Empirical Determinants of Inflation

Inaugural-Dissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaft des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von
Rolf Scheufele
aus Ludwigsburg

2010

Erstgutachter:	Prof. Dr. Dieter Nautz
Zweitgutachter:	Prof. Dr. Oliver Holtemöller
Tag der Disputation:	03. Mai 2011

Acknowledgments

This dissertation thesis is the summary of research on inflation issues I conducted at the Department of Macroeconomics at the Halle Institute for Economic Research (IWH) during the period 2006-2010. First of all I would like to express my gratitude to the former head of department Prof. Dr. Udo Ludwig who introduced me into empirical business cycle issues and gave me the opportunity for writing this thesis. Furthermore, I want to thank the institute for all the support I received within these years. I am very grateful to all members of the Macroeconomics Department in general and, in particular, to Diemo Dietrich, Katja Drechsel, Makram El-Shagi, Sebastian Giesen, Tobias Knedlik and Axel Lindner.

I am also very thankful to my supervisor Prof. Dr. Dieter Nautz for his interest in this work and his invaluable support. A special word of thanks goes to Prof. Jean-Marie Dufour, PhD, for his guidance and for promoting this thesis during his stays at the IWH and by providing me a research stay at McGill University. The members of the dissertation committee, Prof. Dr. Dieter Nautz, Prof. Dr. Oliver Holtemöller, Prof. Dr. Jürgen Wolters and Prof. Dr. Chistian Offermanns are truly acknowledged for their comments on this thesis. Additionally, I benefited from discussions and comments from participants at numerous international workshops and conferences.

I am greatly indebted to my mother and my wife for their continuous support and encouragement.

Rolf Scheufele

Zusammenfassung

Die Phillipskurve wird seit ihrer Einführung von A. W. Phillips Mitte der fünfziger Jahre als das theoretische und empirische makroökonomische Modell zwischen Inflation und realwirtschaftlicher Entwicklung für die kurze Frist angesehen. In ihrer ursprünglichen Form ist die Phillipskurve lediglich ein empirischer Zusammenhang zwischen Inflation und Arbeitslosigkeit. Allerdings wurde sehr schnell versucht, diesen Zusammenhang in der Wirtschaftspolitik auszunutzen. So wurde in der Phillipskurve eine Wahlmöglichkeit zwischen zwei ökonomischen Ergebnissen gesehen (hohe Inflation und niedrige Arbeitslosigkeit oder niedrige Inflation und hohe Arbeitslosigkeit). Seit den bahnbrechenden Arbeiten von Friedman und Phelps in den späten sechziger Jahren und der Einbeziehung von rationalen Erwartungen durch Lucas, gehört der Glaube an einen mechanistischen Trade-off zwischen Arbeitslosigkeit und Inflation der Vergangenheit an.

Trotz zwischenzeitlich verstärkter Kritik hat die Phillipskurve in den letzten Jahren eine Renaissance in der modernen Makroökonomik erfahren. Als Neu-Keynesianische Phillipskurve (NKPC) ist sie ein unverzichtbarer Bestandteil der Neu-Keynesianischen Makroökonomik. Vergleichbar mit traditionellen Spezifikationen der Phillipskurve, setzt die NKPC Inflation in Beziehung zu realwirtschaftlicher Aktivität. Auch Erwartungen, wie bereits von Friedman und Phelps proklamiert, spielen eine wichtige Rolle für die Inflationsmodellierung. Ein zusätzliches Charakteristikum der NKPC ist, dass ihr theoretischer Zusammenhang direkt aus dem Optimierungskalkül von Unternehmen unter den Annahmen von monopolistischer Konkurrenz und nominalen Preisrigiditäten abgeleitet werden kann. Somit hängt ihre Struktur von „tiefen Parametern“ ab und kann aus diesem Grund auch für Politikanalysen verwendet werden (Lucas Kritik).

Die neue Phillipskurve stellt jedoch auch neue Herausforderungen für die empirische Forschung dar. Da die NKPC auch (rationale) Erwartungen zukünftiger Variablen beinhaltet, die grundsätzlich unbeobachtbar sind, sind Standardansätze der Ökonometrie für ihre Untersuchung ungeeignet. Allerdings kann man zukünftige Erwartungen als latente Variable behandeln und mithilfe einer Instrumentvariablenschätzung lassen sich dann die Parameter der Phillipskurve schätzten. Eine wesentliche Annahme bei der Schätzung struktureller makroökonometrischer Beziehungen ist aber die Identifikationsbedingung, die für den Erhalt sinnvoller Schätzerergebnisse und aussagekräftiger Inferenz notwendig ist. Im Rahmen von

Methoden bei beschränkter Information (bspw. Instrumentvariablenschätzer) ist dies gleichbedeutend mit der Fragestellung, ob die verwendeten Instrumente stark genug sind um alle strukturellen Parameter des Modells problemlos zu bestimmen. Dies impliziert, dass bei Vorliegen von Identifikationsproblemen, bedingt durch schwache Instrumente, die Standardverfahren (verallgemeinerte Momentenmethode bzw. Instrumentvariablenschätzer) unzuverlässig sein können und damit alternative Methoden gefragt sind.

Diese Dissertation stellt einen Beitrag zur Analyse und Diskussion neuerer Aspekte der empirischen Modellierung von Inflation dar. Die drei Kapitel dieser Arbeit setzen sich mit der Anwendung und Erweiterung ökonometrischer Methoden zur Beurteilung von theoretischen Inflationsmodellen auseinander, die in engem Zusammenhang mit der Neu-Keynesianischen Phillipskurve stehen. Im ersten Kapitel wird die NKPC für Deutschland empirisch untersucht. Im darauf folgenden Kapitel wird eine Erweiterung des Standardmodells betrachtet und getestet. Abschließend wird in Kapitel 3 untersucht, inwieweit qualitative Inflationseinschätzungen von Unternehmen für die Prognose zukünftiger Inflationsraten verwendet werden können. In den folgenden Absätzen werden die einzelnen Themen dieser Arbeit näher beleuchtet.

