

ESSAYS ON
PREFERENCES AND NOMINAL RIGIDITIES
AND ON
MACROECONOMIC FORECASTING

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Diplom-Volkswirtin

Inske Pirschel

aus Salzwedel

Zürich, 2016

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Prof. Dr. Till Requate
Erstberichterstattender: Prof. Dennis J. Snower, Ph.D.
Zweitberichterstattender: Prof. Dr. Kai Carstensen

Tag der Abgabe der Arbeit: 29. Februar 2016
Tag der mündlichen Prüfung: 10. November 2016

Acknowledgments

I would like to thank my first supervisor, Prof. Dennis J. Snower, for giving me guidance and advise during my time as his doctoral student and for approving me the financial resources to present my research on various national and international conferences and to visit a summer school.

I also want to thank Prof. Kai Carstensen for taking over the duty to act as my second supervisor and for offering me the opportunity to present and discuss my research in his research seminar.

Moreover, I am greatly indebted to Prof. Maik Wolters, the chair of my examination committee, who has invited me to work together with him on my first empirical research project and who has supported me with valuable comments on my subsequent research ideas.

Large parts of the final paper of this thesis were written during my stay at the Research Center of the Deutsche Bundesbank and I want to thank Dr. Christian Schumacher, the deputy head of the Research Center, for giving me the opportunity to collaborate with him. I greatly benefited not only from working together with him on our joint projects but also from all our inspiring discussions and his advice.

I would also like to thank all my former colleagues at Kiel University and the Kiel Institute for the World Economy and in particular Dr. Steffen Ahrens for very helpful comments and interesting exchange of ideas.

Finally, I want to thank my fiancé Dr. Tim Schwarzmüller, my parents and my sister for their support over the last years.

Contents

Introduction	1
1 A Theory of Price Adjustment under Loss Aversion	3
1.1 Introduction	4
1.2 Relation to the Literature	6
1.3 Model	10
1.3.1 Consumers	11
1.3.2 Monopolistic Firms	14
1.3.3 Demand Shocks	16
1.3.4 Calibration	17
1.4 Results	18
1.4.1 Intuition	19
1.4.2 Sensitivities	24
1.5 Conclusion	25
References	25
2 Path-Dependent Wage Responsiveness	37
2.1 Introduction	38
2.2 Relation to the Literature	40
2.3 Model	43
2.3.1 Labor Supply Curve of the Loss Averse Worker	44
2.3.2 The Firm's optimization problem	46
2.3.3 Demand Shocks	48
2.3.4 Calibration	49
2.4 Results	50
2.4.1 Intuition	51
2.4.2 Sensitivities	55
2.5 Conclusion	59
References	59
3 Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?	69
3.1 Introduction	70
3.2 Forecasting Models	74
3.2.1 Large Bayesian VAR (LBVAR)	75
3.2.2 Factor Models (FAAR, FAVAR, BFAVAR, DF)	76
3.2.3 Model Averaging (EWA, BMA)	78
3.2.4 Benchmark Model (AR)	78

3.3	Data	78
3.4	Forecasting Approach	79
3.5	Results	81
3.5.1	Forecasting 11 German Key Macroeconomic Variables Jointly	81
3.5.2	Forecasting Performance for the Individual Variables	84
3.5.3	Robustness with Respect to Alternative Model Specifi- cations	90
3.6	Conclusion	96
	References	97
4	Forecasting Euro Area Recessions in Real-Time	102
4.1	Introduction	103
4.2	The Euro Area Business Cycle	107
4.3	Data	110
4.4	Forecasting Approaches	111
4.4.1	The Bayesian Mixed-Frequency VAR	111
4.4.2	A Quarterly Bayesian VAR	114
4.4.3	Markov-Switching Models	115
4.4.4	A Google Trends Real-Time Recession Indicator for the Euro Area	118
4.5	Evaluation of Probabilistic Recession Forecasts	118
4.6	Results	121
4.7	Robustness	130
4.8	Conclusion	134
	References	136
	Appendix	142
	Affirmation/Eidesstattliche Erklärung	149

List of Figures

1.1	Initial steady state.	16
1.2	Shock-arc elasticities to temporary and permanent demand shocks.	19
1.3	Shock-arc elasticities to positive and negative demand shocks.	22
1.4	Sensitivity with respect to the loss aversion parameter.	24
2.1	Initial problem of the monopsonistic firm.	48
2.2	Shock-arc elasticities to positive and negative demand shocks.	50
2.3	Classification of negative labor demand shocks.	53
2.4	Classification of positive labor demand shocks.	55
2.5	Sensitivity with respect to the loss aversion parameter.	56
2.6	Sluggishness from the direct <i>demand and supply effect</i>	57
2.7	Sensitivity with respect to the Frisch elasticity of labor supply.	58
3.1	German key macro variables.	82
3.2	Great Recession GDP growth forecasts.	89
4.1	Euro area economic activity.	109
4.2	Ideal <i>ROC</i> curve.	121
4.3	Real-time recession signals for the euro area.	123
4.4	<i>ROC</i> curves and scores.	127
4.5	Real-time recession signals for the euro area, MFBVAR robustness.	147
4.6	Real-time recession signals for the euro area, Pool robustness.	148

List of Tables

1.1	Base calibration.	17
2.1	Base calibration.	50
3.1	Forecasting a Set of 11 German Key Variables.	83
3.2	Univariate MSEs.	86
3.3	Forecasting German GDP growth, excluding the Great Recession of 2008/2009.	90
3.4	Parameter range to determine ex post optimized specification of forecasting models.	91
3.5	Absolute multivariate MSEs with IC, PBC and PBTV.	92
3.6	Multivariate MSEs with IC, PBRT and Pooling.	95
4.1	CEPR euro area business cycle phases since 2000.	107
4.2	Euro area monthly indicators with respective reporting lag in early January 2008.	111
4.3	Evaluation of real-time recession probabilities, <i>QPS</i>	122
4.4	Evaluation of real-time recession probabilities, <i>PS</i>	128
4.5	Contingency table, MFBVAR and INDCONF.	129
4.6	Robustness of MFBVAR performance.	132
4.7	Robustness of the performance of the linear opinion pools.	133
4.8	RMSFE.	143
4.9	Log score.	143
4.10	CRPS.	144

Introduction

This dissertation thesis consists of four academic papers. Each paper represents a single chapter.

Chapter 1

This chapter is based on my joint paper with Steffen Ahrens and Dennis J. Snower entitled *A Theory of Price Adjustment under Loss Aversion*, *Kiel Working Paper No. 1915, 2014 (updated in November 2015)*. My contribution to this paper includes the mathematical formulation of large parts of the theoretical model. Moreover, I contributed substantially to the interpretation of the theoretical results, the literature review and to writing the paper. The permanent download link to the original working paper is:

<https://www.ifw-members.ifw-kiel.de/publications/a-theory-of-price-adjustment-under-loss-aversion>

Chapter 2

This chapter is based on my joint paper with Steffen Ahrens and Dennis J. Snower entitled *Path-Dependent Wage Responsiveness*, *Kiel Working Paper No. 1977, 2014 (updated in November 2015)*. My contribution to this project lies in the derivation of the theoretical model and in performing the numerical simulation exercise. Moreover, I also contributed substantially to writing the paper. The original working paper is available at:

<https://www.ifw-members.ifw-kiel.de/publications/a-theory-of-wage-adjustment-under-loss-aversion>

Chapter 3

This chapter is based on my joint paper with Maik H. Wolters entitled *Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?*, Kiel Working Paper No. 1925, 2014 (updated in November 2015). My contribution to this paper includes collecting and processing the data and performing the empirical estimation of the different forecasting models, various testing procedures and the forecast evaluation. Besides, I also contributed substantially to writing the paper. The original working paper can be found under:

<https://www.ifw-members.ifw-kiel.de/publications/forecasting-german-key-macroeconomic-variables-using-large-dataset-methods>

Chapter 4

This chapter is based on my single-authored paper published under the title *Forecasting Euro Area Recessions in Real-Time*, Kiel Working Paper No. 2020, 2015. The original working paper can be downloaded at:

<https://www.ifw-members.ifw-kiel.de/publications/forecasting-euro-area-recessions-in-real-time-1>

Chapter 1

A Theory of Price Adjustment under Loss Aversion*

Abstract

We present a new partial equilibrium theory of price adjustment, based on consumer loss aversion. In line with prospect theory, the consumers' perceived utility losses from price increases are weighted more heavily than the perceived utility gains from price decreases of equal magnitude. Price changes are evaluated relative to an endogenous reference price, which depends on the consumers' rational price expectations from the recent past. By implication, demand responses are more elastic for price increases than for price decreases and thus firms face a downward-sloping demand curve that is kinked at the consumers' reference price. Firms adjust their prices flexibly in response to variations in this demand curve, in the context of an otherwise standard dynamic neoclassical model of monopolistic competition. The resulting theory of price adjustment is starkly at variance with past theories. We find that - in line with the empirical evidence - prices are more sluggish upwards than downwards in response to temporary demand shocks, while they are more sluggish downwards than upwards in response to permanent demand shocks.

Keywords: Price sluggishness, Loss aversion, State-dependent pricing.

JEL-Codes: D03, D21, E31, E50.

*This chapter is based on my joint paper with Steffen Ahrens and Dennis J. Snower entitled *A Theory of Price Adjustment under Loss Aversion*, Kiel Working Paper No. 1915, 2014 (updated in November 2015).

1.1 Introduction

This paper presents a theory of price sluggishness based on consumer loss aversion, along the lines of prospect theory (Kahneman and Tversky, 1979). The theory has distinctive implications, which are starkly at variance with major existing theories of price adjustment. In particular, the theory implies that prices are more sluggish upwards than downwards in response to temporary demand shocks, while they are more sluggish downwards than upwards in response to permanent demand shocks.

These implications turn out to be consonant with recent empirical evidence. Though this evidence has not thus far attracted much explicit attention, it is clearly implicit in a range of influential empirical results. For instance, Hall et al. (2000) document that firms mostly accommodate negative temporary demand shifts by temporary price cuts, yet they are reluctant to temporarily increase their prices in response to positive temporary demand shifts. Furthermore, the empirical evidence provided by Kehoe and Midrigan (2008) indicates that temporary price reductions are - on average - larger and much more frequent than temporary price increases, implying that prices are relatively downward responsive.

By contrast, in the event of a permanent demand shock, the empirical evidence points towards a stronger upward flexibility of prices for a wide variety of industrialized countries (Kandil, 1995, 1996, 1998, 2001, 2002b,a, 2010; Weise, 1999; Karras, 1996; Karras and Stokes, 1999) as well as developing countries (Kandil, 1998).

While current theories of price adjustment (e.g. Taylor, 1979; Rotemberg, 1982; Calvo, 1983, among many others) fail to account for these empirical regularities, this paper offers a possible theoretical rationale.

The basic idea underlying our theory is simple. Price increases are associated with utility losses for consumers, whereas price decreases are associated with utility gains. In the spirit of prospect theory, losses are weighted more heavily than gains of equal magnitude. Consequently, demand responses are more elastic to price increases than to price decreases. The result is a kinked demand curve¹, for which the kink depends on the consumers' reference

¹Modeling price sluggishness by means of a kinked demand curve is of course a well-trodden path. Sweezy (1939) and Hall and Hitch (1939) modeled price rigidity in an oligopolistic framework along these lines. In these models, oligopolistic firms do not change their prices flexibly because of their expected asymmetric competitor's reactions to their

price. In the spirit of Kőszegi and Rabin (2006), we model the reference price as the consumers' rational price expectations. We assume that consumers know, with a one period lag, whether any given demand shock is temporary or permanent. Permanent shocks induce changes in the consumers' rational price expectations and thereby in their reference price, while temporary shocks do not.

Given the demand shock is temporary, the kink of the demand curve implies that sufficiently small shocks do not affect the firm's price. This is the case of price rigidity. For larger shocks, the firm's price will respond temporarily, but the size of the response will be asymmetric for positive and negative shifts of equal magnitude. Since negative shocks move the firm along the relatively steep portion of the demand curve, prices decline stronger to negative shocks than they increase to equiproportionate positive shocks.

By contrast, given the demand shock is permanent, the firm can foresee not only the change in demand following its immediate pricing decision, but also the resulting change in the consumers' reference price. A rise in the reference price raises the firms' long-run profits (since the reference price is located at the kink of the demand curve), whereas a fall in the reference price lowers long-run profits, a phenomenon which we term the reference-price updating effect. On this account, firms are averse to initiating permanent price reductions. By implication, prices are more sluggish downwards than upwards for permanent demand shocks.

These results are extremely important for the conduct of monetary policy, since they imply that the sign of the asymmetry of price adjustment depends on the persistence of the underlying demand shock. In particular, if temporary demand shocks are interpreted as non-persistent and permanent demand shocks as fully persistent, our analysis implies that there exists a balance point (i.e. an intermediate degree of persistence of the shock) at which the asymmetry reverses. For shocks less persistent than the balance point prices are more sluggish upwards than downwards, while they are more sluggish downwards than upwards for more persistent shocks. Whether the degree of persistence at the balance point is relatively high or low depends on the adjustment speed of the reference price and on the firm's discount factor. An increase in the adjustment speed of the reference price, as well as in

pricing decisions. A game theoretic foundation of such model is presented by Maskin and Tirole (1988).

the firm's discount factor, strengthens the role of the reference-price updating effect, increasing upward flexibility and downward sluggishness at any given positive persistence of the shock. Therefore, the balance point will be associated with a lower level of persistence. To the best of our knowledge, there is no other paper studying the ramifications of the persistence of the demand shock for asymmetric price adjustment.

The paper is structured as follows. Section 1.2 reviews the relevant literature. Section 1.3 presents our general model setup and in Section 1.4 we analyze the effects of various demand shocks on prices, both analytically and numerically. Section 1.5 concludes.

1.2 Relation to the Literature

We now consider the empirical evidence suggesting that prices respond imperfectly and asymmetrically to exogenous positive and negative shocks of equal magnitude, and that the implied asymmetry depends on whether the shock is permanent or temporary.

There is much empirical evidence for the proposition that, with regard to permanent demand shocks, prices are generally more responsive to positive shocks than to negative ones. For example, in the context of monetary policy shocks, Kandil (1996, 2002a), Kandil (1995), and Weise (1999) find support for the United States over a large range of different samples. Moreover, Kandil (1995) and Karras and Stokes (1999) supply evidence for large panels of industrialized OECD countries, while Karras (1996) provides evidence for developing countries. In the case of the United States, Kandil (2001, 2002b) shows that the asymmetry also prevails in response to permanent government spending shocks. Kandil (1999, 2006, 2010), on the other hand, looks directly at permanent aggregate demand shocks and also confirms the asymmetry for a large set of industrialized countries as well as for a sample of disaggregated industries in the United States. Comparing a large set of industrialized and developing countries, Kandil (1998) finds that the asymmetry is even stronger for many developing countries compared to industrialized ones.

In addition to the asymmetric price reaction in response to permanent demand shocks, the above studies also find an asymmetric reaction of output. They show that output responds significantly less to permanent pos-

itive demand shocks relative to negative ones. This asymmetry – which is also predicted by our model (as shown below) – is further documented by a large body of empirical literature that explicitly focuses on output. For example, DeLong and Summers (1988), Cover (1992), Thoma (1994), and Ravn and Sola (2004) show for the United States that positive changes in the rate of money growth induce much weaker output reductions than negative changes in the rate of money supply. Morgan (1993) and Ravn and Sola (2004) confirm this asymmetry, when monetary policy is conducted via changes in the federal funds rate. Additional evidence is provided by Tan et al. (2010) for Indonesia, Malaysia, the Philippines, and Thailand and by Mehrara and Karsalari (2011) for Iran.

There is also significant empirical evidence for the proposition that, with regard to temporary demand shocks, prices are generally less responsive to positive shocks than to negative ones. For example, the survey by Hall et al. (2000) indicates that firms regard price increases as response to temporary increases in demand to be among the least favorable options. Instead, firms rather employ more workers, extend overtime work, or increase capacities. By contrast, managers of firms state that a temporary fall in demand is much more likely to lead to a price cut. Further evidence for the asymmetry in response to temporary demand shocks is provided by Kehoe and Midrigan (2008), who analyze temporary price movements at Dominick's Finer Foods retail chain with weekly store-level data from 86 stores in the Chicago area. They find that temporary price reductions are much more frequent than temporary price increases and that, on average, temporary price cuts are larger (by a factor of almost two) than temporary price increases. However neither of these studies empirically analyzes the asymmetry characteristics of the output reaction in the face of temporary demand shocks.

Despite this broad evidence, asymmetric reactions to demand shocks have been unexplored by current theories of price adjustment. Neither time-dependent pricing models (Taylor, 1979; Calvo, 1983), nor state-dependent adjustment cost models of (S, s) type (e.g., Sheshinski and Weiss, 1977; Rotemberg, 1982; Caplin and Spulber, 1987; Caballero and Engel, 1993, 2007; Golosov and Lucas Jr., 2007; Gertler and Leahy, 2008; Dotsey et al., 2009; Midrigan, 2011) are able to account for the asymmetry properties in price

dynamics in response to positive and negative exogenous temporary and permanent shifts in demand.²

In this paper we offer a new theory of firm price setting resting on consumer loss aversion in an otherwise standard model of monopolistic competition. The resulting theory provides a novel rationale for the above empirical evidence on asymmetric price sluggishness. Although there is no hard evidence for a direct link from consumer loss aversion to price sluggishness, to the best of our knowledge, there is ample evidence that firms do not adjust their prices flexibly in order to avoid harming their customer relationships (see, e.g., Fabiani et al. (2006) for a survey of euro area countries, Blinder et al. (1998) for the United States³, and Hall et al. (2000) for the United Kingdom).⁴

Furthermore, there is extensive empirical evidence that customers are indeed loss averse in prices. Kalwani et al. (1990), Mayhew and Winer (1992), Krishnamurthi et al. (1992), Putler (1992), Hardie et al. (1993), Kalyanaram and Little (1994), Raman and Bass (2002), Dossche et al. (2010), and many others find evidence for consumer loss aversion with respect to many different product categories available in supermarkets. Furthermore, loss aversion in prices is also well documented in diverse activities such as restaurant visits (Morgan, 1993), vacation trips (Nicolau, 2008), real estate trade (Genesove and Mayer, 2001), phone calls (Bidwell et al., 1995), and energy use (Griffin and Schulman, 2005; Adeyemi and Hunt, 2007; Ryan and Plourde, 2007).

In our model, loss-averse consumers evaluate prices relative to a reference price. Kőszegi and Rabin (2006, 2007, 2009) and Heidhues and Kőszegi (2005, 2008, 2014) argue that reference points are determined by agents' rational expectations about outcomes from the recent past. There is much empirical evidence suggesting that reference points are determined by expectations, in

²Once trend inflation is considered, menu costs can generally explain that prices are more downward sluggish than upwards (Ball and Mankiw (1994)). By contrast, our model does not rely on the assumption of trend inflation.

³In their survey, Blinder et al. (1998) additionally find clear evidence that the pricing of those firms for which the fear of antagonizing their customers through price changes plays an important role is relatively upward sluggish. Unfortunately, the authors do not distinguish between temporary and permanent shifts in demand in their survey questions.

⁴Further evidence for OECD countries is provided by, for example, Fabiani et al. (2004) for Italy, Louprias and Ricart (2004) for France, Zbaracki et al. (2004) for the United States, Alvarez and Hernando (2005) for Spain, Amirault et al. (2005) for Canada, Aucremanne and Druant (2005) for Belgium, Stahl (2005) for Germany, Lünemann and Mathä (2006) for Luxembourg, Langbraaten et al. (2008) for Norway, Hoerberichts and Stokman (2010) for the Netherlands, Kwapil et al. (2010) for Austria, Martins (2010) for Portugal, Ólafsson et al. (2011) for Iceland, and Greenslade and Parker (2012) for the United Kingdom.

concrete situations such as in police performance after final offer arbitration (Mas, 2006), in the United States TV show “Deal or no Deal” (Post et al., 2008), with respect to domestic violence (Card and Dahl, 2011), in cab drivers’ labor supply decisions (Crawford and Meng, 2011), in the effort choices of professional golf players (Pope and Schweitzer, 2011), or in the aggressiveness of professional soccer players (Bartling et al., forthcoming). In the context of laboratory experiments, Knetsch and Wong (2009) and Marzilli Ericson and Fuster (2011) find supporting evidence from exchange experiments, Abeler et al. (2011) and Gill and Prowse (2012) from effort provision experiments, Banerji and Gupta (2014) from an auction experiment, and Karle et al. (2015) from a consumption choice experiment. Endogenizing consumers’ reference prices in this way allows our model to capture that current price changes influence the consumers’ future reference price and thereby affect the demand functions via what we call the “reference-price updating effect.” This effect rests on the observation that firms tend to increase the demand for their product by raising their consumers’ reference price through, for example, setting a “suggested retail price” that is higher than the price actually charged (Thaler, 1985; Putler, 1992). These pieces of evidence are consonant with the assumptions underlying our analysis. Our analysis works out the implications of these assumptions for state-dependent price sluggishness in the form of asymmetric price adjustment for temporary and permanent demand shocks.

There are only a few other papers that study the implications of consumer loss aversion on firms’ pricing decisions. In an early account of price rigidity in response to demand and cost shocks has been presented by Sibly (2002, 2007). In a static environment, Sibly (2002, 2007) shows that a monopolist may not change prices if she faces loss averse consumers with fixed, exogenously given reference prices. In their particularly insightful contributions, Heidhues and Köszegi (2008) and Spiegler (2012) analyze static monopolistic pricing decisions to cost and demand shocks under the assumption that the reference price is determined as a consumer’s recent rational expectations personal equilibrium in the spirit of Köszegi and Rabin (2006). Spiegler (2012) shows that incentives for price rigidity are even stronger for demand shocks compared to cost shocks. We follow Heidhues and Köszegi (2008) and Spiegler (2012) and assume endogenous rational expectations reference price formation, but, by contrast, consider a dynamic approach to the pricing de-

cision of a monopolistically competitive firm facing loss averse consumers. Our dynamic approach confirms earlier findings that consumer loss aversion engenders price rigidity and allows us to study the asymmetry characteristics of pricing reactions to temporary and permanent demand shocks of different sign. Another study close to ours is Popescu and Wu (2007); although they analyze optimal pricing strategies in repeated market interactions with loss averse consumers and endogenous reference prices, they do not analyze the model's reaction to demand shocks.

Finally, this paper offers a new microfounded rationale for state-dependent pricing. The importance of state-dependence for firms' pricing decisions is well documented. For instance, in the countries of the euro area (Fabi-ani et al., 2006; Nicolitsas, 2013), Scandinavia (Apel et al., 2005; Langbraaten et al., 2008; Ólafsson et al., 2011), the United States (Blinder et al., 1998), and Turkey (Şahinöz and Saraçoğlu, 2008), approximately two third of the firms' pricing decisions are indeed driven by the current state of the environment.⁵ Menu costs, by contrast, are clearly rejected as a significant driver for deferred price adjustments in each of the empirical studies above.

1.3 Model

We incorporate reference-dependent preferences and loss aversion into an otherwise standard model of monopolistic competition. Consumers are price takers and loss averse with respect to prices. They evaluate prices relative to their reference prices, which depend on their lagged rational price expectations from the recent past, i.e. consumers are backward-looking. For simplicity, we abstract from saving, implying that workers become single-period optimizers. Firms are monopolistic competitors, supplying non-durable differentiated goods. Firms can change their prices freely in each period to maximize their total expected discounted profits. Firms' price setting decision is forward-looking, taking into account their influence on the consumers' future reference price. Thus, reference dependence in our model is obviously an intertemporal phenomenon, linking the decisions in one period to the decisions in the next. To analyse the firms' price setting decision in response

⁵However in the United Kingdom (Hall et al., 2000) and Canada (Amirault et al., 2005) state-dependence seems to be somewhat less important for firms' pricing decision.

to demand shocks in such an intertemporal context, we consider a dynamic two-period analysis, for algebraic simplicity.⁶

1.3.1 Consumers

We follow (Sibly, 2007) and assume that the representative consumer's period-utility U_t depends positively on the consumption of n imperfectly substitutable nondurable goods $q_{i,t}$ with $i \in (1, \dots, n)$ and negatively on the "loss-aversion ratio" $(P_{i,t}/R_{i,t})$, i.e. the ratio of the price $P_{i,t}$ of good i to the consumer's respective reference price $R_{i,t}$ of the good.⁷ The loss-aversion ratio, which describes how the phenomenon of loss aversion enters the utility function, may be rationalized in terms of (i) Thaler's transaction utility (whereby the total utility that the consumer derives from a good is in part determined by how the consumer evaluates the quality of the financial terms of the acquisition of the good (Thaler, 1991)), (ii) Okun's implicit firm-customer contracts (whereby firms and customers implicitly agree on fair and stable prices despite fluctuations in demand (Okun, 1981)), or (iii) Rotemberg's customer anger or regret (Rotemberg, 2005, 2010). Further approaches that describe reference-dependence in the consumer's utility function in terms of a ratio of actual prices to reference prices are McDonald and Sibly (2001, 2005) and Ahrens et al. (2014) in the context of loss aversion with respect to wages and Sibly (1996, 2002) in the context of loss aversion with respect to prices and quality.⁸

The consumer's preferences in period t are represented by the following utility function:

$$U_t(q_{1,t}, \dots, q_{n,t}) = \left[\sum_{i=1}^n \left(\left(\frac{P_{i,t}}{R_{i,t}} \right)^{-\mu} q_{i,t} \right)^\rho \right]^{\frac{1}{\rho}}, \quad (1.1)$$

⁶A multi-period analysis with longer time horizons would not affect our qualitative conclusions.

⁷Throughout the model, capital letters denote nominal variables, while small letters denote real variables. Greek letters denote model parameters.

⁸Other examples in which prices directly enter the utility function are, for instance, Rosenkranz (2003) and Rosenkranz and Schmitz (2007) in the context of auctions and Popescu and Wu (2007), Nasiry and Popescu (2011), and Zhou (2011) in the context of customer loss aversion.

where $0 < \rho < 1$ denotes the degree of substitutability between the different goods. The parameter μ is an indicator function of the form

$$\mu = \begin{cases} \Gamma & \text{for } P_{i,t} < R_{i,t}, \text{ i.e. gain domain} \\ \Delta & \text{for } P_{i,t} > R_{i,t}, \text{ i.e. loss domain} \end{cases}, \quad (1.2)$$

which describes the degree of the consumer's loss aversion. For loss averse consumers, $\Delta > \Gamma$, i.e. the utility losses from price increases are larger than the utility gains from price decreases of equal magnitude. The consumer's reference price $R_{i,t}$ is formed at the beginning of each period. In the spirit of Kőszegi and Rabin (2006), we assume that the consumer's reference price depends on her lagged rational price expectation. Demand shocks, which may or may not trigger price adjustment, materialize unexpectedly in the course of the period and therefore do not enter the information set used by the consumer at the beginning of the period to form the reference price. Therefore, there is no instantaneous reaction of the reference price in the shock period even if the firm immediately adjusts its price in response to the shock. At the beginning of the next period, however, consumers update their information set and adjust their price expectation accordingly (since they can now infer about the nature of the demand shock and the corresponding price change). While temporary price changes do not provoke a change in the consumer's reference price⁹, the reference price changes in the period after the occurrence of a permanent shock. Thus the consumer's reference price is given by $R_{i,t} = E [P_{i,t} | I_{t-1}]$. The consumer's budget constraint is given by

$$\sum_{i=1}^n P_{i,t} q_{i,t} = \tilde{P}_t y_t, \quad (1.3)$$

where y_t denotes the consumer's real income in period t which is assumed to be constant and \tilde{P}_t is the aggregate price index. For simplicity, we abstract from saving. This implies that consumers are completely myopic.¹⁰ In each

⁹Support for this assumption can be found in the example of sales, i.e. promotions, characterized by non-permanent price decreases, used by firms to temporarily increase consumers' demand for their product (see e.g. Eichenbaum et al., 2011). Sales do not affect the consumers' reference price. Otherwise firms would not conduct sales because any downward adjustment of the consumer's reference price reduces long-run profits for the firm.

¹⁰Evidence to support this assumption is provided by Elmaghraby and Keskinocak (2003) who show that many purchase decisions of non-durable goods take place in economic environments which are characterized by myopic consumers.

period the consumer maximizes her period-utility function (1.1) with respect to her budget constraint (1.3). The result is the consumer's period t demand for the differentiated good i which is given by

$$q_{i,t}(P_{i,t}, R_{i,t}, \mu) = \tilde{P}_t^\eta \left(\frac{P_{i,t}}{R_{i,t}} \right)^{-\mu(\eta-1)} \frac{y_t}{P_{i,t}^\eta}, \quad (1.4)$$

where $\eta = \frac{1}{1-\rho}$ denotes the elasticity of substitution between the different product varieties. The aggregate price index \tilde{P}_t is given by

$$\tilde{P}_t = \left[\sum_{i=1}^n \left(P_{i,t} / \left(\frac{P_{i,t}}{R_{i,t}} \right)^{-\mu} \right)^{1-\eta} \right]^{\frac{1}{1-\eta}}. \quad (1.5)$$

We assume that the number of firms n is sufficiently large so that the pricing decision of a single firm does not affect the aggregate price index. Defining $\lambda = \eta(1 + \mu) - \mu$, we can simplify equation (1.4) to

$$q_{i,t}(P_{i,t}, R_{i,t}, \lambda) = R_{i,t}^{\lambda-\eta} P_{i,t}^{-\lambda} \tilde{P}_t^\eta y_t, \quad (1.6)$$

where the parameter λ denotes the price elasticity of demand, which depends on μ and therefore takes different values for losses and gains. To simplify notation, we define

$$\lambda = \begin{cases} \gamma & \text{for } P_{i,t} < R_{i,t} \\ \delta & \text{for } P_{i,t} > R_{i,t} \end{cases}, \quad (1.7)$$

with $\delta = \eta(1 + \Delta) - \Delta > \gamma = \eta(1 + \Gamma) - \Gamma$. Equation (1.6) indicates that the consumer's demand function for good i is kinked at the reference price $R_{i,t}$. The kink, lying at the intersection of the two demand curves $q_{i,t}(P_{i,t}, R_{i,t}, \gamma)$ and $q_{i,t}(P_{i,t}, R_{i,t}, \delta)$, is given by the price-quantity combination

$$(\widehat{P}_{i,t}, \widehat{q}_{i,t}) = (R_{i,t}, R_{i,t}^{-\eta} \tilde{P}_t^\eta y_t), \quad (1.8)$$

where " $\widehat{}$ " denotes the value of a variable at the kink. Changes in the reference price $R_{i,t}$ give rise to a change of the position of the kink and also shift the demand curve as a whole. The direction of this shift depends on the sign of the difference $\lambda - \eta$. We restrict our analysis to $\lambda \geq \eta$, i.e. we assume that

an increase in the reference price shifts the demand curve outwards and vice versa.¹¹

Needless to say, abstracting from reference-dependence and loss aversion in the consumer's preferences represented by utility function (1.1), restores the standard textbook consumer demand function for a differentiated good i , given by

$$q_{i,t}(P_{i,t}) = P_{i,t}^{-\eta} \bar{P}_t^\eta y_t, \quad (1.9)$$

where \bar{P}_t is the aggregate price index in an economy without loss averse consumers. In what follows, we will use this standard model as a benchmark case, against which we compare the pricing decisions of a monopolistic competitive firm facing loss averse consumers.

1.3.2 Monopolistic Firms

Firms seek to maximize the discounted stream of current and future profits,

$$\Pi_{t,i}^{Total} = \Pi_{i,t} + \beta \Pi_{i,t+1} \quad (1.10)$$

where $\Pi_{i,t} = P_{i,t}q_{i,t} - C(q_{i,t})$ are period t profits, β is the discount factor and $C_t(q_t)$ are firm's total costs. For simplicity, we assume a two period time horizon. (This can serve as a rough approximation for forms of short-sightedness, such as hyperbolic discounting, when the first-period discount rate exceeds the second-period one.¹²) The firm takes into account its period t individual downward-sloping demand function (1.6) and the implications of its current pricing decision for the costumers' reference price. The resulting first order

¹¹The positive relationship between reference price and demand has become a common feature in the marketing sciences (e.g., Thaler, 1985; Putler, 1992; Greenleaf, 1995). It manifests itself, e.g., through the "suggested retail price," by which raising the consumers' reference price causes increases in demand (Thaler, 1985). Furthermore, Putler (1992) provides evidence that an extensive use of promotional pricing in the late 80's had lead to an erosion in demand by lowering consumers' reference prices.

¹²Many authors have shown that consumers' discount rates are generally much higher in the short run than in the long run (Loewenstein and Thaler, 1989; Ainslie, 1992; Loewenstein and Prelec, 1992; Laibson, 1996, 1997). Firm behavior is also often found to be short-sighted for the same reason. The theory of managerial myopia argues that managers often almost exclusively focus on short-term earnings (either because they have to meet certain goals or because their career advancement and compensation structure depends on the firm's current performance), even if this has adverse long-run effects (Jacobson and Aaker, 1993; Graham et al., 2005; Mizik and Jacobson, 2007; Mizik, 2010). For a review of the early literature refer to Grant et al. (1996).

condition of the firm's optimization problem reads as

$$\frac{\partial \Pi_{i,t}^{Total}}{\partial P_{i,t}} = q_{i,t} + P_{i,t} \frac{\partial q_{i,t}}{\partial P_{i,t}} - \frac{\partial C(q_{i,t})}{\partial P_{i,t}} + \beta \frac{\partial \Pi_{i,t+1}}{\partial R_{i,t+1}} \frac{\partial R_{i,t+1}}{\partial P_{i,t}} = 0, \quad (1.11)$$

which is equivalent to

$$\underbrace{q_{i,t} + P_{i,t} \frac{\partial q_{i,t}}{\partial P_{i,t}}}_{MR_t} - \underbrace{\frac{\partial C(q_{i,t})}{\partial P_{i,t}}}_{MC_t} = -\beta \frac{\partial \Pi_{i,t+1}}{\partial R_{i,t+1}} \frac{\partial R_{i,t+1}}{\partial P_{i,t}}. \quad (1.12)$$

The term on the left hand side is the current period marginal revenue MR_t minus the current period marginal cost MC_t . The term on the right hand side measures the influence of the price setting decision on the reference price and thereby future profits. Note that in the absence of reference-price-updating the standard optimality condition of a firm holds, i.e. $MR_t = MC_t$. Only if the firm's price setting decision has an influence on the reference price the firm faces a tradeoff between current period optimality (determined by the left hand side of equation (1.12)) and future ramifications of the current decision (determined by the right hand side of equation (1.12)).

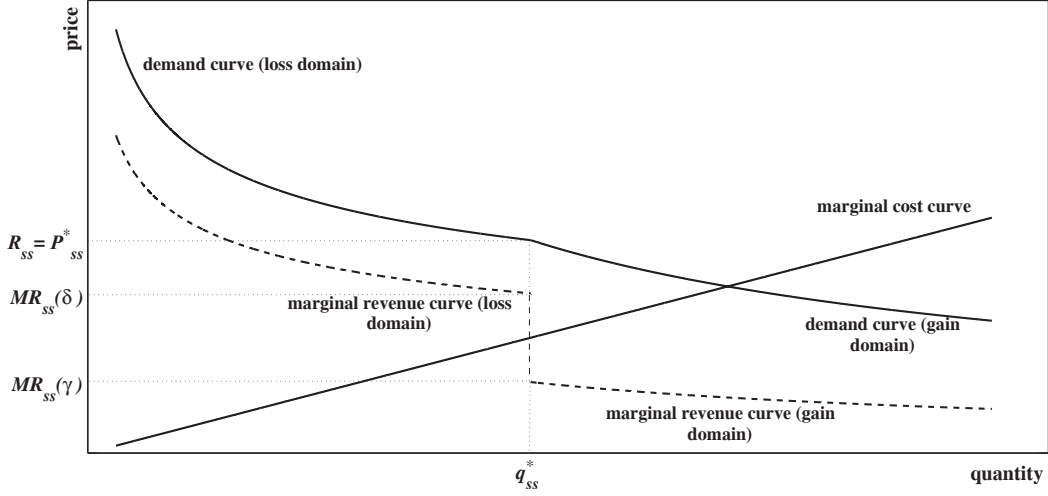
All n firms are identical, enabling us to drop the subscript i . In what follows we assume that the firm's total costs are given by $C_t(q_t) = \frac{c}{2}q_t^2$, where c is a constant, implying that marginal costs are linear in output: $MC_t(q_t) = cq_t$. In the presence of loss aversion ($\delta > \gamma$), the downward-sloping demand curve has a concave kink at the current reference price: $\hat{P}_t = R_t$. Thus the firm's marginal revenue curve is discontinuous at the kink:

$$MR_t(q_t, R_t, \lambda) = \left(1 - \frac{1}{\lambda}\right) \left(\frac{q_t}{R_t^{(\lambda-\eta)} \hat{P}_t^\eta y_t}\right)^{-\frac{1}{\lambda}}, \quad (1.13)$$

with $\lambda = \gamma$ for the gain domain and $\lambda = \delta$ for the loss domain, respectively. The interval $[MR_t(\hat{q}_t, R_t, \gamma), MR_t(\hat{q}_t, R_t, \delta)]$, where $MR_t(\hat{q}_t, R_t, \gamma) < MR_t(\hat{q}_t, R_t, \delta)$, we call "marginal revenue discontinuity" $MRD_t(\hat{q}_t, R_t, \gamma, \delta)$.

We assume that in the initial steady state, the exogenously given reference price is R_{ss} . Furthermore, in the steady state the firm's marginal cost curve intersects the marginal revenue discontinuity, as depicted in Figure 1.1. To fix ideas, we assume that initially the marginal cost curve crosses the mid-

Figure 1.1: Initial steady state.



point of the discontinuity in the marginal revenue curve.¹³ This assumption permits us to derive the symmetry characteristics of the responses to positive and negative demand shocks. This implies that the firm's optimal price in the initial steady state P_{ss}^* is equal to R_{ss} .¹⁴

1.3.3 Demand Shocks

The demand for each product i is subject to exogenous shocks, which may be temporary or permanent. These demand shocks, represented by ε_t , are unexpected and enter the demand function multiplicatively:

$$q_t(P_t, R_t, \lambda, \varepsilon_t) = R_t^{(\lambda-\eta)} P_t^{-\lambda} \tilde{P}_t^\eta y_t \varepsilon_t. \quad (1.14)$$

The corresponding marginal revenue functions of the firm are

$$MR_t(q_t, R_t, \lambda, \varepsilon_t) = \left(1 - \frac{1}{\lambda}\right) \left(\frac{q_t}{R_t^{(\lambda-\eta)} \tilde{P}_t^\eta y_t \varepsilon_t}\right)^{-\frac{1}{\lambda}}. \quad (1.15)$$

¹³To satisfy this condition, the slope parameter c of the marginal cost curve has to take the value $c = \frac{1}{2q_{ss}} [MR_t(q_{ss}, R_{ss}, \gamma) + MR_t(q_{ss}, R_{ss}, \delta)]$

¹⁴The proof is straightforward: Let ν be an arbitrarily small number. Then for prices equal to $R_{ss} + \nu$ the firm faces a situation in which marginal revenue is higher than marginal costs and decreasing the price would raise the firm's profit, while for prices equal to $R_{ss} - \nu$ the firm faces a situation in which marginal revenue is lower than marginal costs and increasing the price would raise the firm's profit. Thus $P_{ss}^* = R_{ss}$ has to be the profit maximizing price in the initial steady state.

