P) \in CS_T(1-\alpha;\mu)\Big) \geq 1-\alpha$.

Assumption 1 holds if the class $\mathscr{P}$ under consideration contains only *uniformly* stable VARs where the error distributions, F, have *uniformly* bounded fourth moments.[8] Assumption 1 turns out to be sufficient to conduct frequentist inference on the structural parameters of a set-identified SVAR, defined as follows.

Coefficients of the Structural Impulse-Response Function: Given the autoregressive coefficients $A \equiv (A_1, A_2, \ldots, A_p)$ define, recursively, the nonlinear transformation

$$C_k(A) \equiv \sum_{m=1}^{k} C_{k-m}(A)\, A_m, \quad k \in \mathbb{N},$$

where $C_0 = \mathbb{I}_n$ and $A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116.

Definition (Coefficients of the Structural IRF). The (k,i,j)*-coefficient of the structural impulse-response function* is defined as the scalar parameter:

$$\lambda_{k,i,j}(A,B) \equiv e_i' C_k(A) B e_j,$$

where e_i and e_j denote the i-th and j-th column of the identity matrix $\mathbb{I}_n$.

---

[7] The radius $\chi^2_{d,1-\alpha}$ in equation (6.4) denotes the $1-\alpha$ quantile of a central χ^2 distribution with d degrees of freedom.

[8] A class $\mathscr{P}$ that satisfies Assumption 1 could be written by using a uniform version of the conditions in Lütkepohl (2007), p. 73. This is, there are positive constants c_1, c_2, c_3, c_4 such that:

$$\begin{aligned}\mathscr{P} \quad = \quad & \{(A_1, A_2, \ldots, A_p, B, F) \mid \det(\mathbb{I}_n - A_1 z - \ldots A_p z^p) \notin (-c_1, c_1) \text{ for } z \in \mathbb{C}, |z| \leq 1; \\ & B \text{ is such that } 0 < c_2 < \text{eigmin}(BB') < \text{eigmax}(BB') < c_3, \\ & \text{and } \mathbb{E}_F[|\varepsilon_{n_1,t}\varepsilon_{n_2,t}\varepsilon_{n_3,t}\varepsilon_{n_4,t}|] < c_4 \text{ for all } t, \text{ and } n_1, n_2, n_3, n_4 \in \{1, \ldots n\}, \\ & \text{and } \mathbb{E}_F[\varepsilon_t] = \mathbf{0_{n\times 1}},\ \mathbb{E}_F[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t}'] = \mathbb{I}_{\mathbf{n}} \quad \}.\end{aligned}$$

Other possible definitions of $\mathscr{P}$ can be given by generalizing Theorem 3.5 in Chen and Fang (2015) to either multivariate linear processes with i.i.d. innovations or to martingale difference sequences.

6.3.3 Main result concerning frequentist inference

In this section we show that, under Assumption 1, it is possible to 'project' the $1-\alpha$ Wald confidence set for μ to conduct frequentist inference about the coefficients of the structural impulse-response function and the function itself in set-identified models.

Set-Identified SVARs: As mentioned in the introduction, the SVAR allows researchers to transform the reduced-form parameters, $\mu \equiv (\text{vec}(A)', \text{vech}(\Sigma)')'$, into the structural parameters of interest, $\lambda_{k,i,j}(A,B)$. The parameter μ determines a unique value of A; however, several values of B are compatible with Σ (any B such that $BB' = \Sigma$). This indeterminacy of B implies there are multiple values of $\lambda_{k,i,j}(A,B)$ that are compatible with one value of μ.

The Identified Set and its Bounds: It is common in applied macroeconomic work to impose restrictions on the matrix $B \in \mathbb{R}^{n\times n}$ in order to limit the range of a structural coefficient of interest, $\lambda_{k,i,j}$ (taking μ as given). Mathematically, a set of restrictions on B—that we denote as $\mathscr{R}(\mu)$—can be interpreted as a subset of $\mathbb{R}^{n\times n}$. This leads to the following definition:

Definition (**Identified Set and its bounds**)**.** Fix a vector of reduced-form parameters, μ, and a set of restrictions $\mathscr{R}(\mu)$ on B.

1. The *identified set* for the structural parameter $\lambda_{k,i,j}(A,B)$ is defined as:

$$\mathscr{I}^{\mathscr{R}}_{k,i,j}(\mu) \equiv \Big\{ v \in \mathbb{R} \,\Big|\, v = \lambda_{k,i,j}(A,B),\ BB' = \Sigma,\ \text{and } B \in \mathscr{R}(\mu) \Big\}.$$

2. The *upper bound* of the identified set $\overline{v}_{k,i,j}(\mu)$ is defined as the value function of the program:

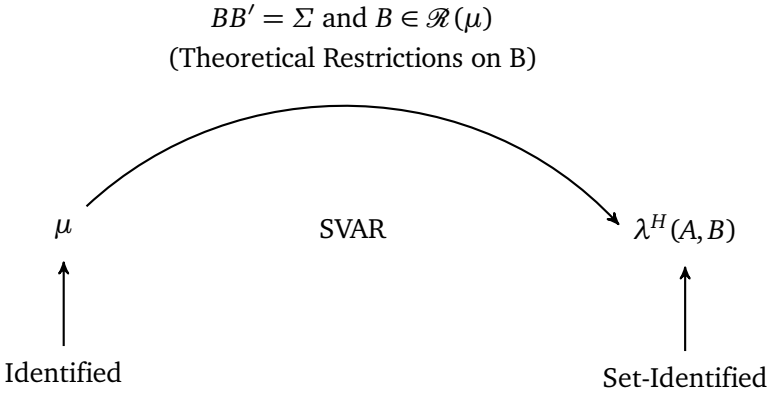
$$\overline{v}_{k,i,j}(\mu) \equiv \sup_{B\in\mathbb{R}^{n\times n}} e_i' C_k(A) B e_j, \quad \text{s.t.} \quad BB' = \Sigma,\ \text{and } B \in \mathscr{R}(\mu). \tag{6.5}$$

   The *lower bound*, $\underline{v}_{k,i,j}(\mu)$, is defined analogously.

3. Consider any collection $\lambda^H \equiv \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ of structural coefficients and let its identified set be given by:

$$\mathscr{I}^{\mathscr{R}}_H(\mu) \equiv \Big\{ (v_1,\dots,v_H) \in \mathbb{R}^H \,\Big|\, v_h = \lambda_{k_h,i_h,j_h}(A,B),\ BB' = \Sigma,\ \text{and } B \in \mathscr{R}(\mu) \Big\}.$$

The main elements in the previous definition can be illustrated as follows:



$$\mathscr{I}_H^{\mathscr{R}}(\mu) \subseteq \mathbb{R}^H : \text{ The identified-Set for } \lambda^H(A,B).$$

Table 6.1 presents a list of the most common restrictions, $\mathscr{R}(\mu)$, used in SVAR analysis (all of which can be handled by our frequentist approach described below).

Projection Approach: A key feature of set-identified SVARs is that the bounds of the identified set depend on a finite-dimensional parameter. 'Projecting' down the $1-\alpha$ Wald ellipsoid for μ seems a natural approach to conduct inference on the structural impulse response function. The first result in this paper establishes the frequentist uniform validity of projection inference.

Result 1 (**Frequentist Coverage of Projection Inference for λ^H**). *Consider the projection region for the collection of structural coefficients $\lambda^H \equiv \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ given by:*

$$CS_T(1-\alpha,\lambda^H) \;\equiv\; CS_T(1-\alpha,\lambda_{k_1,i_1,j_1}) \times \ldots \times CS_T(1-\alpha,\lambda_{k_H,i_H,j_H}) \;\subseteq\; \mathbb{R}^H, \tag{6.6}$$

where

$$CS_T(1-\alpha;\lambda_{k,i,j}) \;\equiv\; \Big[\inf_{\mu\in CS_T(1-\alpha,\mu)} \underline{v}_{k,i,j}(\mu)\;,\; \sup_{\mu\in CS_T(1-\alpha,\mu)} \overline{v}_{k,i,j}(\mu)\Big], \tag{6.7}$$

and $CS_T(1-\alpha;\mu)$ is the $1-\alpha$ Wald confidence ellipsoid for μ. If the class of data generating processes $\mathscr{P}$ satisfies Assumption 1, then:

$$\liminf_{T\to\infty} \inf_{P\in\mathscr{P}} \inf_{\lambda^H\in\mathscr{I}_H^R(\mu(P))} P\Big(\lambda^H \in CS_T(1-\alpha;\lambda^H)\Big) \geq 1-\alpha.$$

That is, the projected confidence interval in (6.6) covers the vector of structural coefficients λ^H with probability at least $1-\alpha$, uniformly over the class $\mathscr{P}$.

Proof. The proof of Result 1 uses a standard and conceptually straightforward projection argument. Take an element $P \in \mathscr{P}$ and let $\lambda^H \in \mathbb{R}^H$ be any given element of the identified set $\mathscr{I}_H^{\mathscr{P}}(\mu(P))$. Note that:

$$
\begin{aligned}
& P\Big(\lambda^H \in CS_T(1-\alpha;\lambda^H)\Big) \\
= \;& P\Big((\lambda_{k_1,i_1,j_1},\ldots,\lambda_{k_H,i_H,j_H}) \\
& \qquad \in \; CS_T(1-\alpha;\lambda_{k_1,i_1,j_1}) \times \ldots \times CS_T(1-\alpha;\lambda_{k_H,i_H,j_H})\Big) \\
& \Big(\text{by definition of our confidence interval for } \lambda^H\Big) \\
\geq \;& P\Big([\underline{\nu}_{k_h,i_h,j_h}(\mu(P)) \;,\; \overline{\nu}_{k_h,i_h,j_h}(\mu(P))] \\
& \qquad \subseteq \Big[\inf_{\mu\in \mathrm{CS}_T(1-\alpha,\mu)} \underline{\nu}_{k_h,i_h,j_h}(\mu) \;,\; \sup_{\mu\in \mathrm{CS}_T(1-\alpha,\mu)} \overline{\nu}_{k_h,i_h,j_h}(\mu)\Big] \quad \forall h \;=\; 1,\ldots,H\Big), \\
& \Big(\text{since } \lambda_{k_h,i_h,j_h} \;\in\; [\underline{\nu}_{k_h,i_h,j_h}(\mu(P),\overline{\nu}_{k_h,i_h,j_h}(\mu(P))]\Big) \\
\geq \;& P\Big(\mu(P) \;\in\; CS_T(1-\alpha;\mu)\Big).
\end{aligned}
$$

The desired result follows directly from Assumption 1. This shows that the projection region for λ^H is uniformly consistent in level. □

Table 6.1. Common restrictions used in set-identified SVARs

Restrictions	**Description**	**Notation**	**Examples**
Short-run	Exclusion restrictions imposed on B or B'^{-1}	$e_i'Be_j = 0$ or $e_i'B'^{-1}e_j = 0$ (Note that $B'^{-1} = \Sigma^{-1}B$)	Sims (1980) Christiano et al. (1996) Rubio-Ramirez et al. (2015)
Long-run	A zero constraint on the long-run impact matrix	$e_i'(\mathbb{I}_n - A_1 - A_2 - \ldots A_p)^{-1}Be_j = 0$	Blanchard and Quah (1989)
Sign	Sign restrictions on IRFs	$e_i'C_k(A)Be_j \geq, \leq 0$	Uhlig (2005) Mountford and Uhlig (2009)
Elasticity Bounds	Bounds on the elasticity of a variable	$\frac{e_i'C_k(A)Be_j}{e_{\tilde{i}}'C_k(A)Be_j} \geq, \leq c, \ \tilde{i} \neq i$	Kilian and Murphy (2012b)
Shape Constraints	Shape constraints on IRFs (e.g., monotonicity)	e.g., $e_i'C_k(A)Be_j \leq e_i'C_{k+1}(A)Be_j$	Scholl and Uhlig (2008)
Other	Sign Restrictions on Long-Run Impacts Noncontemporaneous Zero Restrictions General equalities/inequalities on B	$e_i'(\mathbb{I}_n - A_1 - A_2 - \ldots A_p)^{-1}Be_j \geq, \leq 0$ $e_i'C_k(A)Be_j = 0$ $g(B, \mu) \geq, \leq, = 0$	

The projection approach can handle SVAR models with any of the restrictions described on this table (imposed on one or multiple shocks).
(i denotes the variable, j denotes the shock, and k the horizon)

Remark 1: The idea of 'projecting' a confidence set for a parameter μ to conduct inference about a lower dimensional parameter λ has been used extensively in econometrics; see Scheffé (1953), Dufour (1990), and Dufour and Taamouti (2005, 2007) for some examples. In addition to its conceptual simplicity, one advantage of the projection approach is that its validity does not require special conditions on the identifying restrictions that can be imposed by practitioners.[9]

Remark 2: The problem of conducting inference on the whole impulse-response function (and not only on one specific coefficient) has been a topic of recent interest, both from the Bayesian and frequentist perspective.

For Bayesian set-identified SVARs with only sign restrictions, Inoue and Kilian (2013) report the vector of structural impulse-response coefficients with highest posterior density (based on a prior on reduced-form parameters and a uniform prior on rotation matrices). They propose a Bayesian credible set (represented by *shotgun* plots) that characterizes the joint uncertainty about a given collection of structural impulse-response coefficients.

For frequentist point-identified SVARs, Inoue and Kilian (2016) propose a bootstrap procedure that allows the construction of asymptotically valid confidence regions for any subset of structural impulse responses. To the best of our knowledge, our projection approach is the first frequentist procedure for set-identified SVARs that provide confidence regions for any collection of structural coefficients (response of different variables, to different shocks, over different horizons).

It is important to note that Uhlig (2005)'s approach to conduct inference on set-identified SVARs does not provide credible sets for vectors of the structural parameters. The same is true for the Bayesian approaches described in the recent work of Arias et al. (2014) and Baumeister and Hamilton (2015a), as well as the approaches of Moon et al. (2013) and Giacomini and Kitagawa (2015).

Remark 3: A common concern in set-identified models is whether the suggested inference approach is valid only for the identified parameter, λ^H, or also for its identified set $\mathscr{I}_H^{\mathscr{R}}(\mu)$. Note that the second to last inequality in the proof of Result 2 imply that our projection region covers the identified set of any vector of coefficients λ^H.

---

[9] For instance, we do not need to assume that $\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\cdot)$ are continuous or differentiable functions of the reduced-form parameters.

6.4 Robust Bayesian credibility

This section analyzes the *robust credibility* of projection as the sample size grows large.

Bayesian Set-up: In a Bayesian SVAR the distribution of the structural innovations is fixed and treated as a known object. A common choice—which we follow in this section—is to assume that $F \sim \mathcal{N}_n(0, \mathbb{I}_n)$. We discuss how to relax this restriction after stating Assumption 5.

Let P^* denote some prior for the structural parameters $(A_1, \ldots, A_p, B)$ and let $\lambda^H(A, B) \in \mathbb{R}^H$ denote the vector of structural coefficients of interest. For a given square root of $\Sigma \equiv BB'$ define the 'rotation' matrix $Q \equiv \Sigma^{-1/2}B$. It is well known that a prior P^* can be written as $(P^*_\mu, P^*_{Q|\mu})$, where P^*_μ is a prior on the reduced-form parameters, and $P^*_{Q|\mu}$ is a prior on the rotation matrix, conditional on μ.[10] Following this notation, let $\mathscr{P}(P^*_\mu)$ denote the class of prior distributions such that $\mu \sim P^*_\mu$.

We are interested in characterizing the smallest posterior probability that the set $\text{CS}_T(1-\alpha; \lambda^H)$ could receive, allowing the researcher to vary the prior for Q:

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \ \in \ CS_T(1-\alpha; \lambda^H) \ \Big| \ Y_1, \ldots, Y_T\Big). \tag{6.8}$$

The event of interest is whether the structural coefficients $\lambda^H(A, B)$ (treated as random variables in the Bayesian Set-up) belong to the projection region, after conditioning on the data. This event would typically be referred to as the credibility of $\text{CS}_T(1-\alpha; \lambda^H)$ (see Berger (1985), p. 140). We would like to find the smallest credibility of projection when different priors over Q are considered as in the pioneering work of Kitagawa (2012). We follow the recent work of Giacomini and Kitagawa (2015) and refer to (6.8) as the *robust Bayesian credibility* of the set $\text{CS}_T(1-\alpha, \lambda^H)$.

Let $f(Y_1, ..., Y_T|\mu)$ denote the Gaussian statistical model for the data (which depends solely on the reduced-form parameters) and let $o_p(1; Y_1, \ldots Y_T|\mu)$ denote a random variable such that $\lim_{T\to\infty} P_{Y_1,...,Y_T|\mu}(|o_p(1; Y_1, \ldots Y_T|\mu)| > \epsilon) = 0$ for all $\epsilon > 0$ when the distribution of the data is conditioned on μ.

Main Assumption for Bayesians: Robust credibility can be viewed as a random variable (as it depends on $Y_1, \ldots, Y_T$). We use the following high-level assumption to characterize its asymptotic behavior:

---

[10] Arias et al. (2014) refer to this parameterization of the SVAR model as the orthogonal reduced-form.

Assumption 5. Whenever $Y_1, \ldots, Y_T \sim f(Y_1, \ldots, Y_T|\mu_0)$, the prior P^* is such that:

$$P^*\Big(\mu(A,B) \in CS_T(1-\alpha;\mu) \,\Big|\, Y_1, \ldots, Y_T\Big) = 1-\alpha+o_p(1; Y_1, \ldots, Y_T|\mu_0).$$

Assumption 5 requires the prior over the reduced-form parameters (and the statistical model) to be regular enough to guarantee that the asymptotic Bayesian credibility of the $1-\alpha$ Wald ellipsoid converges in probability to $1-\alpha$. Thus, our high-level assumption is implied by the Bernstein von-Mises Theorem (DasGupta (2008), p. 291) for the reduced-form parameter μ.

Since the Gaussian statistical model $f(Y_1, \ldots Y_T|\mu_0)$ can be shown to be Locally Asymptotically Normal (LAN) whenever A_0 is stable and Σ_0 has full rank, Theorem 1 and 2 in Ghosal et al. (1995) (GGS) imply that Assumption 5 will be satisfied whenever P^*_μ has a continuous density at μ_0 with polynomial majorants.[11] In fact, the same theorems could be used to establish Assumption 5 for non-Gaussian SVARs that are LAN and satisfy the regularity conditions of Ibragimov and Has' minskii (2013) (IH), as long as $\text{CS}_T(1-\alpha;\mu)$ is centered at the Maximum Likelihood estimator of μ and $\widehat{\Omega}_T$ is replaced by the model's inverse information matrix. An alternative approach to establish Assumption 5 using a different set of primitive conditions can be found in the recent work of Connault (2016).

We now establish the robust Bayesian credibility of projection as $T \to \infty$.