In Kapitel 1 wird mit Hilfe eines Schätzansatzes bei beschränkter Information die Neu-Keynesianische Phillipskurve und ihre hybride Erweiterung für Deutschland untersucht. Das Hauptinteresse liegt dabei auf der durchschnittlichen Preissetzungsfrequenz der Unternehmen. Die Grenzkosten der Firmen werden durch die Lohnquote (Verhältnis von Lohnsumme am nominalen Bruttoinlandsprodukt) approximiert. Zusätzlich werden reale Rigiditäten berücksichtigt, die durch einen fixen, firmenspezifischen Kapitalstock modelliert werden. Die verallgemeinerte Momentenmethode (GMM) sowie ein identifikationsrobustes Verfahren, basierend auf der Anderson-Rubin Statistik, werden als Schätzverfahren verwendet. Die Schätzergebnisse zeigen, dass die deutsche Phillipskurve ausschließlich als vorausschauend zu charakterisieren ist. Die Punktschätzungen deuten darauf hin, dass Firmen ihre Preise alle zwei bis drei Quartale neu optimieren. Obwohl diese Schätzungen ökonomisch plausibel sind und im Einklang mit den mikroökonomischen Untersuchen stehen, ist die Schätzunsicherheit sehr groß, so dass auch der Fall perfekter Preisrigidität statistisch nicht verworfen werden kann. Die vorliegende Analyse bietet zudem eine Erklärung für die Tatsache, dass frühere Untersuchungen zur deutschen Phillipskurve zu sehr unterschiedlichen Ergebnissen gekommen sind. So hängen die GMM-Ergebnisse sehr stark von der Formulierung der Orthogonalitätsbedingungen ab. Zusätzlich scheinen Fehlspezifikationen des Modells durch J Tests nicht erkannt zu werden. Darum unterstreicht die Analyse in Kapitel 1 die Bedeutung der Verwendung von identifikationsrobusten Methoden um verlässliche Schätzungen für die Neu-Keynesianische Phillipskurve zu erhalten.

Kapitel 2 geht der Frage nach, ob Zinsänderungen einen Einfluss auf die Grenzkosten von Unternehmen haben und somit auch auf das Preissetzungsverhalten wirken, was wiederum einen Effekt auf die gesamtwirtschaftliche Inflationsrate hätte. Um auf die Existenz dieses sogenannten Cost-Channel-Effekts zu testen wird ein strukturell-ökonometrischer Ansatz gewählt. Die Schätzungen und Inferenz werden mit identifikationsrobusten Verfahren durchgeführt. Es wird gezeigt, dass Identifikationsprobleme für bestimmte Parameter existieren; besonders dann wenn das Gesamtmodell geschätzt wird. Es zeigt sich, dass eine ausschließlich vorausschauende Modellvariante der Phillipskurve, erweitert um einen Zinseffekt, am besten mit den Daten vereinbar ist. Dies impliziert, dass der Cost Channel Effekt eine Relevanz für die Geldpolitik hat.

In Kapitel 3 werden die Eigenschaften von qualitativen Inflationserwartungen für Deutschland untersucht. Zunächst werden die Eigenschaften dieses Indikators bezüglich Rationalität und Granger-Kausalität beschrieben. Anhand einer Out-of-Sample Analyse wird geprüft, inwieweit dieser Indikator geeignet ist zukünftige Inflationsraten zu prognostizieren. Dafür werden alternative Prognosemodelle herangezogen und mit den Prognosen verglichen, die auf Erwartungsmaßen basieren. Die Ergebnisse zeigen, dass Modelle, die qualitative Inflationserwartungen berücksichtigen, alternativen Modelle bezüglich der Prognosegüte zumeist überlegen sind. Außerdem zeigt die Analyse, dass die Umfragedaten bereits die Informationen anderer Modelltypen beinhalten.

Als Schussfolgerung der drei Aufsätze kann festgehalten werden, dass Preiserwartungen eine wichtige Rolle spielen und somit ein wichtiges Element bei der Modellierung von Inflationsbewegungen darstellen. Methodisch wird gezeigt, dass die verallgemeinerte Momentenmethode für die Schätzung dynamischer makroökonomischer Phänomene (wenn ohne weitere Sensitivitätsanalysen angewendet) oftmals zu Fehleinschätzungen führen kann. Nur wenn auf zusätzliche Methoden wie identifikationsrobuste Verfahren, Simulationstechniken und die sorgfältige Auswahl von Momentenbedingungen Wert gelegt wird, kann man robuste und glaubwürdige Ergebnisse erwarten.

List of Original Working Papers

This thesis consists of the following three papers:

1. Rolf Scheufele (2010):
   Evaluating the German (New Keynesian) Phillips Curve
   The North American Journal of Economics and Finance, Vol. 21 (2),
   Special Issue: 50 Years of the Phillips Curve, 145–164.
   http://dx.doi.org/10.1016/j.najef.2009.06.001

2. Rolf Scheufele (2010):
   Does the cost channel matter for inflation dynamics? An identification robust structural econometric analysis,
   Halle Institute for Economic Research, Germany

3. Rolf Scheufele (2011):
   Are Qualitative Inflation Expectations useful to Predict Inflation?
   Journal of Business Cycle Measurement and Analysis, Vol. 2011 (1), 29–53.
   http://dx.doi.org/doi:10.1787/jbcma-2011-5kgg5k52xb5c

Contents

List of Figures

List of Tables

Overview

Since its discovery by A. W. Phillips[1] in the mid 1950s, the Phillips curve has been regarded as the theoretical and empirical macroeconomic link between inflation and real economic activity in the short run. While the original Phillips curve simply reflected the negative correlation between inflation and unemployment, it became a relation used by politicians when choosing between two economic outcomes. However, since Friedman and Phelps's natural rate hypothesis in the late 1960s and Lucas's rational expectation revolution in the 1970s, the belief in a mechanical trade-off between unemployment and inflation has been overcome.