Table 1.1: Base calibration.

Parameter	Symbol	Value
Discount rate	β	0.99
Elasticity of substitution	η	5
implying substitutability	ρ	0.8
Price elasticity (gain domain)	γ	6
Price elasticity (loss domain)	δ	12
Loss aversion	κ	2
Exogenous nominal income	Y	1
Exogenous price index	P_t	1

We consider the effects of a demand shock that hits the economy in period t . The demand shock shifts the marginal revenue curve, along with the marginal revenue discontinuity $MRD_t(\hat{q}_t, R_t, \gamma, \delta, \varepsilon_t)$. We define a “small” shock as one that leaves the marginal cost curve passing through the marginal revenue discontinuity, and a “large” shock as one that shifts the marginal revenue curve sufficiently so that the marginal cost curve no longer passes through the marginal revenue discontinuity.

The maximum size of a small shock for the demand function (1.14) is

$$\bar{\varepsilon}_t(\lambda) = \left(1 - \frac{1}{\lambda}\right) \frac{R_t^{1+\eta}}{c\bar{P}_t^\eta y_t}, \quad (1.16)$$

i.e. $\bar{\varepsilon}_t(\lambda)$ is the shock size for which the marginal cost curve lies exactly on the boundaries of the shifted marginal revenue discontinuity $MRD_t(\hat{q}_t, R_t, \gamma, \delta, \bar{\varepsilon}_t(\lambda))$.¹⁵ In the analysis that follows, we will distinguish both between small and large demand shocks and between temporary and permanent demand shocks. To qualitatively and quantitatively assess the price setting reaction of the firm we calibrate the model and simulate it numerically.

1.3.4 Calibration

We calibrate the model for a quarterly frequency in accordance with standard values in the literature. We assume an annual interest rate of 4 percent, which yields a discount factor $\beta = 0.99$. We follow Schmitt-Grohé and Uribe (2007)

¹⁵For $\bar{\varepsilon}(\delta)$, the marginal cost curve intersects the marginal revenue gap on the upper bound, whereas for $\bar{\varepsilon}(\gamma)$ it intersects it on the lower bound.

and set the monopolistic markup to 25 percent, i.e. $\eta = 5$, which is also close to the value supported by Erceg et al. (2000) and which implies that goods are only little substitutable, i.e. $\rho = 0.8$. Loss aversion is measured by the relative slopes of the demand curves in the gain and loss domain, i.e. $\kappa = \frac{\delta}{\gamma}$. The empirical literature on loss aversion in prices finds that losses induce demand reactions approximately twice as large as gains (Tversky and Kahneman, 1991; Putler, 1992; Kalyanaram and Little, 1994; Griffin and Schulman, 2005; Adeyemi and Hunt, 2007). Therefore, we set $\kappa = 2$. The loss aversion parameters from the utility function Γ and Δ are chosen to match specific price elasticities. Price elasticities are commonly values as low as 5 (e.g., Klenow and Willis, 2006) up to values as high as 11 (e.g., Kimball, 1995; Chari et al., 2000; Eichenbaum and Fisher, 2004; Woodford, 2005). Therefore, we set the price elasticity in the gain domain to $\gamma = 5.5$, which given $\kappa = 2$ implies a price elasticity in the loss domain of $\delta = 11$. Furthermore, this calibration satisfies the restriction that $\lambda \geq \eta$. The exogenous nominal income Y and price index P_t are normalized to unity.¹⁶ The base calibration is summarized in Table 1.1.

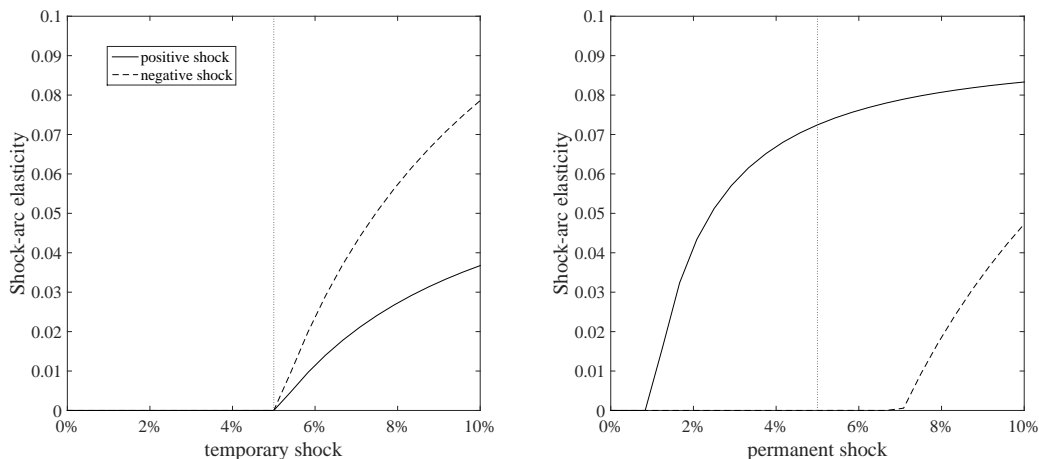
1.4 Results

Figure 2.2 present the numerical results of our base calibration in the two-period model. In the figure we show the shock-arc-elasticity of price $\left(\tilde{\eta}_{\varepsilon,P} = \frac{\% \Delta P}{\% \Delta \varepsilon}\right)$ in the period of the shock t for positive and negative temporary (left panel) and permanent (right panel) demand shocks. On the vertical axis we show the shock-arc elasticities of price, which measure the relative strength of the price reaction in response to demand shocks. The horizontal axis measures the size of the shock in percent. The vertical, dotted lines denote the thresholds between small and the large demand shocks.

Our numerical analysis finds that the firm's price reaction in response to demand shocks depends crucially on the size, the sign, and the persistence (temporary vs. permanent) of the shock. The left panel of Figure 2.2 shows that prices are completely rigid for small positive and negative temporary demand shocks (to the left of the dotted lines), while they are relatively upward sluggish for large shocks (to the right of the dotted lines). By contrast,

¹⁶All results are completely robust to variations of these numerical values.

Figure 1.2: Shock-arc elasticities to temporary and permanent demand shocks.



the right panel of Figure 2.2 shows that in response to permanent positive demand shocks prices are generally downward sluggish. The range of full price rigidity is small, as even many small positive permanent demand shocks induce a positive price reaction. By contrast, in response to permanent negative demand shocks, prices remain fully rigid also for a considerable range of large shocks. In the following we will analyse the intuition for these results.

1.4.1 Intuition

Temporary demand shocks

For a temporary (one-period) demand shock, the consumers' reference price is not affected (since information reaches them with a one-period lag and they have rational expectations). This implies that $\frac{\partial R_{t+1}}{\partial p_t} = 0$ and thus the firm's price response to the shock is the same as that of a myopic firm (which maximizes its current period profit). According to the optimality condition (1.12), the new profit-maximizing price is determined by the standard condition according to which $MR_t = MC_t$.

Result 1: *In response to a small temporary shock, prices remain rigid.*

As noted, for a sufficiently small demand shock $\varepsilon_t^s \leq \bar{\varepsilon}_t(\lambda)$ the marginal cost curve still intersects the marginal revenue discontinuity, i.e. $MC_t(\hat{q}_t) \in$

$MRD_t(\hat{q}_t, R_{ss}, \gamma, \delta, \varepsilon_t^s)$. Therefore, the prevailing steady state price remains the firm's profit-maximizing price,¹⁷ i.e. $P_t^* = P_{ss}^*$, and we have complete price rigidity. By contrast, the profit-maximizing quantity changes to $q_t^* = R_{ss}^{-\eta} \tilde{P}_t^\eta y_t \varepsilon_t^s$, thus the percentage change of quantity is given by

$$\Delta q_t^* = \frac{q_t^* - q_{ss}^*}{q_{ss}^*} = \frac{\varepsilon_t^s - 1}{1} = \varepsilon_t^s - 1 \neq 0. \quad (1.17)$$

This holds true irrespective of the sign of the small temporary demand shock. Consequently, the quantity reaction is symmetric for positive and negative demand shocks of equal magnitude.

Result 2: *In response to a large temporary shock, prices are more sluggish upwards than downwards.*

For a large shock, i.e. $\varepsilon_t^l > \bar{\varepsilon}_t(\lambda)$, the marginal cost curve intersects the marginal revenue curve outside the discontinuity of the latter. Consequently both, a price and a quantity reaction are induced. The new profit-maximizing price of the firm is

$$P_t^* = \left(\frac{R_{ss}^{(\lambda-\eta)} \tilde{P}_t^\eta y_t \varepsilon_t^l}{q_t^*} \right)^{\frac{1}{\lambda}}, \quad (1.18)$$

while its corresponding profit-maximizing quantity is

$$q_t^* = \left(\frac{1}{c} \left(1 - \frac{1}{\lambda} \right) \right)^{\frac{\lambda}{\lambda+1}} \left(R_{ss}^{(\lambda-\eta)} \tilde{P}_t^\eta y_t \varepsilon_t^l \right)^{\frac{1}{\lambda+1}}, \quad (1.19)$$

where $\lambda = \delta$ for positive and $\lambda = \gamma$ for negative shocks, respectively.

In comparison to the standard firm the price reaction of the firm facing loss-averse consumers in response to a large temporary demand shock is always smaller, whereas the quantity reaction is always larger. Additionally, prices and quantities are less responsive to positive than to negative shocks. The intuition is obvious once we decompose the demand shock into the maximum small shock and the remainder:

$$\varepsilon_t^l = \bar{\varepsilon}_t(\lambda) + \varepsilon_t^{rem}. \quad (1.20)$$

¹⁷Compare the proof from Section 1.3.2.

From our theoretical analysis above, the maximum small shock $\bar{\varepsilon}_t(\lambda)$ has no price effects, but feeds one-to-one into demand. This holds true irrespective of the sign of the shock. By contrast, the remaining shock ε_t^{rem} has asymmetric effects. Let \bar{q}_t be the quantity corresponding to $\bar{\varepsilon}_t(\lambda)$. Then the percentage change in quantity in response to ε_t^{rem} is given by

$$\Delta q_t^{rem} = \frac{q_t^* - \bar{q}_t}{\bar{q}_t} = \left(1 + \frac{\varepsilon_t^{rem}}{\bar{\varepsilon}_t(\lambda)}\right)^{\frac{1}{\lambda+1}} - 1. \quad (1.21)$$

As can be seen from equation 1.21, the change of quantity in response to ε_t^{rem} depends negatively on λ , the price elasticity of demand. Since by definition $\delta > \gamma$, the quantity reaction of the firm facing loss-averse consumers is smaller in response to large positive temporary demand shocks than to large negative ones. This however implies that prices are also less responsive to positive than to negative large temporary demand shocks, because the former move the firm along the relatively flat portion of the demand curve, whereas the latter move it along the relatively steep portion of the demand curve. This asymmetric sluggishness in the reaction to positive and negative large temporary demand shocks is a distinct feature of consumer loss aversion and stands in obvious contrast to the standard textbook case of monopoly pricing.

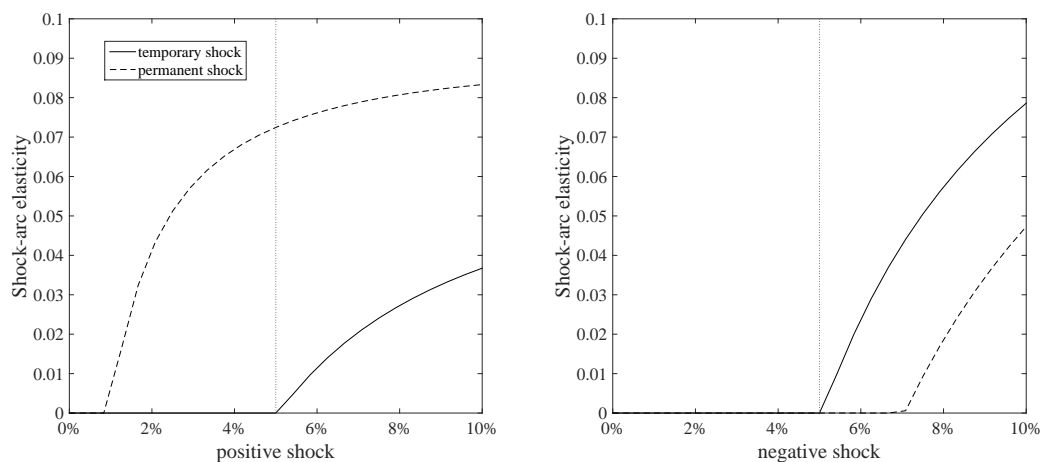
Permanent demand shocks

Now consider a permanent demand shock that occurs in period t . For a permanent (two-period) demand shock, the consumers' period $t + 1$ reference price is now affected by the firms' period t pricing decision, i.e. $\frac{\partial R_{t+1}}{\partial P_t} \neq 0$. Thus the firm's price response to the shock differs from that of a myopic firm (which maximizes only its current period profit), as according to the optimality condition (1.12) the effect of the pricing decision on the reference price drives a wedge between MR_t and MC_t in optimality.

Result 3: *For all permanent shocks, prices are less sluggish upwards than downwards.*

The intuition for this result is as follows: Whereas the firm is assumed to change its price immediately in response to this shock, consumers update

Figure 1.3: Shock-arc elasticities to positive and negative demand shocks.



their reference price in the following period $t + 1$, i.e. $R_{t+1} = E_t[P_{t+1}|I_t]$. Consequently, for price increases (decreases) the demand curve shifts outwards (inwards) and the kink moves to

$$(\widehat{P}_{t+1}, \widehat{q}_{t+1}) = \left(R_{t+1}, (\tilde{P}_{t+1}/R_{t+1})^\eta y_{t+1} \varepsilon_{t+1} \right). \quad (1.22)$$

An outward shift of the demand curve (initiated by an upward adjustment in the reference price) increases the firm's long-run profits, whereas an inward shift (initiated by a downward adjustment of the reference price) lowers them. We term this phenomenon the "reference-price updating effect." The firm can anticipate this. Thus, it may have an incentive to set its price above the level that maximizes its profits in the shock period $P'_t > P_t^*$, thereby exploiting (dampening) the outward (inward) shift of the demand curve resulting from the upward (downward) adjustment of the consumers' reference price for positive (negative) permanent shocks.¹⁸ The firm exploits this effect, as long as the gain from a price rise relative to P_t^* in terms of future profits ($\Pi_{t+1}(R_{t+1} = P'_t) > \Pi_{t+1}(R_{t+1} = P_t^*)$, due to the relative rise in the reference price) exceeds the firm's loss in terms of present profits ($\Pi_t(P'_t) < \Pi_t(P_t^*)$, since the price P'_t is not appropriate for maximizing current profit).

The extend, to which the firm exploits the reference-price-updating effect can be seen in Figure 1.3. The figure shows the shock-arc elasticities of price for temporary (solid lines) and permanent (dashed lines) shocks, given that

¹⁸Needless to say, setting a price lower than optimal in the shock period with the aim to decrease the reference price permanently is not a preferable option for the firm.

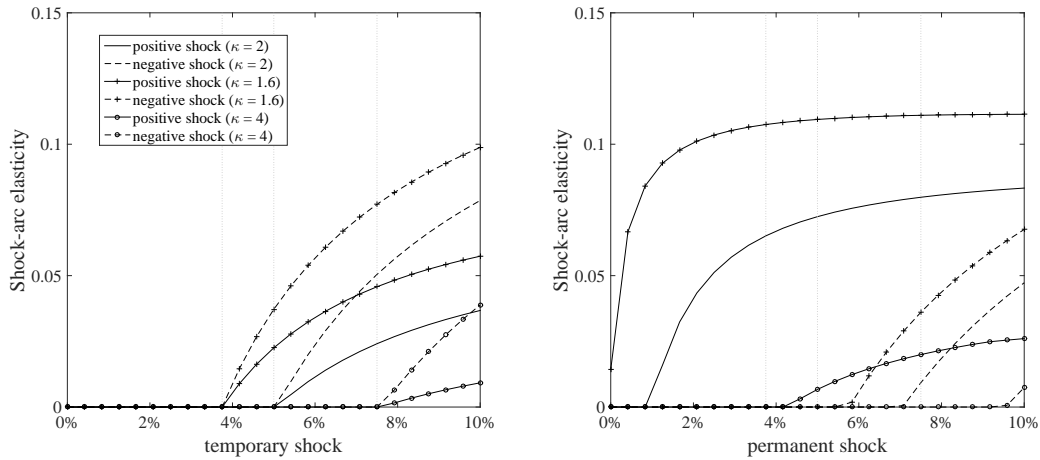
a shock is positive (left panel) and negative (right panel). The reference-price-updating effect is measured by vertical difference between the dashed lines and the solid lines. Figure 1.3 indicates that in response to a permanent shock the firm significantly exploits the “reference-price updating effect” and thus generally sets a price that is higher than the price it would optimally set in response to a temporary shock, i.e. $P'_t > P_t^*$. For positive permanent demand shocks this implies that the pricing reaction of the firm is always stronger than for positive temporary demand shocks for both, small and large shocks¹⁹. This can be seen in the left panel of Figure 1.3, as the dashed line is always on or above the solid line. By contrast, for negative permanent demand shocks the firm adjust its price downward to a considerably lower extent than for negative temporary shocks, as shown in the right panel of Figure 1.3, where the dashed line is always on or below the solid line.

As a consequence, price sluggishness is considerably less pronounced for positive than for negative permanent demand shocks. The asymmetry of the price reaction to positive and negative shocks therefore reverses, when moving from temporary to permanent shocks. While this result may seem surprising at first glance, it is straightforward intuitively: As noted, for temporary shocks, consumers abstract from updating their reference price. Therefore, the firm does not risk to suffer from a downward adjustment of the consumers’ reference price, when encountering a temporary drop in demand with a price reduction. On the other hand, for positive temporary shocks, the firm cannot generate permanent increases in demand due to upward-adjustments of the reference price. Since consumers react more sensitive to price increases relative to price decreases, the price and quantity reactions are smaller for positive temporary shocks compared to negative ones. By contrast, for permanent demand shocks, the firm exploits the positive “reference-price updating effect” which follows from price increases in response to positive shocks, whereas it tries to avoid the negative “reference-price updating effect” which follows from price decrease in response to negative shocks.²⁰

¹⁹Our numerical analysis indicates, however, that the positive reference-price updating effect is never strong enough to invalidate the general result that the pricing reaction of the firm facing loss averse consumers is more sluggish compared to the standard firm.

²⁰Since the firm avoids price reductions, which lead to downward-adjustments in the reference price, but conducts price reductions, which do not influence the reference price, loss aversion offers a simple rationale for the firm’s practice of “sales”(see e.g. Eichenbaum et al., 2011).

Figure 1.4: Sensitivity with respect to the loss aversion parameter.



1.4.2 Sensitivities

Figure 2.5 shows the shock-arc elasticities of the price for the following values of the loss aversion ratio: $\kappa \in (1.6; 2; 4)$, where our base case is $\kappa = 2$. A loss aversion ratio as low as approximately $\kappa = 1.6$ was estimated by Kalwani et al. (1990), Hardie et al. (1993), and Kalyanaram and Little (1994), whereas the higher value was estimated by Raman and Bass (2002). All estimates are based on a wide variety of frequently used non-durable supermarket products.

Figure 2.5 shows that the higher the loss aversion ratio, ceteris paribus, the more sluggish is price adjustment in response to demand shocks, both upwards and downwards. The shock-arc elasticity curves for higher parameter values always lie below the curves from lower parameter values. This result is independent of the persistence of the shock. The critical shocks, positive and negative, increase in the degree of loss aversion and thereby widen the marginal revenue discontinuity. This implies that for both temporary demand shocks (left panel) and permanent demand shocks (right panel), the range of full price rigidity increases as consumers become more loss averse.

The right panel of Figure 2.5 shows that the firms incentive to avoid permanent price cuts increases substantially the more loss averse consumers are. The intuition for this result is straightforward: The higher the price elasticity of demand in the loss domain, the stronger is the resulting decrease of demand due to a downward adjustment of the reference price. Consequently, period $t + 1$ profits decrease. Therefore, the firm's incentive to deviate up-

wards in response to permanent negative demand shocks increases in the degree of loss aversion. By contrast, for permanent positive demand shocks the firm's incentive to deviate upwards decreases. Because the reference-price-updating effect is stronger, the firm does not necessitate to deviate by as much in order to generate the profit maximizing quantity in period $t + 1$. These responses to permanent positive and negative demand shocks imply that the degree of downward-sluggishness decreases in the degree of loss aversion. This is also apparent from the right panel of Figure 2.5, where the degree of downward sluggishness is measured by the vertical difference between the shock-arc elasticity curves for positive demand shocks and for negative demand shocks.

Our sensitivity analysis confirms that over the range of reasonable parameter values for the loss aversion parameter, our theory implies that - in line with the empirical evidence - prices are more sluggish upwards than downwards in response to temporary demand shocks, while they are more sluggish downwards than upwards in response to permanent demand shocks.

1.5 Conclusion

In contrast to the standard time-dependent and state-dependent models of price sluggishness, our theory of price adjustment is able to account for asymmetric price and quantity responses to positive and negative temporary and permanent shocks of equal magnitude. In contrast to the New Keynesian literature, our explanation of price adjustment is thoroughly microfounded, without recourse to ad hoc assumptions concerning the frequency of price changes or physical costs of price adjustments.

There are many avenues of future research. Consideration of heterogeneous firms and multi-product firms will enable this model to generate asynchronous price changes, as well as the simultaneous occurrence of large and small price changes, and heterogeneous frequency of price changes across products. Extending the model to a stochastic environment will generate testable implications concerning the variability of individual prices. Furthermore, our model needs to be incorporated into a general equilibrium setting to validate the predictions of our theory.

References

- Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference points and effort provision. *American Economic Review* 101(2), 470–492.
- Adeyemi, O. I. and L. C. Hunt (2007). Modelling OECD industrial energy demand: Asymmetric price responses and energy-saving technical change. *Energy Economics* 29(4), 693–709.
- Ahrens, S., I. Pirschel, and D. J. Snower (2014). A theory of wage adjustment under loss aversion. *CEPR Discussion Papers No. 10288*, Centre for Economic Policy Research.
- Ainslie, G. W. (1992). *Picoeconomics*. Cambridge: Cambridge University Press.
- Alvarez, L. J. and I. Hernando (2005). The price setting behavior of spanish firms: Evidence from survey data. *Working Paper Series No. 0538*, European Central Bank.
- Amirault, D., C. Kwan, and G. Wilkinson (2005). Survey of price-setting behaviour of canadian companies. *Bank of Canada Review, Winter 2004-2005*, 29-40.
- Apel, M., R. Friberg, and K. Hallsten (2005). Microfoundations of macroeconomic price adjustment: Survey evidence from swedish firms. *Journal of Money, Credit and Banking* 37(2), 313–338.
- Aucremanne, L. and M. Druant (2005). Price-setting behaviour in belgium: What can be learned from an ad hoc survey? *Working Paper Series No. 0448*, European Central Bank.
- Ball, L. and N. G. Mankiw (1994). Asymmetric price adjustment and economic fluctuations. *Economic Journal* 104(423), 247–261.
- Banerji, A. and N. Gupta (2014). Detection, identification, and estimation of loss aversion: Evidence from an auction experiment. *American Economic Journal: Microeconomics* 6(1), 91–133.
- Bartling, B., L. Brandes, and D. Schunk (forthcoming). Expectations as reference points: Field evidence from professional soccer. *Management Science*.

- Bidwell, M. O., B. R. Wang, and J. D. Zona (1995). An analysis of asymmetric demand response to price changes: The case of local telephone calls. *Journal of Regulatory Economics* 8(3), 285–298.
- Blinder, A., E. R. D. Canetti, D. E. Lebow, and J. B. Rudd (1998). *Asking about prices: A new approach to understanding price stickiness*. New York: Russell Sage Foundation.
- Caballero, R. J. and E. M. R. A. Engel (1993). Heterogeneity and output fluctuations in a dynamic menu-cost economy. *Review of Economic Studies* 60(1), 95–119.
- Caballero, R. J. and E. M. R. A. Engel (2007). Price stickiness in ss models: New interpretations of old results. *Journal of Monetary Economics* 54(Supplement), 100–121.
- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 12(3), 383–398.
- Caplin, A. S. and D. F. Spulber (1987). Menu costs and the neutrality of money. *The Quarterly Journal of Economics* 102(4), 703–725.
- Card, D. and G. B. Dahl (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics* 126(1), 103–143.
- Chari, V., P. Kehoe, and E. McGrattan (2000). Sticky price models of the business cycle: Can the contract multiplier solve the persistence problem? *Econometrica* 68(5), 1151–1179.
- Cover, J. P. (1992). Asymmetric effects of positive and negative money-supply shocks. *The Quarterly Journal of Economics* 107(4), 1261–1282.
- Crawford, V. P. and J. Meng (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review* 101(5), 1912–1932.
- DeLong, J. B. and L. H. Summers (1988). How does macroeconomic policy affect output? *Brookings Papers on Economic Activity* 19(2), 433–494.

- Dossche, M., F. Heylen, and D. Van den Poel (2010). The kinked demand curve and price rigidity: Evidence from scanner data. *Scandinavian Journal of Economics* 112(4), 723–752.
- Dotsey, M., R. G. King, and A. L. Wolman (2009). Inflation and real activity with firm level productivity shocks. *2009 Meeting Papers 367*, Society for Economic Dynamics.
- Eichenbaum, M. and J. Fisher (2004). Evaluating the calvo model of sticky prices. *NBER Working Papers No. 10617*, National Bureau of Economic Research, Inc.
- Eichenbaum, M., N. Jaimovich, and S. Rebelo (2011). Reference prices, costs, and nominal rigidities. *American Economic Review* 101(1), 234–262.
- Elmaghraby, W. and P. Keskinocak (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science* 49(10), 1287–1309.
- Erceg, C. J., D. W. Henderson, and A. T. Levin (2000). Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics* 46(2), 281–313.
- Fabiani, S., M. Druant, I. Hernando, C. Kwapil, B. Landau, C. Loupias, F. Martins, T. Mathä, R. Sabbatini, H. Stahl, and A. Stockman (2006). What firm's surveys tell us about price-setting behavior in the euro area. *International Journal of Central Banking* 2(3), 1–45.
- Fabiani, S., A. Gattulli, and R. Sabbatini (2004). The pricing behaviour of italian firms: New survey evidence on price stickiness. *Working Paper Series No. 0333*, European Central Bank.
- Genesove, D. and C. Mayer (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics* 116(4), 1233–1260.
- Gertler, M. and J. Leahy (2008). A phillips curve with an ss foundation. *Journal of Political Economy* 116(3), 533–572.
- Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review* 102(1), 469–503.

- Golosov, M. and R. E. Lucas Jr. (2007). Menu costs and phillips curves. *Journal of Political Economy* 115(2), 171–199.
- Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40(1-3), 3–73.
- Grant, S., S. King, and B. Polak (1996). Information externalities, share-price based incentives and managerial behaviour. *Journal of Economic Surveys* 10(1), 1–21.
- Greenleaf, E. A. (1995). The impact of reference-price effects on the profitability of price promotions. *Marketing Science* 14(1), 82–104.
- Greenslade, J. V. and M. Parker (2012). New insights into price-setting behaviour in the UK: Introduction and survey results. *Economic Journal* 122(558), F1–F15.
- Griffin, J. M. and C. T. Schulman (2005). Price asymmetry in energy demand models: A proxy for energy-saving technical change? *The Energy Journal* 0(2), 1–22.
- Hall, R. and C. Hitch (1939). Price theory and business behaviour. *Oxford Economic Papers* 2(1), 12–45.
- Hall, S., M. Walsh, and A. Yates (2000). Are UK companies' prices sticky? *Oxford Economic Papers* 52(3), 425–446.
- Hardie, B. G. S., E. J. Johnson, and P. S. Fader (1993). Modeling loss aversion and reference dependence effects on brand choice. *Marketing Science* 12(4), 378–394.
- Heidhues, P. and B. Köszegi (2005). The impact of consumer loss aversion on pricing. *CEPR Discussion Papers No. 4849*, Centre for Economic Policy Research.
- Heidhues, P. and B. Köszegi (2008). Competition and price variation when consumers are loss averse. *American Economic Review* 98(4), 1245–1268.
- Heidhues, P. and B. Köszegi (2014). Regular prices and sales. *Theoretical Economics* 9(1), 217–251.

- Hoeberichts, M. and A. Stokman (2010). Price setting behaviour in the netherlands: Results of a survey. *Managerial and Decision Economics* 31(2-3), 135–149.
- Jacobson, R. and D. Aaker (1993). Myopic management behavior with efficient, but imperfect, financial markets: A comparison of information asymmetries in the U.S. and Japan. *Journal of Accounting and Economics* 16(4), 383–405.
- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–291.
- Kalwani, M. U., C. K. Yim, H. J. Rinne, and Y. Sugita (1990). A price expectations model of customer brand choice. *Journal of Marketing Research* 27(3), 251–262.
- Kalyanaram, G. and L. D. C. Little (1994). An empirical analysis of latitude of price acceptance in consumer package goods. *Journal of Consumer Research* 21(3), 408–418.
- Kandil, M. (1995). Asymmetric nominal flexibility and economic fluctuations. *Southern Economic Journal* 61(3), 674–695.
- Kandil, M. (1996). Sticky wage or sticky price? analysis of the cyclical behavior of the real wage. *Southern Economic Journal* 63(2), 440–459.
- Kandil, M. (1998). Supply-side asymmetry and the non-neutrality of demand fluctuations. *Journal of Macroeconomics* 20(4), 785–809.
- Kandil, M. (1999). The asymmetric stabilizing effects of price flexibility: Historical evidence and implications. *Applied Economics* 31(7), 825–839.
- Kandil, M. (2001). Asymmetry in the effects of US government spending shocks: Evidence and implications. *The Quarterly Review of Economics and Finance* 41(2), 137–165.
- Kandil, M. (2002a). Asymmetry in economic fluctuations in the US economy: The pre-war and the 1946-1991 periods compared. *International Economic Journal* 16(1), 21–42.

- Kandil, M. (2002b). Asymmetry in the effects of monetary and government spending shocks: Contrasting evidence and implications. *Economic Inquiry* 40(2), 288–313.
- Kandil, M. (2006). Asymmetric effects of aggregate demand shocks across U.S. industries: Evidence and implications. *Eastern Economic Journal* 32(2), 259–283.
- Kandil, M. (2010). The asymmetric effects of demand shocks: international evidence on determinants and implications. *Applied Economics* 42(17), 2127–2145.
- Karle, H., G. Kirchsteiger, and M. Peitz (2015). Loss aversion and consumption choice: Theory and experimental evidence. *American Economic Journal: Microeconomics* 7(2), 101–120.
- Karras, G. (1996). Why are the effects of money-supply shocks asymmetric? convex aggregate supply or “pushing on a string”? *Journal of Macroeconomics* 18(4), 605–619.
- Karras, G. and H. H. Stokes (1999). On the asymmetric effects of money-supply shocks: International evidence from a panel of OECD countries. *Applied Economics* 31(2), 227–235.
- Kehoe, P. J. and V. Midrigan (2008). Temporary price changes and the real effects of monetary policy. *NBER Working Papers No. 14392*, National Bureau of Economic Research, Inc.
- Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121(4), 1133–1165.
- Kőszegi, B. and M. Rabin (2007). Reference-dependent risk attitudes. *American Economic Review* 97(4), 1047–1073.
- Kőszegi, B. and M. Rabin (2009). Reference-dependent consumption plans. *American Economic Review* 99(3), 909–936.
- Kimball, M. (1995). The quantitative analytics of the basic monetarist model. *Journal of Money, Credit and Banking* 27(4), 1241–1277.
- Klenow, P. and J. Willis (2006). Real rigidities and nominal price changes. *Federal Reserve Bank of Kansas City Working Paper No. 06-03*.

- Knetsch, J. L. and W. K. Wong (2009). The endowment effect and the reference state: Evidence and manipulations. *Journal of Economic Behavior & Organization* 71(2), 407–413.
- Krishnamurthi, L., T. Mazumdar, and S. P. Raj (1992). Asymmetric response to price in consumer brand choice and purchase quantity decisions. *Journal of Consumer Research* 19(3), 387–400.
- Kwapil, C., J. Scharler, and J. Baumgartner (2010). How are prices adjusted in response to shocks? survey evidence from austrian firms. *Managerial and Decision Economics* 31(2-3), 151–160.
- Laibson, D. (1996). Hyperbolic discount functions, undersaving, and savings policy. *NBER Working Papers No 5635*, National Bureau of Economic Research, Inc.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112(2), 443–477.
- Langbraaten, N., E. W. Nordbø, and F. Wulfsberg (2008). Price-setting behaviour of norwegian firms - results of a survey. *Norges Bank Economic Bulletin* 79(2), 13–34.
- Loewenstein, G. and D. Prelec (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics* 107(2), 573–597.
- Loewenstein, G. and R. H. Thaler (1989). Anomalies: Intertemporal choice. *The Journal of Economic Perspectives* 3(4), 181–193.
- Loupias, C. and R. Ricart (2004). Price setting in france: New evidence from survey data. *Working Paper Series No. 0423*, European Central Bank.
- Lünnemann, P. and T. Y. Mathä (2006). New survey evidence on the pricing behaviour of luxembourg firms. *Working Paper Series No. 0617*, European Central Bank.
- Martins, F. (2010). Price stickiness in portugal evidence from survey data. *Managerial and Decision Economics* 31(2-3), 123–134.

- Marzilli Ericson, K. M. and A. Fuster (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *The Quarterly Journal of Economics* 126(4), 1879–1907.
- Mas, A. (2006). Pay, reference pay and police performance. *The Quarterly Journal of Economics* 121(3), 783–821.
- Maskin, E. and J. Tirole (1988). A theory of dynamic oligopoly, II: Price competition, kinked demand curves, and edgeworth cycles. *Econometrica* 56(3), 571–599.
- Mayhew, G. E. and R. S. Winer (1992). An empirical analysis of internal and external reference prices using scanner data. *Journal of Consumer Research* 19(1), 62–70.
- McDonald, I. M. and H. Sibly (2001). How monetary policy can have permanent real effects with only temporary nominal rigidity. *Scottish Journal of Political Economy* 48(5), 532–546.
- McDonald, I. M. and H. Sibly (2005). The diamond of macroeconomic equilibria and non-inflationary expansion. *Metroeconomica* 56(3), 393–409.
- Mehrara, M. and A. R. Karsalari (2011). Asymmetric effects of monetary shocks on economic activities: The case of Iran. *Journal of Money, Investment and Banking* 20, 62–74.
- Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica* 79(4), 1139–1180.
- Mizik, N. (2010). The theory and practice of myopic management. *Journal of Marketing Research* 47(4), 594–611.
- Mizik, N. and R. Jacobson (2007). Myopic marketing management: Evidence of the phenomenon and its long-term performance consequences in the seo context. *Marketing Science* 26(3), 361–379.
- Morgan, D. P. (1993). Asymmetric effects of monetary policy. *Federal Reserve Bank of Kansas City Economic Review QII*, 21-33.
- Nasiry, N. and I. Popescu (2011). Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations Research* 59(6), 1361–1368.

- Nicolau, J. L. (2008). Testing reference dependence, loss aversion and diminishing sensitivity in spanish tourism. *Investigaciones Económicas* 32(2), 231–255.
- Nicolitsas, D. (2013). Price setting practices in greece: Evidence from a small-scale firm-level survey. *Working Papers No. 156*, Bank of Greece.
- Okun, A. M. (1981). Prices and quantities: A macroeconomic analysis. *Brookings Institution*, Washington, DC.
- Ólafsson, T. T., Á. Pétursdóttir, and K. Á. Vignisdóttir (2011). Price setting in turbulent times: Survey evidence from icelandic firms. *Working Paper No. 54*, Central Bank of Iceland.
- Pope, D. G. and M. E. Schweitzer (2011). Is tiger woods loss averse? persistent bias in the face of experience, competition, and high stakes. *American Economic Review* 101(1), 129–157.
- Popescu, I. and Y. Wu (2007). Dynamic pricing strategies with reference effects. *Operations Research* 55(3), 413–429.
- Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008). Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review* 98(1), 38–71.
- Putler, D. S. (1992). Incorporating reference price effects into a theory of consumer choice. *Marketing Science* 11(3), 287–309.
- Raman, K. and F. M. Bass (2002). A general test of reference price theory in the presence of threshold effects. *Tijdschrift voor Economie en Managerriënt XLVII(2)*, 205-226.
- Ravn, M. O. and M. Sola (2004). Asymmetric effects of monetary policy in the United States. *Federal Reserve Bank of St. Louis Review* 86(5), 41–60.
- Rosenkranz, S. (2003). The manufacturer’s suggested retail price. *CEPR Discussion Papers No 3954*, Centre for Economic Policy Research.
- Rosenkranz, S. and P. W. Schmitz (2007). Reserve prices in auctions as reference points. *Economic Journal* 117(520), 637–653.

- Rotemberg, J. J. (1982). Monopolistic price adjustment and aggregate output. *Review of Economic Studies* 49(4), 517–531.
- Rotemberg, J. J. (2005). Customer anger at price increases, changes in the frequency of price adjustment and monetary policy. *Journal of Monetary Economics* 52(4), 829–852.
- Rotemberg, J. J. (2010). Altruistic dynamic pricing with customer regret. *Scandinavian Journal of Economics* 112(4), 646–672.
- Ryan, D. L. and A. Plourde (2007). *A systems approach to modelling asymmetric demand responses to energy price changes*, Volume 18 of *International Symposia in Economic Theory and Econometrics*.
- Şahinöz, S. and B. Saraçoğlu (2008). Price-setting behavior in turkish industries: Evidence from survey data. *The Developing Economies* 46(4), 363–385.
- Schmitt-Grohé, S. and M. Uribe (2007). Optimal simple and implementable monetary and fiscal rules. *Journal of Monetary Economics* 54(6), 1702–1725.
- Sheshinski, E. and Y. Weiss (1977). Inflation and costs of price adjustment. *Review of Economic Studies* 44(2), 287–303.
- Sibly, H. (1996). Consumer disenchantment, loss aversion and price rigidity. *Papers 1996-12, Tasmania - Department of Economics*.
- Sibly, H. (2002). Loss averse customers and price inflexibility. *Journal of Economic Psychology* 23(4), 521–538.
- Sibly, H. (2007). Loss aversion, price and quality. *The Journal of Socio-Economics* 36(5), 771–788.
- Spiegler, R. (2012). Monopoly pricing when consumers are antagonized by unexpected price increases: A “cover version” of the Heidhues-Kőszegi-Rabin model. *Economic Theory* 51(3), 695–711.
- Stahl, H. (2005). Price setting in german manufacturing: New evidence from new survey data. *Working Paper Series No. 0561*, European Central Bank.
- Sweezy, P. (1939). Demand under conditions of oligopoly. *The Journal of Political Economy* 47(4), 568–573.