Result 2. *[**Asymptotic Robust Bayesian Credibility of Projection**] Suppose that the prior P^* for (A,B) satisfies Assumption 5 at μ_0. Then:*

$$\inf_{P^* \in \mathscr{P}^*(\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big) \geq 1-\alpha+o_p(1; Y_1, \ldots Y_T|\mu_0).$$

Proof. Note that:

$$P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \,\Big|\, Y_1, \ldots Y_T\Big)$$

$$= P^*\Big(\lambda_{k_h,i_h,j_h}(A,B) \in CS_T(1-\alpha;\lambda_{k_h,i_h,j_h}) \;\forall\, h = 1\ldots,H \,\Big|\, Y_1, \ldots, Y_T\Big)$$
$$\Big(\text{by definition of the projection region for } \lambda^H\Big)$$

[11] In Appendix 6.A.1.1 we verify an 'almost sure' version of Assumption 5 for a Gaussian SVAR for the Normal-Wishart priors suggested in Uhlig (1994) and Uhlig (2005) and a confidence set for μ based on the formula for the asymptotic variance $\widehat{\Omega}_T$ that obtains in the Gaussian model [Lütkepohl (2007) p. 93].

$$
\begin{aligned}
&\geq P^*\Big([\underline{\nu}_{k_h,i_h,j_h}(\mu(A,B)) \ , \ \overline{\nu}_{k_h,i_h,j_h}(\mu(A,B))] \ \in \ CS_T(1-\alpha;\lambda_{k_h,i_h,j_h}) \\
&\qquad \forall \ h \ = \ 1\ldots,H \ \Big| \ Y_1,\ldots,Y_T \Big), \\
&\quad \Big(\text{since } \lambda_{k_h,i_h,j_h}(A,B) \ \in \ [\underline{\nu}_{k_h,i_h,j_h}(\mu(A,B)), \overline{\nu}_{k_h,i_h,j_h}(\mu(A,B))] \text{ for any } A,B\Big) \\
&\geq P^*\Big(\mu(A,B) \ \in \ CS_T(1-\alpha;\mu) \ \Big| \ Y_1,\ldots,Y_T\Big).
\end{aligned}
$$

This implies that in any finite sample:

$$
\inf_{P^*\in\mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \ \Big| \ Y_1,\ldots,Y_T\Big)
$$

is at least as large as

$$
P^*\Big(\mu(A,B) \in CS_T(1-\alpha;\mu) \ \Big| \ Y_1,\ldots,Y_T\Big).
$$

Assumption 5 gives the desired result. □

This means that—given any prior that satisfies Assumption 5—our projection region can be interpreted, in large samples, as a robust $1-\alpha$ credible region for the impulse-response function and its coefficients.

6.5 Calibrated projection for a Robust Bayesian

The projection approach generates *conservative* regions for both a frequentist and a robust Bayesian. For a frequentist, the large-sample coverage may be strictly above the desired confidence level. For a robust Bayesian, the asymptotic robust credibility of the nominal $1-\alpha$ projection region may be strictly above $1-\alpha$.

This section applies the approach in Kaido et al. (2016) to eliminate the excess of robust Bayesian credibility in a computationally tractable way. We focus on calibrating the robust credibility of our projection region to be exactly equal to $1-\alpha$ (either in a finite sample for a given prior on μ, or in large samples for a large class of priors on μ).[12]

Given a vector $\Lambda^H = \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ of structural coefficients of interest and its corresponding nominal $1-\alpha$ projection region, the calibration exercise is based on the following result:

[12] We also discuss the calibration of projection in SVARs from the frequentist perspective (see Appendix 6.A.2). We argue that the computational feasibility of the frequentist calibration might be compromised when μ is of large dimension.

Result 3. *(**Calibration of Robust Credibility**) Let P^*_μ denote a prior for the reduced-form parameters. Suppose there is a nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$ such that for every data realization:*

$$P^*_\mu\Big(\times_{h=1}^{H}[\underline{\nu}_{k_h,i_h,j_h}(\mu),\overline{\nu}_{k_h,i_h,j_h}(\mu)]\subseteq CS_T(1-\alpha^*(Y_1,\ldots,Y_T),\lambda^H)|Y_1,\ldots,Y_T\Big)$$

equals $1-\alpha$. Then, for every data realization:

$$\inf_{P^*\in\mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B)\in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H)\;\Big|\;Y_1,\ldots,Y_T\Big)=1-\alpha.$$

Proof. See Appendix 6.A.1.2. □

This means that in order to calibrate the robust credibility of projection, it is sufficient to choose $1-\alpha^*(Y_1,\ldots,Y_T)$ to guarantee that exactly α% of the bounds of the identified set for the different structural coefficients in λ^H fall outside the projection region.

Calibration Algorithm: The calibration algorithm we propose consists in finding a nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$ such that the posterior probability of the event:

$$[\underline{\nu}_{k_1,i_1,j_1}(\mu),\overline{\nu}_{k_1,i_1,j_1}(\mu)]\times\ldots\times[\underline{\nu}_{k_h,i_h,j_h}(\mu),\overline{\nu}_{k_h,i_h,j_h}(\mu)]\subseteq \mathrm{CS}_T(1-\alpha^*,\lambda^H)$$

equals $1-\alpha$ under the posterior distribution associated to the prior P^*_μ or under a suitable large-sample approximation for the posterior such as $\mu|Y_1,\ldots Y_T\sim\mathscr{N}_d(\widehat{\mu}_T,\widehat{\Omega}_T/T)$.[13]

The calibration algorithm is the following:

1. Generate M draws (for example, $M=1,000$) from the posterior of the reduced-form parameters. If desired, one could use the large-sample approximation of the posterior given by:
$$\mu^*_m\sim\mathscr{N}_d(\widehat{\mu}_T,\widehat{\Omega}_T/T).$$

2. Let $\lambda^H=\{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ denote the structural coefficients of interest. For each $h=1,\ldots H$ and for each $m=1,\ldots M$ evaluate:
$$[\underline{\nu}_{k_h,i_h,j_h}(\mu^*_m),\overline{\nu}_{k_h,i_h,j_h}(\mu^*_m)],$$
as defined in equation (6.5). We provide Matlab code to evaluate these bounds.

[13] The Gaussian approximation for the posterior will eliminate projection bias asymptotically provided a Berstein von-Mises Theorem for μ holds. We establish this result in Appendix 6.A.1.4.

3. Fix an element α_s on the interval $(\alpha, 1)$. Set a tolerance level $\eta > 0$.

4. For each $m = 1, \dots M$ generate the indicator function z_m that takes the value of 0 whenever there exists an index $h \in \{1, \dots H\}$ such that:

$$[\underline{v}_{k_h,i_h,j_h}(\mu^*_m), \overline{v}_{k_h,i_h,j_h}(\mu^*_m)] \notin \text{CS}_T(1-\alpha_s, \lambda_{k_h,i_h,j_h}).$$

   The projection region $\text{CS}_T(1-\alpha_s, \lambda_{k_h,i_h,j_h})$ is defined in equation (6.7) in Result 3 and implemented using the SQP/IP algorithm that will described in the next section (Section 6.6).

5. Compute the robust credibility of the nominal $1-\alpha_s$ projection as:

$$RC_T(\alpha_s) = \frac{1}{M}\sum_{m=1}^{M} z_m.$$

   If such quantity is in the interval $[1-\alpha-\eta, 1-\alpha+\eta]$ stop the algorithm. If $\text{RC}_T(\alpha_s)$ is strictly above $1-\alpha+\eta$, go back to Step 3 and choose a larger value of α_s. If $\text{RC}_T(\alpha_s)$ is strictly below $1-\alpha-\eta$ go back to Step 3 and choose a smaller value of α_s.

We now show that whenever the bounds of the identified set for each λ_h are differentiable, our calibration algorithm also removes the excess of frequentist coverage.

Result 4 (**Robust Bayes Calibration and the Frequentist Coverage of the Identified Set**). *Suppose that for each $h = 1, \dots H$ the bounds of the identified set $\underline{v}_h(\mu)$ and $\overline{v}_h(\mu)$ are differentiable at μ_0. Suppose in addition that at μ_0:*

1. $\sqrt{T}(\widehat{\mu} - \mu_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$,
2. $\widehat{\Omega}_T \xrightarrow{p} \Omega$, *where Ω is positive definite,*
3. *The prior for the reduced-form parameters used in the calibration satisfies the Bernstein von-Mises Theorem in Ghosal et al. (1995):*

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \left| P^*\left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1, \dots, Y_T\right) - \mathbb{P}(Z \in B)\right| \xrightarrow{p} 0,$$

   where $Z \sim \mathcal{N}_d(\mathbf{0}, \Omega)$, and $\mathcal{B}(\mathbb{R}^d)$ is the set of all Borel measurable sets in $\mathbb{R}^d$.

Then:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda_h),\ \forall h = 1,\ldots H\right) \to 1-\alpha.$$

Proof. See Appendix 6.A.1.3. The heuristic argument behind this result is the following. We show that the differentiability of $\underline{v}_h$ and $\overline{v}_h$ at μ_0 implies that:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \mathrm{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda_h),\ \forall h = 1,\ldots H\right)$$

is approximately the same as:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \Big[\underline{v}_h(\widehat{\mu}_T) - \frac{r_T^*\underline{\sigma}_h(\mu_0)}{\sqrt{T}}\,,\ \overline{v}_h(\widehat{\mu}_T) + \frac{r_T^*\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\Big],\ \forall h \leq H\right).$$

where r_T^* be the radius that calibrates robust Bayesian credibility. The Bernstein-von Mises Theorem implies that such probability is approximately the same as:

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \Big[\underline{v}_h(\widehat{\mu}_T) - \frac{r_T^*\underline{\sigma}_h(\mu_0)}{\sqrt{T}}\,,\ \overline{v}_h(\widehat{\mu}_T) + \frac{r_T^*\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\Big],\ \forall h \leq H\right),$$

which, by the calibration of robust Bayesian credibility is approximately $1-\alpha$. □

6.6 Implementation of baseline and calibrated projection

6.6.1 Projection as a mathematical optimization problem

This subsection discusses the implementation of the baseline projection region:

$$\mathrm{CS}_T(1-\alpha;\lambda_{k,i,j}) \equiv \Big[\inf_{\mu\in\mathrm{CS}_T(1-\alpha,\mu)} \underline{v}_{k,i,j}(\mu)\,,\ \sup_{\mu\in\mathrm{CS}_T(1-\alpha,\mu)} \overline{v}_{k,i,j}(\mu)\Big].$$

We note that both the upper bound and lower bound of this confidence interval can be thought of as solutions to a pair of 'nested' optimization problems.

The first optimization problem—that we refer to as the *inner* optimization—solves for $\overline{v}_{k,i,j}(\mu)$ and $\underline{v}_{k,i,j}(\mu)$. These functions correspond to the largest and smallest value of the structural impulse response $\lambda_{k,i,j}$ given a set of restrictions and a vector of reduced-form parameters μ.

The second optimization problem—that we refer to as *outer* optimization—solves for the maximum value of $\overline{v}_{k,i,j}(\cdot)$ and the minimum value of $\underline{v}_{k,i,j}(\cdot)$ over the $(1-\alpha)$ Wald Confidence ellipsoid, $\text{CS}_T(1-\alpha,\mu)$.

Implementation: Our proposal is to combine the inner and outer problem into a *single* mathematical program that gives the bounds of the projection confidence interval directly. The upper bound can be found by solving:

$$\sup_{A,\Sigma,B} e_i' C_k(A) B e_j \quad \text{subject to} \quad BB' = \Sigma, \quad B \in \mathscr{R}(\mu), \text{ and} \tag{6.9}$$
$$T(\widehat{\mu}_T - \mu(A,\Sigma))' \widehat{\Omega}_T^{-1} (\widehat{\mu}_T - \mu(A,\Sigma)) \leq \chi^2_{d,1-\alpha}.$$

The lower bound of the projection confidence interval can be found analogously. Importantly, the simple reformulation in (6.9) allows us to base the implementation of our projection region upon state-of-the-art solution algorithms for optimization problems. Our suggestion is to use a simple SQP/IP algorithm.

6.6.2 Solution algorithms for baseline projection

The nature of the optimization problem: The nonlinear mathematical program in (6.9) has two challenging features. On the one hand, the optimization problem is non-convex; this complicates the task of finding a global minimum with algorithms designed to detect local optima. On the other hand, the number of optimization arguments and constraints increases quadratically in the dimension of the SVAR; this compromises the feasibility of some optimization routines designed to detect global optima (for example, brute-force grid search on $\text{CS}_T(1-\alpha,\mu)$ to optimize $\underline{v}_{k,i,j}(\mu)$ and $\overline{v}_{k,i,j}(\mu)$).

Our Approach: Taking these two features into consideration, we first implemented projection by running a local optimization algorithm followed by a global algorithm that used the local solution as an input. The algorithms and the functions used to implement the projection confidence interval are described below. In the application analyzed in this paper, the global stage of the algorithm did not have any impact on the local solution. We thus suggest researchers to implement our approach using only the SQP/IP routine described below.

Local Algorithms: Although no standard classification exists for local optimization algorithms, the most common procedures are often grouped as follows: penalty and Augmented Lagrangian Methods; Sequential Quadratic Programming (SQP); and Interior Point Methods (IP); see p. 422 of Nocedal and Wright (2006) for more details.

Within this class of algorithms, we focus on the IP and SQP algorithms, both of which are considered as the "*most powerful algorithms for large-scale nonlinear programming*", Nocedal and Wright (2006), p. 563.[14] Conveniently, IP and SQP are included in Matlab®'s `fmincon` function, which comes with the Optimization toolbox. We run the SQP algorithm—which is usually faster than IP—and in case it does not find a solution, we switch to IP, an algorithm which we denote by *SQP/IP*.

Global Algorithms: IP and SQP are well adjusted to handle various degeneracy problems in order to find a local minimum for large-scale non-convex problems. There is now a large body of literature on global optimization strategies; see Horst and Pardalos (1995) and Romeijn and Pardalos (2013). Popular global optimization algorithms include adaptive stochastic search; branch and bound methods; homotopy methods; Genetic algorithms (GA); simulated annealing and two-phase algorithms such as *MultiStart* and *GlobalSearch*.[15]We focus on the two-phase algorithms MultiStart, GlobalSearch and on the genetic algorithm available in Matlab.[16]

6.6.3 Implementing baseline projection in an example

As an example, we consider the demand-supply SVAR model studied in Section 5 of Baumeister and Hamilton (2015a) [henceforth, BH]. We fit a 6-lag VAR to U.S. data on growth rates of real labor compensation, Δw_t, and total employment, Δn_t, from 1970:Q1 to 2014:Q2.[17]

Using our notation, the demand-supply SVAR can be written as:

$$\begin{pmatrix} \Delta w_t \\ \Delta \eta_t \end{pmatrix} = A_1 \begin{pmatrix} \Delta w_{t-1} \\ \Delta \eta_{t-1} \end{pmatrix} + \ldots + A_6 \begin{pmatrix} \Delta w_{t-6} \\ \Delta \eta_{t-6} \end{pmatrix} + B \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \end{pmatrix},$$

---

[14] Furthermore, these algorithms exploit the existence of second-order derivatives which are well-defined in our problem.

[15] For a more detailed list and classification of global methods see p. 519 of Chapter 15 in Romeijn and Pardalos (2013). For a description of two-phase algorithms see Chapter 12 in Romeijn and Pardalos (2013).

[16] Genetic algorithms are a well developed field of computing and they have been used in many applications; see the introduction to Chapter 9 in Romeijn and Pardalos (2013). A very interesting application in economics that motivated our focus on GA is given in Qu and Tkachenko (2015).

[17] Our selection is based on the fact that 6 is the smallest number of lags such that CS(68%; μ) does not contain unstable VAR coefficients and non-invertible reduced-form covariance matrices. 68% confidence sets correspond to a single standard deviation and are frequently used in applied macroeconomic research. The Bayes Information Criteria and the Information Criteria both select less than six lags.

BH set-identify an expansionary demand and supply shock by means of the following sign restrictions:

$$B \equiv \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \quad \text{satisfies} \quad \begin{bmatrix} + & - \\ + & + \end{bmatrix}.$$

The sign restrictions state that a demand shock increases both real labor compensation and total employment, while a supply shock lowers wages but raises employment.

In this model, the short-run wage elasticity of labor supply (identified from a demand shock) is defined as:

$$\alpha \equiv b_2 / b_1$$

Likewise, the short-run wage elasticity of labor demand (identified from a supply shock) is defined as:

$$\beta \equiv b_4 / b_3$$

Finally, the long-run impact of a demand shock over employment is given by:

$$\gamma \equiv e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1.$$

BH impose three additional restrictions. The first two of them are elasticity bounds motivated by the findings of different empirical studies. Hamermesh (1996), Akerlof and Dickens (2007), Lichter et al. (2014) provide bounds on the wage elasticity of labor demand. Chetty et al. (2011), Reichling and Whalen (2012) provide bounds on the wage elasticity of labor supply. The third and final restriction arises from imposing lower and upper bounds on the long-run impact of a demand shock on employment.

BH incorporate the restrictions in the form of priors on the structural parameters, but we treat the constraints as additional sign restrictions. Let t_ν denote the standard t distribution with ν degrees of freedom. The following table summarizes the way in which BH incorporate prior information:

Thus, summarizing, our version of the BH model has 10 sign restrictions:

Table 6.2. Additional identifying restrictions

Restrictions	**Motivation**	**BH**	**This paper**
Bounds on α	Empirical studies report $\alpha \in [.27, 2]$	$\alpha \sim \max\{.6 + .6t_3, 0\}$	$.27 \leq \alpha \leq 2$
Bounds on β	Empirical studies $\beta \in [-2.5, -.15]$	$\beta \sim \min\{-.6 + .6t_3, 0\}$	$-2.5 \leq \beta \leq -.15$
Bounds on γ	$\gamma = 0$ is too strong	$\gamma \sim \mathcal{N}(0, V)$	$-2V \leq \gamma \leq 2V$

$$
\begin{aligned}
\text{Demand and Supply Shocks:} \quad &: \quad b_1 \geq 0, b_2 \geq 0, -b_3 \geq 0, b_4 \geq 0, \\
\text{Elasticity Bounds} \quad &: \quad 2b_1 - b_2 \geq 0, b_2 - .27b_1 \geq 0, \\
&\qquad\quad b_4 + .15b_3 \geq 0, -2.5b_3 - b_4 \geq 0, \\
\text{Long-Run} \quad &: \quad e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1 + 2V \geq 0, \\
- \quad &\quad e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1 + 2V \geq 0,
\end{aligned}
$$

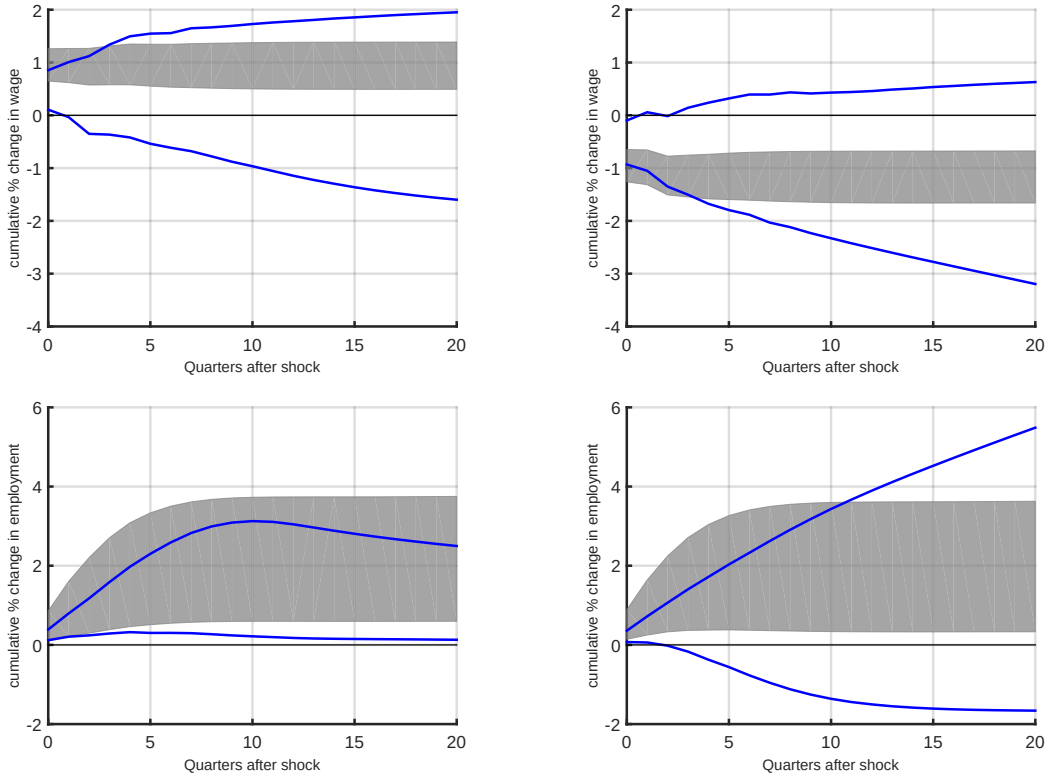
where the parameter V is allowed to take the values $\{.01, .1, 1\}$ as in p. 1992 of BH.