In recent years, the Phillips curve has undoubtedly experienced a great revival in modern macroeconomics. In its current form, it is known as New Keynesian Phillips curve (NKPC), and can be seen as the backbone of New Keynesian Macroeconomics. Like previous Phillips curve specifications, the NKPC relates aggregate inflation dynamics to a measure of real economic activity. Expectations also play a crucial role in the inflation process, thus integrating certain aspects of the Friedman-Phelps-Lucas critics. Most important as far as modern macroeconomics is concerned, the relationship can be derived explicitly from the optimal behavior of firms under the assumption of monopolistic competition and sticky prices. This means that its structure depends on "deep parameters", which makes it suitable for policy analysis (and thus renders it immune to the Lucas critique).

[1] PHILLIPS, A. W. (1958): "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861-1957," *Economica*, 25(100), 283–299.

The new Phillips curve also provides new challenges for empirical research. Since the NKPC includes (rational) expectations of future variables that are generally unobservable to the econometrician, standard techniques turn out to be inappropriate. One approach is to treat future expectations as a latent variable and to use instrumental variables (IV) techniques to estimate the parameters of the NKPC. However, a crucial issue in the estimation of structural macroeconometric relationships is the identification condition necessary for guaranteeing sensible estimates and meaningful inference. In a limited information framework, this translates into the question of whether the instruments employed are strong enough to pin down all the structural parameters of the model. Thus identification difficulties induced by weak instruments may make standard IV methods unreliable and calling for alternative procedures.

This dissertation discusses and analyzes recent issues in the empirics of inflation dynamics. Three different aspects of the NKPC are covered in this volume. In the first chapter, the empirical relevance of the NKPC is investigated for the German economy. The second chapter is devoted to analyzing a recently proposed extension of the NKPC: the cost channel. In Chapter 3, a direct measure of inflation expectations obtained and computed from a qualitative business survey is tested for its ability to forecast future inflation rates. The following paragraphs present the main topics covered in this thesis in more detail.

Chapter 1 evaluates the New Keynesian Phillips curve and its hybrid variant within a limited information framework for Germany. Its main interest resides in the average frequency of price re-optimization by firms. The labor income share is used as the driving variable and real rigidities are considered by allowing for a fixed firm-specific capital stock. A GMM estimation strategy is employed as well as an identification robust method based on the Anderson-Rubin statistic. In this chapter, it is found that the German Phillips curve is purely forward-looking. Moreover, the point estimates are consistent with the view that firms re-optimize prices every two to three quarters. The uncertainties around these estimates are considerable and are also consistent with perfect nominal price rigidity, where firms never re-optimize prices. Further, the analysis offers some explanation as to why previous results for the German NKPC based on GMM differ substantially. First, standard GMM results are very sensitive to the way in which orthogonality conditions are formulated. Further, model mis-specifications may remain undetected by conventional J tests. This analysis points out the need for identification robust methods to get reliable estimates for the NKPC.

Chapter 2 investigates whether interest rate changes impact on firms' marginal costs and whether this has a direct effect on their price setting behavior, translating into aggregate inflation dynamics. Empirical tests for the existence of the cost channel are conducted, using a structural econometric approach. Estimation and inference is conducted using identification robust methods based on the continuous-updating GMM objective function. Identification difficulties are documented for some parameters when estimating the general model structure. For the US, a pure forward-looking interest rate augmented Phillips curve is most compatible with data. This suggests that considering the cost channel is relevant for monetary policy analysis.

Chapter 3 examines the properties of qualitative inflation expectations collected from economic experts for Germany. It describes their characteristics relating to rationality and Granger causality. In an out-of-sample simulation study, this indicator is investigated for its suitability for inflation forecasting. Results from other standard forecasting models are considered and compared with models employing survey measures. It is found that a model using survey expectations outperforms most of the competing models. Moreover, there is evidence that the survey indicator already contains information from other model types (e.g. Phillips curve models). However, the forecast quality may be further improved by taking into account information from some financial indicators.

A conclusion drawn from the essays as a whole makes it evident from the empirical analysis that expectations matter and that they are a key factor in modeling aggregate inflation dynamics. From the point of view of method, it is shown that the blind use of standard GMM procedures for estimating macroeconomic dynamic relationships is often misleading. Only with the use of supplementary techniques, such as identification robust methods, simulation techniques and the careful selection of moment conditions, more robust and reliable results are obtained.

Each of the next three chapters presents one idea as a self-contained unit.

Chapter 1

Evaluating the German (New Keynesian) Phillips curve*

1.1 Introduction

Explaining the evolution of aggregate prices is one of the most prominent issues in empirical macroeconomics. Nowadays, the canonical inflation model is the New Keynesian Phillips curve (NKPC). Like earlier Phillips curve specifications, the NKPC relates price behavior to a measure of real economic activity. However, in contrast to traditional Phillips curves, the NKPC can be derived directly from the optimizing behavior of households and firms, and thus builds on a suitable micro-foundation. The NKPC framework assumes monopolistically competitive firms that face nominal prices rigidities. The standard model of staggered price adjustment by Calvo (1983) has an attractive feature: the coefficients of the NKPC

*All computations used in this Chapter are done in MATLAB programs. Parts of the programs are based on Mike Cliff's computation libraries (`http://www.feweb.vu.nl/econometriclinks/mcliffprogs.html`) and on replication files available on Jim Stock's homepage (`http://www.economics.harvard.edu/faculty/stock`).

depend directly on the average frequency with which prices are adjusted in the economy.

The aim of this chapter is to determine the degree of nominal price rigidity in the German economy. In doing so, we estimate the NKPC and allow for different specifications. We use a generalized version of the model proposed by Christiano, Eichenbaum, and Evans (2005) as our benchmark model, which assumes a dynamic indexation scheme for those firms that do not re-optimize. Further, in spirit of Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001), incorporating "rule-of-thumb" firms. Following Galí, Gertler, and López-Salido (2001) and Sbordone (2002) we allow for some real rigidities derived from the assumption of firm-specific capital.