- Tan, S. H., M. S. Habibullah, and A. Mohamed (2010). Asymmetric effects of monetary policy in ASEAN-4 economies. *International Research Journal of Finance and Economics* 44, 30–42.
- Taylor, J. B. (1979). Staggered wage setting in a macro model. *American Economic Review* 69(2), 108–113.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science* 4(3), 199–214.
- Thaler, R. (1991). *Quasi rational economics*. New York: Russell Sage Foundation.
- Thoma, M. A. (1994). Subsample instability and asymmetries in money-income causality. *Journal of Econometrics* 64(1-2), 279–306.
- Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics* 106(4), 1039–1061.
- Weise, C. L. (1999). The asymmetric effects of monetary policy: A nonlinear vector autoregression approach. *Journal of Money, Credit and Banking* 31(1), 85–108.
- Woodford, M. (2005). Firm-specific capital and the New Keynesian Phillips curve. *International Journal of Central Banking* 1(2), 1–46.
- Zbaracki, M. J., M. Ritson, D. Levy, S. Dutta, and M. Bergen (2004). Managerial and customer costs of price adjustment: Direct evidence from industrial markets. *The Review of Economics and Statistics* 86(2), 514–533.
- Zhou, J. (2011). Reference dependence and market competition. *Journal of Economics & Management Strategy* 20(4), 1073–1097.

Chapter 2

Path-Dependent Wage Responsiveness*

Abstract

We present a theory of nominal wage adjustment based on worker loss aversion, along the lines of prospect theory. Wage changes are evaluated relative to an endogenous reference wage, which depends on the workers' rational wage expectations from the recent past. By implication, firms face an upward-sloping labor supply curve that is convexly kinked at the workers' reference wage. Firms adjust wages flexibly in response to variations in labor demand, incorporating the endogenous response of the reference wage. The resulting theory of wage adjustment is starkly at variance with past theories. In line with the empirical evidence, we find that (1) wages are completely rigid in response to small labor demand shocks, (2) wages are downward rigid but upward flexible for medium-sized labor demand shocks, and (3) wages are relatively downward sluggish for large shocks.

Keywords: Downward wage sluggishness, Loss aversion, Endogenous reference points.

JEL-Codes: D03, D21, E24.

*This chapter is based on my joint paper with Steffen Ahrens and Dennis J. Snower entitled *Path-Dependent Wage Responsiveness*, Kiel Working Paper No. 1977, 2014 (updated in November 2015).

2.1 Introduction

This paper presents a theory of nominal wage adjustment based on worker loss aversion, along the lines of prospect theory (Kahneman and Tversky, 1979). Workers evaluate nominal wage changes relative to an endogenous reference wage, which depends on their rational wage expectations from the recent past. Firms anticipate adjustments of the reference wage when they make their wage setting decisions. The theory has distinctive implications which are starkly at variance with major existing theories of nominal wage adjustment but consonant with the empirical evidence. In particular, the theory implies that (1) for small labor demand shocks, nominal wages are fully rigid, (2) for medium-sized shocks there is upward nominal wage adjustment for positive shocks, but complete downward nominal wage rigidity for negative shocks and (3) for large shocks, nominal wages decline less strongly to negative shocks than they increase to equiproportionate positive shocks. In short, our theory can explain the empirically well documented occurrence of nominal wage rigidity in the presence of small labor demand variations, downward nominal wage rigidity but upward nominal wage adjustment to intermediate labor demand variations, and relative downward nominal wage sluggishness in the presence of large shocks. While current theories of nominal wage adjustment fail to account for all three of these pieces of empirical evidence, this paper offers a theoretical rationale.

The basic idea underlying our theory is simple. In the spirit of prospect theory, the workers' utility losses from nominal wage decreases are weighted more heavily than the utility gains from nominal wage increases of equal magnitude. Consequently, employment responses are more elastic to nominal wage decreases than to nominal wage increases. The result is a kinked labor supply curve, for which the kink depends on the workers' nominal reference wage. The kink of the labor supply curve implies that nominal wages are rigid in response to sufficiently small labor demand shocks, but nominal wages adjust asymmetrically to larger shocks. While it is well-known that wage loss aversion leads to kinked labor supply curves (e.g. Bhaskar, 1990), our contribution lies in combining nominal wage loss aversion with endogenous reference wage dynamics and in investigating the implications of these dynamics for the firm's wage and employment decisions. In the spirit of Kőszegi and Rabin (2006), we model the reference wage as the workers' ratio-

nal wage expectations from the recent past, which are adjusted through time. The reference wage implicitly determines workers' endogenous income target. An increase in the reference wage raises their implicit income target, whereas a decrease in the expected nominal wage lowers it. Workers adjust their labor supply accordingly. Consequently, there is a positive relationship between the workers' reference wage and their labor supply. Therefore, a labor demand shock not only produces a change in employment following the firm's immediate wage setting decision, but also an adjustment in the workers' future reference wage. Firms foresee that their wage setting decision has an effect on the workers' future reference wage and thereby their future labor supply. A rise in the reference wage raises the firms' long-run profits (since the reference wage is located at the kink of the labor supply curve), whereas a fall in the reference wage lowers long-run profits. On this account, medium-sized to large positive labor demand shocks lead to nominal wage increases, while medium-sized to large negative labor demand shocks may lead to relatively little if any downward nominal wage adjustment.¹

From a methodological point of view, this paper is closely related to Ahrens et al. (2014) which considers how consumer loss aversion with endogenous reference price dynamics affects firms' price setting decision.

The remainder of the paper is structured as follows. Section 2.2 reviews the relevant literature. Section 2.3 presents our general model setup. In section 2.4 we analyze the effects of various demand shocks on nominal wages, both numerically and analytically, and check our results for robustness. Section 2.5 concludes.

¹Our theory may help shed light on asymmetric effects of monetary policy, though such implications lie beyond the scope of this paper. First it is relevant to the literature on short-run monetary policy, which has asymmetric effects under downward nominal wage rigidity (e.g. McDonald and Sibly, 2001; Carlsson and Westermarck, 2008; Fahr and Smets, 2010). Second, while symmetric nominal rigidities give rise to a long-run Phillips curve which is virtually vertical (e.g. Goodfriend and King, 1997; Khan et al., 2003), downward nominal wage rigidity leads to a significantly non-vertical long-run Phillips curve, thereby generating substantial long-run real effects of monetary policy on output and employment for negative shocks, as shown by Kim and Ruge-Murcia (2009, 2011), Fagan and Messina (2009), Fahr and Smets (2010), Benigno and Ricci (2011) and Abo-Zaid (2013). In all of these latter contributions, downward nominal wage rigidity is introduced in an ad-hoc way, using a linex function as proposed by Varian (Varian). The only exception to this is Benigno and Ricci (2011), who use a case sensitive approach. Consequently, these models exhibit permanent downward nominal wage rigidity, independent of the size and the sign of the shock. However, since the degree of downward nominal wage rigidity varies with the size of the shock, the short- and long-run Phillips curves are state-dependent, a feature not considered in the studies above.

2.2 Relation to the Literature

In this section, we review the empirical evidence suggesting that nominal wages are (imperfectly) downward rigid, while they are upward flexible. In particular, ample microeconomic evidence points towards three important stylized facts, namely that (i) nominal wage rigidity is common in the presence of minor labor market shocks, (ii) under mid-range shocks (such as those in standard business cycle fluctuations), downward-wage rigidity and upward wage adjustment are common, (iii) nominal wage cuts do take place in severe downturns.

This evidence implies that the distribution of nominal wage changes spikes at zero and contains much fewer observations below zero than above. Such a distribution of nominal wage changes is documented for a wide variety of industrialized countries. For the United States, McLaughlin (1994), Card and Hyslop (1997), Kahn (1997), and Altonji and Devereux (1999) derive such evidence from the Panel Study of Income Dynamics, while Akerlof et al. (1996), Lebow et al. (2003), Gottschalk (2005), and Dickens et al. (2007) find this distribution based on employer reports, social security files, and several different household surveys. Based on national wage and income surveys as well as on employer reports, Smith (2000), Agell and Lundborg (2003), Nickell and Quintini (2003), Fehr and Goette (2005), Bauer et al. (2007), Dickens et al. (2007), Babecký et al. (2010), Böckerman et al. (2010), and Sigurdsson and Sigurdardottir (2011) provide this evidence for a large sample of European economies, while Kimura and Ueda (2001), Cobb and Opazo (2008), and Iregui et al. (2009) find this for Japan, Chile, and Colombia, respectively.

While all these studies find that nominal wage cuts are rare, they do happen and commonly take place in times of severe financial distress, such as long lasting and deep recessions or any other sort of immanent risk of bankruptcy for a firm (Kahneman et al., 1986; Bewley, 1995, 1999; Akerlof et al., 1996; Campbell and Kamlani, 1997; Kimura and Ueda, 2001; Fehr and Goette, 2005; Böckerman et al., 2010). Moreover, there is empirical evidence that extremely large demand shocks induce responses of hours and hourly wages, both for positive and negative shocks.

Furthermore, there is much macroeconomic empirical evidence pointing towards relative downward nominal wage rigidity. Kandil (1995) shows for a sample of 19 industrialized countries that in response to permanent mon-

etary policy shocks nominal wages generally respond stronger to positive shocks than to negative shocks of equal magnitude. Similar evidence in response to permanent aggregate demand shocks is provided by Kandil (2006) for United States industries and Kandil (2010) for a large variety of industrialized countries.²

There is a variety of wage adjustment theories accounting for downward nominal wage rigidity, the most prominent being contract theory (Fischer, 1977; Taylor, 1979), implicit contract theory (Baily, 1974; Azariadis, 1975; Gordon, 1976; Stiglitz, 1986), efficiency wage theory (Weiss, 1980; Akerlof, 1982; Shapiro and Stiglitz, 1984; Weiss, 1990), the fair wage hypothesis (Akerlof and Yellen, 1990), and the insider-outsider theory (Lindbeck and Snower, 1988). These theories aim at explaining, why firms avoid nominal wage cuts or, in the case of the (implicit) contract theory, why nominal wages are sluggish in general. However, none of these theories explain all three of the empirical regularities on nominal wage adjustment as outlined above.

In this paper we offer a new theory of downward nominal wage rigidity resting on worker loss aversion in the wage dimension, endogenous reference wage dynamics, and the implications of these dynamics for the firm's decision making. The resulting theory provides an account of asymmetric nominal wage rigidity in line with the empirical evidence cited above. Although, there is no hard evidence for a direct link of worker loss aversion and downward nominal wage rigidity, there is ample indicative evidence for the existence of such a link. Dunn (1996) presents survey evidence from US labor markets and finds that the behavior of labor supply is consistent with the notion of loss averse workers. Similar evidence is also presented by Goette et al. (2004) and Fehr and Goette (2007).

Furthermore, there is a large literature that documents that relative pay matters for subjective well-being (Clark and Oswald, 1996). Workers evalu-

²In addition to the asymmetric wage reaction in response to the permanent demand shock, Kandil (1995, 2006, 2010) finds an asymmetric reaction of output. Output responds much stronger to permanent negative demand shocks than to positive ones, a feature which is implied by standard theories of downward nominal rigidities and, given standard production technologies, also predicted by our model. This asymmetry in output is further documented by a large empirical literature. While DeLong and Summers (1988) DeLong and Summers (1988), Cover (1992), Kandil (2001), Ravn and Sola (2004) provide evidence for the United States, Karras (1996), Lenz (1997), Kandil (1999), and Karras and Stokes (1999) provide evidence for a wide variety of industrialized countries. Finally, evidence for developing countries is given by Kandil (1998), Tan et al. (2010), and Mehrara and Karsalari (2011).

ate their wages relative to a reference point, e.g. in the form of an implicit wage norm (Jaques, 1956, 1961), past earnings (Clark, 1999; Grund and Sliwka, 2007; Kawaguchi and Ohtake, 2007), or the earnings of others (Clark and Oswald, 1996; Clark et al., 2008). Falling behind reference points lowers life satisfaction and gives rise to negative morale effects. Supportive evidence for such morale effects is provided by, e.g. Kube et al. (2013) who document in a field experiment that there is a highly asymmetric reaction of work morale to positive and negative deviations from a reference wage. Similar evidence is provided by a field experiment by Chemin and Kurmann (2014). Survey evidence for the United States and various European economies suggests that amongst the most important factors for why firms do not adjust wages downward is the risk of negative effect to workers' morale (Campbell and Kamlani, 1997; Du Caju et al., 2015). However, Chen and Horton (2015) show that the effect on work morale vanishes if the wage cut is justified by reasonable arguments such as severe financial stress of the firm. Furthermore, Koch (2015) shows in an laboratory experiment that wage cuts in recessions are stronger in the absence of reference wages. If reference wages exist, wage cuts are smaller by approximately half the amount.

In our model, loss-averse workers evaluate wages relative to a reference wage. Kőszegi and Rabin (2006, 2007, 2009) and Heidhues and Kőszegi (2005, 2008, 2014) argue that reference points are determined by agents' rational expectations about outcomes from the recent past. There is much empirical evidence suggesting that reference points are determined by expectations, in concrete situations such as in police performance after final offer arbitration (Mas, 2006), in the United States TV show "Deal or no Deal" (Post et al., 2008), with respect to domestic violence (Card and Dahl, 2011), in cab drivers' labor supply decisions (Crawford and Meng, 2011), in the effort choices of professional golf players (Pope and Schweitzer, 2011), or in the aggressiveness of professional soccer players (Bartling et al., forthcoming). In the context of laboratory experiments, Knetsch and Wong (2009) and Marzilli Ericson and Fuster (2011) find supporting evidence from exchange experiments, Abeler et al. (2011) and Gill and Prowse (2012) from effort provision experiments, Banerji and Gupta (2014) from an auction experiment, and Karle et al. (2015) from a consumption choice experiment. Endogenizing workers' reference wages in this way allows our model to capture that current nominal wage changes influence the workers' future reference wage and thereby affect la-

bor supply. That reference wages influence reservation wages via this effect is supported by experimental evidence of Falk et al. (2006) who introduce a minimum wage as reference point and show that this introduction leads to an increase in the subjects' reservation wage, whereas the removal of that minimum wage, only leads to a marginal reduction in reservation wage. These pieces of evidence are consonant with the assumptions underlying our analysis. Our analysis works out the implications of these assumptions for state-dependent nominal wage sluggishness in the form of asymmetric nominal wage adjustment for positive and negative labor demand shocks.

While we are not of course the first to explain downward nominal wage rigidity through workers' loss aversion with respect to wages, our innovation lies in accounting for all three of the empirical regularities above through loss aversion. McDonald and Sibly (2001) set up an insider-outsider model with wage bargaining, where workers are loss averse with respect to real wages and where the reference wage equals last period's wage, i.e. the status quo, as suggested by Kahneman et al. (1991). They find that wages are rigid with respect to the reference wage, giving rise to real effects of monetary policy for expansionary monetary shocks. An analogous result is derived by Bhaskar (1990) in a model of union bargaining, where workers are loss averse with respect to their own wages relative to wages paid to members of other unions. Finally, Eliaz and Spiegler (2014) analyze loss averse workers in a restricted search and matching model. They follow Kőszegi and Rabin (2006) and assume that reference points are determined by rational expectations from the recent past. Eliaz and Spiegler (2014) find that in response to productivity shocks, wages of newly hired workers are (imperfectly) flexible, whereas they are downward rigid for existing workers. As noted, none of these papers can explain all three pieces of evidence outlined above.

2.3 Model

We incorporate reference-dependent preferences and loss aversion into an otherwise standard model of monopsony on the labor market. Workers are loss averse with respect to nominal wages. They evaluate nominal wages relative to their reference wage, which depends on their lagged rational wage expectations from the recent past, i.e. workers are backward-looking. For simplicity, we abstract from saving, implying that workers become single-

period optimizers. Firms are monopsonists and can set their wages freely in each period to maximize their total expected discounted profits. Firms' wage setting decision is forward-looking, taking into account their influence on the workers' future reference wage. Thus, reference dependence in our model is obviously an intertemporal phenomenon, linking the decisions in one period to the decisions in the next. To analyse the firms' wage setting decision in response to permanent labor demand shocks in such an intertemporal context we consider a dynamic two-period analysis, for algebraic simplicity. (A multi-period analysis with longer time horizons would not affect our qualitative conclusions.)

2.3.1 Labor Supply Curve of the Loss Averse Worker

We assume that workers are loss averse with respect to nominal wage changes, i.e. the perceived utility losses from nominal wage decreases relative to the reference wage are weighted more heavily than the perceived utility gains from nominal wage increases of equal magnitude. This gives rise to a labor supply curve which is convexly kinked at the reference wage. In what follows, we assume that this labor supply curve is upward sloping, since the substitution effect of a nominal wage change dominates the income effect.³ Consequently the employment increase associated with a nominal wage increase is small relative to the employment decrease associated with a wage decrease of equal magnitude.

The worker's preferences in period t are represented by the following utility function⁴

$$U_t(c_t, n_t) = U_t^c(c_t) - \theta_i \frac{n_t^{\vartheta_i}}{\vartheta_i}, \quad (2.1)$$

where c_t is consumption in period t , θ_i is a shifting parameter that ensures continuity of the worker's preferences at the nominal reference wage⁵ W_t^r

³As long as labor is less responsive to nominal wage increases (relative to the reference wage) than to nominal wage decreases, it can be shown that our model can explain the above outlined three empirical regularities on nominal wage adjustment, irrespective of the sign of the slope of the labor supply curve.

⁴In what follows, we normalize the worker's marginal utility of consumption $\frac{\delta U_t^c}{\delta c_t}$ equal to 1.

⁵Therefore, it must hold that $\theta_g = (W^r)^{1-\frac{\lambda_l}{\lambda_g}} \theta_l^{\frac{\lambda_l}{\lambda_g}}$.

and n_t is hours worked in period t .⁶ The parameter ϑ_i is an indicator function of the form

$$\vartheta_i = \begin{cases} \vartheta_g & \text{for } W_t > W_t^r, \text{ i.e. gain domain} \\ \vartheta_l & \text{for } W_t < W_t^r, \text{ i.e. loss domain} \end{cases}, \quad (2.2)$$

which describes the degree of the worker's loss aversion and where W_t and W_t^r are the workers current nominal wage and reference wage, respectively.

For loss averse workers $\vartheta_g > \vartheta_l$, which implies that the worker's disutility of labor $U_t^n(n_t) = \frac{n_t^{\vartheta_i}}{\vartheta_i}$ is steeper, i.e. the marginal disutility of labor is higher in the gain domain than in the loss domain. Therefore, the workers willingness to adjust hours is lower when the nominal wage is above the worker's reference wage than when it is below. Since labor income is assumed to be the only means to finance nominal consumption, the household's budget constraint is $P_t c_t = W_t n_t$. Without loss of generality, we normalize the price P_t to unity so that real and nominal wages are the same. Maximization of the utility function (2.1) subject to the budget constraint yields the following kinked labor supply function

$$n_t = \begin{cases} \left(\frac{W_t}{\vartheta_g}\right)^{\lambda_g} & \text{for } W_t > W_t^r, \text{ i.e. gain domain} \\ \left(\frac{W_t}{\vartheta_l}\right)^{\lambda_l} & \text{for } W_t < W_t^r, \text{ i.e. loss domain} \end{cases}, \quad (2.3)$$

where $\lambda_i = \frac{1}{\vartheta_i - 1}$ denotes the Frisch elasticity of labor supply. Loss aversion with respect to nominal wage changes implies that $\lambda_g < \lambda_l$, i.e. the worker reacts stronger to wage decreases relative to the reference wage (by reducing employment) than to wage increases relative to the reference wage (by increasing employment).⁷

The worker's nominal reference wage W_t^r is formed at the beginning of each period. In the spirit of Kőszegi and Rabin (2006), we assume that the worker's nominal reference wage depends on her rational nominal wage

⁶Throughout the model, capital letters denote nominal variables, while small letters denote real variables.

⁷While this point is crucial for the predictions of our theory, it is worth pointing out that these results hold irrespective of the sign of the slope of the labor supply curve above the kink as long as the ratio of the absolute slopes above and below the kink remains unchanged (i.e. the labor supply curve is steeper above than below the kink). Thus, our theory does cover the evidence that the substitution effect always outweighs the income effect (upward sloping labor supply curve) as well as the evidence of, e.g., Kőszegi and Rabin (2006) and others according to which we have a backward bending labor supply curve above the reference wage.

expectation from the recent past. Shocks materialize unexpectedly in the course of the period and therefore do not enter I_t , the information set available to the worker at the beginning of the period. In short, the worker observes the shock with a one-period lag. Thus, the worker's nominal reference wage is defined as $W_t^r = E_{t-1}[W_t | I_{t-1}]$. Changes in the reference wage W_t^r change the position of the kink of the worker's labor supply curve and also shift the labor supply curve as a whole. We follow Kőszegi and Rabin (2006) and assume that the worker's expected nominal wage implicitly determines the worker's endogenous income target.⁸ Thus, an increase in the expected nominal wage raises her implicit income target, whereas a decrease in the expected nominal wage lowers it. If, at the beginning of the period, the worker anticipates a higher (lower) nominal wage for the following period, i.e. her reference wage increases (decreases), she will supply relatively more (less) labor in order to reach her new higher (lower) implicit income target. From this, it follows that the worker's labor supply curve shifts outwards (inwards) in response to an upward (downward) adjustment of the worker's reference wage.

The kink, lying at the intersection of the two labor supply curves $n_t(W_t, \lambda_g, \theta_g)$ and $n_t(W_t, \lambda_l, \theta_l)$, is given by the wage-labor combination

$$(\widehat{W}_t, \widehat{n}_t) = \left(W_t^r, \left(\frac{\theta_g}{\theta_l} \right)^{\frac{1}{\lambda_l - \lambda_g}} \right), \quad (2.4)$$

where " $\widehat{}$ " denotes the value of a variable at the kink.

2.3.2 The Firm's optimization problem

The firm maximizes its expected discounted profits

$$\Pi_t^{Total} = \Pi_t + \beta \Pi_{t+1} \quad (2.5)$$

where $\Pi_t = y_t - W_t n_t$ are period t profits and β is the discount factor. The firm takes into account its production function $y_t(n_t)$, the workers' kinked labor supply function $n_t = \left(\frac{W_t}{\theta_i} \right)^{\lambda_i}$, and the influence of its wage decision on

⁸If the labor demand curve is inelastic and the firm faces costs of labor adjustment (a realistic scenario, certainly for the short run), so that the profit-maximizing employment can take place in the inelastic portion of the labor demand curve, then increases in the reference wage translate one-to-one into increases in the reference income.

the workers' future labor supply via changes in the workers' reference wage. The resulting first order condition of the firm's optimization problem reads as

$$\begin{aligned} \frac{\partial \Pi_t^{Total}}{\partial W_t} &= \left[\frac{\partial y_t(n_t)}{\partial n_t(W_t)} \frac{\partial n_t(W_t)}{\partial W_t} - n_t(W_t) - W_t \frac{\partial n_t(W_t)}{\partial W_t} \right] + \beta \left[\frac{\partial \Pi_{t+1}}{\partial W_{t+1}^r} \frac{\partial W_{t+1}^r}{\partial W_t} \right] \\ &= 0, \end{aligned} \quad (2.6)$$

which is equivalent to

$$\underbrace{\frac{\partial y_t(n_t)}{\partial n_t}}_{mpl_t} - \underbrace{\left(W_t(n_t) + \frac{\partial W_t}{\partial n_t} n_t \right)}_{mcl_t} = -\beta \frac{\partial \Pi_{t+1}}{\partial W_{t+1}^r} \frac{\partial W_{t+1}^r}{\partial W_t} \frac{\partial W_t}{\partial n_t}. \quad (2.7)$$

The term on the left hand side is the current period marginal product of labor mpl_t minus the current period marginal cost of labor mcl_t . The term on the right hand side measures the influence of the wage setting decision in period t on the workers' next period reference wage and thereby the firm's future profits. For the partial derivatives, it holds that $\frac{\partial W_{t+1}^r}{\partial W_t} \geq 0$ and $\frac{\partial W_t}{\partial n_t} > 0$. Thus, for $\frac{\partial \Pi_{t+1}}{\partial W_{t+1}^r} > 0$, the reference-wage-updating effect drives a negative wedge between the marginal product of labor and the marginal cost of labor, i.e. $mpl_t \leq mcl_t$. Note that in the absence of reference-wage-updating the standard optimality condition of a firm holds, i.e. $mpl_t = mcl_t$. Only if the firm's wage setting decision has an influence on the future reference wage the firm faces a tradeoff between current period optimality (determined by the left hand side of equation (2.7)) and future ramifications of the current decision (determined by the right hand side of equation (2.7)).

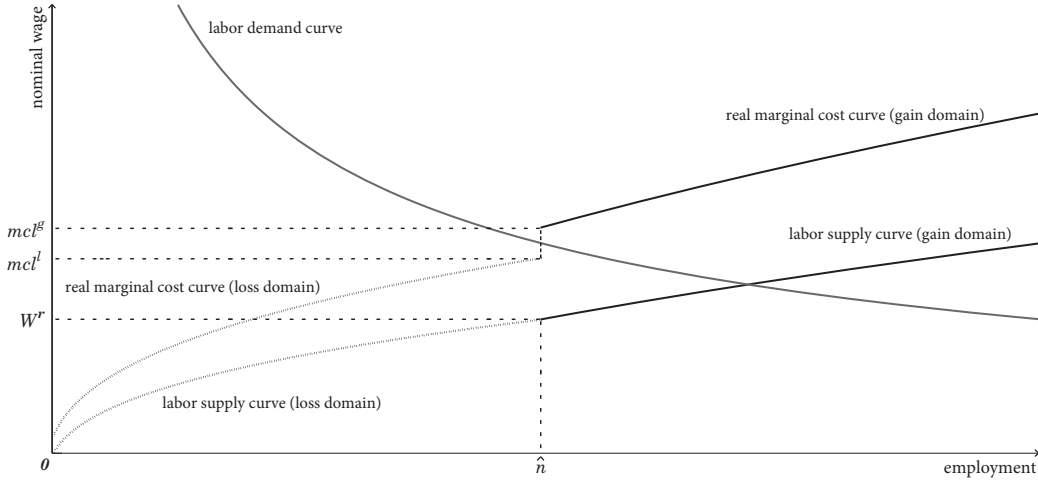
In what follows we assume that the firm's production function is given by $y_t(n_t) = \mu n_t^\alpha$ where $\mu > 0$ and $0 < \alpha < 1$. The firm's current period labor demand function, given by its marginal product of labor (mpl), is downward sloping: $l_t^D = mpl_t = \mu \alpha n_t^{(\alpha-1)}$. Since the labor supply function of the loss averse worker is kinked at the reference wage W^r , the firm's real marginal cost of labor is discontinuous at the kink:

$$mcl_t^i(\hat{n}_t, \lambda_i, \theta_i) = \left(1 + \frac{1}{\lambda_i} \right) \theta_i \hat{n}_t^{\frac{1}{\lambda_i}} \quad \text{for } i = g, l. \quad (2.8)$$

The interval $[mcl_t^l, mcl_t^g]$ we call "marginal cost discontinuity".⁹

⁹Note that $mcl_t^g(\hat{n}_t, \lambda_g, \theta_g) > mcl_t^l(\hat{n}_t, \lambda_l, \theta_l)$. See also Figure 2.1.

Figure 2.1: Initial problem of the monopsonistic firm.



We assume that in the initial steady state, the exogenously given reference wage is W_{ss}^r . Furthermore, in the steady state the firm's labor demand curve (mpl) intersects the marginal cost discontinuity. To fix ideas, we assume that initially the labor supply curve crosses the midpoint of the discontinuity in the marginal cost curve¹⁰, as depicted in Figure 2.1. This assumption permits us to derive the symmetry characteristics of wage and employment responses to positive and negative labor demand shocks. It follows that the firm's optimal wage in the initial steady state W_{ss}^* is equal to W_{ss}^r .¹¹

2.3.3 Demand Shocks

For simplicity, we analyse the firm's wage setting reaction in response to permanent labor demand shocks in a two-period context. These labor demand shocks, represented by ε_t , are unexpected and enter the labor demand function multiplicatively:

$$l_t^D = \mu \alpha n_t^{(\alpha-1)} \varepsilon_t. \quad (2.9)$$

¹⁰This implies that the slope parameter of the firm's labor demand function has to fulfill $\mu = \frac{mcl_{ss}(\widehat{n}_{ss}, \lambda_g, \theta_g) + mcl_{ss}(\widehat{n}_{ss}, \lambda_l, \theta_l)}{2\alpha \widehat{n}_{ss}^{(\alpha-1)}}$, evaluated at the initial steady state.

¹¹The proof is straightforward: Let ν be an arbitrarily small number. Then for wages equal to $W_{ss}^r + \nu$ the firm faces a situation in which marginal cost is higher than marginal revenue product and decreasing the wage would raise the firm's profit, while for wages equal to $W_{ss}^r - \nu$ the firm faces a situation in which marginal cost is lower than marginal revenue product and increasing the wage would raise the firm's profit. Thus $W_{ss}^* = W_{ss}^r$ has to be the profit maximizing wage in the initial steady state.

We consider the effects of a shock that hits the economy in period t . We define a “small” shock as one that leaves the labor demand curve passing through the marginal cost discontinuity, and a “large” shock as one that shifts the labor demand curve sufficiently so that it no longer passes through the marginal cost discontinuity.

The maximum size of a small shock for the labor demand function (2.9) is

$$\bar{\varepsilon}_t(\lambda_i, \theta_i) = \frac{\left(1 + \frac{1}{\lambda_i}\right) \theta_i}{\mu\alpha} \widehat{n}_{ss}^{\frac{1}{\lambda_i} - (\alpha-1)}, \quad (2.10)$$

i.e. $\bar{\varepsilon}_t(\lambda_i, \theta_i)$ is the shock size for which the shifted labor demand curve lies exactly on the upper (for $\bar{\varepsilon}(\lambda_g, \theta_g)$) or lower (for $\bar{\varepsilon}(\lambda_l, \theta_l)$) boundaries of the marginal cost discontinuity. In the analysis that follows, we will distinguish between small and large permanent labor demand shocks. We simulate our model numerically in order to quantitatively assess the wage setting reaction of the firm to small and large labor demand shocks.

2.3.4 Calibration

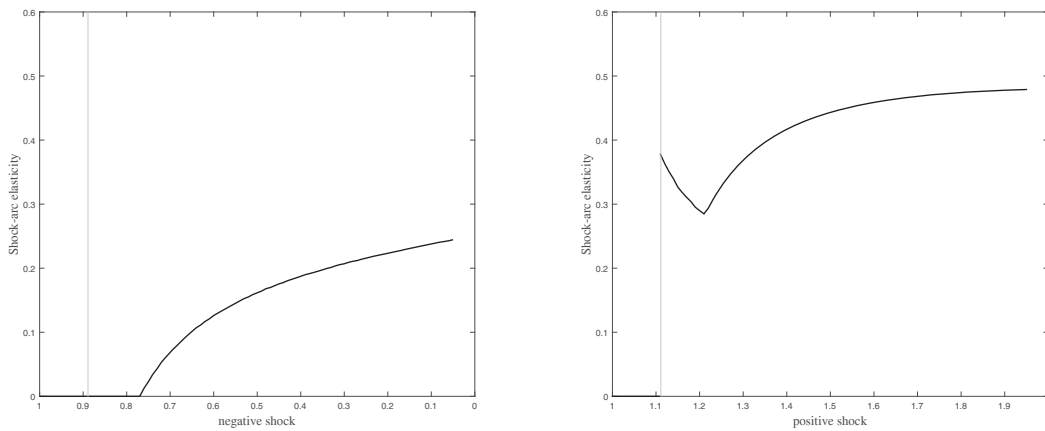
We calibrate the model for a quarterly frequency in accordance with standard values in the literature. We assume an annual interest rate of 4 percent, which yields a discount factor $\beta = 0.99$. Loss aversion is measured by the relative slopes of the demand curves in the gain and loss domain, i.e. $\kappa = \frac{\lambda_l}{\lambda_g}$. The empirical literature finds that the loss aversion ratio is commonly around 2 (e.g. Bleichrodt et al., 2001; Tversky and Kahneman, 1992; Pennings and Smidts, 2003; Booij and van de Kuilen, 2009). Therefore, we set $\kappa = 2$. Following Galí (2008), we set $\alpha = 2/3$. The Frisch elasticity of labor supply is set to $\lambda_g = 1.5$, which ensures that λ_g and λ_l are well between the estimates of Prescott (2004), Chetty et al. (2011), and Fiorito and Zanella (2012), which range from 1.1 to 3.¹² The base calibration is summarized in Table 2.1.

¹²This calibration takes a macro point of view, as micro estimates of the Frisch labor supply elasticity are usually much lower, i.e. between 0 and 1 with a strong tendency towards zero rather than one (for a survey see Fiorito and Zanella (2012)). In section 2.4.2, we show that the results hold true also for a micro approach to our calibration.

Table 2.1: Base calibration.

Parameter	Symbol	Value
Discount rate	β	0.99
Frisch elasticity of labor supply (gain domain)	λ_g	1.5
Frisch elasticity of labor supply (loss domain)	λ_l	3
Loss aversion	κ	2
Output elasticity of labor	α	2/3
Shifting parameter	θ_l	1/2

Figure 2.2: Shock-arc elasticities to positive and negative demand shocks.



2.4 Results

Figure 2.2 presents the shock-arc elasticities of the wage ($\tilde{\eta}_{\varepsilon, W} = \frac{\% \Delta W}{\% \Delta \varepsilon}$) in the period of the shock t for negative and positive labor demand shocks of the two-period model, given the base calibration given in Table 2.1. On the vertical axis we show the shock-arc elasticities of wage, which measure the relative strength of the wage reaction in response to negative and positive labor demand shocks. The horizontal axis measures the shock, where the shock size increases from the left to the right. The vertical, dotted lines denote the thresholds between the small and the large labor demand shocks as defined in section 2.3.3.

According to our numerical analysis the firm's wage reaction in response to permanent labor demand shocks depends crucially on the size and the sign of the shock. Figure 2.2 indicates that wages are completely rigid for small positive and small negative labor demand shocks (region left of the dotted lines in both panels of Figure 2.2), while they are relatively downward

sluggish for larger shocks. Moreover, for a certain range of large shocks, wages are completely downward rigid but upwards flexible.

2.4.1 Intuition

Small labor demand shocks

As noted, for a sufficiently small demand shock $\bar{\varepsilon}_t(\lambda_l, \theta_l) \leq \varepsilon_t^s \leq \bar{\varepsilon}_t(\lambda_g, \theta_g)$ the labor demand curve still intersects the marginal cost discontinuity, i.e. $l_t^D(\hat{n}_t) \in [mcl_t^l, mcl_t^g]$.

Therefore, the prevailing nominal steady state wage, which is equal to the worker's current reference wage, remains the firm's profit-maximizing wage,¹³ i.e. $W_t^* = W_{ss}^* = W_{ss}^r$, and we have *complete wage rigidity*. With rigid wages, labor supply is unaffected by the small labor demand shock. Accordingly, the profit-maximizing amount of labor employed remains unchanged as well: $\Delta n_t^* = 0$. This holds true *irrespective of the sign of the small labor demand shock*.

Large labor demand shocks

In contrast to the small labor demand shock, for a large shock, i.e. $\varepsilon_t^l > \bar{\varepsilon}_t(\lambda_g, \theta_g)$ or $\varepsilon_t^l < \bar{\varepsilon}_t(\lambda_l, \theta_l)$, generally both, a nominal wage and a labor reaction are induced.

In our analysis, there are two channels whereby a large permanent labor demand shock affects nominal wages and employment; a direct demand and supply effect and an indirect reference-wage-updating effect.

The direct demand and supply effect: For the analysis of nominal wage adjustment in response to large variations in labor demand it proves useful to suppose, for the moment, that the worker's reference wage is exogenously fixed and does not change. This implies that $\frac{\partial W_{t+1}^r}{\partial W_t} = 0$ and therefore the firm's profit maximization problem becomes a one-period problem. According to the optimality condition (2.7), the new profit-maximizing wage of the firm is determined by the standard condition according to which $mpl_t = mcl_t$. The new profit maximizing wage of the firm is

¹³Compare the proof from footnote 11.

$$W_t^* = \theta_i \left[\frac{\mu \alpha \varepsilon_t^l}{\left(1 + \frac{1}{\lambda_i}\right) \theta_i} \right]^{\frac{1}{\lambda_i(1-\alpha)+1}}, \quad (2.11)$$

while its corresponding profit-maximizing amount of labor is

$$n_t^* = \left[\frac{\mu \alpha \varepsilon_t^l}{\left(1 + \frac{1}{\lambda_i}\right) \theta_i} \right]^{\frac{1}{(1-\alpha)+\frac{1}{\lambda_i}}}, \quad (2.12)$$

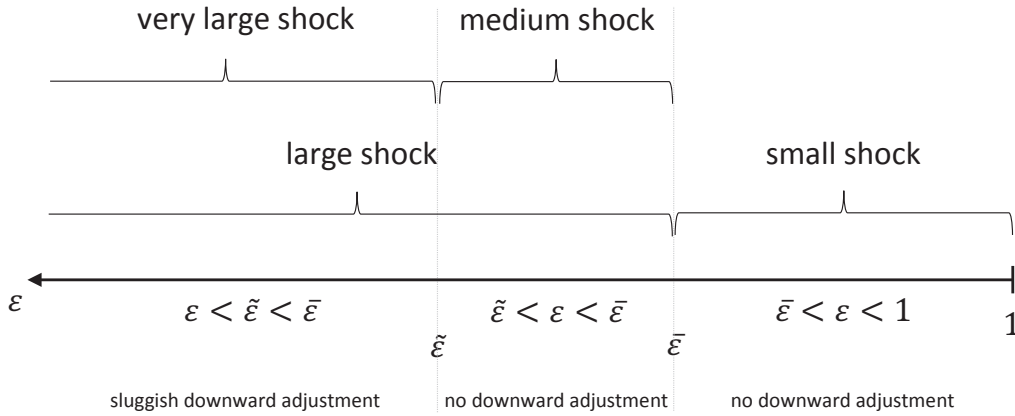
where $\lambda_i = \lambda_g$, $\theta_i = \theta_g$ for positive and $\lambda_i = \lambda_l$, $\theta_i = \theta_l$ for negative shocks, respectively.