6.6.4 Results of the implementation of baseline projection

Using our SQP/IP local solution algorithm, we compute the 68% projection confidence intervals for the cumulative response of wages and employment to the structural shocks in the model (20 consecutive quarters and setting $V = 1$). In addition to the projection region, we compute the 68% Bayesian credible set following the implementation in both Uhlig (2005) and BH.

Figure 6.1 shows the projection region as solid blue line and the standard Bayesian credible set (based on BH priors) as a grey-shaded area.

Figure 6.1. 68% projection region and 68% credible set.
(Baumeister and Hamilton (2015a) priors)



(a) Expansionary Demand Shock

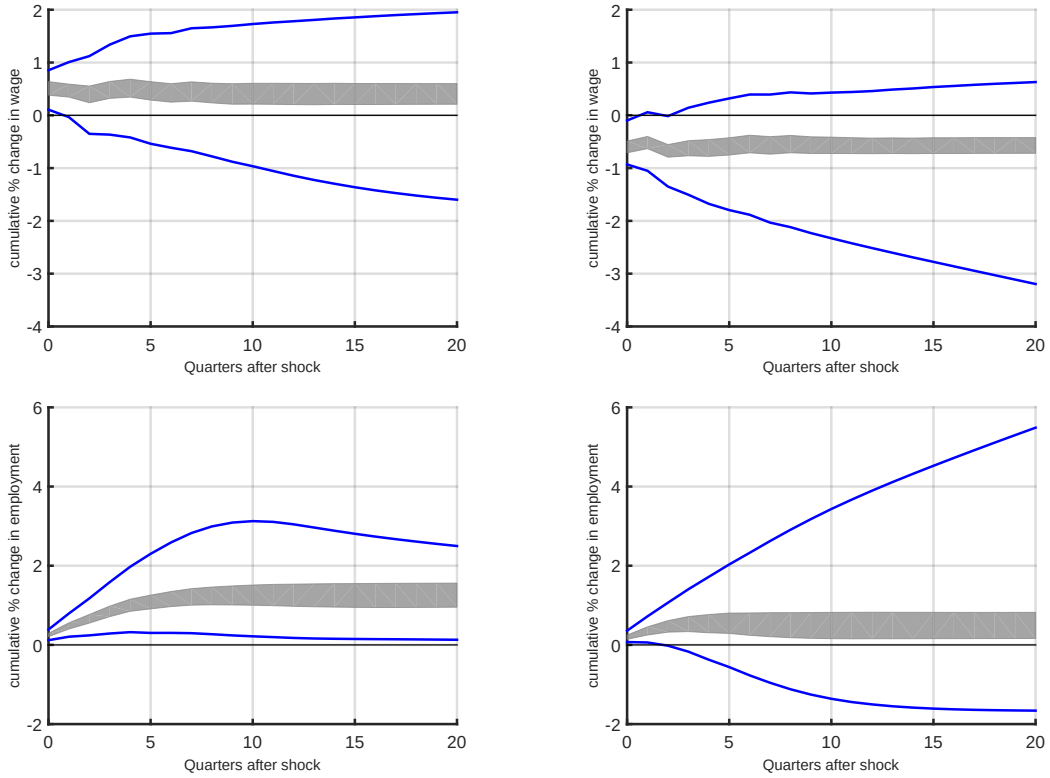
(b) Expansionary Supply Shock

(Solid, Blue Line) 68% Projection Region; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Baumeister and Hamilton (2015a).

Figure 6.2 shows the boundaries of the projection region as solid blue line and the Bayesian credible set based on Uhlig (2005)'s priors as a grey-shaded area.

Comment about Credible Sets: The 68% credible sets differ substantially depending on the specification of prior beliefs. Such sensitivity is the main motivation for our projection approach. In this example, the length of the credible sets for the cumulative response of employment seems to differ by a factor of at least two. The projection region seems quite large compared to the credible sets. This could be a consequence of either the robustness of projection or its conservativeness. To disentangle these effects, we calibrate projection to guarantee that it has exact robust Bayesian credibility in the next subsection.

Figure 6.2. 68% projection region and 68% credible set.
(Uhlig (2005) priors)



(a) Expansionary Demand Shock (b) Expansionary Supply Shock

(Solid, Blue Line) 68% Projection region; (Shaded, Gray Area) 68% Bayesian Credible Set based on the Nornal-Wishart-Haar priors suggested in Uhlig (2005) and the inequality constraints summarized below Table 6.2. The credible set is implemented following Arias et al. (2014).

Concrete comments regarding computational feasibility: Table 6.3 compares computing time for the projection (which has both a frequentist and a Robust Bayes interpretation) and the standard Bayesian methods.[18] Since the global methods are initialized at the local solution, these procedures take as least as much time as SQP/IP. Among the three global methods considered, the Genetic Algorithm takes the longest. Brute-force grid search (which refers to grid search on $CS_T(1-\alpha,\mu)$ to optimize $\underline{\nu}_{k,i,j}(\mu)$ and $\overline{\nu}_{k,i,j}(\mu)$) with only

[18] To get a fair sense of the computational cost, none of the global algorithms were parallelized.

1,000 draws from $\mu \in \mathbb{R}^{27}$ takes about 6 times longer than the baseline SQP/IP and generates substantially smaller bounds (see Appendix 6.A.4.2).[19]

Table 6.3. Computational time in seconds

Algorithm	Details	Time
SQP/IP		734
SQP/IP + MultiStart	100 initial points	33,314
SQP/IP + GlobalSearch	100 trial points (20 in Stage 1)	1,359
Genetic Algorithm	population of 100, 500 generations	76,863
Grid Search on $CS_T(1-\alpha, \mu)$	1,000 draws from μ	4,548
Bayesian, BH	1,000,000 Metropolis-Hastings draws	3,992
Bayesian, Uhlig	100,000 accepted posterior draws	2,338

Notes: Laptop @2.4GHz IntelCore i7.

Comments Regarding Local and Global algorithms: Figure 6.6 in Appendix 6.A.4.1 compares the bounds of the projection confidence interval for the first four algorithms listed in Table 6.3. For this application, it seems that none of the global algorithms improve on the local solution obtained from SQP/IP.[20]

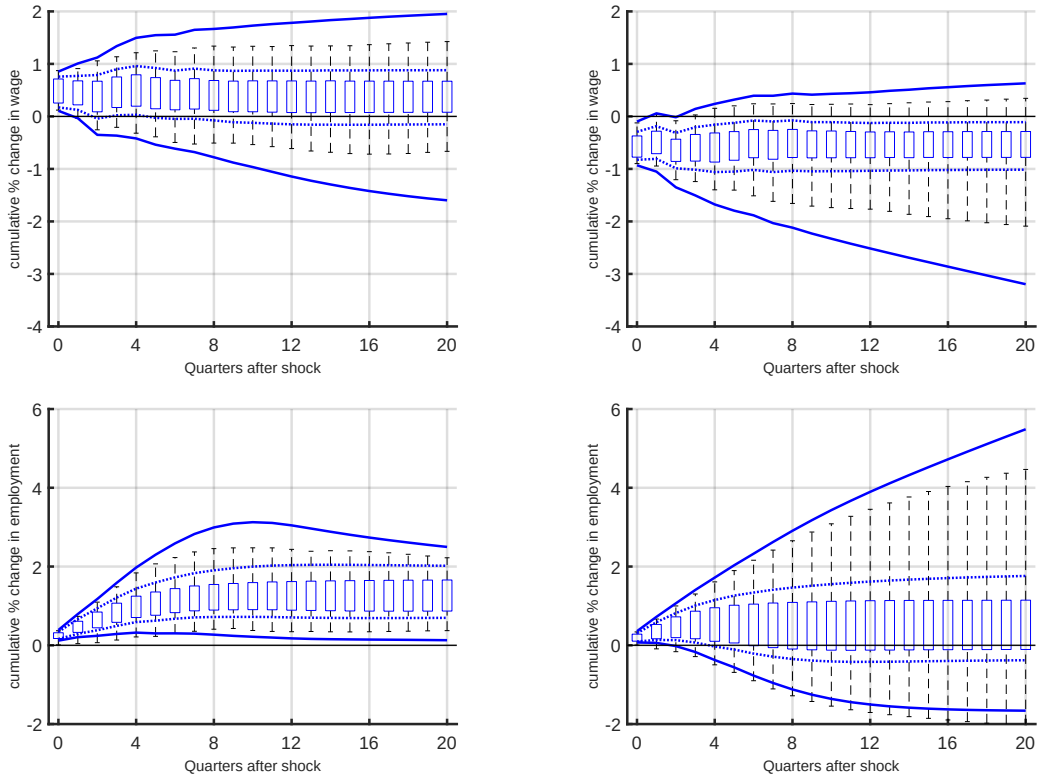
6.6.5 Implementing calibrated projection in our example

The key restriction used to set-identify an expansionary demand shock in the illustrative example is that it must increase wages and employment, upon impact. According to the credible sets in Figures 6.1 and 6.2, the expansionary shock has—in fact—noncontemporaneous effects over these two variables (every quarter over a 5 year horizon). Our calibrated projection confirms that there are medium-run effects of demand shocks over employment, but suggests that the non-zero effects over wages beyond the first two quarters could be an artifact of prior beliefs.

[19] Instead of pseudo-random draws from the multi-variate normal distribution, we use quasi-random Sobol sequences, which have the property of being a low-discrepancy sequence in the hypercube. We translate the sequence into multivariate-normal draws using Cholesky decomposition. In our experience, this improves the performance of grid search substantially for a given number of grid points.

[20] In our Matlab code to implement projection we take SQP/IP as the default algorithm to construct the projection region.

Figure 6.3. 68% projection region and 68% calibrated projection.



(a) Expansionary Demand Shock

(b) Expansionary Supply Shock

(Solid Line) 68% Projection region; (Dotted Line) 68% Projection region *calibrated* to guarantee 68% robust Bayesian credibility of the IRF functions jointly (100,000 draws from the Gaussian approximation to the posterior of μ); (Box) 68% Projection region *calibrated* horizon by horizon and shock by shock; (Black Dashed Line) Support of the bounds of the identified set given the 100,000 posterior draws.

A similar observation is true for supply shocks. Our calibrated projection suggests that the decrease in wages five years after an expansionary supply shock is robust to the choice of prior on the set-identified parameters. The medium-run effects of supply shocks over employment lack this robustness.

Implementation of our Calibrated Projection: We close this subsection providing further details about the computational demands of our calibration exercise.

Instead of working with a specific posterior for μ, we calibrated projection relying on the large-sample approximation $\mu|Y_1,\ldots,Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T / T)$. Taking draws from this model

is straightforward and does not require any special sampling technique (as a Monte-Carlo Markov Chain). Figure 6.3 used M=100,000 draws.

As described in our calibration algorithm, for each of the draws of μ (denoted μ_m^*), and for each horizon $k \in \{0, 1, 2, \dots 20\}$, variable $i \in \{\text{wage}, \text{employment}\}$ and shock $j \in \{\text{demand shock}, \text{supply shock}\}$ we solved two mathematical programs to generate:

$$[\underline{v}_{k,i,j}(\mu_m^*), \overline{v}_{k,i,j}(\mu_m^*)].$$

Computing the bounds of the identified set for all the combinations (k, i, j) given μ_m^* took approximately 9 seconds. Generating the boxes and the black dashed lines in Figure 6.3 took approximately 5 hours using 50 parallel Matlab 'workers' on a computer cluster at the University of Bonn.[21] Notice that we choose M=100,000 for illustrative purposes and the calibration results are barely different for M=1,000, which takes 3 minutes using the same computer cluster (or 2.5 hours not using parallelization at all).

After generating the bounds of the identified set, the calibration exercise adjusts the nominal level of projection to simulatenously contain 68% of the draws from the bounds of the identified set for each combination (k, i, j).[22] The calibrated confidence level for the Wald ellipsoid is $1.85 \cdot 10^{-4}$% instead of the original 68%. This means that instead of projecting a Wald ellipsoid with radius $\chi^2_{68\%,27}$ we are using a $\chi^2_{68\%,4.5}$.

6.7 Conclusion

A practical concern regarding standard Bayesian inference for set-identified Structural Vector Autoregressions is the fact that prior beliefs continue to influence posterior inference even when the sample size is infinite. Motivated by this observation, this paper studied the properties of projection inference for set-identified SVARs.

A nominal $1-\alpha$ projection region collects all the structural parameters of interest that are compatible with the VAR reduced-form parameters in a nominal $1-\alpha$ Wald ellipsoid.

---

[21] Calibrating projection to guarantee frequentist coverage at one point in the parameter space took us 76 hours using the 50 parallel Matlab workers in the same computer cluster.

[22] To do this, we ran the baseline projection SQP/IP algorithm for different nominal confidence levels. An efficient calibration algorithm that requires only few iterations over the nominal level is the combination of bisection with secant and interpolation as provided by Matlab's `fzero` function. For reasonably low tolerance of $\eta = 0.001$, we need 15 iteration steps. With each step taking about 734 seconds, see Table 6.3, steps 3 through 5 take about 1 hour.

By construction, projection inference does not rely on the specification of prior beliefs for set-identified parameters.

We argued that the projection approach is general, computationally feasible, and—under mild assumptions concerning the asymptotic behavior of estimators and posterior distributions for the reduced-form parameters—produces regions with frequentist coverage and asymptotic *robust* Bayesian credibility of at least $1-\alpha$.

The main drawback of our projection region is that it is conservative. For a frequentist, the large-sample coverage is strictly above the desired confidence level. For a robust Bayesian, the asymptotic robust credibility of the nominal $1-\alpha$ projection region is strictly above $1-\alpha$.

We used the calibration idea described in Kaido et al. (2016) to eliminate the excess of robust Bayesian credibility. The calibration procedure consists of drawing the reduced-form parameters, μ, from its posterior distribution (or a suitable large-sample Gaussian approximation); evaluating the functions $\underline{v}(\mu), \overline{v}(\mu)$ for each draw of μ; and, finally, decreasing the nominal level of the projection region until it contains exactly $(1-\alpha)\%$ of the values of $\underline{v}(\mu), \overline{v}(\mu)$. The calibration exercise required more work than the baseline projection, but it is computationally feasible (and easily parallelizable). Moreover, if the bounds of the identified set are differentiable, our calibrated projection covers the identified set with probability $1-\alpha$.

We implemented our projection confidence set in the demand/supply SVAR for the U.S. labor market. The main set-identifying assumptions were sign restrictions on contemporaneous responses. Standard Bayesian credible sets suggested that the medium-run response of wages and employment to structural shocks behave in the same way as the contemporaneous responses. Our projection region (baseline and calibrated) showed that only the qualitative effects of demand shocks over employment and the qualitative effects of supply shocks over wages are robust to the choice of prior. Our projection approach is a natural complement for the Bayesian credible sets that are commonly reported in applied macroeconomic work.

Appendix 6.A Appendix

6.A.1 Proof of main results

6.A.1.1 Verification of Assumption 5 for the Gaussian SVAR with a Normal-Wishart prior.

Consider the SVAR in (6.3) and assume that $F \sim \mathcal{N}(0, \mathbb{I}_n)$. Let P^* denote a prior on the SVAR parameters (A, B).

Note first that Assumption 5 depends only on the distribution that P^* induces over the reduced-form parameters, μ. Thus, we abuse notation and refer to P^* as the prior distribution on (A, Σ).

The analysis in this section focuses on the *Normal-Wishart* prior P^* used in Gaussian SVAR analysis. We establish an almost sure version of Assumption 5.

Prior for μ: Consider the hyper-parameters:

$$\bar{A}_0 \in \mathbb{R}^{n \times np},\ S_0 \in \mathbb{R}^{n \times n},\ N_0 \in \mathbb{R}^{np \times np},\ \nu_0 \in \mathbb{R}.$$

Definition. The Normal-Wishart Prior P^* over the parameters $(\text{vec}(A), \text{vech}(\Sigma))$—defined by hyper parameters $(\bar{A}_0, S_0, N_0, \nu_0)$—is given by:

$$\text{vec}(A)|\Sigma \sim \mathcal{N}\Big(\text{vec}(\bar{A}_0)\ ,\ N_0^{-1} \otimes \Sigma\Big),$$

and

$$\Sigma^{-1} \sim \text{Wishart}_n\Big(S_0^{-1} / \nu_0\ ,\ \nu_0\Big).$$

Posterior in the Gaussian SVAR: Let

$$Q_T \equiv \frac{1}{T} \sum_{t=1}^{T} X_t X_t',$$

and define the updated hyperparameters:

$$\begin{aligned}\bar{A}_T &= \widehat{A}_T Q_T\Big(\frac{N_0}{T}+Q_T\Big)^{-1}+\bar{A}_0\frac{N_0}{T}\Big(\frac{N_0}{T}+Q_T\Big)^{-1}\\ S_T &= \frac{\nu_0}{T+\nu_0}S_0+\frac{T}{T+\nu_0}\widehat{\Sigma}_T+\frac{1}{T+\nu_0}\Big(\bar{A}_T-\bar{A}_0\Big)N_0\Big(\frac{N_0}{T}+Q_T\Big)^{-1}Q_T\Big(\bar{A}_T-\bar{A}_0\Big)'\end{aligned}$$

where $\widehat{A}_T$ and $\widehat{\Sigma}_T$ are ordinary least squares estimators for A and Σ defined in Section 6.3.1. From p. 410 in Uhlig (1994) and p. 410 in Uhlig (2005) the posterior distribution for the vector $(\text{vec}(A)', \text{vech}(\Sigma)')'$ can be written as:

$$\text{vec}(A)|Y_1,\dots,Y_T = \text{vec}(\bar{A}_T)+\Big[\Big(\frac{N_0}{T}+Q_T\Big)^{-1}\otimes\frac{\Sigma}{T}\Big]^{1/2}W,\quad W\sim\mathcal{N}_{n^2p}(\mathbf{0},\mathbb{I}_{n^2p}),$$

$$\Sigma|Y_1,\dots,Y_T = S_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}Z_tZ_t'\Big)^{-1}S_T^{1/2},\quad Z_t\sim\mathcal{N}_n(\mathbf{0}\,,\mathbb{I}_n),\text{ i.i.d},$$

where both random vectors are independent of the data and $\{Z_t\}_{t=1}^T$ independent of W. Note that for a given data realization, the posterior distribution of (A,Σ) is a measurable function of $\mathscr{W}\equiv(W,Z_1,\dots Z_T)$. We use the term $o_{\mathscr{W}}(1)$ to denote any sequence that converges to zero as $T\to\infty$ for almost every realization of $\mathscr{W}$.