Empirical studies assessing the degree of nominal price rigidity in the German economy with estimations from Phillips curves are still rare. The primary evidence stems from cross-country comparisons, for example Banerjee and Batini (2004), Benigno and López-Salido (2006), Leith and Malley (2007) or Rumler (2007). In most cases, this evidence is based on GMM estimation, with additional aspects pertaining to an open economy. While the open economy aspect seems to be unimportant for the German Phillips curve (at least according to Banerjee and Batini, 2004; Leith and Malley, 2007), results vary considerably as far as the degree of nominal price rigidity is concerned. The estimated average frequency of price re-optimization ranges from 2.5 quarters (Banerjee and Batini, 2004) to 13 quarters (Leith and Malley, 2007), there is also disagreement on whether the inflation contains a lagged term (through backward-looking behavior) or whether it is purely forward-looking. A more rigorous treatment of nominal price rigidity in Germany is provided by Coenen, Levin, and Christoffel (2007), who focus on the interaction of real and nominal rigidities. Their estimation technique relies on indirect inference methods. They employ a generalized Calvo model, whit their estimates pointing to a frequency of price re-optimization of roughly 2 quarters. Tillmann (2005) considers present-value implications from the Calvo model within a VAR and concludes that a pure forward-looking model is empirically very fragile.

Apart from Phillips curve estimates, some empirical evidence using disaggregated price data on price stickiness exists. The general view of this line of research is that prices appear to be stickier in the Euro area than they are in the U.S. according to a comparison of the average frequency of price chances (see e.g., Álvarez, Dhyne, Hoeberichts, Kwapil, Bihan, Lünnemann, Martins, Sabbatini, Stahl, Vermeulen, and Vilmunen, 2006; Dhyne, Álvarez, Bihan, Veronese, Dias, Hoffmann, Jonker, Lünnemann, Rumler, and Vilmunen, 2006). While prices in the U.S. are, on average, fixed for about 2 quarters, the average price duration in the Euro area is approximately twice that. For Germany, Hoffmann and Kurz-Kim (2006) report similar findings on consumer prices, the average price duration for these being about a year. However, in analyzing data on producer prices, Stahl (2006) reports an average frequency of price adjustment of about seven months, which is closer to the U.S. experience. A potential shortcoming of the available micro studies for Germany is that they cover only a very short time period which makes them less comparable with macro results, which cover a much longer time span.

Our empirical strategy is as follows. We apply a standard GMM method to estimate the structural parameters of the Phillips curve, paying particular attention to the selection of relevant instruments. Since the choice of a particular instrument set can hardly be justified by theory, we propose a statistical criterion for culling out relevant instruments. We then evaluate the robustness of our results with respect to several parameter restrictions, measures of real rigidity and additional lags of inflation. Next, we conduct an identification robust procedure based on a non-linear Anderson–Rubin (AR) statistic (following Ma, 2002; Mavroeidis, 2006) and compare these results with those obtained from the standard GMM estimation. We contribute to this line of research by applying identification robust estimation techniques to estimate the German NKPC. As long as there are weak instrument problems present, the two procedures should display quite different results. In this case, the GMM results are generally unreliable (see e.g., Stock and Wright, 2000).

For a given economically plausible degree of real rigidity, the estimates of the frequency of price re-optimization point to about 2.5 quarters. But this estimate is surrounded by a considerable degree of uncertainty, since the confidence intervals for this estimate are very large. Unless we avoid restricting other parameter values, the estimated degree of nominal rigidity is consistent with both a very low degree of price stickiness and a situation in which prices are never re-optimized (perfect price rigidity). This also casts doubt on the proxy of marginal cost, the labor share as the driving variable of inflation (a finding that is also obtained by Mavroeidis, 2006;

Kleibergen and Mavroeidis, 2009, for the U.S.). Moreover, we find that backward-looking behavior is unimportant in explaining the German inflation process, so a purely forward-looking specification is more appropriate. The identification robust procedure indicates some problems with the orthogonality conditions not detected by the conventional J statistic.

This chapter is organized as follows. We first present our basic model framework in Section 1.2. Then we turn to the econometric strategies for estimating and testing the different model specifications (Section 1.3). In Section 1.4 we discuss our data set and how we obtain the instrument set. Next, we present our econometric results (Section 1.5). Finally, we draw some conclusions in Section 1.6.

1.2 The modeling framework

This section presents the basic theoretical framework, which includes monopolistically competitive goods markets and price stickiness. These are the two key elements in modern macroeconomic models used for analyzing monetary policy. This model structure attempts to ensure consistency with the behavior of optimizing economic agents. Here, we are interested mainly in the price-setting behavior of firms in order to derive an expression for aggregate inflation. We therefore assume random price contracts due to Calvo (1983), which is now standard in many macroeconomic models (e.g., Christiano, Eichenbaum, and Evans, 2005; Smets and Wouters, 2003). However, we deviate from the standard Calvo model in assuming that capital is firm-specific and subject to a form of real rigidity, so that capital cannot be instantaneously reallocated and is thus a predetermined factor.[1]

1.2.1 The market structure

As is standard in New Keynesian models, we assume a monopolistic competitive environment with a continuum of firms indexed by $i \in [0, 1]$. Each firm i produces a differentiated good $Y_t(i)$ according to a Cobb–Douglas technology

[1] Here we follow Galí, Gertler, and López-Salido (2001) and Sbordone (2002). See also Eichenbaum and Fisher (2007) for a more rigorous treatment of real rigidities in the Calvo price-setting framework

$$Y_t(i) = A_t\overline{K}_t(i)^{\alpha}N_t(i)^{1-\alpha}, \tag{1.1}$$

where A_t is a common country-wide technological factor, $\overline{K}_t(i)$ is the (fixed) firm-specific capital stock and $N_t(i)$ is the labor factor employed by firm i.