Whether the new profit maximizing wage and employment reactions are larger for positive or negative shocks depends on the relative slopes of the demand and supply functions, which differ for negative and positive shocks due to loss aversion. For the base calibration and reasonable shock sizes¹⁴, equations (2.11) and (2.12) imply that abstracting from any adjustment in the workers' reference wage, nominal wages are *relatively downward sluggish* (i.e. less responsive to large negative than to large positive shocks). Intuitively, the change of quantity in response to a large labor demand shock depends positively on λ_i , the Frisch elasticity of labor supply, whereas the change of the wage in response to a large labor demand shock depends negatively on λ_i . Since for the loss averse worker $\lambda_g < \lambda_l$, the labor reaction of the firm facing loss-averse workers is relatively smaller in response to large positive labor demand shocks than to large negative ones of equal magnitude. This however implies that wages are relatively less responsive to negative than to positive large labor demand shocks, since the former move the firm along the relatively flat portion of the labor supply curve, whereas the latter move it along the relatively steep portion of the labor supply curve.

The reference-wage-updating effect: Accounting for the adjustment of the worker's reference wage in response to large labor demand shocks changes the wage setting decision of the firm dramatically. Now $\frac{\partial W_{t+1}^r}{\partial W_t} \neq 0$. The new profit-maximizing wage of the firm is not only determined by the relation between mpl_t and mcl_t , but also by the effect of the wage decision on the reference wage.

¹⁴Refer to Section 2.4.2 for a sensitivity analysis concerning this condition.

Figure 2.3: Classification of negative labor demand shocks.



The threshold shocks in this figure are $\tilde{\varepsilon} = \tilde{\varepsilon}_t(\lambda_l, \theta_l)$ and $\bar{\varepsilon} = \bar{\varepsilon}_t(\lambda_l, \theta_l)$.

As discussed above, a large labor demand shock induces a nominal wage and a labor reaction in the shock period t . Accordingly the worker's reference wage adjusts at the beginning of the following period $t + 1$, i.e. $W'_{t+1} = E_t[W_{t+1} | I_t] = W_t^*$, which triggers an outward shift of the worker's labor supply curve for positive labor demand shocks and an inward shift for negative labor demand shocks. This phenomenon is the *reference-wage-updating effect*. This effect implies that in period $t + 1$ the firm's profit is higher than in the shock period t for positive permanent shocks, while it is lower for negative permanent shocks due to the worker's labor supply reaction in response to the change of her implicit income target¹⁵.

Since the firm anticipates this, the following incentives arise: In response to a large positive labor demand shock, the firm could raise the nominal wage above the optimal current period nominal wage W_t^* in order to induce a stronger outward shift of the worker's labor supply curve in the following period. By contrast, in response to a large negative labor demand shock, the firm could try to dampen or even completely avoid the inward shift of the worker's labor supply curve in the next period by lowering the nominal wage less than otherwise optimal or by not lowering the nominal wage at all¹⁶.

¹⁵Intuitively, the firm can employ more labor for the same optimal nominal wage $W'_{t+1} = W_t^*$ in the case of a large positive labor demand shock, whereas it must employ less labor in the case of a large negative labor demand shock.

¹⁶Note that setting a nominal wage below W_t^* is never an option for the firm since it negatively affects future profits.

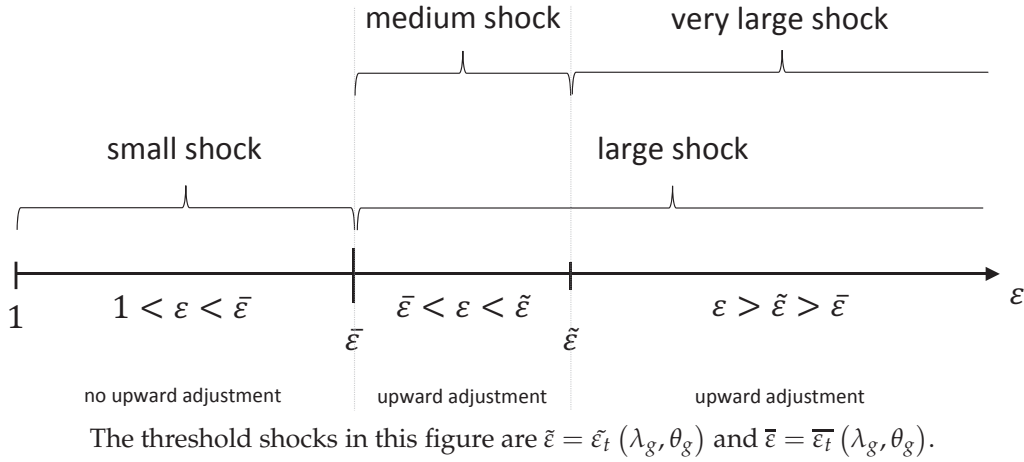
Whether or not this occurs, generally depends on whether the firm's gain from an upward deviation from the optimal nominal wage W_t^* in terms of future profits (due to the relative rise in the reference wage) exceeds the firm's loss in terms of present profits (due to not setting the profit maximizing wage), i.e. whether $\Pi_t(W_t') + \beta\Pi_{t+1}(W_{t+1}^r = W_t') > \Pi_t(W_t^*) + \beta\Pi_{t+1}(W_{t+1}^r = W_t^*)$ where $W_t' > W_t^*$. While period t marginal losses (the left hand side of equation (2.7)) strictly increase in an upward deviations from W_t^* , the discounted period $t + 1$ gains due to the reference-wage-updating effect (the right hand side of equation (2.7)) feature an inverse u-shaped function in an upward deviations from W_t^* . Therefore, the firm exploits the reference-wage-updating effect as long as the discounted marginal gain is larger than its period t marginal loss. Hence, the optimal wage W_t' is set, where marginal gains are equal to marginal losses.¹⁷

Figure 2.2 shows that for large negative labor demand shocks there is some threshold value $\tilde{\varepsilon}_t(\lambda_l, \theta_l)$ up to which there is still complete downward nominal wage rigidity. For shocks larger than $\tilde{\varepsilon}_t(\lambda_l, \theta_l)$ there is some downward nominal wage adjustment. More precisely, we find that for $\varepsilon_t^l < \tilde{\varepsilon}_t(\lambda_l, \theta_l)$, the firm sets the wage W_t' such that $W_{ss}^* > W_t' > W_t^*$, while for $\tilde{\varepsilon}_t(\lambda_l, \theta_l) < \varepsilon_t^l < \bar{\varepsilon}_t(\lambda_l, \theta_l)$ the firm does not adjust the wage, i.e. $W_t' = W_{ss}^*$. Due to these qualitative differences in the response of wages to large negative demand shocks, we divide the class of large shocks into two categories: The range of large shocks which do not induce any downward wage reaction we classify as "medium-sized" shocks, whereas all other large shocks are considered "very large" shocks. Figure 2.3 summarizes the full classification of negative permanent labor demand shocks. These results suggest that the firm's incentive to dampen the inward shift of the worker's labor supply curve due to an adjustment in the worker's reference wage in the following period is very strong. The firm apparently always deviates upwards from W_t^* , the optimal wage given by equation (2.11), and sets W_t' instead.

By contrast, for large positive labor demand shocks $\varepsilon_t^l > \bar{\varepsilon}_t(\lambda_g, \theta_g)$, the firm always adjusts the nominal wage upwards (see Figure 2.2). However, our results also indicate that the firm's incentive to induce a stronger outward shift of the worker's labor supply curve declines as the positive shock

¹⁷Note that the marginal gain from an upward deviation from W_t^* depends negatively on the absolute value of the shock, i.e. the effect declines the larger the positive shock, while it increases the larger the negative shock (the smaller the shock in absolute value).

Figure 2.4: Classification of positive labor demand shocks.



increases. In particular, our numerical results suggest that the firm does not always set a wage W'_t that is higher than W_t^* for large positive labor demand shocks. If the shock exceeds a certain threshold, i.e. $\varepsilon_t^l > \tilde{\varepsilon}_t(\lambda_g, \theta_g)$, the firm's loss in terms of present profits from not setting W_t^* is not compensated by the gain in terms of future profits. This is due to the effect that the marginal gain from an upward deviation declines in the absolute size of the shock. Thus only for medium-sized positive shocks $\tilde{\varepsilon}_t(\lambda_g, \theta_g) > \varepsilon_t^l > \bar{\varepsilon}_t(\lambda_g, \theta_g)$ the firm set the wage W'_t such that $W'_t > W_t^*$. Otherwise, for very large shocks, it just sets W_t^* . Figure 2.4 summarizes the full classification of positive permanent labor demand shocks.

Finally, comparing the left hand panel and the right hand panel from Figure 2.2, our numerical results also confirm that for large labor demand shocks nominal wages always adjust stronger upwards than downwards for equiproportionate shocks as predicted by our theory.

2.4.2 Sensitivities

Figure 2.5 shows the shock-arc elasticities of the wage for the following values of the loss aversion parameters: $\kappa \in (1.43; 2; 4.8)$, where our base case is $\kappa = 2$. The lower value was estimated by Schmidt and Traub (2002). The higher value was estimated by Fishburn and Kochenberger (1979). Intermediate values are supported by Bleichrodt et al. (2001), Tversky and Kahneman (1992), Pennings and Smidts (2003), and Booij and van de Kuilen (2009).

Figure 2.5: Sensitivity with respect to the loss aversion parameter.

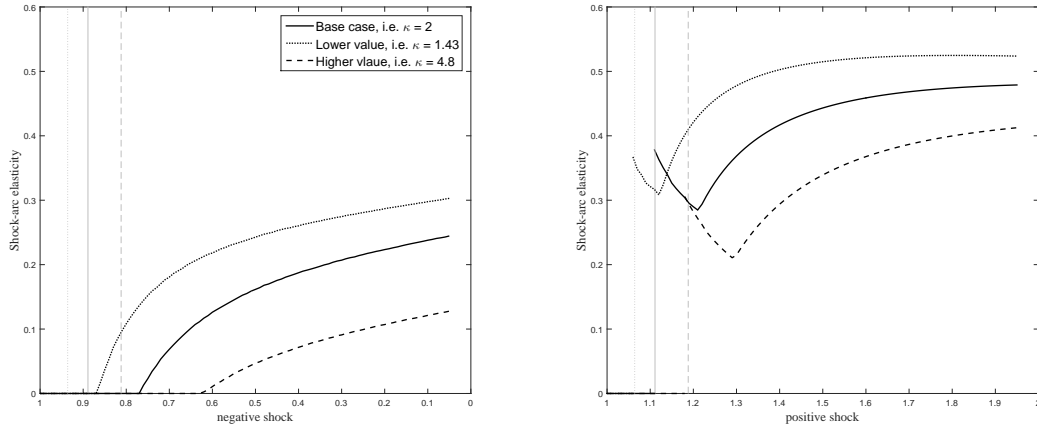


Figure 2.5 shows that the higher the loss aversion parameter, *ceteris paribus*, the more sluggish is wage adjustment in response to labor demand shocks, both upwards and downwards. The shock-arc elasticity curves stemming from higher parameter values always lie below the curves stemming from lower parameter values. Both critical shocks, positive and negative, increase. This implies that the marginal cost discontinuity widens as the loss aversion parameter increases, extending the range of full wage rigidity. The range of medium-sized shocks (i.e. large shocks for which there is upward flexibility but full downward rigidity) is shifted towards larger shocks, as the loss aversion parameter increases. Finally, overall wage sluggishness increases, as the positive and negative shock-arc elasticities are generally lower the higher the loss aversion parameter. It is apparent from the left panel of Figure 2.5 that the firms' incentive to avoid wage cuts increases substantially as the shock-arc elasticities are generally lower for higher loss aversion parameters compared to lower ones. Intuitively, the higher the Frisch elasticity of labor supply for the loss domain, the stronger is the permanent loss in profit due to a decrease in the reference wage. This stems from the fact that, according to the labor supply function (2.3), the reference-wage-updating effect increases in loss aversion. For the positive shock it holds that the higher the loss aversion parameter, the lower is the firm's incentive to deviate upwards. Since the reference-wage-updating effect is stronger, the firm does not necessitate to deviate by as much in order to produce the profit maximizing amount of labor in period $t + 1$.

Figure 2.6: Sluggishness from the direct *demand and supply* effect.

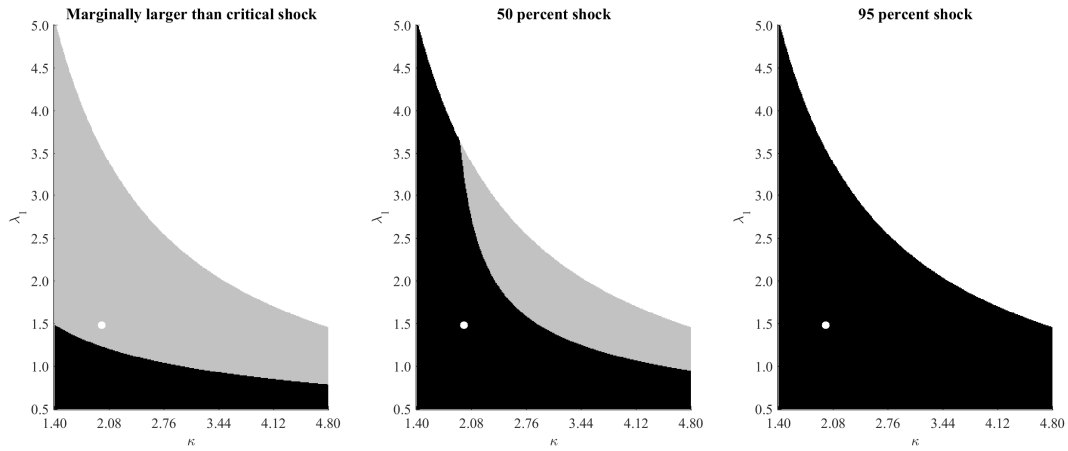
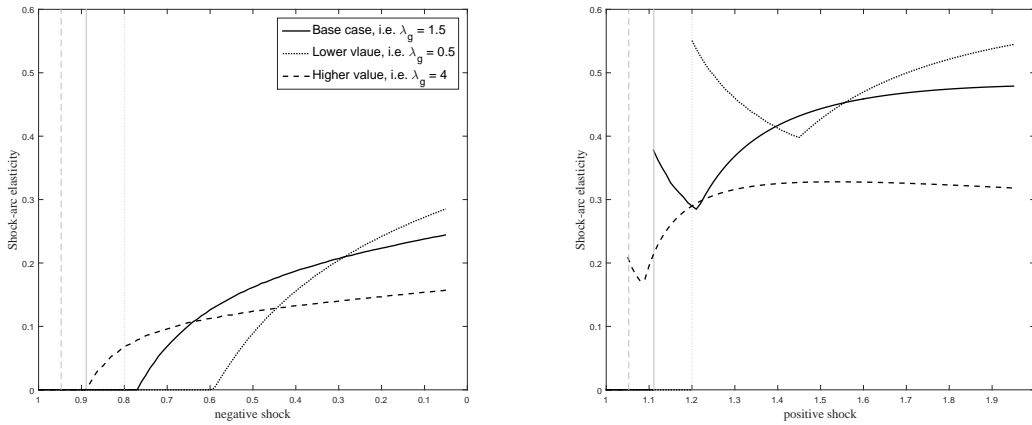


Figure 2.5 shows another interesting fact. Due to the negative reference-wage-updating effect, the firms' wage responses are always downward sluggish. This implies that the reference-wage-updating effect always dominates the direct demand and supply effect. Even though for the base calibration the direct demand and supply effect produces downward sluggishness, the direction of the direct demand and supply effect depends strongly on the calibration of the model. Figure 2.6 shows a comparison of the relative shock arc elasticities for positive and negative demand shocks for different combinations of the Frisch elasticity of labor supply for the gain domain (i.e. λ_g) and the loss aversion coefficient (i.e. κ) and for three different shock sizes. Black shaded areas denote relative upward sluggishness (i.e. shock-arc elasticities of the wage is larger for permanent negative labor demand shocks relative to positive ones of equal magnitude), while gray shaded areas show relative downward sluggishness. Areas in white are not considered, as the resulting parameter value for λ_l exceeds reasonable values.¹⁸ The white dot indicates our base calibration. As is apparent from Figure 2.6, for larger shocks (middle and right panel) the direct demand and supply effect generates upward sluggishness.

From this it follows that in the absence of the reference-wage-updating effect the shock-arc elasticities of wage in response to a permanent labor demand shock should be higher than their positive shock counterparts over a large range of the shocks. From the overall reaction (including both, the

¹⁸As indicated above, λ takes values between 0.5 and rarely above 4. The cutoff value for consideration is set to $\bar{\lambda} = 7.2$, which is the highest value considered in Figure 2.5.

Figure 2.7: Sensitivity with respect to the Frisch elasticity of labor supply.



direct demand and supply effect and the reference-wage-updating effect) it is apparent that the negative reference-wage-updating effect dominates the demand and supply effect over the full range of shocks considered.

Figure 2.7 shows the sensitivity with respect to the following reasonable values for the Frisch labor supply elasticity: $\lambda_g \in (0.5; 2; 4)$, where $\lambda_g = 1.5$ is our base case. We contrast our base with a much lower value $\lambda_g = 0.5$ as often estimated in micro studies (for a survey refer to Chetty et al. (2011)) and a much higher value $\lambda_g = 4$, as estimated by Imai and Keane (2004).

From Figure 2.7 it is apparent that the larger the Frisch elasticity of labor supply, the more sluggish are the responses of wages to positive and negative demand shocks. The intuition for this result is the same as for the loss aversion parameter: According to the labor supply equation (2.3), the larger Frisch elasticity of labor supply, the stronger is the reference-wage-updating effect. Therefore, the firm's incentive to avoid the negative reference-wage-updating effect increases with λ_g , while the necessity to deviate from W_0^* decreases with λ_g . Additionally, the qualitative result that wages are relatively downward sluggish carries over over the full range of elasticities considered. In contrast to the loss aversion parameter, however, the marginal cost gap closes, *ceteris paribus*, the larger the Frisch labor supply elasticities. Both, positive and negative critical shocks decrease and therewith the range of full wage rigidity decreases.

Our sensitivity analysis confirms that over the wide range of reasonable parameter values for the Frisch labor supply elasticity and the loss aversion parameter, our theory implies that (1) for small labor demand shocks, nomi-

nal wages are fully rigid, (2) for medium-sized shocks there is upward nominal wage adjustment for positive shocks, but complete downward nominal wage rigidity for negative shocks and (3) for large shocks, nominal wages decline less strongly to negative shocks than they increase to equiproportionate positive shocks.

2.5 Conclusion

With our theory of wage adjustment under loss aversion we are able to provide an integrated account of the three important empirical regularities concerning wage adjustments to labor market shocks. In particular, we can explain wage rigidity in the face of minor labor market shocks, downward nominal wage rigidity combined with upward wage adjustment in “normal” times, and wage cuts in deep recessions.

In contrast to the New Keynesian literature, our explanation of wage adjustment is thoroughly microfounded, without recourse to ad hoc assumptions. As future work, our model needs to be incorporated into a general equilibrium setting to validate the predictions of our theory.

References

- Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference points and effort provision. *American Economic Review* 101(2), 470–492.
- Abo-Zaid, S. (2013). Optimal monetary policy and downward nominal wage rigidity in frictional labor markets. *Journal of Economic Dynamics and Control* 37, 345–364.
- Agell, J. and P. Lundborg (2003). Survey evidence on wage rigidity and unemployment: Sweden in the 1990s. *Scandinavian Journal of Economics* 105, 15–30.
- Ahrens, S., I. Pirschel, and D. J. Snower (2014). A theory of price adjustment under loss aversion. *CEPR Discussion Papers No 9964*, Centre for Economic Policy Research, London.
- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics* 97, 543–569.

- Akerlof, G. A., W. R. Dickens, and G. L. Perry (1996). The macroeconomics of low inflation. *Brookings Papers on Economic Activity* 27, 1–76.
- Akerlof, G. A. and J. L. Yellen (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics* 105, 255–283.
- Altonji, J. G. and P. J. Devereux (1999). The extent and consequences of downward nominal wage rigidity. *NBER Working Papers No 7236*, National Bureau of Economic Research, Inc.
- Azariadis, A. (1975). Implicit contracts and underemployment equilibria. *Journal of Political Economy* 83, 1183–1202.
- Babecký, J., P. Du Caju, T. Kosma, M. Lawless, J. Messina, and T. Rõõm (2010). Downward nominal and real wage rigidity: Survey evidence from european firms. *Scandinavian Journal of Economics* 112, 884–910.
- Baily, M. N. (1974). Wages and employment under uncertain demand. *Review of Economic Studies* 41, 37–50.
- Banerji, A. and N. Gupta (2014). Detection, identification, and estimation of loss aversion: Evidence from an auction experiment. *American Economic Journal: Microeconomics* 6(1), 91–133.
- Bartling, B., L. Brandes, and D. Schunk (forthcoming). Expectations as reference points: Field evidence from professional soccer. *Management Science*.
- Bauer, T., H. Bonin, L. Goette, and U. Sunde (2007). Real and nominal wage rigidities and the rate of inflation: Evidence from west german micro data. *Economic Journal* 117, F508–F529.
- Benigno, P. and L. A. Ricci (2011). The inflation-output trade-off with downward wage rigidities. *American Economic Review* 101, 1436–1466.
- Bewley, T. F. (1995). A depressed labor market as explained by participants. *American Economic Review* 85, 250–254.
- Bewley, T. F. (1999). *Why wages don't fall during a recession*. Cambridge: Harvard University Press.
- Bhaskar, V. (1990). Wage relativities and the natural range of unemployment. *Economic Journal* 100, 60–66.

- Bleichrodt, H., J. L. Pinto, and P. P. Wakker (2001). Using descriptive findings of prospect theory to improve the prescriptive use of expected utility. *Management Science* 47, 1498–1514.
- Böckerman, P., S. Laaksonen, and J. Vainiomäki (2010). Micro and macro level wage rigidity: Lessons from finland. *Finnish Economic Papers* 23, 27–42.
- Booij, A. S. and G. van de Kuilen (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology* 30, 651–666.
- Campbell, C. M. and K. S. Kamlani (1997). The reasons for wage rigidity: Evidence from a survey of firms. *The Quarterly Journal of Economics* 112, 759–789.
- Card, D. and G. B. Dahl (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics* 126(1), 103–143.
- Card, D. and D. Hyslop (1997). *Does inflation “grease the wheels of the labor market”?* Reducing Inflation: Motivation and Strategy. Boston: National Bureau of Economic Research.
- Carlsson, M. and A. Westermark (2008). Monetary policy under downward nominal wage rigidity. *The B.E. Journal of Macroeconomics* 8, 1–61.
- Chemin, M. and A. Kurmann (2014). Do workers feel entitled to high wages? evidence from a long term field experiment. *mimeo*.
- Chen, D. L. and J. J. Horton (2015). The wages of pay cuts: Evidence from a field experiment. *mimeo*.
- Chetty, R., A. Guren, D. Manoli, and A. 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.
- Clark, A. E. (1999). Are wages habit-forming? Evidence from micro data. *Journal of Economic Behavior & Organization* 39, 179–200.

- Clark, A. E., P. Frijters, and M. A. Shields (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic Literature* 46, 95–144.
- Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics* 61, 359–381.
- Cobb, M. and L. Opazo (2008). Microeconomic evidence of nominal wage rigidity in Chile. *Working Papers Central Bank of Chile No 496*, Central Bank of Chile.
- Cover, J. P. (1992). Asymmetric effects of positive and negative money-supply shocks. *The Quarterly Journal of Economics* 107(4), 1261–1282.
- Crawford, V. P. and J. Meng (2011). New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review* 101(5), 1912–1932.
- DeLong, J. B. and L. H. Summers (1988). How does macroeconomic policy affect output? *Brookings Papers on Economic Activity* 19(2), 433–494.
- Dickens, W. T., L. Goette, E. L. Groshen, S. Holden, J. Messina, M. E. Schweitzer, J. Turunen, and M. E. Ward (2007). How wages change: Micro evidence from the international wage flexibility project. *Journal of Economic Perspectives* 21, 195–214.
- Du Caju, P., T. Kosma, M. Lawless, J. Messina, and T. Rõõm (2015). Why firms avoid cutting wages. *ILR Review* 68, 862–888.
- Dunn, L. F. (1996). Loss aversion and adaptation in the labor market: Empirical indifference functions and labor supply. *The Review of Economics and Statistics* 441-450, 78.
- Eliaz, K. and R. Spiegler (2014). *Reference dependence and labor market fluctuations*, Volume 28 of *NBER Macroeconomics Annual*, 2013. Boston: National Bureau of Economic Research.
- Fagan, G. and J. Messina (2009). Downward nominal wage rigidity and optimal steady-state inflation. *Working Paper Series No 1048*, European Central Bank.

- Fahr, S. and F. Smets (2010). Downward wage rigidities and optimal monetary policy in a monetary union. *Scandinavian Journal of Economics* 112, 812–840.
- Falk, A., E. Fehr, and C. Zehnder (2006). Fairness perceptions and reservation wages - the behavioral effects of minimum wage laws. *The Quarterly Journal of Economics* 121, 1347–1381.
- Fehr, E. and L. Goette (2005). Robustness and real consequences of nominal wage rigidity. *Journal of Monetary Economics* 52, 779–804.
- Fehr, E. and L. Goette (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review* 97, 298–317.
- Fiorito, R. and G. Zanella (2012). The anatomy of the aggregate labor supply elasticity. *Review of Economic Dynamics* 15, 171–187.
- Fischer, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy* 85, 191–205.
- Fishburn, P. C. and G. A. Kochenberger (1979). Two-piece von Neumann Morgenstern utility functions. *Decision Sciences* 10, 503–518.
- Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.
- Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review* 102(1), 469–503.
- Goette, L., D. Huffman, and E. Fehr (2004). Loss aversion and labor supply. *Journal of the European Economic Association* 2, 216–228.
- Goodfriend, M. and R. G. King (1997). The new neoclassical synthesis and the role of monetary policy. In B. S. Bernanke and J. J. Rotemberg (Eds.), *National Bureau of Economic Research Macroeconomics Annual 1997*. Cambridge, MA: MIT Press.
- Gordon, D. F. (1976). A neo-classical theory of keynesian unemployment. *Carnegie-Rochester Conference Series on Public Policy* 1, 65–97.

- Gottschalk, P. (2005). Downward nominal-wage flexibility: Real or measurement error? *The Review of Economics and Statistics* 87, 556–568.
- Grund, C. and D. Sliwka (2007). Reference-dependent preferences and the impact of wage increases on job satisfaction: Theory and evidence. *Journal of Institutional and Theoretical Economics* 163, 313–335.
- Heidhues, P. and B. Kőszegi (2005). The impact of consumer loss aversion on pricing. *CEPR Discussion Papers No. 4849*, Centre for Economic Policy Research.
- Heidhues, P. and B. Kőszegi (2008). Competition and price variation when consumers are loss averse. *American Economic Review* 98(4), 1245–1268.
- Heidhues, P. and B. Kőszegi (2014). Regular prices and sales. *Theoretical Economics* 9(1), 217–251.
- Imai, S. and M. P. Keane (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review* 45, 601–641.
- Iregui, A. M., L. A. M. Melo Becerra, and G. M. T. Ramírez. (2009). Are wages rigid in colombia?: Empirical evidence based on a sample of wages at the firm level. *Borradores de Economía No 571i*, Banco de la Republica de Colombia.
- Jaques, E. (1956). *Measurement of responsibility. A study of work, Payment, and Individual Capacity*. London: Tavistock Publications.
- Jaques, E. (1961). *Equitable payment: A general theory of work, differential payment, and individual progress*. London: Heinemann.
- Kahn, S. (1997). Evidence of nominal wage stickiness from microdata. *American Economic Review* 87, 993–1008.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review* 76, 728–741.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5, 193–206.

- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–291.
- Kandil, M. (1995). Asymmetric nominal flexibility and economic fluctuations. *Southern Economic Journal* 61(3), 674–695.
- Kandil, M. (1998). Supply-side asymmetry and the non-neutrality of demand fluctuations. *Journal of Macroeconomics* 20(4), 785–809.
- Kandil, M. (1999). The asymmetric stabilizing effects of price flexibility: Historical evidence and implications. *Applied Economics* 31(7), 825–839.
- Kandil, M. (2001). Asymmetry in the effects of US government spending shocks: Evidence and implications. *The Quarterly Review of Economics and Finance* 41(2), 137–165.
- Kandil, M. (2006). Asymmetric effects of aggregate demand shocks across U.S. industries: Evidence and implications. *Eastern Economic Journal* 32(2), 259–283.
- Kandil, M. (2010). The asymmetric effects of demand shocks: international evidence on determinants and implications. *Applied Economics* 42(17), 2127–2145.
- Karle, H., G. Kirchsteiger, and M. Peitz (2015). Loss aversion and consumption choice: Theory and experimental evidence. *American Economic Journal: Microeconomics* 7(2), 101–120.
- Karras, G. (1996). Why are the effects of money-supply shocks asymmetric? convex aggregate supply or “pushing on a string”? *Journal of Macroeconomics* 18(4), 605–619.
- Karras, G. and H. H. Stokes (1999). On the asymmetric effects of money-supply shocks: International evidence from a panel of OECD countries. *Applied Economics* 31(2), 227–235.
- Kawaguchi, D. and F. Ohtake (2007). Testing the morale theory of nominal wage rigidity. *Industrial and Labor Relations Review* 61, 59–74.
- Khan, A., R. G. King, and A. L. Wolman (2003). Optimal monetary policy. *Review of Economic Studies* 70, 825–860.

- Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121(4), 1133–1165.
- Kőszegi, B. and M. Rabin (2007). Reference-dependent risk attitudes. *American Economic Review* 97(4), 1047–1073.
- Kőszegi, B. and M. Rabin (2009). Reference-dependent consumption plans. *American Economic Review* 99(3), 909–936.
- Kim, J. and F. J. Ruge-Murcia (2009). How much inflation is necessary to grease the wheels? *Journal of Monetary Economics* 56, 365–377.
- Kim, J. and F. J. Ruge-Murcia (2011). Monetary policy when wages are downwardly rigid: Friedman meets Tobin. *Journal of Economic Dynamics and Control* 35, 2064–2077.
- Kimura, T. and K. Ueda (2001). Downward nominal wage rigidity in Japan. *Journal of the Japanese and International Economies* 15, 50–67.
- Knetsch, J. L. and W. K. Wong (2009). The endowment effect and the reference state: Evidence and manipulations. *Journal of Economic Behavior & Organization* 71(2), 407–413.
- Koch, C. (2015). Can reference-dependent fairness concerns cause wage rigidity? Experimental evidence. *mimeo*.
- Kube, S., M. A. Maréchal, and C. Puppe (2013). Do wage cuts damage work morale? Evidence from a natural field experiment. *Journal of the European Economic Association* 11, 853–870.
- Lebow, D. E., R. E. Saks, and B. A. Wilson (2003). Downward nominal wage rigidity: Evidence from the employment cost index. *The B.E. Journal of Macroeconomics* 3, 1–30.
- Lenz, C. (1997). Asymmetric effects of monetary policy in Switzerland. *Swiss Journal of Economics and Statistics* 133, 441–454.
- Lindbeck, A. and D. J. Snower (1988). *The insider-outsider theory of employment and unemployment*. Cambridge, MA: MIT Press.

- Marzilli Ericson, K. M. and A. Fuster (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *The Quarterly Journal of Economics* 126(4), 1879–1907.
- Mas, A. (2006). Pay, reference pay and police performance. *The Quarterly Journal of Economics* 121(3), 783–821.
- McDonald, I. M. and H. Sibly (2001). How monetary policy can have permanent real effects with only temporary nominal rigidity. *Scottish Journal of Political Economy* 48(5), 532–546.
- McLaughlin, K. J. (1994). Rigid wages? *Journal of Monetary Economics* 34, 383–414.
- Mehrara, M. and A. R. Karsalari (2011). Asymmetric effects of monetary shocks on economic activities: The case of iran. *Journal of Money, Investment and Banking* 20, 62–74.
- Nickell, S. and G. Quintini (2003). Nominal wage rigidity and the rate of inflation. *Economic Journal* 113, 762–781.
- Pennings, J. M. E. and A. Smidts (2003). The shape of utility functions and organizational behavior. *Management Science* 49, 1251–1263.
- Pope, D. G. and M. E. Schweitzer (2011). Is tiger woods loss averse? persistent bias in the face of experience, competition, and high stakes. *American Economic Review* 101(1), 129–157.
- Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008). Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review* 98(1), 38–71.
- Prescott, E. C. (2004). Why do americans work so much more than europeans? *Quarterly Review, Federal Reserve Bank of Minneapolis* 28, 2–13.
- Ravn, M. O. and M. Sola (2004). Asymmetric effects of monetary policy in the United States. *Federal Reserve Bank of St. Louis Review* 86(5), 41–60.
- Schmidt, U. and S. Traub (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty* 25, 233–249.

- Shapiro, C. and J. E. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review* 74, 433–444.
- Sigurdsson, J. and R. Sigurdardottir (2011). Evidence of nominal wage rigidity and wage setting from Icelandic microdata. *Economics WP55*, Department of Economics, Central bank of Iceland.
- Smith, J. C. (2000). Nominal wage rigidity in the United Kingdom. *The Economic Journal* 110, 176–195.
- Stiglitz, J. E. (1986). *Theories of wage rigidity*. Keynes' Economic Legacy: Contemporary Economic Theories. New York: Praeger Publishers.
- Tan, S. H., M. S. Habibullah, and A. Mohamed (2010). Asymmetric effects of monetary policy in ASEAN-4 economies. *International Research Journal of Finance and Economics* 44, 30–42.
- Taylor, J. B. (1979). Staggered wage setting in a macro model. *American Economic Review* 69(2), 108–113.
- Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.
- Varian, H. *A Bayesian approach to real estate assessment*. Studies in Bayesian Economics in Honour of L.J. Savage. Amsterdam: North-Holland.
- Weiss, A. W. (1980). Job queues and layoffs in labor markets with flexible wages. *Journal of Political Economy* 88, 526–538.
- Weiss, A. W. (1990). *Efficiency wages: Models of unemployment, layoffs, and wage dispersion*. Princeton: Princeton University Press.

Chapter 3

Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?*

Abstract

We study the forecasting performance of three alternative large data forecasting approaches. These three approaches handle the dimensionality problem evoked by a large dataset by aggregating its informational content, yet on different levels. We consider different factor models, a large Bayesian vector autoregression and model averaging techniques, where aggregation takes place before, during and after the estimation of the different forecasting models, respectively. We use a dataset for Germany that consists of 123 variables in quarterly frequency and find that overall the large Bayesian VAR and the Bayesian factor augmented VAR provide the most precise forecasts for a set of 11 core macroeconomic variables. Both considerably outperform the remaining large scale forecasting models in terms of joint forecasting accuracy as measured by the multivariate MSE. Further, we find that the performance of these two models is very robust to the exact specification of the forecasting model.

Keywords: Large Bayesian VAR, Model Averaging, Factor Models, Great Recession.

JEL-Codes: C53, C55, E31, E32, E37, E47.

*This chapter is based on my joint paper with Maik H. Wolters entitled *Forecasting with Large Datasets: Aggregating Before, During or After the Estimation?*, Kiel Working Paper No. 1925, 2014 (updated in November 2015).

3.1 Introduction

While forecasters may wish to use as much information as possible to increase the accuracy of their forecasts, the estimation of models with a large number of different time series causes huge technical difficulties as the number of parameters to be estimated quickly becomes very large and in-sample overfitting occurs or estimation becomes even infeasible. To overcome this *curse of dimensionality* several large scale time-series methods have been proposed. The three most prominent of these approaches are factor models, large Bayesian vector autoregressions and model averaging techniques. All of these three approaches handle the dimensionality problem by aggregating the informational content of the large dataset, yet the aggregation takes place on different levels.¹

In particular, with factor models (see e.g. Stock and Watson, 2002a,b; Bernanke and Boivin, 2003; Forni et al., 2000, 2005) the aggregation of the informational content of a large dataset into a small number of static or dynamic factors takes place *prior* to the estimation of small scale forecasting models such as e.g. autoregressive distributed lag models, vector autoregressions or Bayesian vector autoregressions. These small forecasting models then include the factor time series rather than all the time series of the large dataset (see Banerjee, 2013, for an overview of the factor model approach and a survey on recent papers using factor models for large dataset problems).

Large Bayesian vector autoregressions (De Mol et al., 2008; Bańbura et al., 2010), on the other hand, can be estimated with a large number of time series by applying shrinkage to aggregate the information contained in the large dataset *during* the estimation process. The degree of shrinkage thereby increases with the number of times series included in the respective model.

By contrast, when using model averaging techniques (see e.g. Bates and Granger, 1969; Stock and Watson, 2003; Timmermann, 2006; Wright, 2009; Faust and Wright, 2009) the aggregation of the informational content of the large dataset takes place *after* the estimation of a large number of small scale forecasting models. Here, the final forecast is computed as a weighted average over the individual forecasts of all the small scale forecasting models.

¹An alternative approach to solving this problem are variable selection methods such as targeted predictors (Bai and Ng, 2007), Bayesian variable selection (Korobilis, 2013) or the LASSO approach (Tibshirani, 1996).

While De Mol et al. (2008) show that there is a theoretical connection between the factor and the shrinkage approach, it is not clear which method to aggregate the informational content of a large dataset performs best from an empirical perspective. In this paper, we therefore systematically study the performance of all three of these alternative large scale approaches using a dataset for Germany that consists of 123 variables in quarterly frequency.