Asymptotic Behavior of the posterior for μ: We now show that all of the Normal-Wishart priors in the Gaussian model satisfy our Assumption 5. Note first that for almost every data realization $(Y_1,\dots,Y_T)$ and almost every realization of the random vector Z_t we have that

$$\Sigma-\widehat{\Sigma}_T\to 0,$$

by applying the strong law of large numbers to $(1/T)\sum_{t=1}^{T} Z_t Z_t'$. Consequently:

$$
\begin{aligned}
& \sqrt{T}(\text{vec}(A) - \text{vec}(\widehat{A}_T)) \\
= \; & \widehat{A}_T \sqrt{T}\Big(Q_T\Big(\frac{N_0}{T} + Q_T\Big)^{-1} - \mathbb{I}_{n^2p}\Big) + \bar{A}_0 \frac{N_0}{\sqrt{T}}\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \\
& + \Big[\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{P^*|Y_1,\ldots Y_T}(1), \\
= \; & \widehat{A}_T \sqrt{T}\Big(Q_T\Big(Q_T^{-1} + Q_T^{-1}\frac{N_0}{T}Q_T^{-1} + O(1/T^2)\Big) - \mathbb{I}_{n^2p}\Big) \\
& \text{(by a first-order Taylor expansion)} \\
& + \bar{A}_0 \frac{N_0}{\sqrt{T}}\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \\
& + \Big[\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{\mathscr{W}}(1), \\
= \; & \Big[Q_T^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{\mathscr{W}}(1).
\end{aligned}
$$

This implies that the posterior distribution of $\sqrt{T}(\text{vec}(A) - \text{vec}(\widehat{A}_T))$ converges in distribution, for almost every data realization $(Y_1, \ldots, Y_T)$, to the random vector:

$$
[Q_T^{-1/2} \otimes \widehat{\Sigma}_T^{1/2}]W, \text{ where } W \sim \mathcal{N}_{n^2p}(\mathbf{0}, \mathbb{I}_{n^2p}). \tag{6.10}
$$

Note now that

$$
\begin{aligned}
& \sqrt{T}(\text{vech}(\Sigma) - \text{vech}(\widehat{\Sigma}_T)) \\
= \; & \sqrt{T}\text{vech}\Big(S_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} S_T^{1/2} - \widehat{\Sigma}_T\Big), \\
= \; & \sqrt{T}\text{vech}\Big(\widehat{\Sigma}_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} \widehat{\Sigma}_T^{1/2} + O(1/T) - \widehat{\Sigma}_T\Big), \\
= \; & \sqrt{T}\text{vech}\Big(\widehat{\Sigma}_T^{1/2}\Big[\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} - \mathbb{I}_n\Big]\widehat{\Sigma}_T^{1/2}\Big) + o(1).
\end{aligned}
$$

This implies that the posterior distribution of $\sqrt{T}(\text{vech}(\Sigma) - \text{vech}(\widehat{\Sigma}_T))$ converges in distribution, for almost every data realization $(Y_1, \ldots, Y_T)$, to the random vector:

$$
\Big(2D^+(\widehat{\Sigma}_T \otimes \widehat{\Sigma}_T)D^+\Big)^{1/2} Z, \text{ where } Z \sim \mathcal{N}_{n(n+1)/2}(\mathbf{0}, \mathbb{I}_{n(n+1)/2}),\ Z \perp W, \tag{6.11}
$$

and $D^+ \equiv (D'D)^{-1}D'$ is the Moore-Penrose inverse of the duplication matrix D such that $\text{vec}(\Sigma) = D\text{vech}(\Sigma)$.

Now, assume that the confidence set for the reduced-form parameters is constructed using the Gaussian Maximum Likelihood asymptotic variance of $\widehat{\mu}_T$ as in p.93 of Lütkepohl (2007); that is:

$$\widehat{\Omega}_T \equiv \begin{pmatrix} Q_T^{-1} \otimes \widehat{\Sigma}_T & \mathbf{0}_{n^2p \times (n(n+1)/2)} \\ \mathbf{0}_{(n(n+1)/2) \times n^2p} & 2D^+(\widehat{\Sigma}_T \otimes \widehat{\Sigma}_T)D^{+\prime} \end{pmatrix}. \tag{6.12}$$

Let G denote the joint distribution of (W, Z), which is a standard multivariate normal independently of the data. Then, combining (6.10), (6.11), (6.12)

$$\begin{aligned} & P^*\Big(\mu \in \mathrm{CS}_T(1-\alpha,\mu)|(Y_1,\ldots,Y_T)\Big) \\ = \; & P^*\Big(\sqrt{T}(\mu-\widehat{\mu}_T)'\widehat{\Omega}_T^{-1}\sqrt{T}(\mu-\widehat{\mu}_T) \leq \chi^2_{d,1-\alpha}|(Y_1,\ldots Y_T)\Big) \\ \to \; & G\Big(\begin{pmatrix} W \\ Z \end{pmatrix}' \begin{pmatrix} W \\ Z \end{pmatrix} \leq \chi^2_{d,1-\alpha} \mid Y_1,\ldots,Y_T\Big) \text{for a.e. data realization} \\ = \; & G\Big(\begin{pmatrix} W \\ Z \end{pmatrix}' \begin{pmatrix} W \\ Z \end{pmatrix} \leq \chi^2_{d,1-\alpha}\Big) \\ = \; & (1-\alpha). \end{aligned}$$

6.A.1.2 Proof of Result 3 (Finite-sample calibration for a Robust Bayesian)

Proof. The proof of Result 2 has already established that for any data realization:

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big).$$

is at least as large as:

$$P^*_\mu\Big(\times_{h=1}^H [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \mathrm{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T),\lambda^H)|Y_1,\ldots,Y_T\Big).$$

Hence, it is sufficient to show that for any data realization:

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big) \leq 1-\alpha.$$

In order to establish this upper bound for each data realization, we will find a prior on Q (conditional on μ) that gives credibility of exactly $1-\alpha$ to the calibrated projection region. Fix the data, and denote the set $\mathrm{CS}_T(1-\alpha(Y_1,\ldots,Y_T);\lambda^H)$ simply by $\mathscr{C}(Y^T)$. Before the realization of the data, the set $\mathscr{C}(Y^T)$ is just some subset of $\mathbb{R}^H$, so the prior can depend

on this set. Let $\overline{v}_h(\mu)$ abbreviate $\overline{v}_{k_h,i_h,j_h}(\mu)$ and define $\underline{v}_h(\mu)$ analogously. Let $Q_{\max}(\mu;h)$ denote the rotation matrix for which the structural parameter achieves its upper bound; i.e., $\lambda(\mu, Q_{\max}(\mu;h)) = \overline{v}_h(\mu)$ (the matrix $Q_{\min}$ is defined analogously).

For each μ such that $\times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \notin \mathscr{C}(Y^T)$, let $\overline{h}(\mu)$ denote the smallest horizon for which $\overline{v}_{\overline{h}(\mu)}(\mu)$ is not contained in the $h(\mu)$-th coordinate of the region $\mathscr{C}(Y^T)$. If no upperbound falls outside $\mathscr{C}(Y^T)$ set $\overline{h}(\mu) = 0$. Define $\underline{h}(\mu)$ analogously. Consider the following prior for $Q|\mu$ that depends on the set $\mathscr{C}_T(Y^T)$:

$$Q|\mu = \begin{cases} Q_{\max}(\mu;1) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \subseteq \mathscr{C}_T(Y^T), \\ Q_{\max}(\mu, \overline{h}(\mu)) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \nsubseteq \mathscr{C}(Y^T) \text{ and } \overline{h}(\mu) \geq \underline{h}(\mu), \\ Q_{\min}(\mu, \underline{h}(\mu)) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \nsubseteq \mathscr{C}(Y^T) \text{ and } \overline{h}(\mu) < \underline{h}(\mu), \end{cases}$$

Finally, let P^{**} denote the prior induced by P^*_μ and $Q|\mu$ as defined above. Note that for each data realization $(Y_1, \ldots, Y_T)$:

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha(Y_1,\ldots Y_T); \lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big)$$

is—by definition of infimum—smaller than or equal

$$P^{**}\Big(\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T); \lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big).$$

By construction, the prior for $Q|\mu$ is such that $\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T); \lambda^H)$ if and only if $\times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \subseteq \mathscr{C}_T(Y^T)$. To see this, note that whenever the bounds of the identified set $\times_{h=1}^{H}[\underline{v}_h(\mu), \overline{v}_h(\mu)] \nsubseteq \mathscr{C}_T(Y^T)$, either $\overline{h}(\mu) \neq 0$ or $\underline{h}(\mu) \neq 0$ implying that the structural parameter $\lambda_h(\mu,Q)$ takes the value of $\overline{v}_{\overline{h}(\mu)}(\mu)$ or $\underline{v}_{\overline{h}(\mu)}(\mu)$ (whichever horizon is largest). Since these bounds are not contained in $\mathscr{C}_T(Y^T)$:

$$P^{**}\Big(\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T); \lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big).$$

equals

$$P^*_\mu\Big(\times_{h=1}^{H}[\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \in \mathrm{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T), \lambda^H) | Y_1, \ldots, Y_T\Big) = 1-\alpha.$$

This means that:

$$1-\alpha \leq \inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T); \lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big) \leq 1-\alpha.$$

□

6.A.1.3 Proof of Result 4 (Robust Bayesian calibration and frequentist coverage)

Let r_T^* be the radius that calibrates robust Bayesian credibility; i.e., the radius corresponding to the calibrated nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$. In a slight abuse of notation we replace $\text{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda_h)$ by $\text{CS}_T(r_T^*;\lambda_h)$. The proof of this result is based on Lemma 1,2, and 3 in Appendix 6.A.3.

Proof. We show that for every $\epsilon>0$ there is $T(\epsilon)$ such that $T>T(\epsilon)$ implies that:

$$1-\alpha-\epsilon \leq P_{\mu_0}\left([\underline{v}_h(\mu_0),\overline{v}_h(\mu_0)] \subseteq \text{CS}_T(r_T^*;\lambda_h),\ \forall h=1,\ldots H\right) \leq 1-\alpha+\epsilon.$$

Let $M_1(\epsilon)$ and $M_2(\epsilon)$ be two real-valued (nonnegative) functions. Define the events:

$$\begin{aligned}
A_1(\epsilon) &= \{(Y_1,\ldots,Y_T)|\ ||\sqrt{T}(\widehat{\mu}_T-\mu_0)||\ \leq\ M_1(\epsilon)\},\\
A_2(\epsilon) &= \{(Y_1,\ldots Y_T)|\ \text{the largest eigenvalue of } \widehat{\Omega}_T \text{ is smaller than } M_2(\epsilon)\},\\
A_3(\eta) &= \{(Y_1,\ldots Y_T)|\ |\widehat{\overline{\sigma}}_h(\mu_0)-\overline{\sigma}_h(\mu_0)|\ <\ \eta/(2\chi^2_{d,1-\alpha})\\
&\qquad \text{and } |\widehat{\underline{\sigma}}_h(\mu_0)-\underline{\sigma}_h(\mu_0)|\ <\ \eta/(2\chi^2_{d,1-\alpha})\},\\
A_4(\epsilon) &= \{(Y_1,\ldots Y_T)|\ \sup_{B\in\mathscr{B}(\mathbb{R}^d)}|P^*(\sqrt{T}(\mu^*-\widehat{\mu}_T)\ \in\ B\mid Y_1,\ldots,Y_T)\\
&\qquad -\mathbb{P}(Z\ \in\ B)|\ \leq\ \epsilon/8,\ \text{ where } Z\ \sim\ \mathcal{N}_d(\mathbf{0},\Omega),\},\\
A_5(\epsilon) &= \{(Y_1,\ldots Y_T)|\ r_T^*\ <\ \chi^2_{d,1-\alpha}+\epsilon\}.
\end{aligned}$$

where the standard errors $\widehat{\overline{\sigma}}_h(\mu_0)$, $\overline{\sigma}_h(\mu_0)$, $\widehat{\underline{\sigma}}_h(\mu_0),\underline{\sigma}_h(\mu_0)$ are defined in Lemma 1 of Appendix 6.A.3.1. We first show that the probability of these events can be made arbitrarily close to 1 for a large enough sample size.

To see this, note that Assumption 1 of Result 4 (convergence in distribution of $\sqrt{T}(\widehat{\mu}_T-\mu_0)$) implies there exists a function $M_1(\epsilon)$ and a large enough sample size $T_1(\epsilon)$ such that for $T\geq T_1(\epsilon)$, $P_{\mu_0}(A_1(\epsilon))>1-\epsilon/25$. Assumption 2 of Result 4 (convergence in probability of $\widehat{\Omega}_T$) implies there exists a function $M_2(\epsilon)$ and a large enough sample size $T_2(\epsilon,\eta)$ such that for $T\geq T_2(\epsilon,\eta)$, $P_{\mu_0}(A_2(\epsilon))>1-\epsilon/25$ and $P_{\mu_0}(A_3(\eta))>1-\epsilon/25$. Assumption 3 of Result 4 (Bernstein von-Mises Theorem in total variation) implies that there is $T_4(\epsilon)$ such that $T\geq T_4(\epsilon,\eta)$, $P_{\mu_0}(A_4(\epsilon))>1-\epsilon/25$. Finally, since Assumption 3 of Result 4 implies the assumption of Result 2 (the baseline projection has robust credibility of at least $1-\alpha$ with high probability) implies there is $T_5(\epsilon)$ such that $T\geq T_5(\epsilon)$, $P_{\mu_0}(A_5(\epsilon))>1-\epsilon/25$

This means that for $T \geq \max\{T_1(\epsilon), T_2(\epsilon,\eta), T_4(\epsilon), T_5(\epsilon)\}$,

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \mathrm{CS}_T(r_T^*; \lambda_h),\ \forall h = 1, \ldots H\right)$$

$$\leq$$

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \ \subseteq\ \mathrm{CS}_T(r_T^*; \lambda_h),\ \forall h\ =\ 1, \ldots H\ \cap A(\epsilon,\eta)\right) + \epsilon/5, \tag{6.13}$$

where

$$A(\epsilon,\eta) \equiv A_1(\epsilon) \cup A_2(\epsilon) \cup A_3(\eta) \cup A_4(\epsilon) \cup A_5(\epsilon).$$

Define now, given $Z \sim \mathcal{N}(0, \Omega)$, the quantile $r_{1-\alpha+\epsilon}$ by the equation:

$$\begin{aligned}&\mathbb{P}(-\underline{\sigma}_h(\mu_0) r_{1-\alpha+\epsilon}\ \leq\ \underline{\dot{v}}_h(\mu_0)' Z,\\ &\quad \text{and } \overline{\dot{v}}_h(\mu_0)' Z\ \leq\ \overline{\sigma}_h(\mu_0) r_{1-\alpha+\epsilon},\ \forall h\ =\ 1, \ldots H)\ =\ 1 - \alpha + \epsilon.\end{aligned}$$

Lemma 3 in Appendix 6.A.3.3 has shown that there exists $T_6(\epsilon)$ such that :

$$(Y_1, \ldots, Y_T) \in A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})) \implies r_T^* \leq r_{1-\alpha+\epsilon/5} \equiv \overline{r},$$

where the function $\eta(\epsilon/10, \chi^2_{d,1-\alpha}))$ is defined as in Lemma 2 in Appendix 6.A.3.2. This implies that whenever $T \geq \max\{T_1(\epsilon), T_2(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})), T_4(\epsilon), T_5(\epsilon), T_6(\epsilon)\}$, Equation (6.13) is bounded above by:

$$\begin{aligned}&P_{\mu_0}([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \ \subseteq\ \mathrm{CS}_T(\overline{r}; \lambda_h),\\ &\qquad \forall h\ =\ 1, \ldots H\ \cap A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha}))) + \epsilon/5.\end{aligned} \tag{6.14}$$

Lemma 1 in Appendix 6.A.3.1 has shown that $(Y_1, \ldots, Y_T) \in A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha}))$ implies that the projection region of radius $r_{1-\alpha+\epsilon/5}$ is contained in the delta-method interval $DM_T^h(r,\eta)$ (with radius $r = r_{1-\alpha+\epsilon/5}$ and expansion $\eta = \eta(\epsilon/10, \chi^2_{d,1-\alpha})$):

$$\left[\underline{v}_h(\widehat{\mu}_T) - \frac{(\overline{r} + \eta(\epsilon/10, \chi^2_{d,1-\alpha}))\underline{\sigma}_h(\mu_0)}{\sqrt{T}},\ \overline{v}_h(\widehat{\mu}_T) + \frac{(\overline{r} + \eta(\epsilon/10, \chi^2_{d,1-\alpha}))\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\right],$$

for every $h = 1, \ldots H$. This means that Equation (6.14) is bounded above by:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \ \subseteq\ DM_T^h(\overline{r}, \eta(\eta/10, \chi^2_{d,1-\alpha})),\ \forall h\ =\ 1, \ldots H\ \right) + \epsilon/5. \tag{6.15}$$

An application of the delta-method implies there is $T_7(\epsilon)$ larger than

$$T \geq \max\{T_1(\epsilon), T_2(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})), T_4(\epsilon), T_5(\epsilon), T_6(\epsilon)\}$$

such that Equation 6.15 is bounded above by $2\epsilon/5$ plus:

$$\begin{aligned}\mathbb{P}(& -\underline{\sigma}_h(\mu_0)(\overline{r} + \eta(\epsilon/10, \chi^2_{d,1-\alpha})) \\ & \leq \underline{\dot{\nu}}_h(\mu_0)'Z, \text{ and } \overline{\dot{\nu}}_h(\mu_0)'Z \\ & \leq \overline{\sigma}_h(\mu_0)(\overline{r} + \eta(\epsilon/10, \chi^2_{d,1-\alpha}),\ \forall h = 1, \ldots H),\end{aligned} \tag{6.16}$$

which, by definition of $\eta(\epsilon, \chi^2_{d,1-\alpha})$, imploes that the latter equation is bounded above by

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)\overline{r} \leq \underline{\dot{\nu}}_h(\mu_0)'Z, \text{ and } \overline{\dot{\nu}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)\overline{r},\ \forall h = 1, \ldots H\right) + 4\epsilon/5.$$

Using the definition of $\overline{r}$, we conclude that there is $T(\epsilon)$ such that for $T \geq T(\epsilon)$:

$$P_{\mu_0}\left([\underline{\nu}_h(\mu_0), \overline{\nu}_h(\mu_0)] \subseteq \text{CS}_T(r^*_T; \lambda_h),\ \forall h = 1, \ldots H\right) \leq 1 - \alpha + \epsilon.$$

The lower bound is derived analogously.

□

6.A.1.4 Asymptotic calibration for a Robust Bayesian $(\mu|Y_1, \ldots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T))$

We now show that whenever $\alpha^*_T \equiv \alpha(Y_1, \ldots, Y_T)$ is calibrated to guarantee that

$$\begin{aligned}\mathbb{P}_T\Big(& \times_{h=1}^H [\underline{\nu}_{k_1,i_1,j_1}(\mu), \overline{\nu}_{k_1,i_1,j_1}(\mu)] \times \ldots \times [\underline{\nu}_{k_h,i_h,j_h}(\mu), \overline{\nu}_{k_h,i_h,j_h}(\mu)] \\ & \subseteq \text{CS}_T(1 - \alpha^*_T, \lambda^H) \mid Y_1, \ldots Y_T\Big)\end{aligned}$$

equals $1-\alpha$ whenever $\mu|Y_1, \ldots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T)$, then one can guarantee asymptotic robust credibility of $1-\alpha$ for a large class of priors on μ. This is formalized below.

Let $f(Y_1, \ldots Y_T \mid \mu_0)$ denote the Gaussian density for the VAR data and let $\Omega \in \mathbb{R}^{d\times d}$ denote the probability limit of $\widehat{\Omega}_T$. Let G_Ω denote a Gaussian measure centered at $\mathbf{0}_d$ with covariance matrix Ω. Let $\mathscr{B}(d)$ denote Borel sets in $\mathbb{R}^d$.