Each firm i is faced with a demand function with a constant elasticity of substitution that is given by

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} Y_t, \tag{1.2}$$

where Y_t is aggregate output (which equals aggregate demand), P_t is the aggregate price level in the economy and $P_t(i)$ is the price charged by firm i for good $Y_t(i)$. The price elasticity of demand for good i is equal to ϵ (with $\epsilon > 1$).[2]

Without any price frictions the price of the differentiated good is set as a constant mark-up over nominal marginal costs

$$P_t(i) = \mu\frac{W_t}{(1-\alpha)Y_t(i)/N_t(i)} = \mu MC_t(i), \tag{1.3}$$

with $\mu = \epsilon/(\epsilon - 1)$. In a symmetric equilibrium, all firms produce the same output, employ the same labor inputs and charge the same price. In this situation $p_t(i) = p_t$ (expressed in logs) and the optimal price under perfect price flexibility is equal to $p_t = \log(\mu) + mc_t$.

1.2.2 The Calvo model

The second essential element of New Keynesian macroeconomics is nominal rigidities. Sticky price models are now frequently employed in studying the monetary transmission process. In the following analysis, we concentrate solely on time-dependent models where we use in particular a Calvo (1983) style model.[3] This

[2] According to Dixit and Stiglitz (1977) aggregate output Y_t is a constant-elasticity-of-substitution aggregator $Y_t = \left[\int_0^1 Y_t(i)^{(\epsilon-1)/\epsilon}di\right]^{\epsilon/(\epsilon-1)}$. This expression abstracts from investment and foreign trade, so output Y_t equals consumption C_t and P_t is the corresponding aggregate price index $P_t = \left[\int_0^1 P_t(i)^{1-\epsilon}di\right]^{1/(1-\epsilon)}$.

[3] Another model class is that of state-dependent sticky prices models where the number of firms that changes prices in a given period is determined endogenously (e.g., Dotsey, King, and

framework assumes that each firm optimizes its price only from time to time. This is motivated by costs associated with information gathering. The frequency of price re-optimization is thus a stochastic process with a constant probability that a firm sets its prices in an optimal way at each point in time. This means that there is always a fraction of firms $1-\theta$ in the economy that are optimally adjusting their prices. This arrival rate can be described by an exogenous stochastic process with the expected waiting time between price changes given by $1/(1-\theta)$.

A firm that re-optimizes sets its price $P_t^*(i)$ in order to maximize the expected discounted sum of profits

$$E_t \sum_{k=0}^{\infty} (\beta\theta)^k \upsilon_{t,t+k} \left[P_t^*(i) X_{t,t+k} - MC_{t,t+k}(i)\right] \frac{Y_{t+k}(i)}{P_{t+k}}, \tag{1.4}$$

subject to the demand constraints eq. (1.2) and

$$X_{t,t+k} = \begin{cases} \prod_{l=0}^{k-1} \overline{\pi}^{1-\xi} \pi_{t+l}^{\xi} & \text{for } k > 0 \\ 1 & \text{for } k = 0. \end{cases} \tag{1.5}$$

where β is the subjective discount factor and $\upsilon_{t,t+k} = U^{'}(C_t)/U^{'}(C_{t+k})$ with $U^{'}(C_{t+k})$ the marginal utility of consumption in period $t+k$. $\overline{\pi}$ is the long-run average gross rate of inflation. When a firm does not re-optimize its price, it is assumed to have reset it according to some sort of indexation scheme. Our baseline specification is the partial indexation scheme used in Smets and Wouters's (2003) model, and further discussed by Sahuc (2004) with $\xi \in [0,1]$ which measures the degree of indexation to past inflation. This is a further generalization of Christiano, Eichenbaum, and Evans's (2005) dynamic indexation scheme with $\xi = 1$, where prices are reset according to $P_t(i) = \pi_{t-1} P_{t-1}(i)$ during periods when firms are re-optimizing.

After solving the maximization problem in eq. (1.4) and some further manipulations,[4] an expression for aggregate inflation can be derived as

$$\widehat{\pi}_t = \frac{\xi}{1+\beta\xi}\widehat{\pi}_{t-1} + \frac{\beta}{1+\beta\xi} E_t \widehat{\pi}_{t+1} + \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta} A\widehat{s}_t, \tag{1.6}$$

Wolman, 1999). Another popular model other than that of Calvo (1983) was developed by Taylor (1980).

[4] See e.g. Sahuc (2004) or Walsh (2003, Ch. 5) for a derivation.

where $\widehat{s}_t$ is the percentage deviation of average marginal cost MC_t/P_t from its steady state. This type of equation is often referred to as the New Keynesian Phillips curve.[5] Note that a particular feature of this inflation equation is its sound microeconomic foundation, i.e., it depends on structural parameters with a direct economic interpretation. With $\xi = 0$ the expression reduces to the pure forward-looking Phillips curve that coincides with a static indexation scheme.[6]

The parameter A measures the degree to which inflation responds to changes in the current and future values of real marginal costs. In contrast to a situation in which all firms face the same marginal cost ($A = 1$), firm specific marginal cost may differ across firms on account of differences in the output level. These differences are generated by the assumption of a fixed stock of firm-specific capital.[7] As shown by Sbordone (2002) and Galí, Gertler, and López-Salido (2001) A also depends on structural parameters with

$$A = \frac{1-\alpha}{1+\alpha(\epsilon-1)},$$

ϵ the elasticity of substitution among different goods from eq. (1.2) and α the technology parameter from the Cobb–Douglas production function eq. (1.1), whereas $\epsilon > 1$ and $0 < \alpha < 1$.

An additional way of modeling a lesser reaction of prices to marginal cost is proposed by Eichenbaum and Fisher (2007) and Coenen, Levin, and Christoffel (2007). They assume a varying elasticity of demand, but, as shown by Coenen, Levin, and Christoffel (2007), this assumption does not lead to a substantial reduction of the sensitivity of prices to marginal cost for reasonable values of α. In order to remain straightforward, we do not consider this type of additional friction.

[5] This expression is an augmented version of a specific relation that does not include the lagged inflation term. The version with an additional inflation lag is sometimes called the "hybrid" Phillips curve.

[6] A static indexation scheme implies that firms set prices according to $P_{it} = \overline{\pi} P_{it-1}$ during periods when they are not re-optimizing (e.g., Erceg, Henderson, and Levin, 2000).

[7] A more comprehensive description for the role of firm-specific capital is given by Eichenbaum and Fisher (2007) where firms face convex capital adjustment costs. Our specification of A can be seen as a special case of this framework, where the adjustment costs are very high.