Previous literature has so far only focused on evaluating the forecasting performance of one or two of these large scale approaches relative to several small benchmark models, to the Federal Reserve's Greenbook projections (for US data) or to each other. For example, Bernanke and Boivin (2003) study the performance of a factor augmented autoregressions and vector autoregression relative to the Greenbook forecasts, Faust and Wright (2009) evaluate static factor models as well as model averaging models relative to a number of benchmark models and the Greenbook projections, Bańbura et al. (2010) study a large Bayesian vector autoregression and a Bayesian factor augmented vector autoregression and Berg and Henzel (2013) focus on the same models, but study euro area instead of US data and additionally evaluate the different models' density forecasts.²

By contrast, our analysis includes all three large scale approaches outlined above. It focuses on Germany, the largest economy in the euro area, which is considerably smaller but also much more open than the US or the euro area. For Germany, several authors have investigated the forecasting performance of factor models estimated on large datasets relative to small benchmark models (see e.g. Schumacher and Dreger, 2004; Kholodilin and Siliverstovs, 2006; Schumacher, 2007, 2010, 2011). However, so far no comparison of the forecasting performance of factor models to alternative large scale approaches has been provided. Moreover, most existing empirical forecasting applications for Germany focus almost exclusively on forecasting real GDP (see e.g. Drechsel and Scheufele, 2012a,b) rather than a set of key macroeconomic variables as is the case in this study. To our knowledge, the only two exceptions to this are Müller-Dröge et al. (2014) and Buchen and Wohlrabe (2014) who evaluate the forecasts for a larger set of German key macroeco-

²Beyond pure reduced form forecasting models, Wolters (2014) compares the forecasting accuracy of a large Bayesian vector autoregression to Dynamic Stochastic General Equilibrium (DSGE) models and the Fed's Greenbook projections.

conomic variables as well. However, both papers have a different methodological focus than this paper.

With our comprehensive analysis we provide an assessment of the relative joint and also univariate forecasting performance of the different large scale forecasting methods for GDP growth, CPI and PPI inflation, a short- and a long-term interest rate, the unemployment rate, industrial production, real wages, consumption, investment and the current account balance. We deem these 11 variables of special interest to forecasters and policy makers because they are covered, for example, in the monthly survey of *Consensus Economics* among professional forecasters. Moreover, we test whether the forecasts obtained with the different models are unbiased and check whether the relative performance of the different forecasting models is robust against various alternative model specifications.

Our dataset consists of 123 variables in quarterly frequency covering a sample period from 1978 until 2013. We include indicators from the following categories: composition of GDP and gross value added by sectors, prices, labor market, financial market, industry, construction and surveys. Different variants of the three large scale forecasting models as well as a number of small benchmark models are estimated using a moving window of 15 years of data, while the forecasts obtained by the different models are evaluated from 1994 through 2013. To assess the relative (joint) forecasting performance of the different models we compare (multivariate) mean squared forecast errors, while we compute Mincer-Zarnowitz regressions (see Mincer and Zarnowitz, 1969) to test for forecast bias. All forecasting models are specified according to various information criteria. As a robustness check we also specify the models based on their ex post best forecasting performance and implement forecast pooling over a variety of specifications.

Our results indicate that the large Bayesian vector autoregression and the Bayesian factor augmented vector autoregression deliver forecasts that are more precise than those obtained by a univariate autoregressive benchmark or the remaining large scale forecasting models. This holds for both, measures for the joint forecasting performance for the set of 11 variables as well as univariate performance measures for the individual series. We find that in contrast to the remaining factor approaches and the model averaging approaches both models can efficiently exploit the correlation structure between

the series of the large dataset to provide relatively accurate forecasts, even for longer forecasting horizons.

With respect to the robustness of the relative forecasting performance of the different models our findings indicate that the forecasting performance of the large Bayesian vector autoregression and the Bayesian factor augmented vector autoregression is very robust to the specific model specification, i.e. the number of lags or factors and the degree of shrinkage. By contrast, the dynamic factor model outperforms all other forecasting models by far if one chooses the ex-post optimal specification. However, in the quasi real-time exercise, where the number of lags and factors is chosen based on information criteria, on past forecasting accuracy or where the forecasts are obtained by pooling over a large set models with different specifications, we find that this performance is unattainable.

Finally, our results indicate that overall the gains in forecasting accuracy obtained by the large scale approaches relative to an autoregressive benchmark are only modest for most variables considered and are in many cases statistically insignificant. We also find that using a large amount of data would not have helped in forecasting the great slump of German GDP growth in 2008 and that a small forecasting model that only includes the ifo business climate index, which is often cited by professional forecasters as *the* single most important predictor for German GDP growth,³ clearly dominates even the best large scale approaches in terms of short-term GDP growth forecasting performance.

The moderate gains of the large scale approaches can be explained with the extremely low persistence of some of the time series. Moreover, many of the time series seem to be characterized by common components which implies that parsimonious univariate models are often sufficient to capture the most important information contained in the data. Efficient multivariate modelling therefore becomes a hard task so that improvements of the large data forecasting methods are rather small (see also Carriero et al., 2011; Bernardini and Cubadda, 2014).

³The ifo business climate index is based on a monthly survey among about 7000 firms which report their assessments of the current business situation and their expectations for the next six months. From these two assessments the overall ifo index is calculated. The out-of-sample predictive ability of the ifo index for German GDP has been widely studied, see for example Dreger and Schumacher (2005), Kholodilin and Siliverstovs (2006), Abberger (2007), Drechsel and Scheufele (2012b) or Henzel and Rast (2013).

Still, when forecasters are interested in simultaneously predicting a larger number of variables, large-scale forecasting models have the advantage that they can be used to coherently forecast many variables at the same time. This might be an advantage when it comes to the interpretation of forecasts.

The remainder of this paper is structured as follows. In section 3.2 we outline the different forecasting models. In section 3.3 we describe the dataset that we use, while in section 3.4 we describe our forecasting approach. In section 3.5 we evaluate the absolute and relative (joint) forecasting performance of the different models and check for robustness against model misspecification. Finally, in section 3.6 we conclude.

3.2 Forecasting Models

In the following, we provide a brief overview of the different forecasting models.⁴ Let $\{y_{i,t}\}_{i=1}^n$ denote the set of variables to be forecast in log-levels and $\{x_{j,t}\}_{j=1}^m$ the set of possible predictors in log-levels. Variables expressed in rates such as the unemployment rate or interest rates are included in $\{y_{i,t}\}_{i=1}^n$ and $\{x_{j,t}\}_{j=1}^m$ in levels rather than log-levels. The total number of variables in our dataset is given by $n + m = k$. We compute annualized quarter-on-quarter growth rates of all variables, denoted by $\{\Delta y_{i,t}\}_{i=1}^n$ and $\{\Delta x_{j,t}\}_{j=1}^m$, respectively. To avoid overly complicated notation, variables expressed in rates are included in levels in the respective Δ terms as well. Given the information available at time t , we estimate all forecasting models and construct forecasts $\{\Delta y_{i,t+h}\}_{i=1}^n$ with h being the forecast horizon ranging from one to eight quarters ahead. While some of the forecasting models directly yield growth rate forecasts, we obtain log-level forecasts from the other models and use these to compute implied quarter-on-quarter growth rate forecasts. For forecasting models that include lags of the dependent variable the number of lags p included in the estimation of each model is obtained via the Bayesian information criterion unless otherwise stated.

⁴For a more detailed description of the different forecasting models we refer the reader to the working paper version Pirschel and Wolters (2014).

3.2.1 Large Bayesian VAR (LBVAR)

Consider the following VAR $Z_t = c + A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \epsilon_t$, where the vector $Z_t = (y_{1,t}, \dots, y_{n,t}, x_{1,t}, \dots, x_{m,t})'$ contains all the k time series in the dataset. Following Bańbura et al. (2010) we include the variables in log-levels rather than growth rates to not lose information that might possibly be contained in the trends. c is a $(k \times 1)$ vector of constants, A_1, \dots, A_p are $(k \times k)$ -dimensional parameter matrices and ϵ_t is a $(k \times 1)$ vector of independently identically distributed white noise error terms with zero mean and covariance matrix Ψ .

We use Bayesian techniques to estimate the large VAR outlined above. Since the number of variables that we want to include in the estimation is fairly large ($k = 123$), we follow Bańbura et al. (2010) and implement a prior that shrinks the parameters of the VAR. This allows for the aggregation of the information contained in the large dataset *during* the estimation process. The degree of shrinkage thereby increases with the size of the cross-section, thus allowing the estimation of a model where the number of parameters exceeds the number of observations by far.

We implement the Bayesian shrinkage approach by using a version of the Normal inverse Wishart prior (see e.g. Kadiyala and Karlsson, 1997) that retains the main principles of the widely used Minnesota prior (Litterman, 1986). According to this prior specification each equation of the VAR is centered around a random walk with drift or an autoregressive process, respectively. In contrast to Bańbura et al. (2010), we do not set δ_i , the prior coefficient means for the first lag of each variable, equal to zero for stationary variables. Instead, we run a univariate autoregression of order p for each of the k elements in $Z_{i,t}$ and set δ_i equal to the sum of the therewith obtained coefficient estimates defined as $\mu_i = \sum_{\ell=1}^p \beta_\ell$ if $\mu_i < 1$. For $\mu_i \geq 1$ we set $\delta_i = 1$. This approach allows us to capture the different degrees of persistence in the dataset.

The shrinkage of the VAR coefficients towards the prior is achieved through the hyperparameter λ which enters the prior variance of each coefficient. Bańbura et al. (2010) suggest to set the tightness of the prior, so that the LBVAR achieves the same in-sample fit as an unrestricted small VAR without shrinkage.⁵ We slightly depart from this approach and set λ such that the LB-

⁵Of course, this approach is merely an ad-hoc rule of thumb. Alternatively, λ could also be chosen to maximize the out-of sample forecasting performance over a pre-sample as for

VAR achieves the same in-sample fit as a small BVAR containing GDP, prices, the unemployment rate and a short-term interest rate, because the respective unrestricted VAR seems to be severely overparameterized.

We set the lag length $p = 4$, however the forecasting performance of the LBVAR proves to be remarkably robust with respect to the number of lags included in the estimation (see section 3.5.3). Following Bańbura et al. (2010) we implement the prior using dummy variables and augment it to constrain the sum of coefficients of the VAR (see e.g. Sims and Zha, 1998).

3.2.2 Factor Models (FAAR, FAVAR, BFAVAR, DF)

Assume that $\Delta X_{i,t}^*$, the standardized set of potential predictors for each variable of interest, can be represented by two components which are mutually orthogonal to each other and unobservable. These are the common component $\chi_{i,t}$ and the idiosyncratic component $\zeta_{i,t}$, so that we have $\Delta X_{i,t}^* = \chi_{i,t} + \zeta_{i,t}$.

The basic idea of factor models is that the information contained in the common component $\chi_{i,t}$ can be aggregated into a vector of factors $F_{i,t}$ of dimension $\kappa \leq (k - 1)$ which are able to explain most of the variance of the predictor matrix $\Delta X_{i,t}^*$. With these factors the dimension of a large dataset can thus be reduced *prior* to the estimation of the forecasting model.

In general the common component relates to the factors as $\chi_{i,t} = \sum_{l=0}^s \eta_l F_{i,t-l}$. Depending on the lag structure that is assumed we can distinguish two model variants: the static factor model with $s = 0$ and the dynamic factor model with $s > 0$.

Static Factor Models (FAAR, FAVAR, BFAVAR)

From the standardized set of predictors $\Delta X_{i,t}^*$ we extract the $(r \times 1)$ -dimensional vector of factors $F_{i,t} = (f_{i,t}^1, \dots, f_{i,t}^r)'$ via static principal component analysis.

example in Litterman (1986). Giannone et al. (2012) suggest a more sophisticated hierarchical approach to specifying λ which relies on maximizing the marginal likelihood, i.e. the density of the data conditional on λ after integrating out the uncertainty about the parameters of the VAR. However, since we find that the forecasting performance of the large BVAR is very robust to the exact specification of λ , we do stick to the rule of thumb.

Following Stock and Watson (2002a) we use these static factors to estimate a factor augmented direct autoregression (FAAR).⁶

Moreover, we implement a factor augmented vector autoregression as proposed by Bernanke et al. (2005) which allows for a more dynamic structure. Following Faust and Wright (2009) we include the variable to be predicted in log-levels and the factors extracted from the set of predictors in the estimation.⁷

We estimate the factor augmented vector autoregression via ordinary least squares (FAVAR) as well as with Bayesian techniques (BFAVAR), however the FAVAR performs very poorly so we do not report the results for this model. For the BFAVAR, the prior is set in a manner analogous to the large Bayesian VAR with the following two exceptions. First, we set the prior coefficient mean for the first lag of the factors $\delta = 0$ to account for the fact that the factors have been extracted from the standardized predictor matrix $\Delta X_{i,t}^*$. Second, we set the hyperparameter $\lambda = 0.2$. For the determination of the optimal number of factors r we use the information criterion IC_{p2} proposed by Bai and Ng (2002).

Dynamic Factor Models (DF)

We set up a dynamic factor model in the spirit of Forni et al. (2003, 2005). This implies extracting the $(q \times 1)$ -dimensional vector of dynamic factors $\tilde{F}_{i,t}$ from the standardized set of predictors $\Delta X_{i,t}^*$ via dynamic principal component analysis in the frequency domain. Defining $\tilde{F}_{i,t}^* = (\tilde{F}_{i,t}', \tilde{F}_{i,t-1}', \dots, \tilde{F}_{i,t-s}')'$ as a vector of contemporaneous and lagged factors with dimension $r = q(s + 1)$, the dynamic factor model can be rewritten as a static factor model $\chi_{i,t} = \eta \tilde{F}_{i,t}^*$.

The factors $\tilde{F}_{i,t}^*$ are used to augment a direct autoregression, analogously to the FAAR outlined above. For the determination of the optimal number of dynamic factors q we apply the information criterion proposed by Bai and Ng (2007).

⁶According to common practice, we chose the direct version of the autoregressive model because the iterated model variant would require the specification of a subsidiary model for the factors $F_{j,t}$ in order to compute forecasts for horizons $h > 1$.

⁷We also estimate a FAVAR that includes a small set of core variables (including the variable to be predicted) and the factors (see e.g. Bernanke and Boivin, 2003; Bańbura et al., 2010). The forecasting performance of this alternative, however, is considerably worse, so that we do not include this model in the main results.

3.2.3 Model Averaging (EWA, BMA)

For each of the n variables of interest $\Delta y_{i,t}$ we set up $(k - 1)$ direct autoregressive distributed lag models $\Delta y_{i,t} = \rho_0 + \rho_1 \Delta y_{i,t-h} + \dots + \rho_p \Delta y_{i,t-h+1-p} + \beta_j \Delta x_{j,t-h} + \epsilon_{j,t}$, where $\Delta x_{j,t-h}$ is an element of the $(k - 1)$ -dimensional set of potential predictors $\Delta X_{i,t}$.

The general idea of model averaging is to compute a forecast $\Delta y_{i,t+h}^j$ with each of the $(k - 1)$ models and aggregate the model-specific forecasts *afterwards* into one final forecast, i.e. $\Delta y_{i,t+h} = \sum_{j=1}^{(k-1)} \omega_j \Delta y_{i,t+h}^j$, where ω_j denotes the weight given to the model-specific forecast $\Delta y_{i,t+h}^j$.

According to the specification of ω_j we distinguish two model averaging approaches. The first approach is Equal Weighted Averaging (EWA) as in Stock and Watson (2003, 2004), where the $(k - 1)$ models are estimated via OLS and $\omega_j = \omega = \frac{1}{(k-1)}$.

Alternatively, we consider Bayesian Model Averaging (BMA) as laid out in Wright (2009), where each of the model-specific forecasts $\Delta y_{i,t+h}^j$ is weighted with the posterior probability of the respective model $P(M_j)$, i.e. $\omega_j = P(M_j)$.⁸

3.2.4 Benchmark Model (AR)

In order to evaluate the relative forecasting performance of the three large scale approaches described above we implement a univariate autoregression (AR) $\Delta y_{i,t} = c + \sum_{j=1}^p \rho_j \Delta y_{i,t-j} + \epsilon_t$ for each of the variables to be forecast as benchmark.

3.3 Data

Our dataset builds on the one used in Schumacher (2007) which we have slightly modified and updated to cover a sample from 1978Q1 to 2013Q3. Overall, our dataset consists of 123 macroeconomic variables in quarterly frequency. Series that are available at a higher frequency, e.g. monthly, are converted into quarterly frequency by computing the average over the respective quarter. The data can be grouped into the following categories: composition

⁸The model-specific posterior probability $P(M_j)$ is calculated in each estimation period t for each forecasting horizon h . For simplicity however, we omit the respective subscripts.

of GDP and gross value added by sectors, prices, labor market, financial market, industry, construction, surveys and miscellaneous. A detailed list of the different series can be found in Appendix A.

Most of the data is obtained via Thomson Reuters Datastream, while the remaining data is directly obtained from the German Federal Statistical Office. We do not account for data revisions in our quasi real-time forecasting exercise, but use the most recent vintage of the data available in December 2013. The data is seasonally adjusted. Natural logarithms are taken and annualized quarter-on-quarter growth rates are computed for time series not expressed in rates. Following Schumacher (2007) we rescale data which is only available for West Germany prior to 1991 to the pan-German series to avoid regime shifts.

3.4 Forecasting Approach

We estimate the various forecasting models on a moving window consisting of 60 observations to account for possible structural breaks in the estimation sample. For the majority of forecasting models, the forecasts are computed by iterating the forecasting models forward, while for the FAAR, DFM and the two model averaging approaches direct forecasts are computed.

The evaluation sample for our pseudo out-of-sample forecasting exercise, denoted by $T = T_0 + 1, \dots, T_1$, ranges from 1994Q4 until 2013Q3, thus it contains 76 forecasts for each horizon. Forecast errors are computed as $e_{i,T|T-h} = \Delta y_{i,T}^r - \Delta y_{i,T|T-h}^f$, where $\Delta y_{i,T}^r$ denotes the realized quarter-on-quarter growth rate of variable i in period T and $\Delta y_{i,T|T-h}^f$ denotes the quarter-on-quarter period T growth rate forecast of variable i computed h quarters earlier.

For the evaluation of the absolute and relative forecasting performance of the different models we focus on two measures. First, we run Mincer-Zarnowitz regressions (see Mincer and Zarnowitz, 1969) $\Delta y_{i,T}^r = \alpha_{i,h} + \beta_{i,h} \Delta y_{i,T|T-h}^f + \epsilon_{i,T|T-h}$ and conduct F-tests of the joint null hypothesis $\hat{\alpha}_{i,h} = 0$ and $\hat{\beta}_{i,h} = 1$ to check whether the forecasts are unbiased and efficient. This allows us to assess the absolute forecasting accuracy of each model.

Secondly, we compute and analyze (multivariate) mean squared forecasting errors (MSE) to evaluate the relative (joint) predictive ability of the different forecasting models. We report the absolute MSE for the AR forecast

which we use as a benchmark, while for the remaining models we report the MSEs relative to this benchmark. Thus a relative MSE smaller than 1 indicates that the forecasting performance of a specific model is more precise than that of the AR benchmark and vice-versa.

To assess the statistical significance of the forecasting performance of the different models for each individual variable relative to the AR benchmark we implement the test of equal unconditional finite-sample predictive ability (see Giacomini and White, 2006) using a symmetric loss function. This test can be applied to nested models, meaning that one model can be obtained from another model by imposing certain parameter restrictions, as well as non-nested models. It thus provides a coherent framework for comparing a large number of different forecasting models as is the case in this paper. Asymptotic p -values are computed using Newey-West standard errors to account for serial correlation of the forecast errors.

Finally, the multivariate root mean squared forecast error, as proposed by Christoffersen and Diebold (1998), is computed as

$$multMSE_h = \frac{1}{T_1 - T_0 - 1} \sum_{T=T_0+1}^{T_1} e'_{T|T-h} W e_{T|T-h}, \quad (3.1)$$

where the $(1 \times n)$ -dimensional vector $e_{T|T-h}$ contains the forecast errors $e_{i,T|T-h}$ for all n variables of interest and W is an $(n \times n)$ -dimensional diagonal weighting matrix.

We follow Carriero et al. (2011) as well as Buchen and Wohlrabe (2014) and specify $W = W^D$ as diagonal matrix with entries being equal to the inverse of the variances of the variables to be forecast. Müller-Dröge et al. (2014) propose to specify $W = W^C$ as the inverse of the sample variance-covariance matrix and we consider this alternative as well. Both versions of the multivariate MSE aim at assessing the joint predictive ability of the different forecasting models, i.e. their suitability to simultaneously forecast a larger set of variables. In the first version the measure has the advantage that it accounts for the fact that variables with a large variance are generally harder to forecast by attributing them a smaller weight. In addition to that, the second version of the measure compensates for possible correlation of the different series.⁹

⁹The underlying idea is to account for the linear dependence between the different variables that might simultaneously drive their MSEs and thus inflate the measure of joint predictive ability. In principle, this is comparable to the approach of computing the variance of

As for the univariate MSE we report the absolute multivariate MSE for the AR model, while for all remaining models we compute the multivariate MSE relative to this benchmark.

3.5 Results

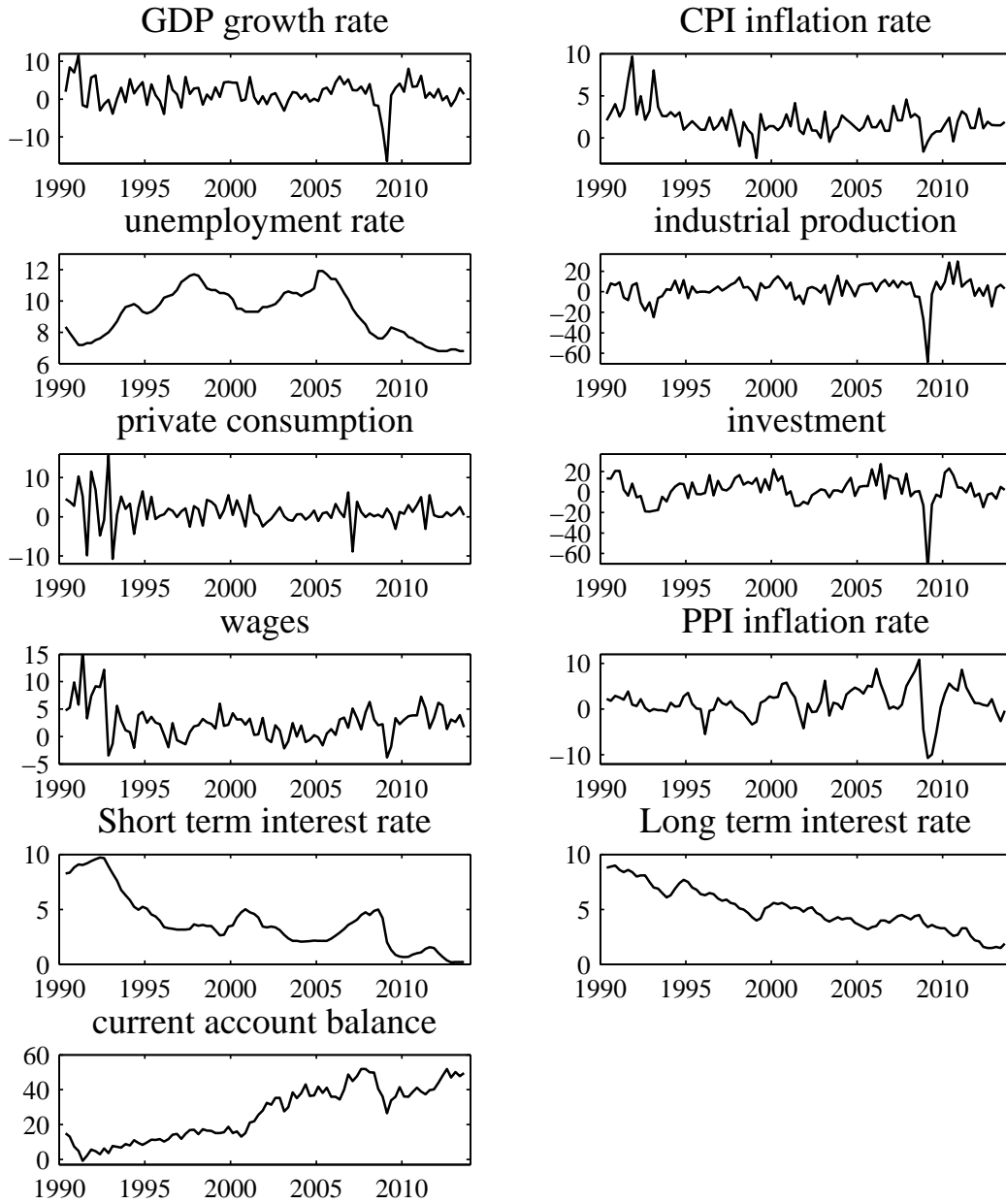
In this section we report the results of our forecasting exercise. We first focus on the joint forecasting performance of each large scale approach for our set of 11 German key variables. Afterwards, we extend the analysis to the performance for the individual variables, with an emphasis on GDP growth. Finally, we check the robustness of our results.

3.5.1 Forecasting 11 German Key Macroeconomic Variables Jointly

Figure 3.1 displays the 11 variables that we consider. It can be seen very clearly that there is considerable variation in the degree of persistence of the different variables. For example, German GDP growth shows extremely little persistence and can thus be expected to be very hard to predict. A comparison of the autocorrelation functions of US and German GDP growth for a sample covering 1978-2013 shows that there is significant autocorrelation of up to two lags for US GDP growth, while there is no significant autocorrelation at all for German GDP growth. The persistence of industrial production, investment and consumption growth is comparable to that of German GDP growth. Thus, we can expect the different forecasting models to have similar problems in predicting these variables. By contrast, CPI inflation is more persistent than GDP growth, but still shows many spikes, which will presumably be hard to predict as well. The persistence of the unemployment rate series is very high, similar to that of the short- and long-term interest rate and the current account balance. The German unemployment rate does not show a clear overall trend, but instead increases until 1998, decreases until 2001, increases again until 2005 and falls from there until the end of

the sum of several random variables where a correction term accounting for the covariance of the pairs of variables is needed as well. Note however, that since we use the inverse of the covariance matrix as correction, the multivariate MSE decreases for positive correlation and increases for negative correlation between the pairs of variables in the dataset.

Figure 3.1: German key macro variables.



Notes: The graph shows the 11 German key macroeconomic variables that we consider from 1990 until 2013. For all variables, except those expressed in rates, annualized quarter-on-quarter growth rates are shown. Data sources are listed in Appendix A and Appendix B contains an exact definition of the different variables.

the sample. Predicting these trend changes might pose another difficulty for most forecasting models.

In Table 3.1, we display both versions of the multivariate MSEs of the different large scale approaches relative to the AR benchmark for horizons $h = 1, 4, 8$ as well as the absolute multivariate MSE for the AR. Both measures indicate that the BFAVAR and to a slightly lesser extent the LBVAR provide the most accurate forecasts for all the variables over all forecasting horizons. For short horizons also the remaining large scale approaches are able to improve upon the AR benchmark, though to a different degree. While the FAAR and the DFM perform almost as good as the BFAVAR and the LBVAR, the model averaging techniques do worse. For $h = 8$, the relative performance of the large scale approaches deteriorates considerably. The BFAVAR is the only model that can clearly outperform the AR benchmark according to both multivariate measures considered here.

Table 3.1: Forecasting a Set of 11 German Key Variables.

(a) Multivariate MSE, $W = W^D$							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	8.03	0.84	0.85	0.80	0.87	0.95	0.94
4	11.16	0.86	1.02	0.82	1.00	0.89	0.91
8	13.52	0.90	1.35	0.86	1.28	1.09	1.10
(b) Multivariate MSE, $W = W^C$							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	11.61	0.79	0.83	0.76	0.82	0.91	0.92
4	15.81	0.86	1.01	0.79	1.06	0.87	0.96
8	18.42	1.00	1.51	0.86	1.46	1.01	1.12

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1994Q4 until 2013Q3, thus for each horizon a total of 76 forecasts is computed. The second column shows the absolute multivariate MSEs for the AR benchmark model, while all other MSEs are computed relative to this benchmark. The two measures differ with respect to the weighting matrix W which is a diagonal matrix with the inverse of the series variance as entries (W^D upper panel) and the inverse of the sample covariance matrix (W^C lower panel).

The entries in the lower part of Table 3.1 indicate to what extent the different forecasting models are able to account for possible correlation of the different series. The absolute multivariate MSE of the AR benchmark is considerably higher for the second version of the measure for all horizons indicating that there is negative correlation in the data that could be useful

for forecasting. Interestingly, the relative performance of the large scale approaches generally improves for short horizons and deteriorates for longer horizons when comparing the upper and the lower part of the table. This indicates that for short horizons the models are able to make use of the correlation structure in the dataset to provide better forecasts than the AR benchmark. For large horizons, however, this seems no longer to be the case. Especially the factor models, FAAR and DFM, display a very poor longer term joint forecasting performance. A notable exception is the BFAVAR which performs equally well under both versions of the multivariate MSE for all horizons. Apparently, the combination of aggregation information of the large dataset into factors and shrinkage enables the model to efficiently use all the information contained in the dataset, even for longer forecasting horizons.

Overall, our results indicate that for short forecast horizons it does not seem to make a very big difference for the joint forecasting performance of our large scale approaches, whether the information of the large dataset is aggregated before or during the estimation process of the forecasting models, as the factor models and the shrinkage approaches perform similarly well. Aggregation after the estimation process (model averaging approach), however, yields somewhat less precise short-horizon forecasts. For obtaining accurate forecasts for longer horizons using a shrinkage approach seems to be essential to extract the relevant information on the longer-run dynamics of the different variables as evidenced by the very good performance of the BFAVAR and, to a slightly lesser extent, the LBVAR for longer forecast horizons.

3.5.2 Forecasting Performance for the Individual Variables

In Table 3.2, panels (a) - (k), we display the univariate MSEs of the different models for the 11 key variables for horizons $h = 1, 4, 8$. Table entries in bold indicate that the null hypothesis of unbiasedness based on the F-test for the coefficients in the Mincer-Zarnowitz regression cannot be rejected at the 5% level. The symbols ●, •, ., indicate that the relative MSE is significantly different from one at the 1, 5, or 10% level, respectively.

As can be seen very clearly from the entries in Table 3.2, the BFAVAR and the LBVAR are the best performing models in most cases. However, the size of the gains in accuracy over the AR benchmark as well as the absolute fore-

casting performance of the different models apparently depend heavily on the specific variable and the respective forecasting horizon.

For *GDP growth* (panel (a) of Table 3.2), the absolute MSEs of the AR benchmark are quite large and flat over the different forecast horizons which is in line with what can be expected for forecasts of a time series with low persistence, (see Del Negro and Schorfheide, 2013, for a detailed exposition). Moreover, the entries in Table 3.2, panel (a) reveal that the gains in forecasting accuracy for German GDP growth obtained by the three large scale approaches are at best moderate and insignificant, while the differences in the relative MSEs between the various forecasting models are rather small. Among the three large scale approaches, the BFAVAR and (to a slightly lesser extent) the LBVAR yield the most accurate forecasts, though for $h = 8$ EWA performs best. In the short-run, the gains of the BFAVAR and the LBVAR over the AR benchmark amount to more than 10%. Yet, for longer horizons, there is almost no improvement upon the AR, which confirms the results of Schumacher (2007, 2010) for factor models and Kholodilin and Siliverstovs (2006) for various alternative leading indicators. Thus, adding more information by using a large dataset for the forecasting process of German GDP growth apparently only leads to marginal improvements in forecasting accuracy over the AR benchmark. The results of the Mincer-Zarnowitz regressions reveal that none of the forecasting models is able to provide unbiased forecasts for all forecasting horizons. With the exception of the LBVAR for $h = 1$, the AR and the BFAVAR for $h = 8$ and the EWA for $h = 1, 8$, for the remaining models, the estimated constant $\hat{\alpha}_{i,h}$ is larger than zero, but the estimate of the slope parameter and $\hat{\beta}_{i,h}$ is smaller than one (and in some cases even negative). This indicates that the forecasts systematically predict less variation than the GDP growth series actually shows.

For the German *CPI inflation rate* (panel (b) in Table 3.2), the absolute MSEs for the AR model are much smaller than those for GDP growth. Still the persistence of quarterly CPI inflation is quite low and thus the MSE does not increase much with the forecast horizon h . In terms of relative forecasting performance for the CPI inflation rate, only the BFAVAR significantly outperforms the AR benchmark over all forecasting horizons with gains in accuracy ranging between 15% to 20%. For $h = 1$ also the LBVAR and the two model averaging approaches significantly beat the AR benchmark. However, none of the forecasting models is able to yield unbiased forecasts. The esti-

Table 3.2: Univariate MSEs.

(a) GDP growth							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	12.09	0.88	0.95	0.86	0.95	0.98	1.03
4	12.64	1.00	1.27●	0.98	1.09	1.01	1.07
8	12.12	1.04	1.42●	0.98	1.22	0.94	1.10
(b) CPI Inflation Rate							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	1.94	0.78●	0.89	0.80●	0.92	0.93●	0.87●
4	1.86	0.95	1.18	0.85●	1.21	0.93	0.92
8	2.07	0.89	1.33	0.82●	1.33	0.92	0.93
(c) Unemployment Rate							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	0.06	0.98	0.98	0.94	0.99	0.95●	0.95●
4	0.76	0.85	1.14	0.81	1.27	1.06	1.07
8	2.17	0.95	1.18	0.83	1.37	1.67	1.79
(d) Industrial Production							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	173.25	0.71	0.64●	0.76●	0.72	0.95	0.97
4	224.63	0.66	0.59	0.64	0.57	0.66	0.63
8	140.27	1.02	1.06	1.00	0.97	1.01	1.07
(e) Private Consumption							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.06	0.86	1.04	0.91	1.05	0.99	0.97
4	7.37	0.84●	1.21	0.85●	1.20●	0.96	0.94
8	6.78	0.93	2.27●	0.95	2.02●	1.04	1.03
(f) Machinery and Equipment Investment							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	167.63	0.86	0.86	0.76	0.93	0.95	0.97
4	221.91	0.84	0.85	0.80	0.85	0.80	0.76
8	196.33	0.95	1.04	0.89	1.03	0.95	1.16
(g) Wages							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	5.99	0.76●	0.71	0.56●	0.64●	0.91●	0.83●
4	7.19	0.87●	1.16	0.84●	1.10	0.88●	0.94
8	8.75	0.85	1.39●	0.86●	1.33●	1.10	1.06
(h) PPI Inflation Rate							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	9.22	1.20	0.82	1.03	0.85	0.91●	0.96
4	15.79	1.02	0.98	0.96	1.04	0.94	0.96
8	14.13	1.05	1.21	1.02	1.02	1.10●	1.22●
(i) Short Term Interest Rate							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	0.14	0.83●	0.88	0.73●	0.92	0.91	0.94
4	1.64	0.77	0.94	0.72●	0.98	1.00	1.07
8	4.58	0.77	1.21	0.72●	1.10	1.18	0.97
(j) Long Term Interest Rate							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	0.10	0.90	1.12	0.96	1.11	0.94	0.68
4	0.84	0.73●	1.02	0.74●	1.01	0.93	0.98●
8	1.94	0.62●	1.49	0.63●	1.46	1.07	0.94
(k) Current Account							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	15.31	0.98	1.15	1.03	1.12	0.99	1.00
4	60.89	1.01	1.22	0.99	1.11	0.95	1.07
8	114.59	1.05	1.33	0.96	1.39	1.09	1.25

Notes: See notes on Table 3.1, first part. The symbols ●, ●, ●, indicate that the relative MSE is significantly different from one at the 1, 5, or 10% level, respectively, while bold numbers imply that the null hypothesis of unbiasedness cannot be rejected at the 5 % level.

mated constant in the Mincer-Zarnowitz regressions for all models is larger than zero, while the slope parameter is smaller than one (smaller than zero in most cases). This indicates that the forecasts are systematically larger than the actual data which may be attributed to the higher trend inflation in the first part of the estimation sample compared to the evaluation sample. While CPI inflation is more persistent than GDP growth, the informational content of the CPI forecasts obtained by all models is even smaller. The R^2 from Mincer-Zarnowitz regressions (not shown in the table) never exceeds 5%.

In contrast to that, for the German *unemployment rate* (panel (c) in Table 3.2) the explanatory power of all forecasts is extremely high, especially for short forecasting horizons. For $h = 1$ the R^2 from Mincer-Zarnowitz regressions (not shown in the table) for all forecasting models amounts to 95% or more, while for $h = 4$ it still ranges between 66% and 76%. This must certainly be attributed to the high persistence in the German unemployment rate series, which is also reflected in the small absolute MSE for the AR which increases with the forecasting horizon. Due to the various trend changes in the German unemployment rate series, no model systematically over- or underestimates the unemployment rate. With a few exceptions for $h = 8$, all forecasts are unbiased. However, except for the two model averaging approaches for $h = 1$, no model can significantly outperform the AR benchmark for the prediction of the German unemployment rate.

With some exceptions this also holds for *private consumption* and the *PPI inflation rate* (panel (e) and (h) in Table 3.2). However, while no model is able to yield unbiased forecast for consumption, the forecasts for PPI inflation are unbiased in many cases.

A similar result regarding the relative forecasting performance of the different models can be observed for the German *current account balance* (panel (k) in Table 3.2) and German *machinery and equipment investment* (panel (f) in Table 3.2). We find that for these variables none of the large scale approaches considered can significantly improve upon the AR benchmark which is surprising given the different degrees of persistence of the series.

By contrast, for the *short- and long-term interest rates* (panel (i) and (j) in Table 3.2) as well as for *wages* (panel (g) in Table 3.2) the best performing large scale model, the BFAVAR, almost always significantly outperforms the AR benchmark with sizeable gains in accuracy (45% for wages for $h = 1$).

For *industrial production*, there are big and significant gains in accuracy for the BFAVAR and the FAAR, but only for $h = 1$. For higher horizons the large scale approaches can not significantly improve upon the AR benchmark, though in most cases the forecasts are unbiased.

To sum up, we find that over all variables considered the best performing large scale approaches, namely the LBVAR and the BFAVAR, can clearly improve upon the AR benchmark, especially in the short run. However, the size of the gains in accuracy for each individual variable is highly heterogenous. Moreover, the BFAVAR, the LBVAR and the EWA approach provide generally more often unbiased forecasts compared to the AR benchmark, the two factor models and the BMA.