Result 5. *Let $Y_1, \dots Y_T \sim f(Y_1, \dots Y_T \mid \mu_0)$ and suppose that the prior P^*_μ is such that:*

$$\sup_{A \in \mathscr{B}(d)} \left| P^*_\mu(\sqrt{T}(\mu - \widehat{\mu}_T) \in A \mid Y_1, \dots Y_T) - G_\Omega(A) \right| = o_p(Y_1, \dots Y_T; \mu_0).$$

Then,

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*_T, \lambda^H) \mid Y_1, \dots Y_T\Big) = 1 - \alpha + o_p(Y_1, \dots Y_T; \mu_0).$$

Proof. Result 3 has shown that for any $\alpha(Y_1, \dots, Y_Y)$

$$\inf_{P^* \in \mathscr{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in \mathrm{CS}_T(1-\alpha^*_T, \lambda^H) \mid Y_1, \dots Y_T\Big) = P^*_\mu\big(\mu \in A^*_T \mid Y_1, \dots Y_T\big),$$

where $A^*_T \subseteq \mathbb{R}^d$ is defined as:

$$\{\mu \in \mathbb{R}^d \mid \times_{h=1}^H [\underline{\nu}_{k_h, i_h, j_h}(\mu), \overline{\nu}_{k_h, i_h, j_h}(\mu)] \subseteq \mathrm{CS}_T(1-\alpha^*_T, \lambda^H)\}.$$

Note that

$$\begin{aligned}
&P^*_\mu\big(\mu \in A^*_T \mid Y_1, \dots Y_T\big) \\
&\quad = P^*_\mu\big(\sqrt{T}(\mu - \widehat{\mu}_T) \in \sqrt{T}(A^*_T - \widehat{\mu}_T) \ : \mid Y_1, \dots Y_T\big) \\
&\qquad - G_\Omega(\sqrt{T}(A^*_T - \widehat{\mu}_T)) + G_\Omega(\sqrt{T}(A^*_T - \widehat{\mu}_T)) \\
&\qquad - G_{\widehat{\Omega}_T}(\sqrt{T}(A^*_T - \widehat{\mu}_T)) + G_{\widehat{\Omega}_T}(\sqrt{T}(A^*_T - \widehat{\mu}_T))
\end{aligned}$$

We make three observations:

1. Note first that:

   $$P^*_\mu\big(\sqrt{T}(\mu - \widehat{\mu}_T) \in \sqrt{T}(A^*_T - \widehat{\mu}_T) \ : \mid Y_1, \dots Y_T\big) - G_\Omega(\sqrt{T}(A^*_T - \widehat{\mu}_T))$$

   is smaller than or equal

   $$\sup_{A \in \mathscr{B}(d)} \left| P^*_\mu(\sqrt{T}(\mu - \widehat{\mu}_T) \in A \mid Y_1, \dots Y_T) - G_\Omega(A) \right|,$$

   which is, by assumption, $o_p(Y_1, \dots Y_T; \mu_0)$.

2. Note then that

$$|G_{\widehat{\Omega}_T}(\sqrt{T}(A_T^* - \widehat{\mu}_T)) - G_{\Omega}(\sqrt{T}(A_T^* - \widehat{\mu}_T))| = o_p(Y_1, \ldots, Y_T; \mu_0)$$

since $\widehat{\Omega}_T \xrightarrow{p} \Omega$ and G is the Gaussian measure centered at zero.

3. Finally, note that $G_{\widehat{\Omega}_T}(\sqrt{T}(A_T^* - \widehat{\mu}_T))$ is the same as is the same as

$$\mathbb{P}(N(\widehat{\mu}_T, \widehat{\Omega}_T / T) \in A_T^* | Y_1, \ldots, Y_T),$$

which, by definition of A_T^*, is the same as:

$$\mathbb{P}_T\Big(\times_{h=1}^{H} [\underline{v}_{k_1,i_1,j_1}(\mu), \overline{v}_{k_1,i_1,j_1}(\mu)] \times \ldots \times [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \text{CS}_T(1-\alpha_T^*, \lambda^H) \mid Y_1, \ldots Y_T \Big)$$

where $\mu | Y_1, \ldots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T / T)$.

We conclude that:

$$|\inf_{P^* \in \mathscr{P}(P_\mu^*)} P^*\Big(\lambda^H(A,B) \in \text{CS}_T(1-\alpha_T^*, \lambda^H) \mid Y_1, \ldots Y_T\Big) - (1-\alpha)| \leq o_p(Y_1, \ldots Y_T; \mu_0),$$

which implies the desired result. □

6.A.2 Frequentist calibration of projection

We have shown that projection can be calibrated to achieve exact robust Bayesian credibility for a given prior on the reduced-form parameters. We now discuss the extent to which projection can be calibrated to achieve large-sample frequentist coverage of $1-\alpha$.

Frequentist calibration requires either an exact or an approximate statistical model for the data. We assume that: $\widehat{\mu}_T \sim \mathbb{P}_\mu \equiv \mathcal{N}_d(\mu, \widehat{\Omega}_T / T)$, where μ belongs to some set $\mathscr{M} \subseteq \mathbb{R}^d$ and $\widehat{\Omega}_T$ is treated as a non-stochastic matrix.

Let λ be some structural coefficient of interest. The frequentist calibration exercise consists in finding a radius, $r_T(\alpha)$, for the Wald ellipsoid such that:

$$\inf_{\mu \in \mathscr{M}} \inf_{\lambda \in \mathscr{I}^{\mathscr{R}}(\mu)} \mathbb{P}_\mu\Big(\lambda \in CS_T(r_T(\alpha); \lambda)\Big) = 1 - \alpha.$$

An algorithm to Calibrate Projection over a grid G: Let d denote the dimension of μ and let $1-\alpha$ be the desired confidence level.

1. Generate a grid of S scalars $\{r_1, r_2, \ldots, r_S\}$ on the interval $[0, \sqrt{\chi^2_{d,1-\alpha}}]$. Each of these values will serve as the potential 'radius' of the Wald ellipsoid for μ. Fix one element r_s.

2. Generate a grid of I values $G \equiv \{\mu_1, \mu_2, \ldots, \mu_I\} \in \mathscr{M} \subseteq \mathbb{R}^d$. Fix an element $\mu_i \in G$.

3. Generate M i.i.d. draws from the model

$$\widehat{\mu}^i_{T,m} \sim \mathcal{N}_d(\mu_i, \widehat{\Omega}_T / T).$$

   Let $\text{CS}^m_T(r_s, \lambda)$ denote the confidence interval for λ associated to $\widehat{\mu}^i_{T,m}$ with radius r_s. Note that in order to compute the confidence interval for λ, $\widehat{\Omega}_T$ is fixed across all draws.

4. Generate a grid of size K $\{\lambda^i_1, \lambda^i_2, \ldots, \lambda^i_K\}$ from the identified-set for λ given μ_i, denoted $\mathscr{I}^{\mathscr{R}}(\mu_i)$.

5. For each μ_i compute:

$$CP_T(\mu_i; r_s, \widehat{\Omega}_T) \equiv \min_{k \in K} \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\Big\{\lambda_k \in CS^m_T(r_s; \lambda)\Big\}.$$

6. Report the approximate confidence level of the projection confidence interval with radius r_s as:

$$\text{ApproxCL}_T(r_s) \equiv \min_{i \in I} CP_T(\mu_i; r_s, \widehat{\Omega}_T)$$

7. Find the value in the grid

$$\{\text{ApproxCL}_T(r_1), \ldots \text{ApproxCL}_T(r_S)\}.$$

   that is the closest to the desired confidence level $1 - \alpha$. Denote this value by $r^*_T(\alpha, G)$.

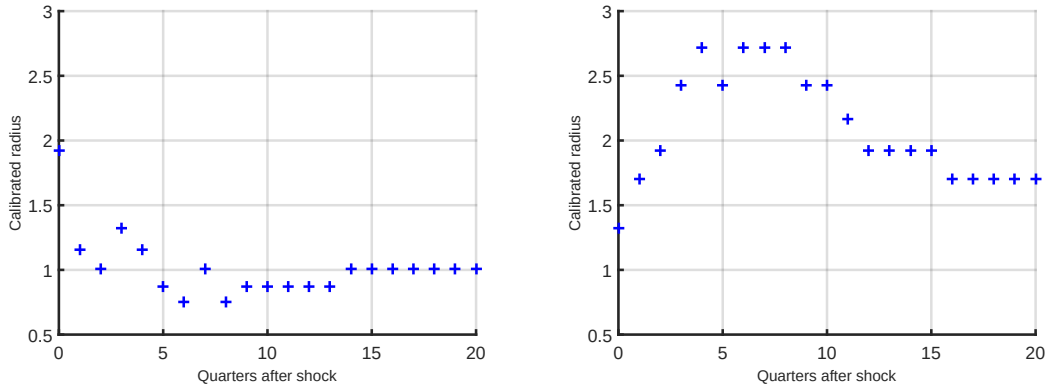
8. The radius $r^*_T(\alpha, G)$ obtained in Step 6 approximates the value $r_T(\alpha)$ that calibrates frequentist projection.

In our application $\mu \in \mathbb{R}^{27}$, which means that constructing an exhaustive grid for μ is computationally infeasible. To illustrate the computational demands of frequentist calibration in the SVAR exercise, consider a grid G that contains only $\widehat{\mu}_T$. We follow Step 1 to 5 to adjust

the confidence set for the responses of wages and employment to a structural demand shock (the first column of Figure 6.1).

Figure 6.4 below reports our calibrated radii, horizon by horizon, for the responses of wages and employment to an expansionary demand shock. Note that the default radius used by our projection method is $\chi^2_{27,68\%} = 29.87$.

Figure 6.4. Calibrated radii for the 68% projection region; $G = \{\widehat{\mu}_T\}$ (responses to an expansionary demand shock)



(a) Radii for Wages (b) Radii for Employment

(Blue Pluses) For each horizon k and each variable i the blue markers in Panel a) and b) correspond to the calibrated radius $r_T(\alpha, G)$ for $\lambda_{k,i,j}$ (as computed in Step 1 to 5). Each radius is computed using a grid of 16 points ranging from .5 to 5 ($S = 16$ in Step 1); a grid G containing only $\widehat{\mu}_T$ ($I = 1$ in Step 2); 1,000 draws for the reduced-form parameters ($J = 1,000$ in Step 3); and a grid of 1,000 points for $\lambda_{k,i,j}$ ($K = 1,000$ in Step 4). Generating this figure took approximately 76 hours using 50 parallel Matlab 'workers' on a computer cluster at Bonn University.

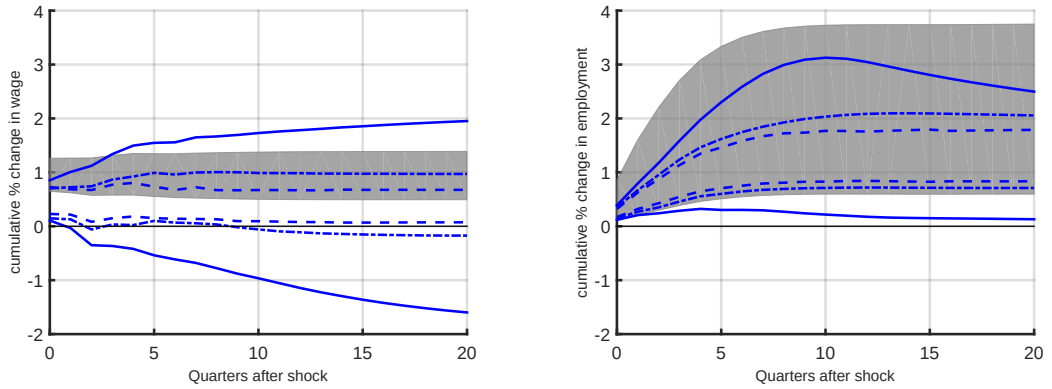
Calibrating coverage for a coefficient or a vector of coefficients: One could modify Step 4 in the algorithm to cover a vector of impulse-response functions, as opposed to one particular coefficient. In our application, this alternative calibrated radius (over the grid that contains only $\widehat{\mu}_T$) is 4.21. This radius is designed to cover the vector of responses for wages and employment to a structural demand shock over the 20 quarters under consideration. Calculating this radius took approximately 57 hours using 50 Matlab workers on a private computer cluster at Bonn University.[23]

The following figure compares the calibrated projection using the horizon by horizon calibrated radii against the calibrated projection using a radius of 4.21. The calibration over $\mathscr{G}$

[23] The cluster consists of 16 worker-nodes, where each node comprises 8 virtual CPUs and 32 GB virtual RAM, that is a maximum of 8 workers. Each virtual CPU is the core of a Xeon E7-8837@2.67GHz-processor.

implies that the true calibrated radius $r_T(\alpha)$ designed to cover the impulse-response functionshould be larger than 4.21.

Figure 6.5. 68% calibrated projection for a frequentist; $G = \{\widehat{\mu}_T\}$
(responses to an expansionary demand shock)



(a) Cumulative Response of Wages

(b) Cumulative Response of Employment

(Solid, Blue Line) 68% Projection region using the default radius $\chi^2_{27,68\%} = 29.87$; (Dash-Dotted, Blue Line) 68% Calibrated Projection Region using the radius 4.21; (Dashed, Blue Line) 68% Calibrated Projection Confidence Region based on the radii in Figure 6.4; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Baumeister and Hamilton (2015a).

6.A.3 Projection region under differentiability

Let h denote some triplet (k, i, j). This section studies the solution to the mathematical program defining the projection region whenever the bounds $\underline{\nu}_h$, $\overline{\nu}_h$ are differentiable at a point μ in the parameter space (and the derivative at this point is bounded away from zero). We show that a projection region for the (k, i, j) coefficient of the impulse-response function—indexed by h—is approximately equal to the delta-method confidence interval suggested in Gafarov et al. (2015):

$$\left[\underline{\nu}_h(\widehat{\mu}_T) - \frac{r\widehat{\underline{\sigma}}_h}{\sqrt{T}}\,,\ \overline{\nu}_h(\widehat{\mu}_T) + \frac{r\widehat{\overline{\sigma}}_h}{\sqrt{T}}\right].$$

This result has two important consequences:

1. Under differentiability, *the frequentist calibration of projection is straightforward*: it is sufficient to use the square of the $(1-\alpha)$ quantile of a standard normal as the radius of the Wald ellipsoid for the reduced-form parameters. For example, if the desired confidence level is 95%, the radius of the Wald ellipsoid can be set to $(1.64)^2$.

2. Under differentiability, *the radius that removes excess robust Bayesian credibility also eliminates excess frequentist coverage*. Thus, the robust Bayes calibration is also a frequentist calibration.

6.A.3.1 Lemma 1: Projection region and delta-method confidence interval

Notation for this Lemma: For $\mu \in \mathbb{R}^p$ define the 'delta-method' type standard errors as:

$$\widehat{\underline{\sigma}}_h(\mu) \equiv \left(\dot{\underline{v}}_h(\mu)'\widehat{\Omega}_T\dot{\underline{v}}_h(\mu)\right)^{1/2}, \quad \widehat{\overline{\sigma}}_h(\mu) \equiv \left(\dot{\overline{v}}_h(\mu)'\widehat{\Omega}_T\dot{\overline{v}}_h(\mu)\right)^{1/2}.$$

Define also their population counterparts as:

$$\underline{\sigma}_h(\mu) \equiv \left(\dot{\underline{v}}_h(\mu)'\Omega\dot{\underline{v}}_h(\mu)\right)^{1/2}, \quad \overline{\sigma}_h(\mu) \equiv \left(\dot{\overline{v}}_h(\mu)'\Omega\dot{\overline{v}}_h(\mu)\right)^{1/2},$$

where $\dot{\overline{v}}_h(\mu)$ denotes the derivative of $\overline{v}_h$ at μ. The function $\dot{\underline{v}}_h$ is defined analogously.

For a positive (or negative) $\delta \in \mathbb{R}$, consider the expansion (or contraction) of the delta-method type confidence interval:

$$\mathrm{DM}_T^h(r,\delta) \equiv \left[\underline{v}_h(\widehat{\mu}_T) - \frac{(r+\delta)\underline{\sigma}_h(\mu_0)}{\sqrt{T}}, \overline{v}_h(\widehat{\mu}_T) + \frac{(r+\delta)\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\right], \tag{6.17}$$

where the standard errors evaluated at $\mu_0 \in \mathbb{R}^d$ (the 'true' reduced-form parameter). Note that—up to the term δ—the interval in (6.17) can be interpreted as a 'delta-method' plug-in version of the Imbens and Manski (2004) confidence interval for a set-identified scalar parameter. The following result—which is taken from the recent working paper of Montiel-Olea and Plagborg-Møller (2016) (henceforth, MOPM16)—establishes the relation between the projection region and the confidence interval in (6.17):

Lemma 1. ***(Projection and delta-method confidence interval)*** *Suppose that $\underline{v}_h$ and $\overline{v}_h$ are differentiable at μ_0 with nonzero derivative. Fix constants $\delta > 0, M_1 > 0$, $M_2 > 0$ and $M_3 > 0$. Suppose that the data $(Y_1, \dots Y_T)$ is such that:*

1. $||\widehat{\mu}_T - \mu_0|| \leq M_1 / \sqrt{T}$,
2. *The largest eigenvalue of $\widehat{\Omega}_T$, denoted $\lambda_{\max}(\widehat{\Omega}_T)$, is smaller than M_2.*
3. $|\widehat{\overline{\sigma}}_h(\mu_0) - \overline{\sigma}_h(\mu_0)| < \delta/(2M_3)$ *and* $|\widehat{\underline{\sigma}}_h(\mu_0) - \underline{\sigma}_h(\mu_0)| < \delta/(2M_3)$.