1.2.3 A variant with rule-of-thumb firms

A variant of the model demonstrated above eq. (1.6) was presented by Galí and Gertler (1999). In this specification there are two types of firms; one fraction $1-\omega$ that re-optimizes prices according to the model of Calvo (as discussed in Section 1.2.2). During periods when firms cannot re-optimize, they set prices according to a static indexation scheme. The other fraction ω of non-re-optimizing firms set prices according to a backward-looking rule-of-thumb. With probability θ they set $P_{it} = \overline{\pi} P_{it-1}$. Otherwise, with probability $1-\theta$, they apply

$$P_t' = \pi_{t-1}\overline{P}_t$$

with $\overline{P}_t = (1-\omega)P_t^* + \omega P_t'$, where P_t^* is the optimized price that is chosen by the fraction of firms that are forward-looking.

In this setting an analog expression of eq. (1.6) can be derived as

$$\widehat{\pi}_t = \frac{\omega}{\phi}\widehat{\pi}_{t-1} + \frac{\beta\theta}{\phi}E_t\widehat{\pi}_{t+1} + \frac{(1-\omega)(1-\theta\beta)(1-\theta)}{\phi}A\widehat{s}_t, \tag{1.7}$$

with $\phi = \theta + \omega\left[1-\theta(1-\beta)\right]$. When $\omega = 0$ this expression is equivalent to the pure forward-looking Phillips curve and is thus equal to eq. (1.6) as long as $\xi = 0$.

Finally, it should be noted that the explanatory variables are the same across the two Phillips curve specifications, the only difference being the way in which the structural parameters appear in the two equations. While the interpretation of θ is the same, the parameters ξ and ω have a different meaning depending on the particular model that both try to rationalize a lagged inflation term in the Phillips curve.

1.3 Econometric methodology

We now present our empirical model and discuss how we can conduct inference about the structural parameters of the Phillips curve model discussed above. In this analysis we take a limited information approach, with its great advantage of not having to specify a complete general equilibrium model, which would include the nature of the forcing variable. Instead, we can leave part of the model unspecified and consider instead a single equation. As is known from the traditional simultaneous equation framework, full information methods may be more efficient, but they can also be more sensitive to specification errors, seeing that errors in one equation affect other equations.[8] We also present some shortcomings of the standard GMM and present an identification robust variant to standard GMM estimation that is valid under much weaker assumptions.[9]

1.3.1 GMM

Our empirical model is given by

$$\widehat{\pi}_t = \gamma_b \widehat{\pi}_{t-1} + \gamma_f \widehat{\pi}_{t+1} + \lambda \widehat{s}_t + u_t, \tag{1.8}$$

where $u_t = \eta_t - \gamma_f \left(\widehat{\pi}_{t+1} - E_t \widehat{\pi}_{t+1}\right)$. Note that expected future inflation $E_t \widehat{\pi}_{t+1}$ has been replaced by its realization $\widehat{\pi}_{t+1}$, whereas the expectation error $(\widehat{\pi}_{t+1} - E_t \widehat{\pi}_{t+1})$ is part of the residual u_t. The coefficients γ_b, γ_f and λ depend in nonlinear form on the structural parameters $(\beta, \theta, \xi, \alpha, \epsilon)$ in the partial indexation model or on $(\beta, \theta, \omega, \alpha, \epsilon)$ in the model including rule-of-thumb firms.

As the residual u_t is correlated with $\widehat{\pi}_{t+1}$ (unless there exist forecast errors of future inflation), an instrumental variables estimator is needed to guarantee unbiased results. We employ a Generalized Method of Moments (GMM) estimator proposed by Hansen (1982) that is suitable for dynamic non-linear models to estimate the structural model parameters. This approach is frequently applied to estimating inter-temporal asset pricing models.[10] Additionally, we use heteroskedastic and autocorrelation consistent (HAC) standard errors due to Newey and West (1987).[11]

The GMM approach is now also very often applied to estimate the parameters of the Calvo model. Examples include Galí and Gertler (1999), Galí, Gertler, and López-Salido (2001) and Eichenbaum and Fisher (2007). First, we set up the

[8] Examples for ML techniques to estimate hybrid Phillips curve specifications include Fuhrer (1997), Lindé (2005) and Jondeau and Le Bihan (2006). See Jondeau and Le Bihan (2008) for discussion of properties of different estimators under misspecifications.

[9] Identification robust methods are currently unavailable for ML estimation. However, as shown by Dufour, Khalaf, and Kichian (2007a), these methods may also suffer from weak instrument problems in the context of the NKPC and may lead to wrong conclusions.

[10] See Hansen and Singleton (1982) for an early example.

[11] Throughout we use a lag length of 5 for the HAC estimator. In our case the estimated standard errors are not very sensitive to the particular choice of the lag length.

orthogonality conditions for the partial indexation model eq. (1.6). We use two specifications that differ in the way in which functions are normalized. These are given by

$$u_t^1 = \widehat{\pi}_t - \frac{\xi}{1+\beta\xi}\widehat{\pi}_{t-1} - \frac{\beta}{1+\beta\xi}\widehat{\pi}_{t+1} - \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}A\widehat{s}_t, \tag{1.9}$$

$$u_t^2 = (1+\beta\xi)\widehat{\pi}_t - \xi\widehat{\pi}_{t-1} - \beta\widehat{\pi}_{t+1} - \frac{(1-\theta\beta)(1-\theta)}{\theta}A\widehat{s}_t, \tag{1.10}$$

with the orthogonality conditions

$$E_{t-1}\left\{u_t^i\left(\beta,\theta,\xi\right)\mathbf{z_{t-1}}\right\} = 0 \tag{1.11}$$

for $i = 1, 2$. $\mathbf{z_{t-1}}$ is the vector of instruments that are assumed to be orthogonal to the error term u_t^i (under the rationality assumption). Note that $\mathbf{z_{t-1}}$ does include only instruments dated $t-1$ or earlier in order to rule out simultaneity issues. This also guarantees that the information is already available at time t owing to a potential publication lag.