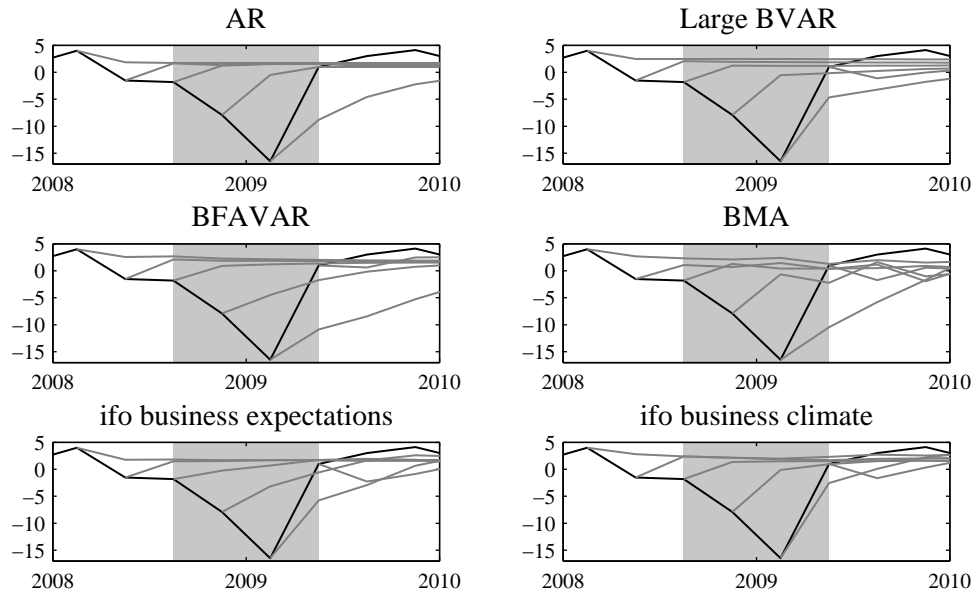
Forecasting the Great Recession. Several studies, for example Kuzin et al. (2013) and Timmermann and van Dijk (2013), indicate that the performance based ranking of different forecasting models may change considerably during the period of the Great Recession of 2008/2009. Therefore, in what follows, we take a closer look at whether the three large scale forecasting methods would have been able to forecast the slump of German GDP growth during the Great Recession.

In addition to the large scale approaches and the small benchmark models outlined in section 3.2, we analyze the predictive content of the ifo business climate index for German GDP growth during the Great Recession. As pointed out before, the ifo index is a leading indicator and often referenced to as *the* most important benchmark when forecasting German GDP growth (see also Dreger and Schumacher, 2005; Kholodilin and Siliverstovs, 2006; Abberger, 2007; Drechsel and Scheufele, 2012b; Henzel and Rast, 2013). We use the ifo business climate index and the subindex covering business expectations for the next six months and regress GDP growth on a constant and the respective lagged indicator as in Henzel and Rast (2013): $\Delta y_t = \alpha_h + \beta_h \text{ifo}_{t-h} + \epsilon_{t,h}$.

Figure 3.2 shows the forecasts of the annualized quarter-on-quarter GDP growth rate obtained by the AR, the LBVAR, the BFAVAR, the BMA and the two ifo indicators considered above computed for the subsample ranging from 2008Q1 to 2009Q2. Generally, the forecasts of all six models look roughly similar and none of them is able to predict the downturn in GDP growth in 2008. Once the recession hits, the models also fail to predict a fur-

ther deepening of the recession, but indicate a relatively quick recovery instead. The only notable exceptions are the one quarter ahead forecasts based

Figure 3.2: Great Recession GDP growth forecasts.



on the ifo expectation index and those obtained with the BFAVAR computed in 2008Q4. As business expectations in Germany already dropped largely in 2008Q3, the ifo expectation index predicts a negative GDP growth rate of -3.17% for 2009Q1.¹⁰ The BFAVAR GDP growth forecast is even slightly more pessimistic and amounts to -4.40%. Still, none of the forecasting models is able to predict the turning point of the Great Recession in 2009Q1. Moreover, once the turning point is reached, the models also considerably underpredict the speed of the recovery for the following quarters.

Given that the performance of the different forecasting models during the Great Recession was more or less equally disappointing, we would not expect that the Great Recession period strongly drives the results reported in Table 3.2, panel (a). Surprisingly, the entries in Table 3.3 which display the MSEs of the different forecasting models relative to the AR benchmark for the pre-Great Recession subsample ranging from 1994Q1 until 2007Q4 indicate that this is not the case. Especially for horizons $h = 4, 8$ the relative performance of the LBVAR and the BFAVAR improves considerably when the

¹⁰However, by construction this model can hardly predict a further deepening of the recession. Since the forecast is computed as $\Delta y_{t+h} = \hat{\alpha}_h + \hat{\beta}_h \text{ifo}_t$, the coefficient $\hat{\beta}_h$ would need to increase strongly with the forecasting horizon h to predict the further deepening of the recession.

Great Recession is excluded from the evaluation sample. In this case both models are able to significantly reduce the relative MSE by approximately 10%. Moreover, the null hypothesis of unbiasedness based on the F-test for the two coefficients in the Mincer-Zarnowitz regression can no longer be rejected for all forecasts obtained by the LBVAR and the BFAVAR.

Table 3.3: Forecasting German GDP growth, excluding the Great Recession of 2008/2009.

Univariate MSE for GDP growth (column 2: absolute; others: relative to AR)							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	6.64	0.86	0.97	0.91	0.96	0.91	0.95
4	7.29	0.87●	1.36●	0.88	1.10	0.92	1.03
8	7.07	0.90●	1.42	0.91●	1.54●	0.99	1.21

Notes: See notes on Table 3.2. In this table the evaluation sample has been adjusted to cover the period from 1994Q4 until 2007Q4.

The ifo business climate index, which performs slightly better than the expectations based index for $h = 1$ (the relative MSEs are equal to 0.84 and 0.94, respectively), yields the most accurate short-run predictions for German GDP growth when we exclude the Great Recession from the evaluation sample. However, the gains upon the best performing large scale approaches are only very small. For $h = 8$ the LBVAR and the BFAVAR clearly beat this important benchmark for the prediction of German GDP growth (the relative MSEs are equal to 1.00 for the ifo business climate and 0.98 for the ifo expectations index, respectively).

3.5.3 Robustness with Respect to Alternative Model Specifications

Next, we want to check the robustness of our results reported thus far with respect to alternative specifications of the different forecasting models. Therefore, we repeat the forecasting exercise of the previous section with an *optimized* specification of each model that is obtained by computing a variety of different specifications for each model and choosing the one that yields ex post the best forecasting performance. For an indication of how robust the model's forecasting performance is against various alternative model specifi-

cations, we then check which of the forecasting models yield similarly accurate forecasts with the information criteria based specification and with the ex post optimized specification.

Table 3.4 shows the ranges of the various parameters of the different forecasting models that we consider for this exercise. For example, for the FAAR

Table 3.4: Parameter range to determine ex post optimized specification of forecasting models.

parameter	range	forecasting model
number of lags p	1, 2, ..., 4	all forecasting models
degree of shrinkage λ	0.01, 0.02, ..., 0.1	LBVAR
degree of shrinkage ϕ	1, 1.1, ..., 2	BMA
number of static factors r	1, 2, ..., 10	FAAR, FAVAR, BFAVAR
number of dynamic factors q	1, 2, ..., 10	DFM
number of lags of the dynamic factors s	1, 2, ..., 4	DFM

the number of static factors r as well as the number of lags p have to be specified. After defining a range for each of these parameters, i.e. $r = 1, \dots, r_{max}$ and $p = 1, \dots, p_{max}$, we estimate the FAAR and compute forecasts for each possible combination of these two parameters. We then choose the specification with the ex post highest forecasting accuracy as the optimized specification for the FAAR model.

We follow Schumacher (2007) and distinguish the following two approaches: *performance based model selection, time-varying model* (PBTv) and *performance based model selection, constant model* (PBC). With PBTv we divide the evaluation sample into subsamples covering 4 quarters each. For each of these subsamples we select the specification for each forecasting model and for each forecasting horizon that minimizes the respective subsample MSE. By contrast, with PBC we choose the specification for each model that minimizes the MSE over the whole evaluation sample for each horizon.

In Table 3.5 we report the results of this exercise for horizons $h = 1, 4$ and 8. Specifically, we display the absolute multivariate MSEs (first version, $W = W^D$) for all 11 key variables for each of the different forecasting models obtained when specified according to the various information criteria (IC) as well as under PBC and PBTv.

Table 3.5: Absolute multivariate MSEs with IC, PBC and PBTV.

(a) IC: Information criteria based model selection (quasi real-time forecasting)							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	8.03	6.74	6.82	6.41	7.01	7.59	7.51
4	11.16	9.60	11.39	9.20	11.18	9.92	10.10
8	13.52	12.14	18.26	11.62	17.32	14.76	14.89
(b) PBC: Performance based model selection, constant model							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.39	6.06	6.16	5.86	6.03	7.01	6.74
4	9.71	9.12	9.70	8.83	9.61	9.86	9.64
8	11.96	11.22	13.61	11.08	13.86	14.19	13.63
(c) PBTV: Performance based model selection, time-varying model							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	6.80	5.15	4.19	4.59	2.53	6.50	6.12
4	8.85	7.15	6.99	7.59	4.48	8.78	8.54
8	10.83	8.39	9.13	9.75	5.78	12.82	11.90

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1994Q4 until 2013Q3, thus for each horizon a total of 76 forecasts is computed. The different specification strategies are described in the text.

Comparing the absolute MSEs reported in Table 3.5 panel (b) and (c) to those reported in panel (a) gives rise to the following observations. First, both ex post performance based model selection approaches generally increase the precision of all forecasting models—which of course is not surprising given that these approaches rely on out-of-sample information. However, while overall PBC leads only to modest gains over the quasi real-time forecasts, the gains obtained with PBTV are very large. This indicates that the optimal specification of the various forecasting models changes over time.

Regarding the relative performance of the different forecasting models with PBC the entries in panel (b) indicate that the LBVAR and the BFAVAR again provide the most accurate forecasts for most horizons. However, especially for higher horizons the gains in accuracy of these two models over the AR benchmark are less pronounced with PBC (6% and 7% respectively for $h = 8$) than with IC (10% and 14% respectively for $h = 8$). By contrast, as the entries in panel (c) reveal with PBTV the best performing large scale methods can improve considerably upon the AR benchmark for all forecasting horizons. The DFM, which now clearly outperforms all remaining models by far, achieves a reduction in the absolute multivariate MSE upon the AR bench-

mark amounting to 60% for $h = 1$ and roughly 50% for higher horizons. The LBVAR and the BFAVAR, which rank lower than the FAAR with PBTv, outperform the AR by 25% and 32%, respectively, for $h = 1$ and 22% and 10%, respectively, for $h = 8$.

Overall the results documented in Table 3.5 allow us to divide the different models into three groups according to the degree of robustness of their forecasting performance against alternative model specifications. First, for the AR benchmark, both model averaging techniques, EWA and BMA, the BFAVAR and (to a slightly lesser extent) the LBVAR we find that the specific model specification does not have a large impact on the models' forecasting performance. For the LBVAR we find that with PBTv the degree of shrinkage λ varies strongly over time, while in our quasi real-time specification λ is very stable over time. This also applies to the number of lags p included in the estimation. However, the optimally specified model reduces the respective MSEs only very little. This finding is in line with Andrea Carriero and Marcellino (2011) who document the robustness of the LBVAR's forecasting performance against the specific choice of λ and p .

Second, according to our results the forecasting performance of the FAAR depends to a moderate degree on the precise model specification. The FAAR's forecasting performance would improve moderately if one could optimally specify the model in real-time. We find that in the PBTv specification the optimal number of static factors r for the FAAR varies largely over time, while the number of factors chosen via the information criterion of Bai and Ng (2002) in the quasi real-time exercise is rather stable (see also Schumacher (2007), p. 288).

Third, we show that the accuracy of the quasi real-time forecasts of the DFM depends to a very large degree on the specific model specification. Choosing the optimal specification and allowing for time heterogeneity rather than specifying the model based on information criteria leads to a considerable improvement in the model's forecasting performance. This confirms the findings of Schumacher (2007) who conducts the same analysis for static and dynamic factor models. One reason for the low degree of robustness of the forecasting performance of the DFM to alternative model specifications is certainly that the optimal number of dynamic factors q seems to vary substantially for different forecasting horizons. Further, it turns out that the number of dynamic factors q chosen according to the Bai and Ng (2007) in-

formation criterion is always considerably smaller than the ex post optimal number of dynamic factors.

Real-Time Performance Based Model Specification and Forecast Pooling

To check whether the principle of performance based model selection can also increase the accuracy of the different forecasting models when applied in a quasi real-time exercise, we specify the models based on past forecasting performance rather than on the various information criteria. We call this approach *performance based model selection, real time* (PBRT). With PBRT, we evaluate the performance of the various specifications of the different forecasting models over a subevaluation sample ranging from $T - s^{eval} + 1$ until T . The best specification of each forecasting model, i.e. the specification that yields the smallest MSE over the subevaluation sample, is then used to estimate the respective model with information up to T and to compute forecasts for $T + h$. We set the length of the subevaluation sample s^{eval} equal to 4 quarters. To be consistent, the various specifications of the different forecasting models for the subsample evaluation as well as for the final forecast are estimated over a rolling window of 60 quarters. This implies that our first subevaluation sample ranges from 1994Q4 until 1995Q3, while the forecasts of the different models for the exercise in this paragraph are evaluated from 1997Q3 until 2013Q3.

Alternatively, we implement forecast pooling, an approach that has been proposed in the literature to overcome the uncertainty related to the selection of the best performing specification of a forecasting model (see for example Kuzin et al., 2013). The basic idea here is, similar to model averaging, to pool over the forecasts obtained with a large set of different specifications of a forecasting model to obtain the final forecast of a variable of interest.¹¹ We implement two versions of forecast pooling: *unweighted pooling* and *MSE-weighted pooling*. According to the first variant, the final forecast of a variable is obtained by averaging over the various forecasts computed with different specifications of a certain forecasting model. By contrast, with the second variant we use a weighted mean to obtain the final forecast, where the weight

¹¹Conceptually, the difference between model averaging and pooling lies in the source of uncertainty. While with model averaging there is uncertainty about the predictor variable to include in the estimation, with pooling there is uncertainty with respect to the best performing specification of a model given a set of predictors.

is the inverse of the MSE of the respective model specification over the sub-evaluation sample ranging from $T - s^{eval} + 1$ until T .

In Table 3.6, panel (b) - (d) we report the absolute multivariate MSEs that result from this exercise. To facilitate the direct comparison we additionally show the absolute MSEs of the different forecasting models with IC for the same evaluation sample in panel (a). The entries indicate that we can

Table 3.6: Multivariate MSEs with IC, PBRT and Pooling.

(a) IC: Information criteria based model selection							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.63	6.56	6.78	6.36	7.01	7.30	7.27
4	11.10	9.68	11.36	9.25	11.40	10.04	10.26
8	13.84	12.64	17.71	11.71	16.79	15.09	15.01
(b) PBRT: Performance based model selection, real time							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.57	7.12	7.37	5.85	9.91	7.27	7.17
4	10.24	14.77	12.54	10.82	17.00	10.18	10.35
8	12.96	33.26	14.89	14.40	22.20	14.12	13.97
(c) Forecast pooling, unweighted mean							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.43	6.18	6.76	6.17	6.48	7.07	6.87
4	10.89	10.11	11.29	9.47	12.25	9.89	10.17
8	13.79	13.47	17.07	11.80	19.55	14.91	14.39
(d) Forecast Pooling, MSE-based mean							
horizon	AR	LBVAR	FAAR	BFAVAR	DFM	EWA	BMA
1	7.45	6.18	6.79	6.19	6.52	7.09	6.88
4	10.95	10.16	11.24	9.48	11.90	9.90	10.17
8	13.74	13.37	16.78	11.81	18.32	14.91	14.40

Notes: All forecasting models are estimated over a rolling window of 60 quarters. The forecasts obtained by the different models are evaluated over the sample ranging from 1997Q3 until 2013Q3, thus for each horizon a total of 65 forecasts is computed. The different specification strategies are described in the text.

increase the accuracy of all forecasting models for $h = 1$ with either PBRT or forecast pooling compared to IC, though in most cases only by very little. Moreover, for higher horizons there is no improvement upon IC for the LBVAR, the BFAVAR and the DFM with either alternative specification approach.

When comparing the entries in panel (b) - (d), we find that, with the exception of the BFAVAR, in the short-run either unweighted or MSE-weighted forecast pooling works best for all models. This confirms previous findings

that pooling is indeed a good alternative to avoid choosing a model specification in real time that does not forecast well (see Kuzin et al. (2013)). By contrast, for $h = 8$ all models for which we find an improvement upon IC perform best with PBRT.

The overall lowest multivariate MSEs for all forecasting horizons are again obtained by the BFAVAR, with PBRT for $h = 1$ and with IC for higher horizons. The gains in accuracy upon the lowest multivariate MSE of the AR benchmark amount to roughly 20% for $h = 1$ and 10% for higher horizons.

Our results also indicate that the extremely good performance of the DFM with the optimized specification cannot be achieved with any model in real time. Therefore, relying on a single forecasting model that delivers a good performance and is not prone to the pitfall of choosing a specification that does not deliver the most precise forecasts, such as the BFAVAR or the LBVAR, seems to be a good choice for applied forecasters who cannot rely on out-of sample information to specify their forecasting models. Whether or not the small gains in forecasting accuracy over information criteria based model specification obtained with PBRT or forecast pooling justify the additional computational burden that comes with these approaches depends of course on the specific forecasting context at hand.

3.6 Conclusion

We have studied three different approaches to aggregating the informational content of a large dataset for forecasting key macroeconomic variables. We find that, overall, the Bayesian factor augmented vector autoregression and the large Bayesian vector autoregression perform best and generally yield more accurate forecasts than a simple AR benchmark model and other large scale approaches. This holds for both, measures for the joint forecasting performance for a set of 11 core variables as well as univariate performance measures for the individual series.

Our assessment of the joint predictive performance of the large scale approaches reveals that in general for short horizons all large scale approaches are able to efficiently use the correlation structure in the dataset to provide better forecasts than the AR benchmark, while for large horizons this is no longer the case. Here, the combination of aggregating the informational content of the large dataset into factors and shrinkage seems to be the most effi-

cient approach to use all the information in the dataset. We also find that the joint forecasting performance of the BFAVAR and the LBVAR is very robust with respect to the precise model specification, e.g. the number of lags of the dependent variable or the degree of shrinkage.

Regarding the size of the gains in forecasting accuracy over the AR benchmark for the individual series of our set of 11 variables, we find considerable differences for the different variables. While there are large increases in forecasting performance for few variables, the gains in accuracy rarely exceed 10% in most cases. One reason for this might be that some time series show very little persistence and are thus very hard to predict by univariate as well as multivariate forecasting models. Yet, even for time series with more persistence, the high collinearity in the large dataset seems to prevent large gains from the large-scale multivariate forecasting models over the AR benchmark. Still, when forecasters are interested in simultaneously predicting a larger number of variables, large-scale forecasting models have the advantage that they can be used to coherently forecast many variables. Finally, this might also be an advantage when it comes to the interpretation of the forecasts.

References

- Abberger, K. (2007). Forecasting quarter-on-quarter changes of German GDP with monthly business tendency survey results. Ifo Working Paper No. 40.
- Andrea Carriero, T. E. C. and M. Marcellino (2011). Bayesian vars: Specification choices and forecast accuracy. *Federal Reserve Bank of Cleveland Working Paper 1112*.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25(1), 52–60.
- Bañbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Banerjee, A. (2013). Editorial introduction to special issue on large data sets. *Oxford Bulletin of Economics and Statistics* 75(1), 1–5.

- Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20(4), 451–468.
- Berg, T. O. and S. R. Henzel (2013). Point and density forecasts for the Euro area using many predictors: are large BVARs really superior? Ifo Working Paper No. 155.
- Bernanke, B., J. Boivin, and P. Elias (2005). Measuring monetary policy: A factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120(1), 387–422.
- Bernanke, B. S. and J. Boivin (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics* 50(3), 525–546.
- Bernardini, E. and G. Cubadda (2014). Macroeconomic forecasting and structural analysis through regularized reduced-rank regression. *International Journal of Forecasting*, forthcoming.
- Buchen, T. and K. Wohlrabe (2014). Assessing the macroeconomic forecasting performance of boosting, evidence for the United States, the Euro Area, and Germany. *Journal of Forecasting* 33(4), 231–242.
- Carriero, A., G. Kapetanios, and M. Marcellino (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics* 26(5), 735–761.
- Christoffersen, P. F. and F. X. Diebold (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics* 16(4), 450–458.
- De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components? *Journal of Econometrics* 146(2), 318–328.
- Del Negro, M. and F. Schorfheide (2013). DSGE model-based forecasting. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2. Elsevier Publications.
- Drechsel, K. and R. Scheufele (2012a). Bottom-up or direct? Forecasting German GDP in a data-rich environment. Swiss National Bank Working Papers 16.

- Drechsel, K. and R. Scheufele (2012b). The performance of short-term forecasts of the German economy before and during the 2008/2009 recession. *International Journal of Forecasting* 28(2), 428–445.
- Dreger, C. and C. Schumacher (2005). Out-of-sample performance of leading indicators for the German business cycle: Single vs. combined forecasts. *Journal of Business Cycle Measurement and Analysis* 1, 71–87.
- Faust, J. and J. H. Wright (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics* 27(4), 468–479.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics* 82(4), 540–554.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). Do financial variables help forecasting inflation and real activity in the Euro area? *Journal of Monetary Economics* 50(6), 1243–1255.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association* 100(471), 830–840.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.
- Giannone, D., M. Lenza, and G. E. Primiceri (2012). Prior selection for vector autoregressions. NBER Working Paper 18467.
- Henzel, S. and S. Rast (2013). Prognoseeigenschaften von Indikatoren zur Vorhersage des Bruttoinlandsprodukts in Deutschland. *ifo Schnelldienst* 66(17), 39–46.
- Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* 12(2), 99–132.
- Kholodilin, K. A. and B. Siliverstovs (2006). On the forecasting properties of the alternative leading indicators for the German GDP: Recent evidence. *Journal of Economics and Statistics* 226(3), 234–259.

- Korobilis, D. (2013). Var forecasting using Bayesian variable selection. *Journal of Applied Econometrics* 28(2), 204–230.
- Kuzin, V., M. Marcellino, and C. Schumacher (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics* 28(3), 392–411.
- Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics* 4(1), 25–38.
- Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic Forecasts and Expectations*. NBER, New York.
- Müller-Dröge, H. C., T. M. Sinclair, and H. O. Stekler (2014). Evaluating forecasts of a vector of variables: a German forecasting competition. CAMA Working Paper 55.
- Pirschel, I. and M. H. Wolters (2014). Forecasting german key macroeconomic variables using large dataset methods. Kiel Working Paper 1925.
- Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting* 26(4), 271–302.
- Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters* 107(2), 95–98.
- Schumacher, C. (2011). Forecasting with factor models estimated on large datasets: A review of the recent literature and evidence for German GDP. *Journal of Economics and Statistics* 231(1), 28–49.
- Schumacher, C. and C. Dreger (2004). Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models? *Journal of Economics and Statistics* 224(6), 732 – 750.
- Sims, C. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review* 39(4), 949–968.
- Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.

- Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20(2), 147–162.
- Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41(3), 788–829.
- Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23(6), 405–430.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* 58(1), 267–288.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 135–196. Amsterdam: North Holland.
- Timmermann, A. and H. K. van Dijk (2013). Dynamic econometric modeling and forecasting in the presence of instability. *Journal of Econometrics* 177(2), 131–133.
- Wolters, M. H. (2014). Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics*, forthcoming.
- Wright, J. (2009). Forecasting U.S. inflation by Bayesian model averaging. *Journal of Forecasting* 28(2), 131–144.

Chapter 4

Forecasting Euro Area Recessions in Real-Time*

Abstract

I present evidence that the linear mixed-frequency Bayesian VAR provides very sharp and well-calibrated monthly real-time recession probabilities for the euro area for the period from 2004 until 2013. The model outperforms not only the univariate regime-switching models for a number of hard and soft economic indicators and their optimal linear combinations, but also a real-time recession index obtained with Google Trends data. This result holds irrespective of whether the joint predictive distribution of several economic indicators or the marginal distribution of real GDP growth is evaluated to extract the real-time recession probabilities of the mixed-frequency Bayesian VAR. The inclusion of the confidence index in industry proves to be crucial for the performance of the model.

Keywords: Density nowcasting, Real-time recession forecasting, Mixed-frequency data, Bayesian VAR, Regime-switching models, Linear opinion pool, Google Trends

JEL-Codes: C53, E32, E37.

*This chapter is based on my paper entitled *Forecasting Euro Area Recessions in Real-Time*, Kiel Working Paper No. 2020, 2016.

4.1 Introduction

Forecasts of macroeconomic activity are highly important for economic policymakers' decision making processes. In addition to precise point forecasts, a reliable and timely prediction of business cycle turning points can be extremely useful for the design of appropriate economic policy, since the effectiveness of monetary and fiscal policy measures can depend heavily on the current phase of the business cycle.¹ In practice, however, there are many problems associated with the real-time availability of many macroeconomic time series. These include mixed data frequencies, the irregular and sometimes varying publication lags of various macroeconomic indicators (often referred to as *ragged edges*) and data revisions. They pose huge challenges to professional forecasters (see Giannone et al., 2008, for a detailed discussion) and should therefore be taken into account when assessing the accuracy of alternative forecasting approaches.²

Researchers looking to separate periods of economic expansion from recessions typically turn to non-linear regime-switching models (for recent applications see Chauvet and Piger, 2008; Nalewaik, 2012; Camacho et al., 2014). As an alternative, Bayesian density forecasting approaches (overviews are provided, for example, in Karlsson, 2013; Geweke and Whiteman, 2006) can be used to compute the probability that the economy is in a specific business cycle phase at a certain point in time. This has been documented, for example, by Österholm (2012), who estimates the probability of a recession in the US in the third and fourth quarter of 2008 with a quarterly linear Bayesian vector autoregression. Dovern and Huber (2015) estimate a linear Bayesian global vector autoregression and show that the model delivers probabilistic recession forecasts that are more precise than those obtained with country-specific models. However, the analyses in both of these papers are not con-

¹Lo and Piger (2005) provide supporting empirical evidence for monetary policy and Auerbach and Gorodnichenko (2012) for fiscal policy.

²Recently, the success of different econometric forecasting methods in providing a reliable assessment of the prevailing economic conditions in terms of GDP growth point forecasts, while at the same time coping with the outlined difficulties, has been demonstrated. These methods include bridge equation models (Baffigi et al., 2004; ECB, 2008), MIDAS-models (Kuzin et al., 2011; Schumacher, 2014) and factor models (Schumacher and Breitung, 2008; Bańbura and Rünstler, 2011) as well as combinations of the aforementioned methods (Angelini et al., 2011; Marcellino and Schumacher, 2010). The relative accuracy of these methods has been studied, for example, in Forni and Marcellino (2014).

ducted in a real-time setting since the models used there do not account explicitly for the aforementioned features of real-time data.

By contrast, the linear mixed-frequency Bayesian vector autoregression (MFBVAR) proposed by Schorfheide and Song (2015) is well-suited to identifying business cycle turning points in real-time, since it can be estimated on mixed-frequency data with ragged edges. The model has been proven to increase the accuracy of short-term point and density forecasts for a number of variables (Schorfheide and Song, 2015), yet it is still an open question whether it can also achieve forecast gains for the real-time detection of business cycle phases.

With this paper, I fill this gap and provide evidence that the MFBVAR provides very accurate monthly real-time recession probabilities for the euro area for the period from 2004 until 2013. The risks of a recession are defined here as the probability that current-quarter GDP growth is part of a sequence of two consecutive quarters, both displaying negative GDP growth rates. They are obtained from the joint predictive distribution of the back-, now- and forecasts for euro area real GDP growth in a real-time forecasting setting. I compare the accuracy of the MFBVAR real-time recession signals with those obtained with univariate regime-switching models for a number of hard and soft economic indicators as well as their optimal linear combinations. Moreover, I consider a real-time recession index based on Google Trends data that is constructed as a population-weighted mean of the Internet query shares for the word "recession" in the eleven largest euro area countries.

Related papers that focus on the real-time detection of recessions (see Hamilton, 2011, for a comprehensive overview) often rely on monthly variables such as industrial production as a proxy of overall economic activity (Chauvet and Piger, 2008; Anas et al., 2008; Bellégo and Ferrara, 2009; Schreiber, 2014). Exceptions to this are Aastveit et al. (2014) and Camacho et al. (2014), who estimate models that account for many of the outlined features of real-time data. In particular, Aastveit et al. (2014) solve the mixed-frequency data issue by applying the Bry-Boschan rule (Bry and Boschan, 1971), an algorithm that detects recessions, to a bridge equation model now-cast and compare the accuracy of the real-time recession probabilities thus obtained to those obtained with an autoregressive Markov-switching model for Norwegian GDP. Camacho et al. (2014) estimate a mixed-frequency

Markov-switching dynamic factor model for the euro area which captures not only co-movements across various economic indicators through a common business cycle factor, but also shifts in the business cycle regime.

In all these studies, the real-time recession signals are compared with an official business cycle chronology such as, for example, that established by the CEPR Euro Area Business Cycle Dating Committee or that of the NBER for the US. Accordingly, in this paper, I use the CEPR euro area business cycle turning points as a benchmark to evaluate the alternative forecasting approaches. However, while most of the aforementioned papers confine their analysis to a comparison of the official business cycle turning points to those obtained with their respective econometric models, I compute formal measures that explicitly assess the calibration as well as the sharpness of the probabilistic recession forecasts obtained with the different methods. An approach is said to deliver well-calibrated probability forecasts if the empirical event probability conditional on a forecast is close to that probability forecast, i.e. that it actually rains in 70% of the times rain was announced with a probability of 70%. Sharpness, on the other hand, refers to the question of whether the probability forecasts are clear-cut, i.e. whether they are clustered around the confident values of zero and one, rather than the ambiguous value of 0.5. The ideal probabilistic forecast maximizes sharpness subject to calibration (Ranjan and Gneiting, 2010). This implies that the real-time recession signals need to be not only very timely but also clear-cut.

Beyond that, I investigate the discriminatory skill of the different approaches. That is to say I explore the extent to which the real-time recession probabilities obtained with the alternative models are useful signals when binary forecasts for the occurrence or non-occurrence of a recession have to be issued. The ad-hoc binary event classifier typically used in related papers is 0.5, and a recession is announced if the recession probability exceeds this threshold (see, for example, Hamilton, 1989; Chauvet and Piger, 2008). However, as it turns out, this threshold is not always optimal in the sense that it maximizes the number of correct recession predictions and, simultaneously, minimizes the number of false alarms. Lahiri and Wang (2013) present a survey of different measures to evaluate probabilistic recession forecasts which take this aspect into account and I apply the receiver operating characteristic and the Peirce skill score to assess the different models' discriminatory skill. Note that these evaluation approaches are closely related to the litera-

ture on the signals approach, where potential indicators for economic crises are analyzed with respect to their early warning properties (see, for example, Reinhart and Kaminsky, 1999; Boysen-Hogrefe et al., 2015).

My findings show that the MFBVAR real-time recession probabilities are very sharp and well-calibrated and that only a univariate Markov-switching model for the confidence index in industry yields probabilistic recession forecasts that perform equally well. Both models also have the highest skill to discriminate between recessions and expansions in real-time, although the optimal binary event classifier used to translate the probabilistic forecasts into binary recession signals varies for both models. By contrast, the real-time recession signals obtained from other soft indicators such as the Economic Sentiment Indicator or the confidence index in retail sales are much less well-calibrated. In fact, these methods deliver many recession signals in non-recession periods, which would suggest that they are potentially driven by more than economic fundamentals. The probabilistic forecasts obtained with the models for the hard economic indicators, in particular for industrial production and real GDP, on the other hand, lack sharpness due to the long publication lag of the respective data. As a consequence, they have no discriminatory skill to distinguish between recession and expansion periods in real-time. The combinations of the probabilistic forecasts of the univariate regime-switching models improve upon most of their components in all dimensions considered here. However, even when an optimal combination scheme is applied, the pooled real-time recession probabilities are outperformed by those of the MFBVAR. The Google Trends real-time recession indicator performs better than most univariate regime-switching models and pools, but it is clearly worse than the MFBVAR in terms of calibration, sharpness and discriminatory skill. The index delivers very ambiguous real-time recession signals particularly between the two recession periods in the sample and proves to be of very limited use.

Finally, in the robustness analysis, I provide evidence that the inclusion of the confidence index in industry is crucial for the good performance of the MFBVAR. Moreover, I investigate the extent to which the MFBVAR real-time recession signals can be improved upon by simultaneously assessing the joint development of several economic indicators through the multivariate predictive distribution of these variables, rather than just the path of GDP growth alone. My findings indicate that no significant gains in accuracy are

obtained compared to the benchmark, where the real-time risks of a recession are defined as the probability that current-quarter GDP growth is part of a sequence of two consecutive quarters both displaying negative GDP growth.

The remainder of this paper is structured as follows. In section 4.2 I give an overview of the euro area business cycle since 2000, while in section 4.3 I describe the dataset used for the empirical application in this paper. In section 4.4 I set out the alternative forecasting approaches, which are evaluated using the formal measures described in section 4.5. In section 4.6 I present the main results, while the results of the robustness checks are shown in section 4.7. Finally, in section 4.8 I conclude.

4.2 The Euro Area Business Cycle

The CEPR Euro Area Business Cycle Dating Committee has been publishing business cycle turning points for the euro area since 2003.³ Table 4.1 displays the euro area business cycle phases since 2000 as stated by the CEPR.

Table 4.1: CEPR euro area business cycle phases since 2000.

Dates	Business cycle phase
Until January 2008	Expansion
February 2008 - April 2009	Recession
May 2009 - July 2011 ⁴	Expansion
August 2011 - January 2013	Recession
Since February 2013	Expansion

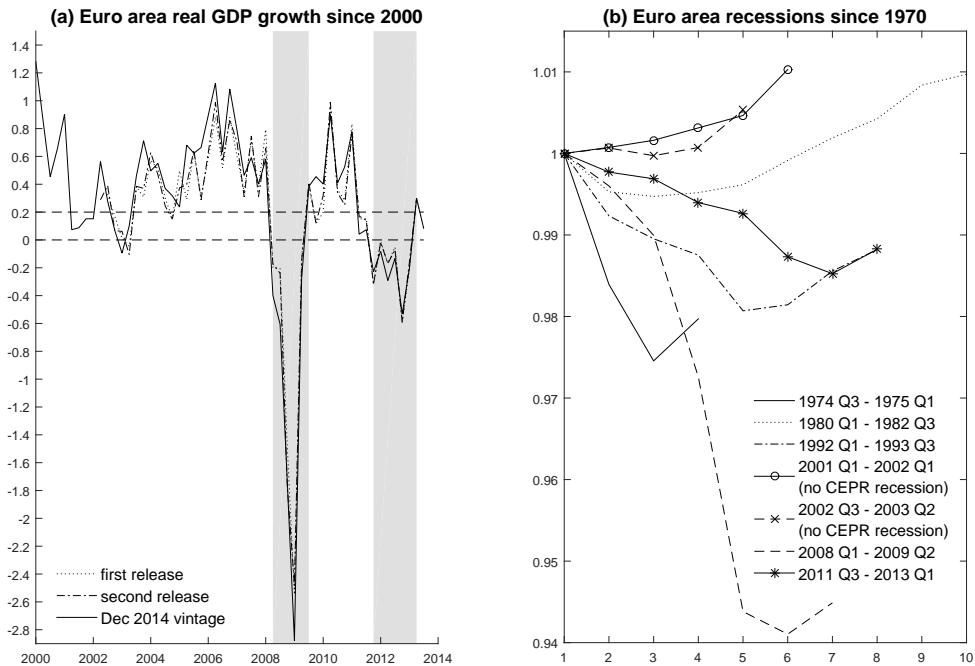
The committee defines a recession as "... a significant decline in the level of economic activity, spread across the economy of the euro area, usually

³The publication lag for the CEPR business cycle turning points is quite substantial. For example, the euro area business cycle peak that occurred in January 2008 was not announced until 31 March 2009 only. Similarly, the trough in April 2009 was identified with a delay of more than 12 months.

⁴The CEPR has recently abandoned its practice of announcing the month of the business cycle turning point. Hence, from July 2011 onwards, I set the first month of the quarter announced as being a business cycle turning point as the month of the respective peak or trough. This assumption is quite conservative and requires the real-time recession signals of the alternative approaches to be very timely. The results of an evaluation where the second or third month of a quarter is set as the turning point are very similar to those presented in section 2.4 and are available upon request.

visible in two or more consecutive quarters of negative growth in GDP, employment and other measures of aggregate economic activity for the euro area as a whole, and reflecting similar developments in most countries. A recession begins just after the economy reaches a peak of activity and ends when the economy reaches its trough." (Artis et al., 2003). In total, the committee has identified two recessions since 2000, namely the Great Recession of 2008-09 and the recession in connection with the European debt crisis of 2011-13. These are marked by the shaded areas in panel (a) of Figure 4.1, which displays quarter-on-quarter euro area real GDP growth since 2000. While the first recession period lasted for 15 months, the second recession in the sample persisted for 18 months in total. During the Great Recession, euro area real GDP growth turned negative in the second quarter of 2008 and remained so until the second quarter of 2009. The strongest decrease in real GDP amounted to -2.5% and occurred in the first quarter of 2009. The European debt crisis, by contrast, was much milder, with real GDP growth dipping by a maximum of -0.6% in the fourth quarter of 2012. In this recession, real GDP growth rates were negative from the fourth quarter of 2011 until the first quarter of 2013. Panel (b) of Figure 4.1 plots the course of euro area real GDP over as many as 10 quarters after all CEPR-dated peaks since 1970 (normalized to one). It can be seen that compared to earlier recessions in the euro area, the Great Recession was by far the most severe in terms of depth, while the recession in connection with the European debt crisis was characterized by a decline in economic activity that was comparably prolonged but only moderate overall.

Figure 4.1: Euro area economic activity.



In addition to the CEPR, other authors have attempted to establish a monthly business cycle chronology for the euro area (see Anas et al., 2008; Billio et al., 2012, for recent examples that also cover the Great Recession of 2008/2009). Their assessment of the Great Recession, which Billio et al. (2012) date from September 2008 until July 2009, differs slightly from that of the CEPR committee. Moreover, there is also disagreement as to whether there was another recession in the euro area between 2000 and 2005. Billio et al. (2012) point to an industrial recession from September 2001 until May 2006, which Anas et al. (2008) date from December 2000 until November 2001. However, the view of the CEPR is that the overall evidence did not support a "fully-fledged recession but rather a prolonged pause in the growth of economic activity" (Artis et al., 2003). This is also confirmed in Panel (b) of Figure 4.1, which includes not only the official CEPR recessions since 1970 but also two periods between 2000 and 2005 with weak real GDP growth rates of less than 0.2%.

Note that there are other formal approaches to identifying business cycle turning points such as the well-known Bry-Broschan rule (Bry and Boschan, 1971; Harding and Pagan, 2002). For the period from 2000 onwards, however,

this rule delivers the same business cycle chronology for the euro area as the CEPR Euro Area Business Cycle Dating Committee.