Then, there exists $T(\delta, M_1, M_2, M_3)$ *such that for any* $T \geq T_1(\delta, M_1, M_2, M_3)$ *and* $0 < r \leq M_3$*:*

$$DM_T^h(r, -\delta) \subseteq CS_T(r; \lambda_h) \subseteq DM_T^h(r, \delta),$$

where $CS_T(r, \lambda_h)$ *is the projection region for the parameter* λ_h *based on the Wald ellipsoid of radius* r^2*. This means that our projection region with radius* r^2 *is approximately equal—in large samples and under differentiability at* μ_0*—to the delta-method confidence interval in equation (6.17).*

Proof of Lemma 1: The proof of this Lemma is based on the theoretical comparison of Bonferroni and (rectangular) Wald projection confidence regions established in MOPM16. We reproduce their argument for the sake of exposition. Note first that one can write $\overline{v}_h(\mu)$ as:

$$\begin{aligned}
\overline{v}_h(\mu) &= \overline{v}_h(\mu) - \overline{v}_h(\widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T) \\
&= \overline{v}_h(\mu) - \overline{v}_h(\mu_0) - (\overline{v}_h(\widehat{\mu}_T) - \overline{v}_h(\mu_0)) + \overline{v}_h(\widehat{\mu}_T).
\end{aligned}$$

For any $\mu \in \mathbb{R}^d$ define the function $\Delta(\mu; \mu_0) \equiv \overline{v}_h(\mu) - \overline{v}_h(\mu_0) - \dot{\overline{v}}_h(\mu_0)'(\mu - \mu_0)$. Therefore:

$$\begin{aligned}
\overline{v}_h(\mu) &= \Delta(\mu; \mu_0) - \Delta(\widehat{\mu}_T; \mu_0) + \dot{\overline{v}}_h(\mu_0)(\mu - \widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T) \\
&= \frac{\Delta(\mu; \mu_0)}{||\mu - \mu_0||}||\mu - \mu_0|| - \frac{\Delta(\widehat{\mu}_T; \mu_0)}{||\widehat{\mu}_T - \mu_0||}||\widehat{\mu}_T - \mu_0|| \\
&\quad + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T).
\end{aligned}$$

Note now that for any μ in the Wald ellipsoid of radius $r \leq M_3$ (i.e., $(\widehat{\mu}_T - \mu)' \widehat{\Omega}_T^{-1} (\widehat{\mu}_T - \mu) \leq r^2 / T$) we have that:

$$\begin{aligned}
||\mu - \mu_0|| &\leq ||\mu - \widehat{\mu}_T|| + ||\widehat{\mu}_T - \mu_0|| \\
&\leq ||\mu - \widehat{\mu}_T|| + M_1 / \sqrt{T} \\
&\quad \text{(by Assumption a) of the Lemma)} \\
&\leq \max_{\{\mu \mid ||\widehat{\Omega}^{-1/2}(\mu - \widehat{\mu}_T)|| \leq r/\sqrt{T}\}} ||\mu - \widehat{\mu}_T|| + M_1 / \sqrt{T} \\
&= \lambda_{\max}(\widehat{\Omega}_T) r / \sqrt{T} + M_1 / \sqrt{T} \\
&\leq (M_2 r + M_1) / \sqrt{T} \\
&\quad \text{(by Assumption b) of the Lemma)} \\
&\leq (M_2 M_3 + M_1) / \sqrt{T} \\
&\quad \text{(since } r \leq M_3\text{)}.
\end{aligned}$$

By the differentiability of $\overline{v}_h$ at μ_0, there exists $T_1(\delta, M_1, M_2, M_3)$ such that $T \geq T_1(\delta, M_1, M_2, M_3)$ implies that:

$$\left| \frac{\Delta(\mu; \mu_0)}{||\mu - \mu_0||} \right| \leq \frac{\delta \overline{\sigma}_h(\mu_0)}{4(M_2 M_3 + M_1)} \quad \text{and} \left| \frac{\Delta(\widehat{\mu}_T; \mu_0)}{||\widehat{\mu}_T - \mu_0||} \right| \leq \frac{\delta \overline{\sigma}_h(\mu_0)}{4M_1}.$$

This implies that for such large T we have that for any μ in the Wald ellipsoid of radius $r \leq M_3$, denoted $\mathrm{CS}_T(r; \mu)$:

$$\begin{aligned}
&\overline{v}_h(\widehat{\mu}_T) + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) - \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&\leq \overline{v}_h(\mu) \\
&\leq \overline{v}_h(\widehat{\mu}_T) + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) + \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T}.
\end{aligned}$$

Consequently:

$$
\begin{aligned}
&\overline{v}_h(\widehat{\mu}_T) + \widehat{\overline{\sigma}}_h(\mu_0) r / \sqrt{T} - \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&= \overline{v}_h(\widehat{\mu}_T) + \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) \right] - \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&\quad \text{(since } \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) \right] = \widehat{\overline{\sigma}}_h(\mu_0) r / \sqrt{T}) \\
&\leq \max_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \\
&\quad \text{(since we have bounded } \overline{v}_h(\mu) \text{ from below)} \\
&\leq \overline{v}_h(\widehat{\mu}_T) + \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) \right] + \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&\quad \text{(since we have bounded } \overline{v}_h(\mu) \text{ from above)} \\
&= \overline{v}_h(\widehat{\mu}_T) + \widehat{\overline{\sigma}}_h(\mu_0) r / \sqrt{T} + \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T},
\end{aligned}
$$

and, likewise,

$$
\begin{aligned}
&\overline{v}_h(\widehat{\mu}_T) - \widehat{\overline{\sigma}}_h(\mu_0) r / \sqrt{T} - \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&= \overline{v}_h(\widehat{\mu}_T) + \min_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) \right] - \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&\leq \min_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \\
&\leq \overline{v}_h(\widehat{\mu}_T) + \min_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) \right] + \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T} \\
&= \overline{v}_h(\widehat{\mu}_T) - \widehat{\overline{\sigma}}_h(\mu_0) r / \sqrt{T} + \delta \overline{\sigma}_h(\mu_0) / 2\sqrt{T}.
\end{aligned}
$$

Finally, note that Assumption c) of the Lemma implies that:

$$
\overline{v}_h(\widehat{\mu}_T) + \overline{\sigma}_h(\mu_0)(r - \delta) / \sqrt{T} \leq \max_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \leq \overline{v}_h(\widehat{\mu}_T) + \overline{\sigma}_h(\mu_0)(r + \delta) / \sqrt{T},
$$

and

$$
\overline{v}_h(\widehat{\mu}_T) - \overline{\sigma}_h(\mu_0)(r + \delta) / \sqrt{T} \leq \min_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \leq \overline{v}_h(\widehat{\mu}_T) - \overline{\sigma}_h(\mu_0)(r - \delta) / \sqrt{T}.
$$

An analogous argument applied to $\underline{v}_h(\mu)$ gives the desired result.

6.A.3.2 Lemma 2: Delta-method interval and a Bernstein-von Mises result

Lemma 1 in the previous subsection will be used to show that calibrating robust Bayesian credibility also calibrates frequentist coverage. We need an additional Lemma before establishing the main result.

We want to show that the posterior probability:

$$P^*\left(\left[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\right] \subseteq \left[\underline{\nu}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}, \overline{\nu}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right], \forall h = 1, \dots H \middle| Y_1, \dots, Y_T\right),$$

can be approximated by the probability that $2H$ correlated normals with unit variance fall below the threshold r. We now show that this approximation result can be applied as a consequence of the Berstein-von Mises Theorem for the reduced-form parameter μ.

Notation for this Lemma: Let P^* denote the prior distribution over μ. Let μ^* denote a random variable with such distribution. For each index $h = 1, \dots H$ define:

$$\overline{Z}_h^* = \sqrt{T}(\overline{\nu}_h(\mu^*) - \overline{\nu}_h(\widehat{\mu}_T)), \quad \underline{Z}_h^* = \sqrt{T}(\underline{\nu}_h(\mu^*) - \underline{\nu}_h(\widehat{\mu}_T)).$$

Let the remainder of the first-order Taylor approximation of $\overline{\nu}_h$ and $\underline{\nu}_h$ be defined as:

$$\overline{\Delta}_h(\mu, \mu_0) \equiv \overline{\nu}_h(\mu) - \overline{\nu}_h(\mu_0) - \dot{\overline{\nu}}_0(\mu_0)'(\mu - \mu_0),$$

and

$$\underline{\Delta}_h(\mu, \mu_0) \equiv \underline{\nu}_h(\mu) - \underline{\nu}_h(\mu_0) - \dot{\underline{\nu}}_0(\mu_0)'(\mu - \mu_0).$$

Let $Z \sim \mathcal{N}_d(\mathbf{0}, \Omega)$. Given a radius $r > 0$ and a constant $\eta > 0$ define the function:

$$\begin{aligned}
&\Gamma_H(r, \eta) \\
&\equiv \mathbb{P}\left(-\dot{\underline{\nu}}_h(\mu_0)'Z / \underline{\sigma}_h(\mu_0) \ \leq \ r + \eta, \text{ and } \dot{\overline{\nu}}_h(\mu_0)'Z / \overline{\sigma}_h(\mu_0) \ \leq \ r + \eta, \ \forall h\right) \\
&\quad - \mathbb{P}\left(-\dot{\underline{\nu}}_h(\mu_0)'Z / \underline{\sigma}_h(\mu_0) \ \leq \ r - \eta, \text{ and } \dot{\overline{\nu}}_h(\mu_0)'Z / \overline{\sigma}_h(\mu_0) \ \leq \ r - \eta, \ \forall h\right).
\end{aligned}$$

For a given $\epsilon > 0$ and $M > 0$ define $\eta(\epsilon, M)$ as the real-valued function such that:

$$\max_{0 \leq r \leq M} \left|\Gamma_H(r, \eta(\epsilon, M))\right| < \epsilon.$$

Lemma 2. *Let $\mathscr{B}(\mathbb{R}^d)$ denote the collection of all Borel sets in $\mathbb{R}^d$. Fix $\epsilon > 0, M > 0$. Let $M_1 \equiv M_1(\epsilon)$ be such that:*

$$\mathbb{P}(||Z|| \leq M_1(\epsilon)) = \epsilon/4, \quad Z \sim \mathcal{N}_d(\boldsymbol{0}, \Omega).$$

Suppose that for every $h = 1, \ldots H$, the bounds $\overline{v}_h$ and $\underline{v}_h$ are differentiable at μ_0 with nonzero derivative. Suppose that the data $(Y_1, Y_2, \ldots, Y_T)$ is such that:

1.
$$\sup_{B \in \mathscr{B}(\mathbb{R}^d)} \left| P^*\left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1, \ldots, Y_T\right) - \mathbb{P}(Z \in B) \right| \leq \epsilon/8,$$

where $Z \sim \mathcal{N}_d(\boldsymbol{0}, \Omega)$. Then for any $0 \leq r \leq M$, there is $T_2(\epsilon, M)$ such that for any $T \geq T_2(\epsilon, M)$ the absolute value of the difference between:

$$P^*\Big(\left[\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)\right] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0) r}{\sqrt{T}}, \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0) r}{\sqrt{T}}\right], \forall h = 1, \ldots H \Big| Y_1, \ldots, Y_T \Big)$$

and for $Z \sim \mathcal{N}_d(\boldsymbol{0}, \Omega)$

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0) r \leq \dot{\underline{v}}_h(\mu_0)' Z, \text{ and } \dot{\overline{v}}_h(\mu_0)' Z \leq \overline{\sigma}_h(\mu_0) r, \forall h = 1, \ldots H\right), \tag{6.18}$$

is smaller than ϵ. This means that the credibility of the delta-method region can be approximated by Equation (6.18).

Proof. We prove the lemma in two parts. The first part establishes an upper bound and the second one establishes a lower bound.

Part 1: We are interested in the posterior probability:

$$P^*\Big(\left[\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)\right] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0) r}{\sqrt{T}}, \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0) r}{\sqrt{T}}\right], \forall h \leq H \text{ and } ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T} \;\Big|\; Y_1, \ldots, Y_T \Big),$$

which is the same as:

$$P^*\left(-r\underline{\sigma}_h(\mu_0) \leq \underline{Z}_h^*, \overline{Z}_h^* \leq r\overline{\sigma}_h(\mu_0), \forall h \leq H, ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T} \;\Big|\; Y_1, \ldots, Y_T\right),$$

Write:

$$\begin{aligned}
&\overline{v}_h(\mu^*) - \overline{v}_h(\mu_0) \\
&= \overline{\Delta}(\mu^*, \mu_0) - \overline{\Delta}(\widehat{\mu}_T, \mu_0) + \dot{\overline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T) \\
&= \frac{\overline{\Delta}(\mu^*, \mu_0)}{||\mu^* - \mu_0||}||\mu^* - \mu_0|| - \frac{\overline{\Delta}(\widehat{\mu}_T, \mu_0)}{||\widehat{\mu}_T - \mu_0||}||\widehat{\mu}_T - \mu_0|| + \dot{\overline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T)
\end{aligned}$$

Note that if $||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T}$, then $||\mu^* - \mu_0|| \leq 2M_1/\sqrt{T}$. The differentiability assumption implies that there is $T_2(\epsilon, M)$ large enough such that for $T \geq T_2(\epsilon, M)$:

$$||\mu - \mu_0|| \leq 2M_1/\sqrt{T} \implies \left|\overline{\Delta}_h(\mu, \mu_0)/||\mu - \mu_0|| \right| < \eta(\epsilon/2, M)\overline{\sigma}_h(\mu_0)/(4M_1)$$

for all $h = 1, \ldots H$, and

$$||\mu - \mu_0|| \leq 2M_1/\sqrt{T} \implies \left|\underline{\Delta}_h(\mu, \mu_0)/||\mu - \mu_0|| \right| < \eta(\epsilon/2, M)\underline{\sigma}_h(\mu_0)/(4M_1)$$

for all $h = 1, \ldots H$. Therefore:

$$-\eta(\epsilon/2, M)\overline{\sigma}_h(\mu_0) + \dot{\overline{v}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T) \leq \overline{Z}_n^*.$$

An analogous argument for $\underline{v}_h(\mu^*)$ implies that

$$\underline{Z}_n^* \leq \eta(\epsilon/2, M)\underline{\sigma}_h(\mu_0) + \dot{\underline{v}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T).$$

Consequently the posterior probability we are interested in, which can be written as:

$$P^*\left(-r\underline{\sigma}_h(\mu_0) \leq \underline{Z}_h^* \text{ and } \overline{Z}_h^* \leq r\overline{\sigma}_h(\mu_0),\ \forall h \leq H,\ ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T} \,\middle|\, Y_1, \ldots, Y_T\right),$$

is smaller than or equal:

$$\begin{aligned}
P^*\Big((-r - \eta(\epsilon/2, M))\underline{\sigma}_h(\mu_0) \ &\leq \ \dot{\underline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T) \\
\text{and } \dot{\overline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T) \ &\leq \ (r + \eta(\epsilon/2, M))\overline{\sigma}_h(\mu_0),\ \forall h \ \leq \ H \,\Big|\, Y_1, \ldots, Y_T\Big).
\end{aligned}$$

By Assumption 1 of the Lemma 2, the latter probability is at most:

$$\mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \leq r + \eta\left(\frac{\epsilon}{2}, M\right) \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \leq r + \eta\left(\frac{\epsilon}{2}, M\right)\right) + \frac{\epsilon}{8}.$$

Assumption 4 of the Lemma (and a further application of Assumption 3 to $P^*(||\sqrt{T}\mu^* - \widehat{\mu}_T|| \leq M_1 \mid Y_1, \ldots, Y_T)$) implies that:

$$P^*\left(\left[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\right] \subseteq \left[\underline{\nu}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\,,\, \overline{\nu}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right], \forall h \leq H \;\middle|\; Y_1, \ldots, Y_T\right),$$

$$\leq$$

$$\mathbb{P}\Big(-\dot{\underline{\nu}}_h(\mu_0)'Z / \underline{\sigma}_h(\mu_0) \;\leq\; r + \eta(\epsilon/2, M), \\ \text{and } \dot{\overline{\nu}}_h(\mu_0)'Z / \overline{\sigma}_h(\mu_0) \;\leq\; r + \eta(\epsilon/2, M)\Big) + \epsilon/2. \tag{6.19}$$

This gives an upper bound to the 'credibility'of the delta-method region.

Part 2: We now derive a lower bound. Start with the probability

$$\mathbb{P}\Big(-\dot{\underline{\nu}}_h(\mu_0)'Z / \underline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M), \\ \text{and } \dot{\overline{\nu}}_h(\mu_0)'Z / \overline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M), \forall h \;\leq\; H\Big) - \epsilon/2.$$

Note that the latter probability is smaller than or equal $-3\epsilon/8$ plus:

$$P^*\Big(-\frac{\dot{\underline{\nu}}_h(\mu_0)'}{\underline{\sigma}_h(\mu_0)}\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \\ \text{and } \frac{\dot{\overline{\nu}}_h(\mu_0)'}{\overline{\sigma}_h(\mu_0)}\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \forall h \;\leq\; H \;\Big|\; Y_1, \ldots, Y_T\Big),$$

by an application of Assumption 3 of the Lemma. Moreover, the latter probability is bounded above by:

$$P^*\Bigg(-\frac{\dot{\underline{\nu}}_h(\mu_0)'}{\underline{\sigma}_h(\mu_0)}\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \\ \text{and } \frac{\dot{\overline{\nu}}_h(\mu_0)'}{\overline{\sigma}_h(\mu_0)}\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \forall h \;\leq\; H, \text{ and} \\ ||\mu^* - \widehat{\mu}_T|| \;\leq\; \frac{M_1}{\sqrt{T}} \;\Big|\; Y_1, \ldots, Y_T\Bigg),$$

by an application of Assumptions 3 and 4 of the Lemma and the monotonicity of probability measures. Finally, Assumption 1 and the differentiability of $\overline{\nu}_h$ at μ_0 implies that

$$\overline{Z}_n^* \leq \dot{\overline{\nu}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T) + \eta(\epsilon/2, M)\overline{\sigma}_h(\mu_0).$$

Assumption 2 and analogous argument for $\underline{v}_h(\mu^*)$ implies that

$$-\eta(\epsilon/2, M)\underline{\sigma}_h(\mu_0) + \dot{\underline{v}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T) \leq \underline{Z}_n^*.$$

Consequently,

$$\begin{aligned} P^*\Bigg(& -\frac{\dot{\underline{v}}_h(\mu_0)}{\underline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \\ & \text{and } \frac{\dot{\overline{v}}_h(\mu_0)}{\overline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^* - \widehat{\mu}_T) \;\leq\; (r - \eta(\epsilon/2, M)), \\ & \forall h \;\leq\; H,\; ||\mu^* - \widehat{\mu}_T|| \;\leq\; \frac{M_1}{\sqrt{T}} \;\Big|\; Y_1, \ldots, Y_T\Bigg), \end{aligned}$$

is bounded above by:

$$P^*\Big(-r\underline{\sigma}_h(\mu_0) \leq \underline{Z}_h^* \text{ and } \overline{Z}_h^* \leq r\overline{\sigma}_h(\mu_0),\; \forall h \leq H \;\Big|\; Y_1, \ldots, Y_T\Big).$$

We conclude that:

$$\begin{aligned} & \mathbb{P}\big(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M), \\ & \quad \text{and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M)\big) - \epsilon/2 \qquad (6.20) \\ & \qquad\qquad \leq \\ & P^*\Big(\Big[\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)\Big] \subseteq \Big[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}},\; \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\Big], \forall h \leq H \;\Big|\; Y_1, \ldots, Y_T\Big). \end{aligned}$$

Conclusion: Equations (6.19) and (6.20) imply that

$$\begin{aligned} & \mathbb{P}\big(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M), \\ & \quad \text{and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \;\leq\; r - \eta(\epsilon/2, M)\big) - \epsilon/2 \\ & \qquad\qquad \leq \\ & P^*\Big(\Big[\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)\Big] \subseteq \Big[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}},\; \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\Big], \forall h \leq H \;\Big|\; Y_1, \ldots, Y_T\Big) \\ & \qquad\qquad \leq \\ & \mathbb{P}\big(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \;\leq\; r + \eta(\epsilon/2, M), \\ & \quad \text{and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \;\leq\; r + \eta(\epsilon/2, M)\big) + \epsilon/2. \end{aligned}$$

These bounds, the monotonicity of the probability measure, and the definition of $\eta(\epsilon/2, M)$ imply the desired result. □

6.A.3.3 Lemma 3: Asymptotic behavior of the radius that calibrates r_T^* Bayesian credibility

The previous lemma showed that for any $0 \leq r \leq M$ the probability

$$P^*\Bigg(\Big[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\Big] \subseteq \left[\underline{\nu}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\,,\ \overline{\nu}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right], \\ \forall h = 1,\ldots,H \Big| Y_1,\ldots,Y_T\Bigg)$$

is approximately the same as

$$\mathbb{P}\big(-\underline{\sigma}_h(\mu_0)r \leq \dot{\underline{\nu}}_h(\mu_0)'Z, \text{ and } \dot{\overline{\nu}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r,\ \forall h = 1,\ldots H\big), Z \sim \mathcal{N}_d(\mathbf{0}, \Omega).$$

Let $r_{1-\alpha-\epsilon}$ denote the 'critical value' such that:

$$\mathbb{P}\big(-\underline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon} \leq \dot{\underline{\nu}}_h(\mu_0)'Z, \text{ and } \dot{\overline{\nu}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon},\ \forall h = 1,\ldots H\big) \\ = 1-\alpha-\epsilon,$$

and let $r_{1-\alpha+\epsilon}$ be defined analogously. We now show that if the sample size is large enough and data satisfies the Assumptions of Lemma 1 and Lemma 2 then:

$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon}.$$

Lemma 3. *Fix $\epsilon > 0$, $M_1, M_2, M_3 > 0$. Suppose that the data $(Y_1,\ldots,Y_T)$ is such that the Assumptions of Lemma 1 and Lemma 2 are satisfied:*

1. $||\widehat{\mu}_T - \mu_0|| \leq M_1/\sqrt{T}$,
2. *The largest eigenvalue of $\widehat{\Omega}_T$, denoted $\lambda_{\max}(\widehat{\Omega}_T)$, is smaller than M_2.*
3. $|\widehat{\overline{\sigma}}_h(\mu_0) - \overline{\sigma}_h(\mu_0)| < \eta(\epsilon/2, M_3)/(2M_3)$ *and* $|\widehat{\underline{\sigma}}_h(\mu_0) - \underline{\sigma}_h(\mu_0)| < \eta(\epsilon/2, M_3)/(2M_3)$, *for all $h = 1,\ldots H$; where η is defined as in Lemma 2.*
4. *For $Z \sim \mathcal{N}_d(\mathbf{0}, \Omega)$,*