The standard two-step GMM estimates are obtained by minimizing

$$J(\vartheta^i) = \left[\frac{1}{T}\sum_{t=1}^{T}\phi_t^i(\vartheta^i)\right]' V(\vartheta_T^{i,1})^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\phi_t^i(\vartheta^i)\right], \tag{1.12}$$

where ϑ^i denotes the structural parameters of the model; $\phi_t(\vartheta^i) = u_t^i(\beta^i, \theta^i, \xi^i)\mathbf{z}_{t-1}$. This objective function is evaluated given an initial estimate $\vartheta_T^{i,1}$ for the weighting matrix. This initial estimator may be obtained by using the identity matrix or the instrument matrix as a weighting matrix for the first step. As we consider two different transformations of the orthogonality conditions u_t^1 and u_t^2 the estimates for the structural parameter may differ across these specifications.[12]

From eqs. (1.9) and (1.10) it follows that θ and A (and thus also α and ϵ) cannot be separately identified. So we are able to estimate only θ, the parameter

[12]It is well known that for finite samples the two-step GMM and the iterated GMM estimator may be sensitive to transformations of the orthogonality conditions (e.g., Hall, 2005). Unless the model is not mis-specified, the two different normalizations should lead to approximately similar results.

that is of most interest, given reasonable values of α and ϵ which cannot be tested explicitly. To identify the remaining parameters β, θ and ξ we need at least three valid instruments.

For the model with rule-of-thumb firms, similar orthogonality conditions can be formulated. They differ only in respect of the functional form of the parameters. The two normalizations are given by

$$u_t^{'1} = \widehat{\pi}_t - \frac{\omega}{\phi}\widehat{\pi}_{t-1} - \frac{\beta\theta}{\phi}\widehat{\pi}_{t+1} - \frac{(1-\omega)(1-\theta\beta)(1-\theta)}{\phi}A\widehat{s}_t, \quad (1.13)$$

$$u_t^{'2} = \phi\widehat{\pi}_t - \omega\widehat{\pi}_{t-1} - \beta\theta\widehat{\pi}_{t+1} - (1-\omega)(1-\theta\beta)(1-\theta)A\widehat{s}_t. \quad (1.14)$$

Again, the orthogonality conditions can be formulated as

$$E_{t-1}\left\{u_t^{'i}\left(\beta,\theta,\omega\right)\mathbf{z_{t-1}}\right\} = 0 \quad (1.15)$$

for $i = 1, 2$ with $\phi = \theta + \omega\left[1 - \theta(1-\beta)\right]$. Everything else is comparable with the partial indexation model.

1.3.2 An identification robust alternative

So far our analysis rests on the implicit assumption that our instrument set is sufficiently correlated with the endogenous variables under consideration. We have thus assumed that our regression analysis does not suffer from weak instrument problems. As shown in a wide range of relevant literature, the presence of weak instruments may cause serious distortions in standard IV point estimates, hypothesis tests and confidence intervals (see Stock, Wright, and Yogo, 2002, for an overview of problems caused by weak instruments and some recommendations for dealing with this problem). Several authors, including Ma (2002), Dufour, Khalaf, and Kichian (2006), Mavroeidis (2006), Nason and Smith (2008), Martins and Gabriel (2009), Kleibergen and Mavroeidis (2009) provide evidence that weak instrument problems may be present in standard GMM estimations of the New Keynesian Phillips curve.

For this reason, potential problems with weak instruments or weak identification in our estimation strategy must be highlighted. As shown by Mavroeidis

(2005) standard pre-tests of identification (or weak instrument problems) are inappropriate in this context. We therefore re-evaluate our GMM results with an identification robust method that is fully robust to problems induced by weak instruments and weak identification. We have therefore adhered to a non-linear variant of the Anderson–Rubin Statistic as suggested by Stock and Wright (2000). They show that identification robust confidence sets can be obtained from the continuous-updating GMM (CUE) objective function.[13] In the linear simultaneous equations model these so called S-sets are asymptotically equivalent to confidence sets constructed by inverting the Anderson–Rubin test statistic.[14]

As shown by Dufour (2003), the AR statistic is suitable for validating a structural model, as it is not only robust in the eventuality of weak instruments, but also in the case of model mis-specifications like over-identification. It thus provides an alternative to the standard J test. S-sets, too, share the characteristic of identification-robust procedures, as described in Dufour (1997). These require that whenever parameters are unidentified, the results should lead to uninformative and thus unbounded confidence sets. S-sets contain all parameter values for which the joint hypothesis $\vartheta = \vartheta_0$ and for which the over-identifying conditions are valid. So, whenever the model is mis-specified and the over-identifying conditions are invalid, the S-sets can be null. On the other hand, with weak instruments or irrelevant instruments, the S-sets can contain the entire parameter space. This is a favored property of the test because it ensures robustness in the face of many pitfalls, the J test having very low power for estimations of the NKPC (Mavroeidis, 2005). It also requires some cautious interpretion of the model's results. In particular, when S-sets are small, it can be because the model is correctly specified or mis-specified, but it does not lead to a full rejection.[15]

[13] The continuous-updating GMM estimator was invented by Hansen, Heaton, and Yaron (1996). As opposed to the standard two-step GMM estimator, the CUE evaluates the weight matrix at the same parameter value as the orthogonality conditions.

[14] Dufour, Khalaf, and Kichian (2006) evaluate the NKPC with the standard AR test which is closer related to 2SLS than GMM. They extended this framework in Dufour, Khalaf, and Kichian (2007b) and in Dufour, Khalaf, and Kichian (2008) to allow for heteroskedastic and autocorrelated residuals. Kleibergen and Mavroeidis (2009) consider not only Stock and Wright's (2000) approach, but also a Lagrange Multiplier statistic (as discussed in Kleibergen, 2007) and a Likelihood Ratio statistic (see Kleibergen, 2005) to construct identification robust confidence intervals for the NKPC. Martins and Gabriel (2009) evaluate the NKPC with generalized empirical likelihood (GEL) methods.