4.3 Data

For the empirical application in this paper I use a real-time dataset that consists of 123 monthly data vintages for October 2003 until December 2014, all of which start in January 1991.⁵ Each of these data vintages provides a historical snapshot of the data at the beginning of each month, as it was available at the time. This implies that the dataset reflects not only the publication lag of each variable with respect to the reference date, i.e. the date at which the snapshot was taken, but also changes in the data flow over time driven by recent improvements in the timeliness of various indicators. The dataset was obtained from the real-time database of the European Central Bank's Statistical Data Warehouse in early December 2014. A detailed description of the database, the variables included as well as the treatment of issues such as data revisions, changing variable definitions and the composition of the euro area over time can be found in Giannone et al. (2010). All series are seasonally adjusted, and natural logarithms are taken for all variables not expressed in rates.

The dataset includes eleven monthly indicators for the euro area which are summarized in Table 4.2. The reporting lag for each indicator, which is displayed in the second column of the table, illustrates the heterogeneity in the timeliness of the publication of different series. For example, in early January 2008, the most recently available observation for industrial production excluding construction was dated October 2007, while for the Economic Sentiment indicator, the figure for December 2007 was already available.

For quarterly euro area real GDP, which is also included in the analysis, the first official release is usually published about 45 days after the end of the reference quarter. Hence, the respective observation is included for the first time in the data vintage of the third month of the following quarter. This implies not only that the figure for current-quarter GDP is unknown throughout the quarter, but also that the figure for previous-quarter GDP is not available

⁵The very small number of vintages that were unavailable for some variables, were replaced by the data vintage for the previous month. Moreover, since the data vintages for the unemployment rate only start after 1991, all vintages were augmented with data taken from the OECD database.

Table 4.2: Euro area monthly indicators with respective reporting lag in early January 2008.

Euro area monthly indicators	Reporting lag in January 2008
Industrial production excluding construction	3 months
CPI	2 months
Unemployment	2 months
New passenger car registrations	2 months
Money supply M1	2 months
Money supply M3	2 months
Economic sentiment indicator	1 month
Stock market index	1 month
Oil price	1 month
Confidence index industry	1 month
Confidence index retail sales	1 month

in the first and second month of a given quarter. Hence, in each quarter,, an estimate of current-quarter GDP (referred to as the *nowcast*) and in some cases of previous-quarter GDP (referred to as the *backcast*) as well have to be computed. To increase the estimation sample for the univariate quarterly Markov-switching model for real GDP growth (see section 4.4.3), each of the data vintages for GDP is augmented with data from the 14th update of the area-wide model database (Fagan et al., 2001) covering the period from 1970 Q1 until 1990 Q4.

4.4 Forecasting Approaches

To assess the real-time probabilities of a recession in the euro area, I implement a number of different econometric models which are described below in sections 4.4.1 to 4.4.3. In addition, I consider a real-time recession indicator based on Internet search data that is described in section 4.4.4.

4.4.1 The Bayesian Mixed-Frequency VAR

Consider the following monthly VAR

$$X_t = C + A_1 X_{t-1} + \dots + A_p X_{t-p} + \epsilon_t, \quad (4.1)$$

where the vector $X_t = (x_{1,t}^m, \dots, x_{11,t}^m, x_t^q)'$ contains the 11 observable monthly indicators listed in Table 4.2 and latent monthly real GDP x_t^q . Following Bańbura et al. (2010) I include the variables in log-levels rather than growth rates so as to not lose information that might possibly be contained in the trends. p denotes the number of lags included in the estimation and is set to $p = 6$. C is a vector of constants, A_1, \dots, A_p are parameter matrices, and ϵ_t is a vector of independently identically distributed white noise error terms with zero mean and covariance matrix Σ .

To account for the mixed frequencies and the ragged edges of the dataset, the VAR outlined in equation (4.1) is rewritten in state-space form with a time-varying measurement equation (Schorfheide and Song, 2015) that reads

$$Y_t = S_t \Lambda Z_t. \quad (4.2)$$

The corresponding transition equation for the states $Z_t = (X_t', \dots, X_{t-p+1}')'$ is simply the companion form of the monthly VAR described in equation (4.1). In equation (4.2) the time-varying diagonal selection matrix S_t governs that the states contained in Z_t are included in the observation vector Y_t only if they are truly observable, while the matrix Λ aggregates latent monthly real GDP into its observed quarterly counterpart. In particular, following Schorfheide and Song (2015) the log of quarterly real GDP is assumed to be observable every third month only and to be equal to the average over the three unobserved monthly GDP figures in the respective quarter, i.e. $y_t^q = \frac{1}{3}(x_t^q + x_{t-1}^q + x_{t-2}^q)$. Hence, for $t = 3, 6, 9, \dots, T^b$, where T^b is the last month in which a quarterly GDP figure is observable, the observation vector reads $Y_t = (y_{1,t}^m, \dots, y_{11,t}^m, y_t^q)'$, where $y_{j,t}^m$ are the $j = 1, \dots, 11$ monthly indicators and y_t^q denotes observed quarterly real GDP. By contrast, in the first and second month of each quarter y_t^q is dropped from Y_t . Moreover, at the current edge, e.g. for $t > T^b$, y_t^q is never included and depending on their publication lags some of the $y_{j,t}^m$ are dropped from Y_t as well.

The mixed-frequency state-space model outlined above is estimated with Bayesian techniques using data up to month $T > T^b$. This involves the estimation of the marginal posterior distributions of the unknown VAR parameters A_1, \dots, A_p, C and Σ as well as the estimation of the unknown state vector $Z_{1:T}$. Following Schorfheide and Song (2015), I rely on a version of the normal inverse Wishart prior that retains the main principles of the widely used

Minnesota prior (Litterman, 1986; Kadiyala and Karlsson, 1997). The prior is augmented to constrain the sum of coefficients of the VAR (Sims and Zha, 1998) as well as to incorporate the belief that the variables in the VAR follow a common stochastic trend. I implement this prior using the dummy variable approach outlined in Bańbura et al. (2010).⁶

The initial values of the state vector Z_0 are sampled conditional on a pre-sample ranging from April 1991 until December 1994. A Gibbs sampler then iteratively samples the VAR parameters A_1, \dots, A_p, C and Σ as well as the unknown states $Z_{1:T}$ from their respective conditional posterior distributions.

For each of the retained Gibbs draws of the VAR coefficients $A_1^i, \dots, A_p^i, C^i, \Sigma^i$ and the vector of states $Z_{1:T}^i$ a shock vector ϵ_{T+h}^i is drawn from $N(0, \Sigma^i)$ and equation (4.1) is iterated forward to compute forecasts for the monthly observable and unobservable variables in \hat{X}_{T+h}^i with $h = 1, \dots, 12$. The forecasts for unobservable monthly GDP are transformed into their quarterly counterparts based on equation (4.2). From these I compute the implied forecasts for quarterly GDP growth $\Delta \hat{y}_{T^B+\tilde{h}}^i$ where T^B denotes the last quarter for which GDP was observable and $\tilde{h} = 1, \dots, 3$. The set of $\{\Delta \hat{y}_{T^B+\tilde{h}}^i\}_{i=1}^N$ approximates the predictive distribution of the back-, now- and forecasts of quarterly euro area GDP growth that can be used to compute pointforecasts as the mean or median of the distribution and real-time recession probabilities.

Note that depending on the current information set, i.e. the month of the quarter in which the prediction is made, $\Delta \hat{y}_{T^B+1}$ could either denote a backcast (implying that we are in the first or second month of a quarter when last quarter GDP is not available yet) or a nowcast. Correspondingly, $\Delta \hat{y}_{T^B+2}$ refers to a nowcast if it is computed in the first two months of a quarter and to a 1-quarter ahead forecast in every third month of a quarter, and so on. For example, in January 2008 the most recently available observation for GDP refers to the third quarter of 2007 ($T^B = 2007Q3$) and $T^B + 1$ denotes the backcast for the fourth quarter of 2007, while $T^B + 2$ refers to the nowcast for the first quarter of 2008. By contrast, two months later, in March 2008, the figure for the fourth quarter of 2007 has been released ($T^B = 2007Q4$), and the nowcast for the first quarter of 2008 is denoted as $T^B + 1$. This has to be taken into account in the following when computing the MFBVAR real-time recession probabilities.

⁶A detailed outline of the prior is provided in the appendix.

According to a widely used (approximate) definition, the economy is in a recession if real GDP growth is negative for at least two consecutive quarters. I will therefore define the real-time risks of a recession as the probability that the nowcast for current-quarter GDP growth (i.e. either $\Delta\hat{y}_{T^B+1}$ or $\Delta\hat{y}_{T^B+2}$, depending on the current information set) is part of a sequence of two consecutive quarters, both displaying negative GDP growth rates. This criterion implies that the GDP growth nowcast could be either the first or the second period of a two-quarter recession sequence. Hence, taking into account the data availability in month T , the real-time recession probabilities implied by the MF-BVAR can be computed as $\pi_T^{MFBVAR} =$

$$\begin{cases} Pr(\Delta\hat{y}_{T^B} < 0, \Delta\hat{y}_{T^B+1} < 0 \cap \Delta\hat{y}_{T^B+1} < 0, \Delta\hat{y}_{T^B+2} < 0 | Y_T) & \text{for } T = 3, 6, \dots \\ Pr(\Delta\hat{y}_{T^B+1} < 0, \Delta\hat{y}_{T^B+2} < 0 \cap \Delta\hat{y}_{T^B+2} < 0, \Delta\hat{y}_{T^B+3} < 0 | Y_T) & \text{otherwise.} \end{cases} \quad (4.3)$$

From the Gibbs sampler output π_T^{MFBVAR} can be easily obtained as

$$\pi_T^{MFBVAR} = N^{-1} \sum_{i=1}^N \mathcal{I}(\Delta\hat{y}_{T^B:T^B+3}^i), \quad (4.4)$$

where $\mathcal{I}(\cdot)$ denotes an indicator function that is equal to one if, and only if, the GDP growth nowcast for the current quarter is part of a consecutive sequence of two quarters both displaying negative GDP growth. Note that in section 4.7 I consider alternative recession definitions to extract the real-time recession probabilities from the predictive distribution of the MFBVAR to see how far the model's performance is robust to the definition in equation (4.3).

4.4.2 A Quarterly Bayesian VAR

As a benchmark, I estimate a quarterly version of the model outlined in section 4.4.1 for each of the monthly data vintages. This implies that all monthly observations beyond T_b , i.e. the last month for which real GDP is available, are dropped and that all monthly indicators are aggregated to a quarterly frequency. Since the quarterly BVAR does not include any latent variables, there is no need to set up a state-space system as described above. However, apart from that the estimation procedure, the prior specification and the computation of the predictive densities and real-time recession probabilities are equivalent to those of the MFBVAR.

4.4.3 Markov-Switching Models

Univariate Markov-Switching Models

For selected indicators in the dataset, I set up the following univariate model:

$$\Delta y_t = \mu_{s_t} + \psi_{s_t} \Delta y_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \sim N(0, \sigma_{s_t}), \quad (4.5)$$

where Δy_t denotes the first difference of the respective indicator.⁷ The latent discrete variable s_t is assumed to evolve as a two-state, 1st order Markov-switching process, i.e. $s_t = \{E, R\}$, with transition probabilities

$$P(s_t = j | s_{t-1} = i) = p_{ij}, \quad i, j = E, R. \quad (4.6)$$

This model implies that the dynamics of the process described in equation (4.5) may differ between the two regimes E and R , thus allowing for structural breaks in the time series which can be estimated. Assuming that $\mu_E > \mu_R$ and that E are expansionary business cycle phases, while R stands for recession periods, the model can be used to identify business cycle turning points and to compute recession probabilities. In particular, the probability that the economy is in a recession in period t given the observations $y_{1:t}$ can be obtained as

$$P(s_t = R | y_{1:t}) = \frac{P(y_t | s_t = R) P(s_t = R | y_{1:t-1})}{\sum_{j=E}^R P(y_t | s_t = j) P(s_t = j | y_{1:t-1})} \quad (4.7)$$

where $P(s_t = j | y_{1:t})$ $j = E, R$ are referred to as filtered probabilities and $P(y_t | s_t = j)$ is the likelihood of the data in period t conditional on state j .

I estimate the model in equation (4.5) for all monthly indicators listed in Table 4.2, except the price indices and the monetary aggregates.⁸ In particular, following Anas et al. (2008) I consider industrial production, the unemployment rate and new passenger car registrations. In addition, I include the sentiment indices in the dataset, i.e. the Economic Sentiment indicator and the indices for confidence in industry and retail sales, and the stock market index, since these could potentially provide even timelier recession signals

⁷For the sake of simplicity I use the subscript t for both, the model in monthly frequency for the monthly indicators and the model in quarterly frequency for real GDP growth.

⁸For these variables, it is not intuitively clear that the assumption $\mu_E > \mu_R$ identifies a "high" state E and a "low" state R which correspond to phases of economic expansion and recession, respectively.

than the aforementioned hard economic indicators. For quarterly real GDP growth I estimate the well-established modified version of equation (4.5) proposed by Hamilton (1989), which only allows for regime shifts in the mean μ_{s_t} but not in the coefficient on the lagged dependent variable or the variance.

I estimate the univariate Markov-Switching models for the selected indicators with Bayesian techniques as described in Kim and Nelson (1999). This involves setting up a Gibbs sampler that iteratively draws the states $S_{1:T}$, the probabilities p_{EE} and p_{RR} and the remaining unknown parameters $\{\mu_j, \psi_j, \sigma_j\}$ for $j = E, R$ from their respective conditional posterior distributions using the filter proposed by Hamilton (1989) and the multi-move sampler suggested by Carter and Kohn (1994). A normal inverse Wishart prior, that is assumed to be symmetrical across the two states, is used for the coefficients and the variance in equation (4.5), while the probabilities p_{EE} and p_{RR} are assumed to a priori follow a beta distribution.

In this real-time application, I estimate the univariate Markov-switching models for each indicator m with data up to period T_m^* , i.e. the period for which the most recent observation for that indicator is available. The real-time recession probability for the current period $T = T_m^* + k_m$ can thus be obtained as

$$\pi_T^m = P(s_T = R | y_{1:T_m^*}) = P^{k_m} (P(s_{T_m^*} = R | y_{1:T_m^*}) P(s_{T_m^*} = E | y_{1:T_m^*}))', \quad (4.8)$$

where k_m is the publication lag of indicator m and the (2×2) matrix P contains the estimated transition probabilities p_{ij} , $i, j = E, R$.

Markov-Switching Linear Opinion Pool

The combination of forecasts from different sources is a very popular way of increasing the accuracy of point forecasts (see, for example, Bates and Granger, 1969; Stock and Watson, 2003; Kuzin et al., 2011; Schwarzmüller, 2015). As shown by Clements and Harvey (2011) and Ranjan and Gneiting (2010), among many others, the concept of forecast pooling can also be extended to probabilistic forecasts.

Following Anas et al. (2008), who construct a business cycle coincident indicator (BCCI) as a linear opinion pool of the probabilistic forecasts obtained with several univariate Markov-switching models to assess recession risks in the euro area, I implement a Markov-switching linear opinion pool with

equal weights as

$$\pi_T^{pool} = M^{-1} \sum_{m=1}^M \pi_T^m. \quad (4.9)$$

While Anas et al. (2008) chose their pooling weights to minimize first and second order forecast errors, I opt for an equal-weight pool. The main reason for this is technical, because in this real-time application, where the realized business cycle phases are observed with a substantial delay, it would be very hard to compute meaningful pooling weights based on the recent forecast performance of the alternative forecast approaches. In addition, the limited sample size impedes the calculation of performance-based weights over a presample. Moreover, for point forecasts, equally weighted forecast pools have proven to be extremely competitive in comparison to pools with performance-based pooling weights (see, for example, Stock and Watson, 2004; Timmermann, 2006).

However, Ranjan and Gneiting (2010) show that for probabilistic forecasts, in general, the linear opinion approach is suboptimal, since it yields pools that are uncalibrated and lack sharpness. They propose to recalibrate the linear opinion pool by applying a beta transform which is given as

$$\pi_T^{pool,opt} = H_{\alpha,\beta} \left(M^{-1} \sum_{m=1}^M \pi_T^m \right), \quad (4.10)$$

where $H_{\alpha,\beta}$ is the cumulative distribution function of a beta density with parameters α and β for which $\alpha = \beta \geq 1$. I apply the beta transform in the robustness analysis in section 4.7 to assess ex post the degree to which the performance of the linear equal-weight pool is inferior to that of the optimal pool.

In total, I implement three linear equal-weight pools. The first combines the probabilistic forecasts of all considered univariate Markov-switching models, while the second combines only the real-time recession probabilities obtained with the models for the soft indicators, i.e. the two confidence indices in industry and retail sales, the Economic Sentiment indicator and the stock market index. Finally, in the spirit of the BCCI proposed by Anas et al. (2008), I implement a pool that aggregates the predictions of the univariate Markov-switching models for industrial production, the unemployment rate and new passenger car registrations.

4.4.4 A Google Trends Real-Time Recession Indicator for the Euro Area

Google Trends (<https://www.google.de/trends/>) provides real-time indices of the relative volume of Internet search queries for specific terms in a predefined geographic area starting from January 2004.⁹ A growing body of literature has documented the usefulness of these data to predict variables such as unemployment (Choi and Varian, 2009a; Askitas and Zimmermann, 2009), consumer demand and sales (Vosen and Schmidt, 2011; Yan and Labbé, 2013; Fantazzini and Toktamysova, 2015) as well as tourism flows (Concha and Galán, 2012) and influenza outbreaks Ginsberg and Brilliant (2009).

As an alternative to the econometric real-time recession indices discussed above, I construct a euro area real-time recession indicator based on Google Trends data for the search query share of the word *recession*. In particular, the indicator is built as a population-weighted mean over the indices for the eleven largest euro area countries for which a query series is available. The list of countries includes Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal and Spain.

The idea behind this approach is very similar to the *R-word index* introduced by *The Economist* magazine in the early 1990s, which tracks the number of newspaper articles that use the word *recession* in a given quarter. The *R-word index* has been found to be a reliable source of early signals for pending recessions in the US (Doms and Morin, 2004), Germany (Mayr and Grossarth-Maticek, 2008) and Switzerland (Iselin and Siliverstovs, 2013).

4.5 Evaluation of Probabilistic Recession Forecasts

I evaluate the real-time recession probabilities of the alternative approaches outlined in section 4.4 with formal scoring rules for the period ranging from January 2004 until December 2013, i.e. a total of 120 recession predictions are considered for the evaluation. In particular, the recession probability forecasts π_t are compared to a binary indicator variable bc_t that is equal to one for periods that were declared recessions by the CEPR and zero otherwise (see Table 4.1).

⁹See Choi and Varian (2009b) for a description of the Google Trends interface and potential uses of the data.

The first scoring rule that I compute to assess the accuracy of the alternative approaches is the widely used quadratic probability score (*QPS*) which is given as

$$QPS = T^{-1} \sum_{t=1}^T (\pi_t - bc_t)^2. \quad (4.11)$$

Gneiting et al. (2007) show that this score is proper, meaning that the forecaster has no incentive to state anything but his or her true beliefs. The *QPS* corresponds to the common notion of mean squared error loss that is typically used to evaluate point forecasts. That implies that the score explicitly accounts for the strength of false signals, meaning that a recession probability $\pi_t^1 = 0.8$ in a month where $bc_t = 0$ is considered to be worse than $\pi_t^2 = 0.6$.

The *QPS* simultaneously addresses the sharpness and calibration of the probabilistic forecasts π_t . It can be decomposed to make the performance in both dimensions visible. The negatively-oriented component that assesses the calibration of the probabilistic forecast is given as

$$CAL = T^{-1} \sum_{j=1}^J T_j^{-1} (\bar{\pi}_j - \bar{bc}_j)^2, \quad (4.12)$$

while the positively-oriented sharpness component reads

$$SHARP = T^{-1} \sum_{j=1}^J T_j^{-1} (\bar{bc}_j - \bar{bc}_t)^2. \quad (4.13)$$

$\bar{\pi}_j \in [0, 1]$ are $j = 1, \dots, J$ discrete probability values used to define probability bins. T_j is the number of times π_t falls into bin j . \bar{bc}_j is the respective empirical conditional event frequency and \bar{bc}_t is the unconditional mean of bc_t (Ranjan and Gneiting, 2010). It holds that

$$QPS = CAL - SHARP + Var(bc_t). \quad (4.14)$$

I assess the statistical significance of the difference between the *QPS* scores for the alternative forecasting approaches with a version of the Diebold- Mariano test (Diebold and Mariano, 1995) that accounts for serial correlation of the forecast errors using Newey-West standard errors as proposed by Lopez (2001).

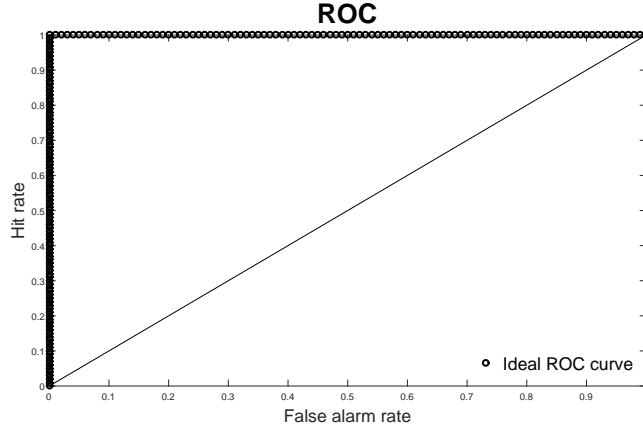
Lahiri and Wang (2013) survey a number of alternative methods that are well-suited to evaluating probabilistic forecasts for a decline in GDP. Unlike global measures of forecast quality such as the *QPS*, these measures explicitly take into account the ability of a forecasting approach to assess the odds for the occurrence of an event against its non-occurrence. This could be particularly important, for example, in the policy process when clear signals for the predicted occurrence or non-occurrence of an event have to be issued.

I apply two of the evaluation methods outlined in Lahiri and Wang (2013), namely the receiver operating characteristic (*ROC*) and the Peirce skill (*PS*) score. Both of these measures are based on (2x2) contingency tables which classify $\{\widehat{bc}_t\}_{t=1}^T$, the binary forecasts for the occurrence or non-occurrence of an event, into *Hits* ($\widehat{bc}_t = bc_t = 1$), *False Alarms* ($\widehat{bc}_t = 1, bc_t = 0$), *Misses* ($\widehat{bc}_t = 0, bc_t = 1$) and *Correct rejections* ($\widehat{bc}_t = bc_t = 0$) for a given period of observations $\{bc_t\}_{t=1}^T$. These binary event forecasts \widehat{bc}_t can be obtained from the probabilistic forecasts π_t via a binary event classifier w , such that $\widehat{bc}_t = 1$ if $\pi_t > w$ and $\widehat{bc}_t = 0$ otherwise.

The *ROC* is calculated for a range of thresholds w and thus explicitly accounts for the role of the binary event classifier for the accuracy of the binary forecast signal. The *ROC* is commonly depicted as a curve of the rates of *Hits* against the corresponding rates of *False alarms* over a range of thresholds w for a given period of observations $\{bc_t\}_{t=1}^T$. Ideally, for high values of w , the rate of *Hits* should increase monotonically from zero to one as w decreases, while the rate of *False alarms* should remain constant at zero. For further decreases in w , the ideal *ROC* curve would indicate increasing *False alarm* rates but a constant *Hit* rate of one (see Figure 4.2). By contrast, a *ROC* curve along the 45 degree line in the unit square indicates no discriminatory skill for the occurrence and non-occurrence of an event.

Alternatively, the *ROC* score can also be expressed as the area above the *ROC* curve. From the description of the ideal *ROC* curve, it is clear the *ROC* score $\in \{0, 1\}$ and that it is zero for the ideal forecasting method which perfectly discriminates between the occurrence and non-occurrence of an event.

Figure 4.2: Ideal ROC curve.



Finally, the PS score is computed as the difference between the rate of *Hits* (H) and the rate of *False alarms* (F) for a given threshold w , i.e.

$$PS(w) = \frac{\sum_{t=1}^T (bc_t \hat{bc}_t)}{\sum_{t=1}^T bc_t} - \frac{\sum_{t=1}^T \hat{bc}_t - \sum_{t=1}^T (bc_t \hat{bc}_t)}{T - \sum_{t=1}^T bc_t} = H - F. \quad (4.15)$$

For an ideal forecasting approach $PS(w) = 1$ or $PS(w) = -1$, whereby the latter value indicates that the binary signals are perfectly mislabeled. By contrast, $PS(w) = 0$ indicates no discriminatory skill at all. Following Lahiri and Wang (2013), I assess the statistical significance of the PS scores for the alternative forecasting approaches using the following standard error formula:

$$SE(w) = \sqrt{\frac{H(1-H)}{\sum_{t=1}^T bc_t} + \frac{F(1-F)}{T - \sum_{t=1}^T bc_t}} \quad (4.16)$$

4.6 Results

The monthly real-time recession signals obtained with the different methods described in section 4.4 are depicted in Figure 4.3. Since the original real-time recession probabilities obtained with the alternative approaches are very noisy, the real-time signals displayed are obtained as three-month weighted moving averages over the original probabilities.¹⁰ The two recessions in the evaluation period from January 2004 until December 2013, as

¹⁰In particular, the real-time recession signals are obtained as $\tilde{\pi}_t = \frac{1}{6}\pi_{t-2} + \frac{2}{6}\pi_{t-1} + \frac{3}{6}\pi_t$, where π_t is the original recession probability for period t .

dated by the CEPR (see Table 4.1), are again indicated by the shaded areas in each of the panels.

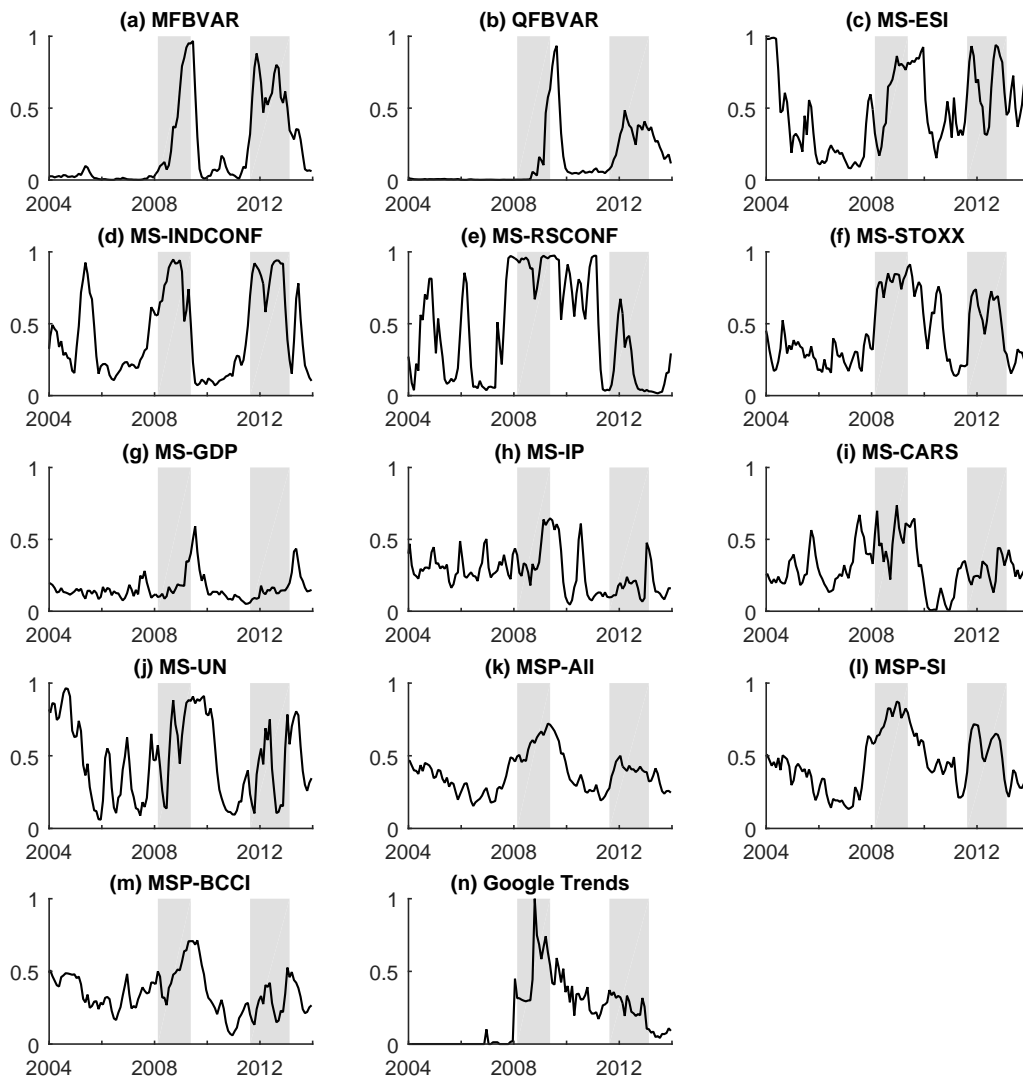
In general, all the approaches shown in Figure 4.3, show increased real-time recession signals during the Great Recession and the recession related to the European debt crisis. However, there are considerable differences between the alternative approaches with respect to the timeliness of recession signals as well as the amount of *False alarms*, i.e. recession signals in non-recession periods. This is also reflected in Table 4.3, which contains the *QPS* for the alternative models in the first column as well as the results of the formal assessment carried out to measure the calibration (*CAL*) and sharpness (*SHARP*) of the probabilistic forecasts of the different approaches in the second and third column. Small values for *QPS* and *CAL* reflect a high overall level of accuracy and a good calibration, respectively, while high values for *SHARP* indicate that the probabilistic forecasts of the alternative approaches are sharp.

Table 4.3: Evaluation of real-time recession probabilities, *QPS*.

	QPS	CAL	SHARP
MFBVAR	0.108	0.028	0.108
QFBVAR	0.211***	0.046	0.048
MS-ESI	0.231***	0.123	0.040
MS-INDCONF	0.110	0.074	0.114
MS-RSCONF	0.317***	0.161	0.005
MS-STOXX	0.148	0.075	0.087
MS-GDP	0.228***	0.033	0.006
MS-IP	0.227***	0.059	0.006
MS-CARS	0.188***	0.022	0.016
MS-UN	0.315***	0.173	0.009
MSP-All	0.168**	0.057	0.056
MSP-SI	0.154	0.100	0.090
MSP-BICC	0.215***	0.053	0.002
Google Trends	0.148	0.021	0.047

Notes: The real-time recession signals of the alternative approaches are evaluated over the sample from January 2004 until December 2013 using the CEPR business cycle chronology as a benchmark. *QPS*: quadratic probability score, *CAL*: calibration score, *SHARP*: sharpness score. ***(**,*) denote that the *QPS* is significantly different from the *QPS* of the MFBVAR at the 1% (5%,10%) level. For the model abbreviations see the notes to Figure 4.3.

Figure 4.3: Real-time recession signals for the euro area.



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas indicate euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. MFBVAR: mixed-frequency Bayesian vector autoregression, QFBVAR: quarterly BVAR. MS: univariate Markov-switching model for ESI: the Economic Sentiment indicator, INDCONF: index for confidence in industry, RSCONF: index for confidence in retail sales, STOXX: stock market indicator, GDP: real gross domestic product, IP: industrial production, CARS: new passenger car registrations, UN: unemployment rate. MSP: combination of probabilistic forecasts from univariate Markov-switching models for ALL: all univariate MS models, SI: the Economic Sentiment index, the confidence indices in industry and retail sale and the stock market index, BCCI: industrial production, the unemployment rate and new passenger car registrations.

Overall, the results presented in Table 4.3 suggest that MFBVAR performs best, as it achieves the lowest *QPS* among all the approaches considered here. However, the improvements of the MFBVAR over the univariate Markov-switching models for industry confidence, the stock market, the pool of models for the sentiment indices and the Google Trends real-time recession index are not statistically significant, as indicated by the results of the respective pairwise Diebold-Mariano tests.

The Markov-switching models for industry confidence achieves a *QPS* that is only slightly higher than that of the MFBVAR, although its real-time recession signals are less well-calibrated. This is confirmed in Figure 4.3, which shows that the small number of *False alarms* produced by the MFBVAR in panel (a) are considerably less pronounced than those of the model for the industry confidence index in panel (d). On the other hand, the latter model performs better than the MFBVAR in terms of sharpness, reflecting the fact that the real-time recession signals of the MFBVAR at the onset of the Great Recession are only very muted.

The univariate Markov-switching models for the stock market index, the pool of all sentiment indices and the Google Trends real-time recession indicator perform more or less equally well in terms of *QPS*. However, for the former two, this is due to the high sharpness of their forecasts, while the Google Trends real-time recession signals are better calibrated. Again, this is confirmed in Figure 4.3, which shows that the real-time recession signals obtained with the models for the stock market index in panel (f) and the sentiment pool in panel (k) are concentrated at the confident values of zero and one. On the other hand, both models issue many *False alarms*, pushing down their relative performance in terms of calibration. By contrast, the Google Trends indicator depicted in panel (n) is almost always equal to zero in the non-recession period prior to 2008 and also very low after the end of the second recession in the sample. This improves the calibration score of the index considerably. Its moderate performance in terms of sharpness can be explained by the fact that, apart from two drastic increases in euro area Internet users' interest in the word *recession* in early 2008 and in September 2008, the signals were mostly not very clear-cut, especially for the period between the two recessions in the sample. One reason for the latter observation could be the general high level of uncertainty in that period about the sustainability

of early signs of economic relief and fears of a double-dip recession (Camacho et al., 2014).

The results for the univariate Markov-switching models for real GDP growth, industrial production and new passenger car registrations (panels (g) - (i) in Figure 4.3) illustrate that a well-calibrated probabilistic forecast can be of minimal use in practice if it lacks sharpness. The *QPS* for these models is very high. To a lesser extent, this also applies to the QFBVAR in panel (b), the pool of all Markov-switching models in panel (k) and the BCCI pool considered in Anas et al. (2008) in panel (m). What all these models have in common is that they deliver real-time recession signals that are not very clear-cut and are, for the most part, also heavily delayed. However, given that these models are estimated with the series that have the highest publication lag (see Table 4.2) this result is not particularly surprising.

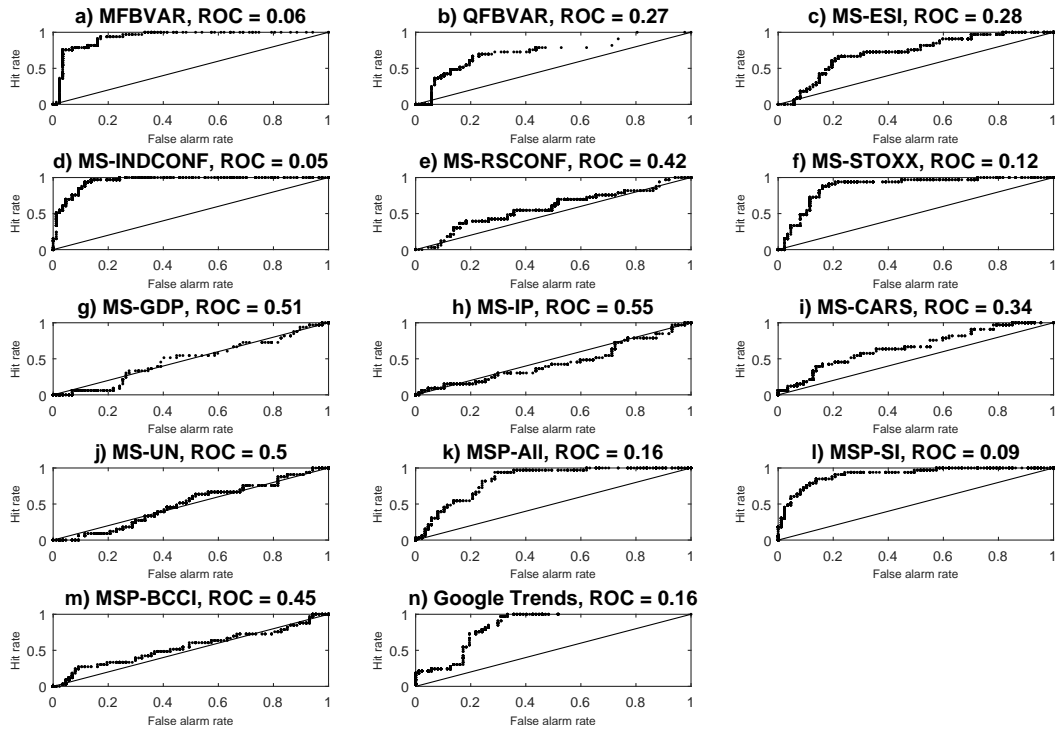
By contrast, the poor performance of the two sentiment indices, namely the Economic Sentiment indicator and the confidence index in retail sales, might come as something of a surprise. Both models apparently not only lack sharpness but are also very poorly calibrated. Indeed, from panel (c) and (e) in Figure 4.3 it can be seen that these two sentiment indices deliver many pronounced *False alarms*. One possible reason for this could be that these sentiment indices might not only be driven by hard economic fundamentals, but also by other factors. These could possibly be unrelated contagious waves of optimism and pessimism which are often referred to as animal spirits or noise shocks (Akerlof and Shiller, 2008; De Grauwe, 2011).

Finally, the univariate Markov-switching model for the unemployment rate achieves by far the highest *QPS* of all models considered here. The pattern of the model's real-time recession signals depicted in panel (j) suggests that the unemployment rate is likely to increase only with a certain lag after the beginning of a recession. In fact, this confirms that this variable is typically regarded as a lagging rather than a contemporaneous or even leading indicator for the state of the economy. In addition, the model also clearly reflects the steady increase in euro area unemployment until early 2005, which was not accompanied by a recession. On closer inspection, panel (j) and panel (c) of Figure 4.3 reveal some similarities between the real-time recession signals obtained with the unemployment rate and those delivered by the model for the Economic Sentiment indicator. This could suggest that the latter is

driven by news about the former. However, much more research would be needed to establish a solid causal link here, of course.

Figure 4.4 depicts the *ROC* curves for the alternative approaches as well as the respective *ROC* scores, i.e. the areas above the *ROC* curve in the unit square. It is plain to see that the *ROC* curves for the MFBVAR in panel (a) and for the model used for the confidence index in industry in panel (d) are far away from the 45-degree line and that they achieve the lowest *ROC* scores. This supports the notion that the discriminatory skill of the two approaches is very high overall, independent of the selected binary event classifier w . The *ROC* scores for the model used for the stock market index in panel (f), the Markov-switching pool of sentiment indices in panel (l) and, to a slightly lesser extent, the pool of all univariate models in panel (k) and the Google Trends indicator in panel (n) are moderately higher but still considerably lower than those for the remaining models. In particular, the models for the hard indicators depicted in panels (g) - (j) and the BCCI pool considered by Anas et al. (2008) in panel (m) produce very flat *ROC* curves, which indicates that the models' real-time recession signals are unable to discriminate between recession and non-recession periods.