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \Big|P^*\big(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1,\ldots,Y_T\big) - \mathbb{P}(Z \in B)\Big| \leq \epsilon/8,,$$

Then, there is $T_3(\epsilon, M_1, M_2, M_3)$ *such that for* $T \geq T_3(\epsilon, M_1, M_2, M_3)$*:*

$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon},$$

provided $r_{1-\alpha+\epsilon} < M_3$*.*

Proof. Without loss of generality, we can assume that $r_{1-\alpha+\epsilon} + \eta(\epsilon/2, M_3) < M_3$. Note that Lemma 1 implies that for $T \geq T_1(\eta(\epsilon/2, M_3), M_1, M_2, M_3)$:

$$P^*\left(\times_{h=1}^H \left[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\right] \subseteq \mathrm{CS}_T(r_{1-\alpha-\epsilon}, \lambda^H) \,\middle|\, Y_1, \ldots, Y_T\right)$$

minus

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0) r_{1-\alpha-\epsilon} \leq \dot{\underline{\nu}}_h(\mu_0)' Z, \text{ and } \dot{\overline{\nu}}_h(\mu_0)' Z \leq \overline{\sigma}_h(\mu_0) r_{1-\alpha-\epsilon}, \ \forall h = 1, \ldots H\right)$$

is bounded above by the sum of two terms. The first term is the difference between

$$P^*\left(\left[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\right] \subseteq DM_T^h(r_{1-\alpha-\epsilon}\,,\ \eta(\epsilon/2\,,\ M_3)),\ \forall h \leq H \,\middle|\, Y_1, \ldots, Y_T\right)$$

and

$$\begin{aligned}\mathbb{P}\big(&-\underline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon} + \eta(\epsilon/2, M_3)) \ \leq\ \dot{\underline{\nu}}_h(\mu_0)' Z, \\ &\text{and } \dot{\overline{\nu}}_h(\mu_0)' Z \ \leq\ \overline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon} + \eta(\epsilon/2, M_3)),\ \forall h \ =\ 1, \ldots H\big)\end{aligned}$$

The second term is the difference between the latter probability and

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}) \leq \dot{\underline{\nu}}_h(\mu_0)' Z, \text{ and } \dot{\overline{\nu}}_h(\mu_0)' Z \leq \overline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}),\ \forall h = 1, \ldots H\right).$$

Lemma 2 implies that the magnitude of the first term is bounded above by $\epsilon/2$ if $T \geq T_2(\epsilon/2, M_3)$. The definition of $\eta(\cdot)$ implies that the second term is bounded above by $\epsilon/2$. Therefore, we conclude that:

$$P^*\left(\times_{h=1}^H \left[\underline{\nu}_h(\mu^*), \overline{\nu}_h(\mu^*)\right] \subseteq \mathrm{CS}_T(r_{1-\alpha-\epsilon}, \lambda^H) \,\middle|\, Y_1, \ldots, Y_T\right) - (1 - \alpha - \epsilon) \leq \epsilon,$$

which implies that for $T \geq \max\{T_1(\eta(\epsilon/2, M_3), M_1, M_2, M_3), T_2(\epsilon/2, M_3)\}$:

$$r_{1-\alpha-\epsilon} \leq r_T^*,$$

as the credibility of the projection region is monotone in its radius. An analogous argument implies that for $T \geq T_1(\eta(\epsilon/2, M_3), M_1, M_2, M_3)$:

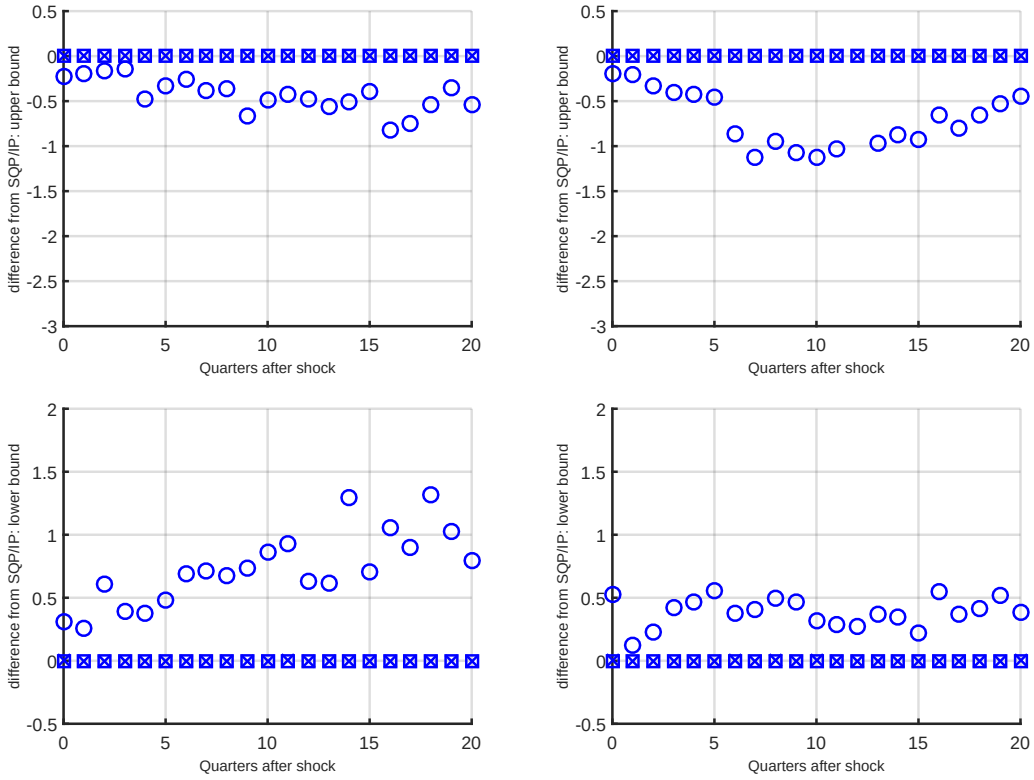
$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon}.$$

□

6.A.4 Addenda for implementation

6.A.4.1 SQP/IP vs. global methods

Figure 6.6. Accuracy of SQP/IP for a demand shock



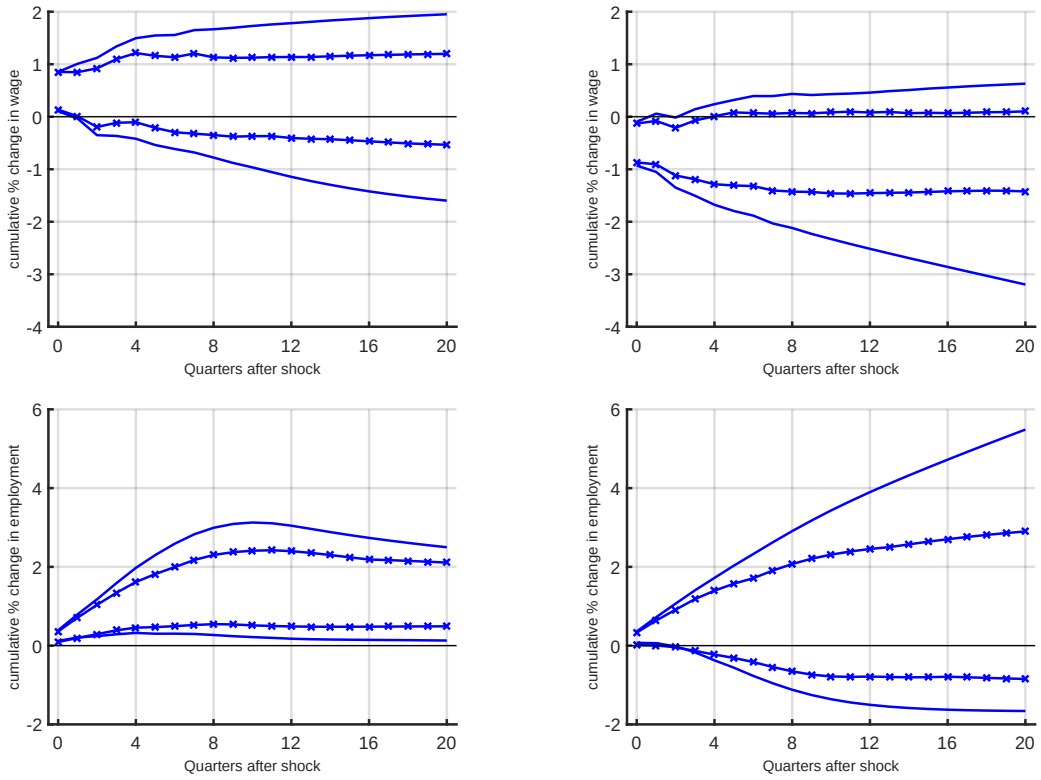
(a) Wage Response

(b) Employment Response

(Square, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by SQP/IP + `Multistart`; (Cross, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by SQP/IP + `Global Search`; (Circle, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by `ga`.

6.A.4.2 SQP/IP vs. grid search on $CS_T(1-\alpha,\mu)$

Figure 6.7. Simulation error in projection region.



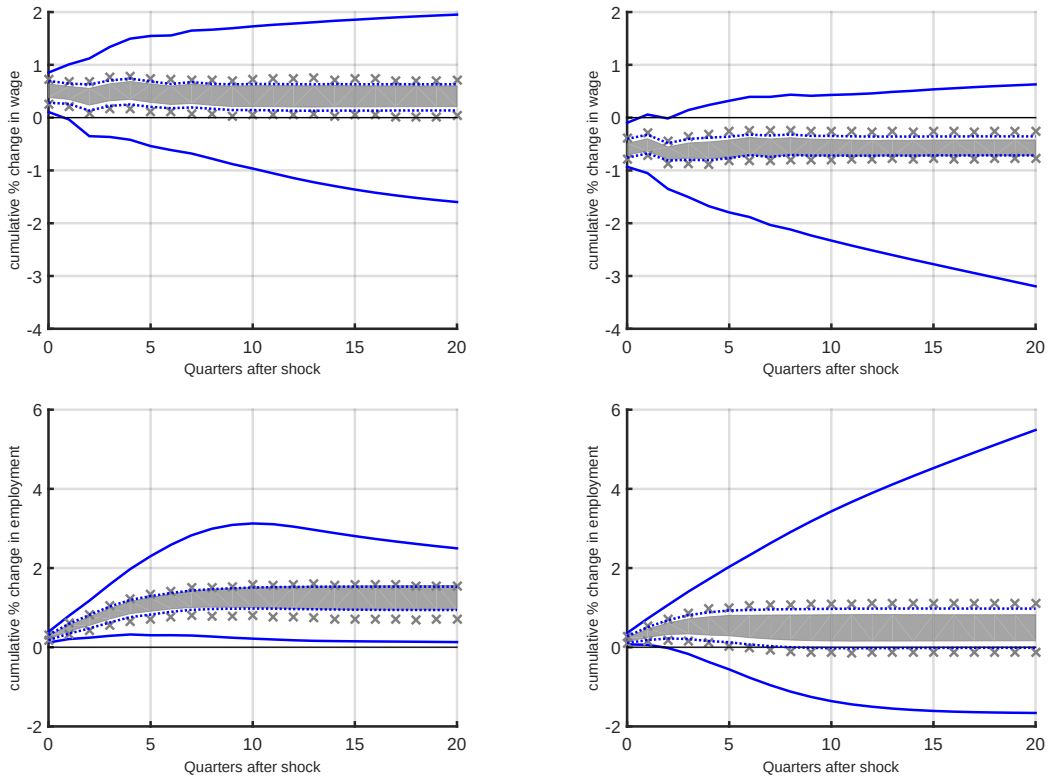
(a) Expansionary Demand Shock

(b) Expansionary Supply Shock

(Solid Line) 68% Projection region using the SQP/IP algorithm described in Section 4; (Connected, Solid Line) 68% Projection region using a two-step algorithm: 1) Sample M=100,000 reduced form parameters that satisfy the 68% Wald ellipsoid constraint. 2) For each draw, solve for the identified set. The smallest and largest value of the identified set is the simulation-based approximation of the Projection region.

6.A.4.3 Comparison with the credible set in Giacomini and Kitagawa (2015)

Figure 6.8. 68% Differentiable projection and 68% GK robust credible Set. (Uhlig (2005) priors)



(a) Expansionary Demand Shock

(b) Expansionary Supply Shock

(Solid, Blue Line) 68% Frequentist Projection Confidence Interval; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Uhlig (2005); (Dotted, Blue Line) 68% Calibrated Projection Confidence Interval. The calibration is implemented assuming differentiability of the bounds of the identified set and strict set-identification of the structural parameter; (Crosses, Gray) 68% Robust Credible Set based on Giacomini and Kitagawa (2015) using the priors for the reduced-form parameters described in Uhlig (2005).

7

References

Abel, Andrew B. and Janice C. Eberly (1996): "Optimal Investment with Costly Reversibility." *The Review of Economic Studies*, 63 (4), 581–593.

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, and Tom Winberry (2016): "No More Excuses! A Toolbox for Solving Heterogeneous Agent Models with Aggregate Shocks." SED presentation slides.

Akerlof, George A and William T Dickens (2007): "Unfinished Business in the Macroeconomics of Low Inflation: A Tribute to George and Bill by Bill and George." *Brookings Papers on Economic Activity*, 31–46.

Alfaro, Ivan, Bloom Nicholas, and Lin Xiaoji (2016): "The Finance-Uncertainty Multiplier." Mimeo.

An, Lian and Jian Wang (2012): "Exchange rate pass-through: Evidence based on vector autoregression with sign restrictions." *Open Economies Review*, 23 (2), 359–380.

Andrews, Donald WK and Gustavo Soares (2010): "Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection." *Econometrica*, 78 (1), 119–157.

Arellano, Cristina, Yan Bai, and Patrick J. Kehoe (2012): "Financial frictions and fluctuations in volatility." Staff Report 466. Federal Reserve Bank of Minneapolis.

Arias, Jonas, Juan Francisco Rubio-Ramirez, and Daniel F Waggoner (2014): "Inference Based on SVAR Identified with Sign and Zero Restrictions: Theory and Applications." *Working paper, Duke University*.

Asker, John, Allan Collard-Wexler, and Jan De Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation." *Journal of Political Economy*, 122 (5), 1013 –1063.

Bachmann, RÃijdiger and Christian Bayer (2013): "'Wait-and-See' business cycles?" *Journal of Monetary Economics*, 60 (6), 704 –719.

Bachmann, R?diger and Christian Bayer (2014): "Investment Dispersion and the Business Cycle." *American Economic Review*, 104 (4), 1392–1416.

Bachmann, R?diger, Steffen Elstner, and Eric R. Sims (2013): "Uncertainty and Economic Activity: Evidence from Business Survey Data." *American Economic Journal: Macroeconomics*, 5 (2), 217–49.

Bachmann, Ruediger and Giuseppe Moscarini (2011): "Business Cycles and Endogenous Uncertainty." 2011 Meeting Papers 36. Society for Economic Dynamics.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016): "Measuring Economic Policy Uncertainty*." *The Quarterly Journal of Economics*, 131 (4), 1593. eprint: /oup/backfile/Content_public/Journal/qje/131/4/10.1093_qje_qjw024/2/qjw024.pdf.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cunat (2012): "Search, Design, and Market Structure." *American Economic Review*, 102 (2), 1140–60.

Basu, Susanto and Brent Bundick (2015): "Uncertainty Shocks in a Model of Effective Demand." Boston College Working Papers in Economics 774. Boston College Department of Economics.

Baumeister, Christiane and Luca Benati (2013): "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound." *International Journal of Central Banking*, 9 (2), 165–212.

Baumeister, Christiane and J Hamilton (2015a): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information." *Econometrica*, 5 (83), 1963–1999.

Baumeister, Christiane and James D. Hamilton (2015b): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information." *Econometrica*, 83 (5), 1963–1999.

Beaudry, Paul, Deokwoo Nam, and Jian Wang (2014): "Do mood swings drive business cycles and is it rational?" *NBER Workind Paper (w17651)*.

Berger, J.O. (1985): *Statistical decision theory and Bayesian analysis*. Springer.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist (1999): "The financial accelerator in a quantitative business cycle framework." In *Handbook of Macroeconomics*. Ed. by J. B. Taylor and M. Woodford. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 21, 1341–1393.

Bester, Helmut and Emmanuel Petrakis (1993): "The incentives for cost reduction in a differentiated industry." *International Journal of Industrial Organization*, 11 (4), 519–534.

Bhattarai, Saroj, Gauti B Eggertsson, and Bulat Gafarov (2014): "Time Consistency and the Duration of Government Debt: A Signalling Theory of Quantitative Easing." *Working paper, Penn State University*.

Blalock, Garrick and Paul J. Gertler (2009): "How firm capabilities affect who benefits from foreign technology." *Journal of Development Economics*, 90 (2), 192–199.

Blanchard, Olivier J and Danny Quah (1989): "The dynamic effects of aggregate demand and supply disturbances." *The American Economic Review*, 79 (4), 655–673.

Bloom, Nicholas (2009): "The Impact of Uncertainty Shocks." *Econometrica*, 77 (3), 623–685.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry (2014): "Really Uncertain Business Cycles." Working Paper 14-18. US Census Bureau Center for Economic Studies.

Bonnans, J Frédéric and Alexander Shapiro (2000): *Perturbation analysis of optimization problems*. Springer.

Born, Benjamin and Johannes Pfeifer (2016): "Uncertainty-driven business cycles: assessing the markup channel." Mimeo.

Botero, Juan, Simeon Djankov, Rafael Porta, and Florencio C. Lopez-De-Silanes (2004): "The Regulation of Labor." *The Quarterly Journal of Economics*, 119 (4), 1339–1382.

Bugni, Federico, Ivan Canay, and Xiaoxia Shi (2014): "Inference for functions of partially identified parameters in moment inequality models." *Working paper, Duke University*.

Bundick, Brent and Susanto Basu (2014): "Uncertainty shocks in a model of effective demand." Research Working Paper RWP 14-15. Federal Reserve Bank of Kansas City.

Caballero, Ricardo J. and Eduardo M. R. A. Engel (1999): "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach." *Econometrica*, 67 (4), 783–826.

Caballero, Ricardo J., Eduardo M. R. A. Engel, and John Haltiwanger (1997): "Aggregate Employment Dynamics: Building from Microeconomic Evidence." *The American Economic Review*, 87 (1), 115–137.

Caggiano, Giovanni, Efrem Castelnuovo, and Nicolas Groshenny (2014): "Uncertainty shocks and unemployment dynamics in U.S. recessions." *Journal of Monetary Economics*, 67, 78–92.

Calvo, G.A. (1983): "Staggered prices in a utility-maximizing framework." *Journal of Monetary Economics*, 12 (3), 383–398.

Campbell, Jeffrey (1998): "Entry, Exit, Embodied Technology, and Business Cycles." *Review of Economic Dynamics*, 1 (2), 371–408.

Canova, Fabio and Gianni De Nicoló (2002): "Monetary disturbances matter for business fluctuations in the G-7." *Journal of Monetary Economics*, 49 (6), 1131–1159.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan (2007): "Business Cycle Accounting." *Econometrica*, 75 (3), 781–836.

Chen, Louis HY and Xiao Fang (2015): "Multivariate normal approximation by Stein's method: The concentration inequality approach." *arXiv preprint arXiv:1111.4073*.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins." *The American Economic Review, Papers and Proceedings*, 101 (3), 471–475.

Chirinko, Bob (2008): "[sigma]: The long and short of it." *Journal of Macroeconomics*, 30 (2), 671–686.

Christiano, Lawrence J, Martin Eichenbaum, and Charles Evans (1996): "The effects of monetary policy shocks: some evidence from the flow of funds." *The Review of Economics and Statistics*, 78 (1), 16–34.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy*, 113 (1), 1–45.

Christiano, Lawrence J., Massimo Rostagno, and Roberto Motto (2010): "Financial Factors in Economic Fluctuations." 2010 Meeting Papers 141. Society for Economic Dynamics.

Coibion, Olivier (2012): "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 4 (2), 1–32.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia (2012): "Innocent Bystanders? Monetary Policy and Inequality in the U.S." Working Paper 18170. National Bureau of Economic Research.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012): "Plausibly Exogenous." *The Review of Economics and Statistics*, 94 (1), 260–272.

Connault, Benjamin (2016): "A weakly-dependent Bernstein-von Mises Theorem." *Working Paper, University of Pennsylvania*.

Cooper, Russell and Jonathan Willis (2009): "The Cost of Labor Adjustment: Inferences from the Gap." *Review of Economic Dynamics*, 12 (4), 632–647.

Cooper, Russell W. and John C. Haltiwanger (2006): "On the Nature of Capital Adjustment Costs." *Review of Economic Studies*, 73 (3), 611–633.

DasGupta, A. (2008): *Asymptotic theory of statistics and probability*. Springer Verlag.

Davis, Steven J. and John Haltiwanger (1992): "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *The Quarterly Journal of Economics*, 107 (3), 819–863.

Davis, Steven J. and John Haltiwanger (2001): "Sectoral job creation and destruction responses to oil price changes." *Journal of Monetary Economics*, 48 (3), 465–512.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh (1998): *Job Creation and Destruction*. Vol. 1. MIT Press Books 0262540932. The MIT Press.

Doms, Mark E. and Timothy Dunne (1998): "Capital Adjustment Patterns in Manufacturing Plants." *Review of Economic Dynamics*, 1 (2), 409–429.

Dufour, Jean-Marie (1990): "Exact tests and confidence sets in linear regressions with autocorrelated errors." *Econometrica: Journal of the Econometric Society*, 475–494.

Dufour, Jean-Marie and Mohamed Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments." *Econometrica*, 73 (4), 1351–1365.

Dufour, Jean-Marie and Mohamed Taamouti (2007): "Further results on projection-based inference in IV regressions with weak, collinear or missing instruments." *Journal of Econometrics*, 139 (1), 133–153.

Dümbgen, Lutz (1993): "On nondifferentiable functions and the bootstrap." *Probability Theory and Related Fields*, 95 (1), 125–140.

Dyrda, Sebastian (2015): "Fluctuations in uncertainty, efficient borrowing constraints and firm dynamics." 2015 Meeting Papers 1243. Society for Economic Dynamics.

Fajgelbaum, Pablo, Edouard Schaal, and Mathieu Taschereau-Dumouchel (2014): "Uncertainty Traps." Working Paper 19973. National Bureau of Economic Research.

Fang, Zheng and Andres Santos (2015): "Inference on Directionally Differentiable Functions." *Working paper, University of California at San Diego*.

Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money." In *Carnegie-Rochester Conference Series on Public Policy*. Vol. 49. Elsevier, 207–244.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015): "The Next Generation of the Penn World Table." *American Economic Review*, 105 (10), 3150–82.

Fernald, John G. (2014): "A quarterly, utilization-adjusted series on total factor productivity." Working Paper Series 2012-19. Federal Reserve Bank of San Francisco.

Fernandez-Villaverde, JesÃžs, Pablo Guerron-Quintana, Keith Kuester, and Juan Rubio-RamÃŋrez (2015): "Fiscal Volatility Shocks and Economic Activity." *American Economic Review*, 105 (11), 3352–84.

Fiacco, Anthony V and Yo Ishizuka (1990): "Sensitivity and stability analysis for nonlinear programming." *Annals of Operations Research*, 27 (1), 215–235.

Fort, Teresa C, John Haltiwanger, Ron S Jarmin, and Javier Miranda (2013): "How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size." *IMF Economic Review*, 61 (3), 520–559.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *The American Economic Review*, 98 (1), 394–425.

Fujita, Shigeru (2011): "Dynamics of worker flows and vacancies: evidence from the sign restriction approach." *Journal of Applied Econometrics*, 26 (1), 89–121.

Gafarov, Bulat, Matthias Meier, and José Luis Montiel Olea (2015): "Delta-Method inference for a class of set-identified SVARs." Tech. rep.

Gafarov, Bulat, Matthias Meier, and Josè Luis Montiel Olea (2016): "Delta-Method Inference for a Class of Set-Identified SVARs." mimeo. Columbia University.

Gafarov, Bulat, Matthias Meier, and José Luis Montiel Olea (2016): "Projection Inference for Set-Identified SVARs." Tech. rep.

Gali, Jordi (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89 (1), 249–271.

Gertler, Mark and Peter Karadi (2015): "Monetary Policy Surprises, Credit Costs and Economic Activity." *American Economic Journal: Macroeconomics*, 7 (1), 44–76.

Gertler, Mark and Nobuhiro Kiyotaki (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis." In *Handbook of Monetary Economics*. Ed. by Benjamin M. Friedman and Michael Woodford. 1st ed. Vol. 3. Elsevier. Chap. 11, 547–599.

Ghosal, Subhashis, Jayanta K Ghosh, and Tapas Samanta (1995): "On convergence of posterior distributions." *The Annals of Statistics*, 2145–2152.

Giacomini, Rafaella and Toru Kitagawa (2015): "Robust Inference about partially identified SVARs." *Working Paper, University College London*.

Gilchrist, Simon and John C. Williams (2000): "Putty-Clay and Investment: A Business Cycle Analysis." *Journal of Political Economy*, 108 (5), 928–960.

Gilchrist, Simon and John C. Williams (2005): "Investment, Capacity, and Uncertainty: A Putty-Clay Approach." *Review of Economic Dynamics*, 8 (1), 1–27.

Gilchrist, Simon, Jae W. Sim, and Egon Zakrajšek (2013): "Misallocation and financial market frictions: Some direct evidence from the dispersion in borrowing costs." *Review of Economic Dynamics*, 16 (1). <ce:title>Special issue: Misallocation and Productivity</ce:title>, 159–176.

Gilchrist, Simon, Jae W. Sim, and Egon Zakrajšek (2014): "Uncertainty, Financial Frictions, and Investment Dynamics." Working Paper 20038. National Bureau of Economic Research.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez (2015): "Capital Allocation and Productivity in South Europe." Working Paper 21453. National Bureau of Economic Research.

Gourio, FranÃŕÂ£Â¡ois (2011): "Putty-clay technology and stock market volatility." *Journal of Monetary Economics*, 58 (2), 117–131.

Gourio, FranÃğois and Leena Rudanko (2014): "Customer Capital." *Review of Economic Studies*, 81 (3), 1102–1136.

Guglielminetti, Elisa (2013): "The Effects of Uncertainty Shocks on the Labor Market: A Search Approach." Mimeo.

Gustafson, Paul (2009): "What are the limits of posterior distributions arising from nonidentified models, and why should we care?" *Journal of the American Statistical Association*, 104 (488), 1682–1695.

Hamermesh, Daniel S (1996): *Labor demand*. Princeton University Press.

Hansen, Gary D. (1985): "Indivisible labor and the business cycle." *Journal of Monetary Economics*, 16 (3), 309 –327.

Henriquez, Claudia (2008): "Stock de Capital en Chile (1985-2005): Metodologia y Resultados." *Central Bank of Chile*, -, –.

Horst, Reiner and Panos M Pardalos (1995): *Handbook of global optimization, Nonconvex Optimization and its Applications*. Springer Science & Business Media.

Hsieh, Chang-Tai and Peter J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124 (4), 1403–1448.

Hurst, Erik and Benjamin Wild Pugsley (2011): "What Do Small Businesses Do?" NBER Working Papers 17041. National Bureau of Economic Research, Inc.

Ibragimov, Il dar Abdulovič and Rafail Z Has' minskii (2013): *Statistical estimation: asymptotic theory*. Vol. 16. Springer Science & Business Media.

Imbens, Guido W and Charles F Manski (2004): "Confidence intervals for partially identified parameters." *Econometrica*, 72 (6), 1845–1857.

Inoue, Atsushi and Lutz Kilian (2013): "Inference on impulse response functions in Structural VAR models." *Journal of Econometrics*, 177 (1), 1–13.

Inoue, Atsushi and Lutz Kilian (2016): "Joint confidence sets for structural impulse responses." *Journal of Econometrics*, 192 (2), 421–432.

Johansen, Leif (1959): "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis." English. *Econometrica*, 27 (2), pp. 157–176.

Jorda, Oscar (2005): "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95 (1), 161–182.

Jordà, Òscar (2009): "Simultaneous confidence regions for impulse responses." *The Review of Economics and Statistics*, 91 (3), 629–647.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (2015): "Measuring Uncertainty." *American Economic Review*, 105 (3), 1177–1216.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2010): "Investment shocks and business cycles." *Journal of Monetary Economics*, 57 (2), 132 –145.

Kaboski, Joseph P. (2005): "Factor price uncertainty, technology choice and investment delay." *Journal of Economic Dynamics and Control*, 29 (3), 509–527.

Kaido, Hiroaki, Francesca Molinari, and Joerg. Stoye (2016): "Inference for Projections of Identified Sets." *Working Paper, Boston University*.

Kalouptsidi, Myrto (2014): "Time to Build and Fluctuations in Bulk Shipping." *American Economic Review*, 104 (2), 564–608.

Karabarbounis, Loukas and Brent Neiman (2013): "The Global Decline of the Labor Share." *The Quarterly Journal of Economics*. eprint: http://qje.oxfordjournals.org/content/early/2013/10/24/qje.qjt032.full.pdf+html.

Kehrig, Matthias (2015): "The Cyclical Nature of the Productivity Distribution." Working Papers. Center for Economic Studies, U.S. Census Bureau.

Kehrig, Matthias and Nicolas Vincent (2016): "Do Firms Mitigate or Magnify Capital Misallocation? Evidence from Plant-Level Data." mimeo. University of Texas at Austin.

Khan, Aubhik and Julia K. Thomas (2008): "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics." *Econometrica*, 76 (2), 395–436.

Kilian, Lutz and Yun Jung Kim (2009): "Do Local Projections Solve the Bias Problem in Impulse Response Inference?" CEPR Discussion Papers 7266. C.E.P.R. Discussion Papers.

Kilian, Lutz and Daniel P. Murphy (2012a): "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models." *Journal of the European Economic Association*, 10 (5), 1166–1188.

Kilian, Lutz and Daniel P Murphy (2012b): "Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models." *Journal of the European Economic Association*, 10 (5), 1166–1188.

Kitagawa, Toru (2012): "Estimation and inference for set-identified models using lower posterior probability." *Working Paper, University College London*.

Kitagawa, Toru; Jonathan Payne, and José Luis Montiel Olea (2016): "Calibrated Inference for nondifferentiable functions." *Working paper, New York University*.

Kline, Brendan and Elie Tamer (2015): "Bayesian Inference in a class of Partially Identified Models." *Working paper, Harvard University*.

Krause, Michael U., David Lopez-Salido, and Thomas A. Lubik (2008): "Inflation dynamics with search frictions: A structural econometric analysis." *Journal of Monetary Economics*, 55 (5), 892–916.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011): "The effects of quantitative easing on interest rates: channels and implications for policy." *Brookings Papers on Economic Activity*.

Krusell, Per, Anthony A. Smith, and Jr. (1998): "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106 (5), 867–896.

Kurmann, AndrÃĽ and Nicolas Petrosky-Nadeau (2007): "Search Frictions in Physical Capital Markets as a Propagation Mechanism." Cahiers de recherche 0712. CIRPEE.

Kydland, Finn E and Edward C Prescott (1982): "Time to Build and Aggregate Fluctuations." *Econometrica*, 50 (6), 1345–70.

Leduc, Sylvain and Zheng Liu (2014): "Uncertainty shocks are aggregate demand shocks." Working Paper Series 2012-10. Federal Reserve Bank of San Francisco.

Leibenstein, Harvey (1966): "Allocative efficiency vs." X-efficiency"." *The American Economic Review*, 392–415.

Lichter, Andreas, Andreas Peichl, and Sebastian Siegloch (2014): "The own-wage elasticity of labor demand: A meta-regression analysis." *ZEW-Centre for European Economic Research Discussion Paper* (14-016).

Ludvigson, Sydney C., Sai Ma, and Serena Ng (2015): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" Working Paper 21803. National Bureau of Economic Research.

Lütkepohl, Helmut (1990): "Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models." *The Review of Economics and Statistics*, 116–125.

Lütkepohl, Helmut (2007): *New Introduction to Multiple Time Series Analysis*. Springer.

Lütkepohl, Helmut, Anna Staszewska-Bystrova, and Peter Winker (2015): "Confidence Bands for Impulse Responses: Bonferroni vs. Wald." *Oxford Bulletin of Economics and Statistics*, 77 (6), 800–821.

Mertens, Karel and Morten O. Ravn (2011): "Understanding the aggregate effects of anticipated and unanticipated tax policy shocks." *Review of Economic Dynamics*, 14 (1). Special issue: Sources of Business Cycles, 27 –54.

Mertens, Karel and Morten O. Ravn (2014): "A reconciliation of SVAR and narrative estimates of tax multipliers." *Journal of Monetary Economics*, 68 (S), S1–S19.

Moon, Hyungsik Roger and Frank Schorfheide (2012): "Bayesian and frequentist inference in partially identified models." *Econometrica*, 80 (2), 755–782.

Moon, Hyungsik Roger, Frank Schorfheide, and Eleonora Granziera (2013): "Inference for VARs Identified with Sign Restrictions." *University of Southern California*.

Mortensen, Dale and Christopher A. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies*, 61 (3), 397–415.

Mountford, Andrew and Harald Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics*, 24 (6), 960–992.

Nalewaik, Jeremy and Eugenio Pinto (2015): "The response of capital goods shipments to demand over the business cycle." *Journal of Macroeconomics*, 43, 62 –80.

Nocedal, Jorge and Stephen Wright (2006): *Numerical optimization*. Second Edition. Springer Science & Business Media.

Oberfield, Ezra and Devesh Raval (2014): "Micro Data and Macro Technology." Working paper. Princeton University.

Oh, Hyunseung and Chamna Yoon (2016): "Time to build and the real-options channel of residential investment." mimeo. Vanderbilt University.

Ottonello, Pablo (2015): "Capital Unemployment, Financial Shocks, and Investment Slumps." Tech. rep. University of Michigan.

Peters, Michael (2013): "Heterogeneous Mark-Ups and Endogenous Misallocation." mimeo. LSE.

Petrella, Ivan and Emiliano Santoro (2012): "Inflation dynamics and real marginal costs: New evidence from U.S. manufacturing industries." *Journal of Economic Dynamics and Control*, 36 (5), 779 –794.

Piketty, T. (2014): *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, Thomas (2011): "On the Long-Run Evolution of Inheritance: France 1820âĂŞ2050." *The Quarterly Journal of Economics*, 126 (3), 1071–1131. eprint: http://qje.oxfordjournals.org/content/126/3/1071.full.pdf+html.

Poirier, Dale J (1998): "Revising beliefs in nonidentified models." *Econometric Theory*, 14 (04), 483–509.

Prescott, Edward C. (1986): "Theory ahead of business-cycle measurement." *Carnegie-Rochester Conference Series on Public Policy*, 25, 11 –44.

Qu, Zhongjun and Denis Tkachenko (2015): "Global Identification in DSGE Models Allowing for Indeterminacy." *Working Paper, Boston University*.

Raddatz, Claudio (2006): "Liquidity needs and vulnerability to financial underdevelopment." *Journal of Financial Economics*, 80 (3), 677–722.

Ramey, Valerie A. and Matthew D. Shapiro (1998): "Costly capital reallocation and the effects of government spending." *Carnegie-Rochester Conference Series on Public Policy*, 48 (1), 145–194.

Ramey, Valerie A. and Daniel J. Vine (2010): "Oil, Automobiles, and the U.S. Economy: How Much have Things Really Changed?" Working Paper 16067. National Bureau of Economic Research.

Raval, Devesh (2014): "The Micro Elasticity of Substitution and Non Neutral Technology." Working paper. Federal Trade Commission.

Reichling, Felix and Charles Whalen (2012): "Review of estimates of the frisch elasticity of labor supply." *Working Paper, Congressional Budget Office*.

Reiter, Michael (2009): "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control*, 33 (3), 649 –665.

Restuccia, Diego and Richard Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics*, 11 (4), 707–720.

Restuccia, Diego and Raul Santaeulalia-Llopis (2015): "Land Misallocation and Productivity." Working Papers tecipa-541. University of Toronto, Department of Economics.

Riegler, Markus (2014): "The Impact of Uncertainty Shocks on the Job-Finding Rate and Separation Rate." 2014 Papers pri337. Job Market Papers.

Rogerson, Richard (1988): "Indivisible labor, lotteries and equilibrium." *Journal of Monetary Economics*, 21 (1), 3 –16.

Romeijn, H. Edwin and Panos M Pardalos (2013): *Handbook of global optimization*. Vol. 2. Springer Science & Business Media.

Romer, Christina D. and David H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94 (4), 1055–1084.

Rubio-Ramirez, Juan, Dario Caldara, and Jonas Arias (2015): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure." *Working Paper, Board of Governors of the Federal Reserve*.

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

Schaal, Edouard (2012): "Uncertainty, Productivity and Unemployment in the Great Recession." Mimeo.

Scheffé, Henry (1953): "A method for judging all contrasts in the analysis of variance*." *Biometrika*, 40 (1-2), 87–110.

Schmitt-Grohe, Stephanie and Martin Uribe (2004): "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control*, 28 (4), 755–775.

Scholl, Almuth and Harald Uhlig (2008): "New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates." *Journal of International Economics*, 76 (1), 1–13.

Sedláček, Petr (2014): "Match efficiency and firms' hiring standards." *Journal of Monetary Economics*, 62 (C), 123–133.

Sedláček, Petr (2016): "The aggregate matching function and job search from employment and out of the labor force." *Review of Economic Dynamics*, 21, 16–28.

Shapiro, A (1990): "On concepts of directional differentiability." *Journal of optimization theory and applications*, 66 (3), 477–487.

Shapiro, Alexander (1991): "Asymptotic analysis of stochastic programs." *Annals of Operations Research*, 30 (1), 169–186.

Sims, Christopher A (1980): "Macroeconomics and reality." *Econometrica*, 48 (1), 1–48.

Sims, Christopher A (1986): "Are forecasting models usable for policy analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 10 (1), 2–16.

Solow, Robert M. (1956): "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics*, 70 (1), 65–94. eprint: http://qje.oxfordjournals.org/content/70/1/65.full.pdf+html.

Terry, Stephen J. (2015): "Alternative Methods for Solving Heterogeneous Firm Models." mimeo. Boston University.

Uhlig, Harald (1994): "What macroeconomists should know about unit roots: a Bayesian perspective." *Econometric Theory*, 10 (3-4), 645–671.

Uhlig, Harald (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics*, 52 (2), 381–419.

United States Bureau of the Census (2015): "Quarterly Workforce Indicators Release R2013q4." [Computer file]. Ithaca, NY, USA: Cornell University, Labor Dynamics Institute [distributor].

Vargas-Silva, Carlos (2008): "Monetary policy and the US housing market: A VAR analysis imposing sign restrictions." *Journal of Macroeconomics*, 30 (3), 977–990.

Vavra, Joseph (2014): "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation." *The Quarterly Journal of Economics*, 129 (1), 215–258.

Vial, Virginie (2006): "New Estimates of Total Factor Productivity Growth in Indonesian Manufacturing." *Bulletin of Indonesian Economic Studies*, 42 (3), 357–369. eprint: http://www.tandfonline.com/doi/pdf/10.1080/00074910601053227.

Winberry, Thomas (2016a): "A Toolbox for Solving and Estimating Heterogeneous Agent Macro Models." mimeo. Chicago Booth.

Winberry, Thomas (2016b): "Lumpy Investment, Business Cycles, and Stimulus Policy." mimeo. Chicago Booth.