Figure 4.4: ROC curves and scores.



Notes: The curves displayed show the rate of *Hits* and the corresponding rate of *False alarms* for varying thresholds used to transform the monthly real-time recession probabilities displayed in Figure 4.3 into binary signals for the occurrence and non-occurrence of a recession. The evaluation sample ranges from January 2004 until December 2013. ROC: receiver operating characteristic (area above the depicted curve). A small ROC denotes high discriminatory skill. For the model abbreviations see the notes to Figure 4.3.

Table 4.4: Evaluation of real-time recession probabilities, PS .

	PS				
	$w =$ 0.1	0.2	0.3	0.4	0.5
MFBVAR	0.75***	0.66***	0.71***	0.66***	0.60***
QFBVAR	0.43***	0.35***	0.32***	0.09	-0.03
MS-ESI	0.03*	0.22***	0.28***	0.26**	0.39***
MS-INDCONF	0.09***	0.38***	0.67***	0.73***	0.82***
MS-RSCONF	0.06	0.11	0.18	0.11	0.10
MS-STOXX	0.00	0.21***	0.48***	0.70***	0.72***
MS-GDP	-0.04	-0.12	-0.02	-0.07	-0.02
MS-IP	0.05**	0.15*	0.19*	0.08	0.04
MS-CARS	0.13***	0.20***	0.25***	0.24***	0.09
MS-UN	0.06**	-0.00	0.05	0.11	0.03
MSP-All	0.00	0.10	0.39***	0.58***	0.28***
MSP-SI	0.00	0.15	0.38***	0.48*	0.68***
MSP-BICC	0.05	-0.02	0.03	0.08	0.13*
Google Trends	0.64***	0.58***	0.41***	0.15*	0.17**

Notes: The Peirce skill score (PS) is calculated as the difference between the rate of *Hits* and the corresponding rate of *False alarms* for the binary event classifier w . The evaluation sample ranges from January 2004 until December 2013. For the model abbreviations see the notes to Figure 4.3.

The entries in Table 4.4 reveal the usefulness of the alternative recession signals for selected binary event classifiers w and could give valuable guidance to practitioners. For example, the MFBVAR achieves the highest significant PS score for a very small threshold, while the other high-performing models, i.e. those for the indices of the stock market and industry confidence and the pool of models for the sentiment indices, perform best for the common threshold of $w = 0.5$. By contrast, the PS score for the Markov-switching models for the confidence index in retail sales, real GDP, indus-

trial production and the unemployment rate is not statistically different from zero, indicating that these models do not have any discriminatory skill for the real-time detection of recession periods. For the remaining approaches, the threshold for which they are most useful varies. For the QFBVAR and the Google Trends indicator, a low threshold works best, while the Markov-switching pools and the univariate models for new passenger car registrations yield the highest *PS* score for an intermediate threshold. Overall, these results illustrate that the absolute values of the real-time recession signals issued by the alternative approaches should be interpreted very carefully. In particular, signals obtained with different approaches which are equal in absolute value cannot necessarily be interpreted as being equally strong.

Table 4.5: Contingency table, MFBVAR and INDCONF.

a) MFBVAR, $w = 0.1$		
	$bc_t = 1$	$bc_t = 0$
$\widehat{bc}_t = 1$	30	14
$\widehat{bc}_t = 0$	3	73
b) INDCONF, $w = 0.5$		
	$bc_t = 1$	$bc_t = 0$
$\widehat{bc}_t = 1$	32	13
$\widehat{bc}_t = 0$	1	74

Notes: The entries display the number of *Hits* ($\widehat{bc}_t = bc_t = 1$), *False Alarms* ($\widehat{bc}_t = 1, bc_t = 0$), *Misses* ($\widehat{bc}_t = 0, bc_t = 1$) and *Correct rejections* ($\widehat{bc}_t = bc_t = 0$) over the period from January 2004 until December 2013. $bc_t = 1$ ($bc_t = 0$) denotes a recession (expansion) month. $\widehat{bc}_t = 1$ ($\widehat{bc}_t = 0$) denotes a recession (expansion) forecast. For the model abbreviations see the notes to Figure 4.3.

According to the entries in Table 4.4 the MFBVAR achieves a maximum *PS* score that is slightly lower than that of the univariate Markov-switching model for the industry confidence index. To understand the significance of this difference, it might be warranted to take a closer look at the contingency table for both models. Table 4.5 shows the contingency table for the threshold value w that maximizes the *PS* scores of the two models. The columns contain the number of months in which a recession did and did not occur in the period from January 2004 until December 2013, and the rows display

the number of times these events were forecast. Out of a total of 33 recession months in the sample, the MFBVAR misses three months, while the industry confidence index model misses only one. Moreover, for the 87 non-recession months the MFBVAR issues one more *False Alarm* than the model for the industry confidence index. Given the high uncertainty surrounding any attempt to precisely determine the start and end months of a recession, these differences seem to be rather small.

It should nonetheless be noted that, depending on the purpose of the forecast and the forecaster's loss function, missed recessions and *False Alarms* could actually be very costly and even small differences between alternative approaches might be very relevant in practice. It is also important to be aware of the characteristics of the alternative evaluation procedures. For example, all formal measures used above treat *Misses* and *False Alarms* symmetrically, which might be inappropriate when the economic costs of the two differ (see Knedlik, 2014, for an application with asymmetric weights). As pointed out by Lahiri and Wang (2013), for known economic costs of these two types of errors, decision theoretic frameworks could be used to derive binary real-time recession signals. However, this goes beyond the scope of this paper.

In summary, the results of the formal evaluation presented in this section indicate that the MFBVAR and the Markov-switching model for the confidence index in industry deliver the most accurate real-time recession signals in terms of calibration, sharpness and discriminatory skill, and that they perform more or less equally well.

4.7 Robustness

In the following robustness analysis, I first investigate whether the performance of the MFBVAR described in section 2.4 is robust with respect to two aspects, namely the variables that are included in the model and the recession definition used to compute the MFBVAR real-time recession probabilities. Second, I assess the extent to which the performance of the linear opinion pools can be improved by applying the beta transform proposed by Ranjan and Gneiting (2010) and described in section 4.4.3.

Regarding the variables included in the estimation of the MFBVAR, the results of section 2.4 suggest that the sentiment indices could be crucially im-

portant for the accuracy of the model's real-time recession signals. To verify this conjecture, I estimate a version of the model that excludes the Economic Sentiment index and the confidence indices in construction and retail sales ($MFBVAR(9)$) from the list of variables in Table 4.2. Additionally, I assess the performance of a version of the MFBVAR that includes only the most useful sentiment index, i.e. the confidence index for the industry sector, and quarterly real GDP ($MFBVAR(2)$).

To assess the robustness of the performance of the MFBVAR with respect to the recession definition used to extract real-time recession probabilities from the model's predictive distribution, I consider the following alternatives. First, I assess the performance of MFBVAR real-time recession signals based on weak real GDP growth rates below 0.1%, rather than below zero, over a sequence of at least two consecutive quarters ($MFBVAR^{\text{slow growth}}$). This might increase the timeliness of the model's real-time recession signals, since at the onset of most recessions, the GDP decline is often quite muted (see panel (b) of Figure 4.1). Second, I consider a multivariate approach that is closer to the recession definition used by the CEPR (Artis et al., 2003), see also section 4.2, and which, among others, also requires monitoring employment conditions and the state of the industrial sector in the euro area. In particular, I use the MFBVAR's joint predictive distribution of real GDP, industrial production, the unemployment rate, the Economic Sentiment index, the confidence index in industry and the stock market index to compute the odds that several of these variables simultaneously signal a deterioration in economic conditions. I include the soft indicators so as not to run the risk of the recession signals being delayed. However, since my previous results suggest that the soft indicators are prone to delivering *False alarms*, I define this alternative recession criterion as the odds that four out of the six above-listed indicators deteriorate repeatedly for at least three consecutive months ($MFBVAR^{4/6}$). As a final alternative, I use the MFBVAR estimate for latent monthly real GDP to assess recession risks in real-time, which I define as the odds that real GDP growth in the current month is part of a three-month sequence of negative growth rates ($MFBVAR^{\text{monthly GDP}}$).

Table 4.6 presents the results of my robustness analysis for the performance of the MFBVAR, while the corresponding real-time recession signals are depicted in Figure 4.6 in appendix A 3. The first row contains the results of the benchmark model version as presented in section 2.4, while the results for

the model versions that include alternative sets of indicators are presented in the middle part of the table. Finally, the last rows contain the results for the benchmark model including all variables listed in Table 4.2, but here, the aforementioned alternative recession definitions were used for the computation of the real-time recession signals.

Table 4.6: Robustness of MFBVAR performance.

	QPS	ROC	PS*
MFBVAR	0.11	0.06	0.75 ($w = 0.1$)
MFBVAR(2)	0.11	0.10	0.67 ($w = 0.3$)
MFBVAR(9)	0.15***	0.13	0.59 ($w = 0.2$)
MFBVAR ^{slow growth}	0.09**	0.06	0.72 ($w = 0.5$)
MFBVAR ^{4/6}	0.13	0.04	0.67 ($w = 0.1$)
MFBVAR ^{monthly GDP}	0.12	0.06	0.69 ($w = 0.2$)

Notes: The evaluation sample ranges from January 2004 until December 2013. QPS: quadratic probability score, ROC: receiver operating characteristic, PS*: maximum Peirce skill score obtained with the binary event classifier w . ***(**,*) denote that the QPS is significantly different from the QPS for the MFBVAR at the 1% (5%,10%) level. See the text for the model abbreviations.

According to the entries in Table 4.6, the performance of the model clearly deteriorates when all sentiment indices are excluded. The MFBVAR(9) achieves a significantly larger *QPS* than the benchmark, while the *ROC* score and the maximum *PS* score both decrease. The most important variable, however, turns out to be the confidence index in industry. The MFBVAR(2) that only includes this index in addition to real GDP achieves the same *QPS* as the benchmark model, and only slightly higher *ROC* and maximum *PS* scores.¹¹

Regarding the alternative conditions for defining real-time recession risks, only the version based on slow growth rather than negative growth rates can significantly improve upon the benchmark in terms of the *QPS*, while the approaches based on the joint deterioration of four monthly indicators or the estimate for latent monthly real GDP perform just as well. Moreover, the differences between the three alternatives in terms of *ROC* scores and

¹¹An evaluation of the point and density forecasts of the different model versions can be found in appendix A 1.

maximum *PS* scores are very small. Hence, the performance of the MFBVAR real-time recession signals turns out to be quite robust with respect to the exact definition of real time recession risks.

Finally, it remains to be investigated whether the performance of the Markov-switching linear equal-weight pools can be improved upon by applying the beta transformation proposed by Ranjan and Gneiting (2010) and described in section 4.4.3. The upper part of Table 4.7 repeats the results from section 2.4 for the linear equal-weight pools, while the lower part of Table 4.7 presents the results for the ex post best-performing beta transformed pools in terms of the *QPS* score. These were obtained through a grid search over alternative values for α , the parameter of the beta distribution. The real-time recession signals of the original and the optimally transformed are depicted in Figure 4.6 in appendix A 4.

As it turns out, the optimal beta transform can improve the performance of the alternative pools relative to the untransformed linear equal-weight pool, especially in terms of calibration. However, the overall gains in accuracy are rather small.

Table 4.7: Robustness of the performance of the linear opinion pools.

	QPS	CAL	SHARP	ROC	PS*
MSP-All	0.168**	0.057	0.056	0.16	0.58 ($w = 0.4$)
MSP-SI	0.154	0.100	0.090	0.09	0.68 ($w = 0.5$)
MSP-BICC	0.215***	0.053	0.002	0.45	0.13 ($w = 0.5$)
MSP-All, $\alpha^* = 3.4$	0.151*	0.028	0.060	0.16	0.57 ($w = 0.1$)
MSP-SI, $\alpha^* = 6.1$	0.128	0.066	0.093	0.10	0.68 ($w = 0.5$)
MSP-BICC, $\alpha^* = 1$	0.215***	0.053	0.002	0.45	0.13 ($w = 0.5$)

Notes: See the notes to Table 4.6. CAL: calibration score, SHARP: sharpness score. MSP: combination of probabilistic forecasts from \hat{A} univariate Markov-switching models for ALL: all univariate MS models, SI: the Economic Sentiment index, the confidence indices in industry and retail sales and the stock market index, BCCI: industrial production, the unemployment rate and new passenger car registrations. Upper part of the table: linear pools, lower part: optimal beta transform.

Moreover, in comparison to the MFBVAR, the optimal pool that includes all univariate Markov-switching models and the optimal BCCI pool in the spirit of Anas et al. (2008) perform significantly worse still, while for the optimal pool of models for the sentiment indices, there is no statistically significant difference.

4.8 Conclusion

The evidence presented in this paper shows that the predictive distribution of the back-, now- and forecasts obtained with a linear mixed-frequency Bayesian VAR (MFBVAR) can be used to extract very accurate monthly real-time recession signals for the euro area. Evaluated over the period from January 2004 until December 2013, the probabilistic real-time recession forecasts of the MFBVAR outperform those obtained with the univariate regime-switching models for a number of hard and soft monthly economic indicators, their linear combinations and a real-time recession index obtained with Google Trends data as measured by the quadratic probability score, the receiver operating characteristic and the Peirce skill score. Only the univariate Markov-switching model for the confidence index in industry delivers real-time recession signals that are more or less as accurate as those of the MFBVAR.

The real-time recession signals obtained with the remaining soft indicators, namely the Economic sentiment index and the indicator for confidence in retail sales, are very poorly calibrated and yield a high number of recession signals in non-recession periods. This could suggest that these variables are possibly driven by more than just economic fundamentals. The hard economic indices considered here, i.e. industrial production and quarterly real GDP growth, perform particularly poorly in terms of sharpness and thus have no discriminatory skill to separate recessions from periods of economic expansion in real-time. The reason for this is most likely the long publication lag of the respective data. The Google Trends real-time recession index, which is obtained as a population-weighted mean of the query shares for the word "recession" in the eleven largest euro area countries, accurately signals the beginning of the Great Recession in 2008-09 and the end of the recession related to the European debt crisis in 2011-13. However, in between these two recession periods, its signals are not very clear-cut, possibly on account

of the uncertainty prevailing at that time over the occurrence of a double-dip recession.

The robustness analysis indicates that the inclusion of the confidence index in industry is crucial for the good performance of the MFBVAR. Moreover, the results show that considering a multivariate recession definition that requires the joint monitoring of several soft and hard economic indicators does not significantly increase the accuracy of the MFBVAR's real-time recession signals compared to the benchmark case, where only the evolution of real quarterly GDP growth is assessed.

Further, the robustness analysis provides evidence that the performance of the linear opinion pools of the probabilistic forecasts of the various univariate regime-switching models can be increased if an optimal beta transformation as suggested by Ranjan and Gneiting (2010) is applied. However, even the ex-post optimized pools are not more accurate than the MFBVAR or the univariate Markov-Switching model for the confidence index in industry.

Finally, the findings illustrate that the absolute values of the real-time recession signals issued by the alternative approaches should be interpreted very carefully. The size of the optimal binary event classifier used to translate the probabilistic recession forecasts into binary signals for the occurrence or non-occurrence of a recession varies considerably. In particular, the widely used threshold of 0.5 turns out to be suboptimal in many cases.

References

- Aastveit, K. A., A. S. Jore, and F. Ravazzolo (2014). Forecasting recessions in real time. *Norges Bank Working Papers* 02/2014.
- Akerlof, G. and R. Shiller (2008). *Animal Spirits*. Princeton University Press.
- Anas, J., M. Billio, L. Ferrara, and G. L. Mazzi (2008). A system for dating and detecting turning points in the euro area. *The Manchester School* 76, 549 – 577.
- Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area GDP growth. *The Econometrics Journal* 14(1), 25 – 44.
- Artis, M., F. Canova, J. Gali, F. Giavazzi, R. Portes, L. Reichlin, H. Uhlig, and P. Weil (2003). Business Cycle Dating Committee of the Centre for Economic Policy Research, Findings 22 September 2003.
- Askatas, N. and K. Zimmermann (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly* 55, 107–120.
- Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy* 4(2), 1 – 27.
- Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting* 20(3), 447 – 460.
- Bañbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Bañbura, M. and G. Rünstler (2011). A look into the factormodel black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting* 27(2), 333 – 346.
- Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20(4), 451–468.
- Bellégo, C. and L. Ferrara (2009). Forecasting euro-area recessions using time-varying binary response models for financial variables. *Banque de France Working Paper* 259.

- Billio, M., R. Casarina, F. Ravazzolo, and H. K. van Dijk (2012). Combination schemes for turning point predictions. *The Quarterly Review of Economics and Finance* 52, 402 – 412.
- Boysen-Hogrefe, J., N. Jannsen, M. Plödt, and T. Schwarzmüller (2015). An empirical evaluation of macroeconomic surveillance in the european union. *Kiel Working Papers, No. 2014*.
- Bry, G. and C. Boschan (1971). Cyclical analysis of time series: Selected procedures and computer programs. *NBER Technical Paper 20*.
- Camacho, M., G. P. Quiros, and P. Poncela (2014). Green shoots and double dips in the euro area: A real time measure. *International Journal of Forecasting* 30(3), 520 – 535.
- Carter, C. and P. Kohn (1994). On gibbs sampling for state space models. *Biometrika* 81, 541 – 533.
- Chauvet, M. and J. Piger (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics* 26(1), 42 – 49.
- Choi, H. and H. Varian (2009a). Predicting initial claims for unemployment benefits. *Economics Research Group, Google..*
- Choi, H. and H. Varian (2009b). Predicting the present with google trends. *Economics Research Group, Google..*
- Clements, M. and D. Harvey (2011). Combining probability forecasts. *International Journal of Forecasting* 27(2), 208 – 223.
- Concha, A. and E. Galán (2012). Tracking the future on the web: construction of leading indicators using internet searches. *Documentos Ocasionales N.º 1203, Banco d'España*.
- De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory* 47(2-3), 423–457.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.

- Doms, M. and N. Morin (2004). Consumer sentiment, the economy, and the news media. *FRBSF Working Paper 2004-09*.
- Dovern, J. and F. Huber (2015). Global prediction of recessions. *Economics Letters 133*, 81–84.
- ECB (2008). Short-term forecast of economic activity on the euro area. *ECB Monthly Bulletin April*, 69 – 74.
- Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model (AWM) for the euro area. *ECB working paper No. 42*.
- Fantazzini, D. and Z. Toktamysova (2015). Forecasting german car sales using google data and multivariate models. *MPRA Paper No. 67110*.
- Froni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting 30(3)*, 554 – 568.
- Geweke, J. and C. Whiteman (2006). *Handbook of Economic Forecasting*, Volume 1, Chapter 1 - Bayesian Forecasting, pp. 3–80. Elsevier Publications.
- Giannone, D., J. Henry, M. Lalik, and M. Modugno (2010). An area-wide real-time database for the euro area. *ECB Working Paper Series No. 1145*.
- Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55(4)*, 665 – 676.
- Ginsberg, J., M. M. R. P. L. B. M. S. and L. Brilliant (2009). Detecting influenza epidemics using search engine query data. *Nature 457*, 1012–1014.
- Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series: B 69(2)*, 243 – 268.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57(2)*, 357 – 384.
- Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting 27(4)*, 1006 – 1026.

- Harding, D. and A. Pagan (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics* 49, 365–381.
- Iselin, D. and B. Siliverstovs (2013). The r-word index for switzerland. *Applied Economics Letters* 20(11), 1032–1035.
- Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* 12(2), 99–132.
- Karlsson, S. (2013). *Handbook of Economic Forecasting*, Volume 2, Chapter 15 - Forecasting with Bayesian Vector Autoregression, pp. 791–897. Elsevier Publications.
- Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. MIT Press.
- Knedlik, T. (2014). The impact of preferences on early warning systems - the case of the european commission's scoreboard. *European Journal of Political Economy* 34(C), 157–166.
- Kuzin, V., M. Marcellino, and C. Schumacher (2011). Midas vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting* 27(2), 529 – 542.
- Lahiri, K. and J. G. Wang (2013). Evaluating probability forecasts for gdp declines using alternative methodologies. *International Journal of Forecasting* 29, 175–190.
- Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics* 4(1), 25–38.
- Lo, M. C. and J. M. Piger (2005). Is the response of output to monetary policy asymmetric? Evidence from a regime-switching coefficients model. *Journal of Money, Credit, and Banking* 37(5), 865 – 886.
- Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting* 20(2), 87–109.

- Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics* 72(4), 518 – 550.
- Mayr, J. and J. Grossarth-Maticek (2008). Medienberichte als konjunkturindikator. *ifo Schnelldienst* 07, 17–29.
- Nalewaik, J. J. (2012). Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money, Credit and Banking* 44(1), 235 – 253.
- Österholm, P. (2012). The limited usefulness of macroeconomic bayesian vars when estimating the probability of a U.S. recession. *Journal of Macroeconomics* 34(1), 76 – 86.
- Ranjan, R. and T. Gneiting (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(1), 71–91.
- Reinhart, C. M. and G. L. Kaminsky (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review* 89(3), 473–500.
- Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business and Economic Statistics*, (forthcoming).
- Schreiber, S. (2014). Anticipating business-cycle turning points in real time using density forecasts from a VAR. *Free University Berlin, School of Business & Economics Discussion Papers* 2.
- Schumacher, C. (2014). MIDAS and bridge equations. *Deutsche Bundesbank Discussion Paper* 26.
- Schumacher, C. and J. Breitung (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data. *International Journal of Forecasting* 24, 368 – 398.
- Schwarz Müller, T. (2015). Model pooling and changes in the informational content of predictors: an empirical investigation for the euro area. *Kiel Working Papers* 1982.
- Sims, C. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review* 39(4), 949–968.

- Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41(3), 788–829.
- Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23(6), 405–430.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 135–196. Amsterdam: North Holland.
- Vosen, S. and T. Schmidt (2011). Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting* 30(6), 565–578.
- Yan, C.-S. and F. Labbé (2013). Nowcasting with google trends in an emerging market. *Journal of Forecasting* 32, 289 – 298.

Appendix

A 1: Evaluation of MFBVAR and QFBVAR GDP Growth Point and Density Forecasts

The following reports the results of the evaluation of the real GDP growth predictions provided by the MFBVAR and the QFBVAR. The evaluation period ranges from the first quarter of 2004 until the fourth quarter of 2013. To compute the different evaluation measures, the first release of euro area GDP growth is used to capture the real-time environment at the time the forecast was made. In all tables, $h^* = 1$ refers to the one-quarter-ahead forecast, while $h^* = 0$ and $h^* = -1$ denote the nowcast and the backcast, respectively. IS denotes the information set, i.e. the month of the current quarter, in which the back-, now- and forecast is computed.

The accuracy of the point forecasts is assessed with the mean squared forecast error, which is given as

$$MSFE = \frac{1}{T^*} \sum_{t=1}^{T^*} (y_t - \hat{y}_t)^2, \quad (4.17)$$

where y_t is the realized value of variable y in period t and \hat{y}_t denotes the respective point forecast.

The accuracy of the density forecasts is assessed with several measures. The first is the logarithmic score, which is given as

$$LS = \frac{1}{T^*} \sum_{t=1}^{T^*} -\log(F(y_t)), \quad (4.18)$$

where $F(\cdot)$ denotes the predictive distribution.

The continuous ranked probability score (CRPS) is given as the average over

$$CRPS_t = \int_{-\infty}^{\infty} \{F(u) - \mathcal{I}(u \geq y_t)\}^2 du \quad (4.19)$$

or

$$CRPS_t = E_F|Y - y_t| - \frac{1}{2}E_F|Y - Y'|, \quad (4.20)$$

where E_F denotes the respective expectations operator and Y and Y' are random draws from the models' predictive cumulated distribution $F(\cdot)$ (see, for example, Gneiting et al., 2007).

Note that all evaluation measures reported in the following are negatively oriented, i.e. the smaller the score, the better.

Table 4.8: RMSFE.

IS	1	2	3	1	2	3	1	2
	$h^* = 1$			$h^* = 0$			$h^* = -1$	
MFBVAR	0.51	0.32	0.36	0.26	0.25	0.19	0.16	0.12
MFBVAR(2)	0.45	0.37	0.34	0.30	0.32	0.24	0.24	0.28
MFBVAR(9)	0.48	0.34	0.41	0.34	0.27	0.26	0.19	0.14
QFBVAR	0.45	0.45	0.43	0.42	0.42	0.28	0.29	0.29

Table 4.9: Log score.

IS	1	2	3	1	2	3	1	2
	$h^* = 1$			$h^* = 0$			$h^* = -1$	
MFBVAR	1.26	1.14	1.13	1.08	1.07	1.03	1.00	0.97
MFBVAR(2)	1.22	1.16	1.13	1.11	1.12	1.07	1.05	1.06
MFBVAR(9)	1.26	1.16	1.16	1.12	1.08	1.05	1.01	0.98
QFBVAR	1.17	1.18	1.16	1.16	1.16	1.09	1.06	1.06

Table 4.10: CRPS.

IS	1	2	3	1	2	3	1	2
	$h^* = 1$			$h^* = 0$			$h^* = -1$	
MFBVAR	0.35	0.28	0.29	0.25	0.24	0.22	0.20	0.18
MFBVAR(2)	0.33	0.29	0.28	0.27	0.28	0.25	0.25	0.26
MFBVAR(9)	0.34	0.28	0.31	0.27	0.24	0.23	0.21	0.19
QFBVAR	0.30	0.30	0.30	0.30	0.30	0.26	0.26	0.26

A 2: MFBVAR Prior Specification

For the parameters of the VAR A_1, \dots, A_p, C and Σ , I implement a normal inverse Wishart prior that retains the main principles of the widely used Minnesota prior (Kadiyala and Karlsson, 1997; Litterman, 1986). This prior implies that A_1, \dots, A_p are assumed to be a priori independently and normally distributed, while with respect to the constant C the prior is assumed to be diffuse. The residual covariance matrix Σ is assumed to a priori follow an inverse Wishart distribution with scale matrix \bar{S} and degrees of freedom $\bar{\alpha}$.

One of the main principles of the Minnesota prior is to center each equation of the VAR around a random walk with drift. Thus, the prior mean for A_1, \dots, A_p is specified as:

$$E[(A_\ell)_{ij}] = \begin{cases} 1 & \text{for } i = j, \ell = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.21)$$

Moreover, the prior also incorporates the belief that more recent lags of a variable should provide more reliable information for the estimation than less recent lags. The zero coefficient prior on more recent lags is therefore not imposed as tightly as on less recent lags. This is captured by specifying the prior variance as

$$\text{Var}[(A_\ell)_{ij}] = \begin{cases} \frac{\lambda^2}{\ell^2} & \text{for } i = j \\ \frac{\lambda^2 \sigma_i^2}{\ell^2 \sigma_j^2} & \text{otherwise,} \end{cases} \quad (4.22)$$

where $\ell = 1, \dots, p$ is the lag length, $\lambda = 0.2$ is a hyperparameter governing the importance of the prior beliefs relative to the data, and σ_i/σ_j is a scale parameter adjusting the prior for the different scale and variability of the data. For the implementation, σ_i is set equal to the standard deviation of the residuals of a simple univariate autoregression for each variable.

Following Schorfheide and Song (2015) I augment the prior outlined above to constrain the sum of coefficients of the VAR (Sims and Zha, 1998) as well as to incorporate the belief that the variables in the VAR follow a common stochastic trend.

I implement the prior outlined above using dummy variables Y^D and X^D , which are given as

$$Y^D = \begin{pmatrix} \frac{\text{diag}(\sigma_1 \dots \sigma_n)}{\lambda} \\ 0_{n(p-1) \times n} \\ \text{diag}(\sigma_1 \dots \sigma_n) \\ 0_{1 \times n} \\ \text{diag}(\mu_1 \dots \mu_n) \gamma \\ \eta \mu_1 \dots \eta \mu_n \end{pmatrix}, X^D = \begin{pmatrix} \frac{\text{diag}(1 \dots p) \otimes \text{diag}(\sigma_1 \dots \sigma_n)}{\lambda} & 0_{n \times 1} \\ 0_{n \times n} & 0_{n \times 1} \\ 0_{1 \times n} & c \\ P \otimes \text{diag}(\mu_1 \dots \mu_n) \gamma & 0_{n \times 1} \\ (P' \otimes \text{diag}(\mu_1 \dots \mu_n) \eta)' & \eta \end{pmatrix}, \quad (4.23)$$

where P is a $(1 \times p)$ matrix of ones, $c = 10^{-4}$ reflects the diffuse prior for the constant C , μ_1, \dots, μ_n are the variable means and $\gamma = 10^3 \lambda$ and $\eta = \lambda$ govern the tightness of the sum of coefficients constraint and the common stochastic trend prior, respectively.

With these dummy variables, the moments of the prior distributions for the parameters A_1, \dots, A_p, C and the residual covariance matrix Σ can be computed as

$$E[(A_1, \dots, A_p, C)] = \bar{A} = (X^{D'} X^D)^{-1} X^{D'} Y^D, \quad (4.24)$$

$$\text{Var}[(A_1, \dots, A_p, C)] = \bar{V} = (X^{D'} X^D)^{-1}, \quad (4.25)$$

$$\bar{S} = (Y^D - X^D \bar{A})' (Y^D - X^D \bar{A}) \quad (4.26)$$

and

$$\bar{\alpha} = T^D - n(p-1) - 1, \quad (4.27)$$

where T^D is the number of rows of Y^D .

Conditional on the most recent Gibbs draw i of the state vector Z_t^{i-1} , A_1^i, \dots, A_p^i, C^i and Σ^i are sampled from their respective posterior distributions. In particular, the (A_1, \dots, A_p, C) follow a multivariate t-distribution

with mean \tilde{A} , covariance matrix $\tilde{S} \otimes \tilde{V}$ and degrees of freedom $\tilde{\alpha}$, while $\Sigma \sim IW(\tilde{S}, \tilde{\alpha})$. The respective moments of these distributions can also be computed using the dummy variables outlined in equation (4.23). In particular, augment the state vector Z_t^{i-1} with the dummy variables to obtain $Y^* = [(Z_t^{i-1})', Y^{D'}]'$ and $X^* = [(Z_{t-1}^{i-1})', X^{D'}]'$. Then

$$\tilde{A} = (X^* X^*)^{-1} X^* Y^*, \quad (4.28)$$

$$\tilde{V} = (X^* X^*)^{-1}, \quad (4.29)$$

$$\tilde{S} = (Y^* - X^* \tilde{A})'(Y^* - X^* \tilde{A}) \quad (4.30)$$

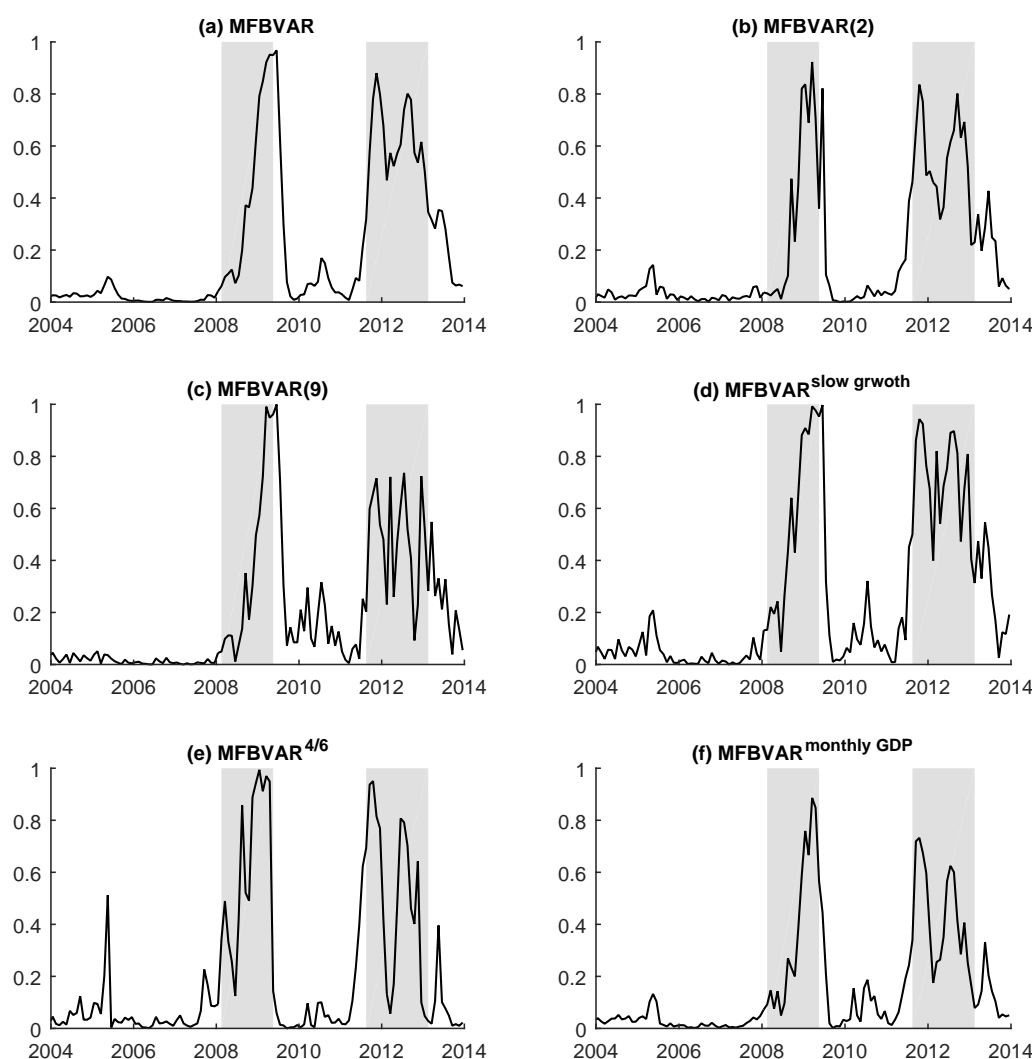
and

$$\tilde{\alpha} = T^* - np + 1, \quad (4.31)$$

where T^* is the number of rows of Y^* .

A 3: Alternative MFBVAR Real-Time Recession Probabilities

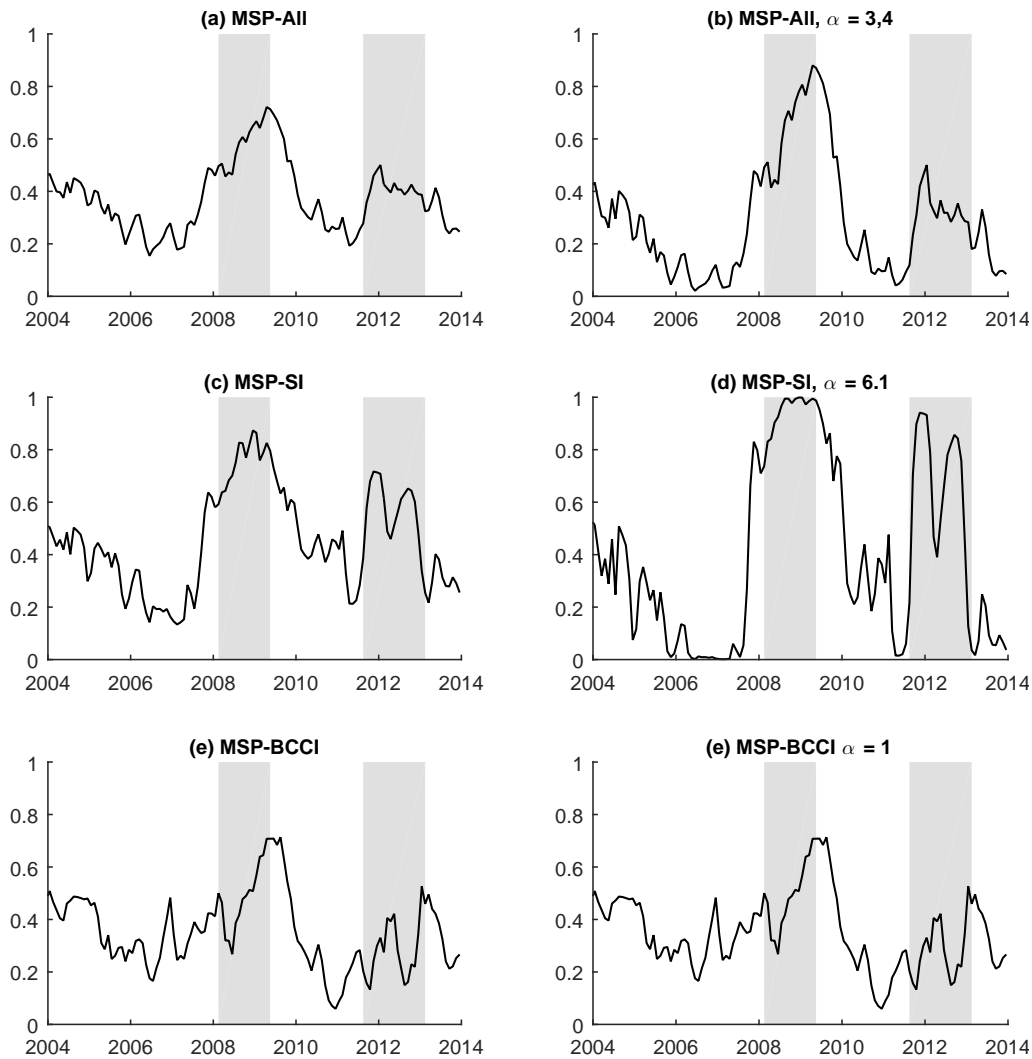
Figure 4.5: Real-time recession signals for the euro area, MFBVAR robustness.



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas denote euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. For the model abbreviations see section 4.7.

A 4: The Beta Transformed Linear Opinion Pools

Figure 4.6: Real-time recession signals for the euro area, Pool robustness.



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas denote euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. For the model abbreviations see section 4.7.

Affirmation/Eidesstattliche Erklärung

I hereby affirm that I have completed my doctoral thesis entitled "Essays on Preferences and Nominal Rigidities and on Macroeconomic Forecasting" entirely on my own and unassisted and that I have contributed substantially to the academic papers this thesis is based on. I have specially marked all of the quotes I have used from other authors as well as those passages in my work that are extremely close to the thoughts presented by other authors, and listed the sources in accordance with the regulations I have been given.

Ich erkläre hiermit, dass ich meine Doktorarbeit "Essays on Preferences and Nominal Rigidities and on Macroeconomic Forecasting" selbstständig und ohne fremde Hilfe angefertigt habe und dass ich als Koautor maßgeblich zu den weiteren Fachartikeln beigetragen habe. Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.

Inske Pirschel