[15] This may become relevant when there are many instruments. In this case the power of the

The objective function of the CUE is given by

$$S(\vartheta) = \left[\frac{1}{T}\sum_{t=1}^{T}\phi_t(\vartheta)\right]' V(\vartheta)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\phi_t(\vartheta)\right] \tag{1.16}$$

with ϑ the parameter vector of interest; $\phi_t(\vartheta) = u_t\mathbf{z_{t-1}}$ with $u_t = u_t\left(\beta,\theta,\xi\right)$ or $u_t = u_t\left(\beta,\theta,\omega\right)$ as defined in eqs. (1.9) and (1.13) and $\mathbf{z_{t-1}}$ the vector of instruments. Note that the CUE is invariant to transformations of the orthogonality condition, so this differentiation does not require consideration. $V(\vartheta)$ is defined as a HAC estimator to allow for serial correlation as well as heteroskedasticity in the residuals. This coincides with the two-step estimator used above.

We now check whether our baseline GMM results hold when we use S-sets, as suggested by Stock and Wright (2000). First, we can examine whether our GMM point estimates are also included in the S-sets. This should be the case when the model is correctly specified and there are no weak instrument problems present. According to Stock and Wright (2000) $S(\vartheta_0) \xrightarrow{D} \chi^2_k$, where $S(\vartheta_0)$ is the objective function as defined above evaluated at the true parameter values $(\beta_0, \theta_0, \xi_0)$ or $(\beta_0, \theta_0, \omega_0)$. Second, we can construct confidence intervals for the parameters of interest (so-called S-sets).[16] Here, we ask what parameter values are comparable with the model. All values which are not rejected form the confidence region.[17].

1.4 Data and empirical implementation

Our sample period is 1973:1 – 2004:4. While data before 1973 are principally available, we take this date as starting point, as it marks the end of the fixed exchange rate regime of the Bretton Woods system. This is also associated with a change in monetary policy that became more independent of external influences. Inflation is measured as the quarterly annualized change in the GDP deflator. From the production function eq. (1.1) it follows that real marginal costs are proportional to the labor income share of output. The labor share is defined as

test might be too low to reject a potentially mis-specified model.

[16]We follow Stock and Wright and construct 90% S-set as they did in their chapter.

[17]The parameter space that we consider involves all possible values in the range of 0 to 1. In the search process all values within this range are evaluated with increments of 0.01. For the measure of real rigidity A we take as given the values for α and ϵ as calibrated in section (1.3.1)

the total wage bill (W_tN_t) divided by nominal GDP (P_tY_t). The variable $\widehat{s}_t$ is constructed as the percentage deviation of the labor share from its sample average (see Figure 1.1).[18]

Because in our Phillips curve specification the term A cannot be separately identified, we have to calibrate α and ϵ in an economically reasonable way. We set α, the output elasticity with respect to capital, equal to 0.3 of how it is usually done for the German economy (e.g., Dreger and Schumacher, 2000). The calibration of the elasticity of substitution among different goods is more controversial. For the definition of the steady state mark-up μ, it follows that the elasticity of substitution can be redefined as $\epsilon = \frac{\mu}{(\mu-1)}$. We consider a steady-state mark-up of 10% ($\mu = 1.1$) as our baseline value (as it was done by Galí, Gertler, and López-Salido, 2001; Eichenbaum and Fisher, 2007). This corresponds to $\epsilon = 11$.

A crucial further issue is concerned with the instrument vector $\mathbf{z_{t-1}}$. To be a valid instrument, variables have to fulfill two important characteristics. First, they have to be uncorrelated with the error term (which is the orthogonality condition). Second, they have to be correlated with the variable they have to instrument (that is the relevance condition). Both conditions have to be fulfilled if reliable point estimates and confidence intervals of the model parameters are to be obtained. So the first practical challenge is to decide which variables should be included into the instrument set. In principle, any variable dated $t-1$ and earlier may be considered as an instrument, seeing that under rational expectations it fulfills the orthogonality conditions (and thus the first condition of a valid instrument). This leaves us with a potentially infinite set of possible variables that could be used as instruments. But as Tauchen (1986) and Kocherlakota (1990) recognized early on, instruments should be used quite sparingly.[19]

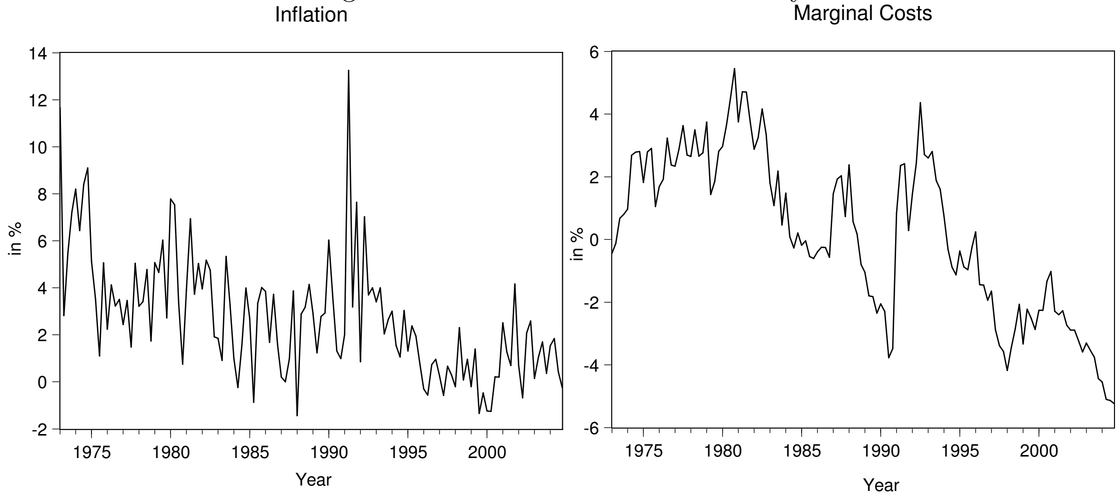
To deal with problems of redundant instruments we apply a unique two-step approach, in which we try to cull out the really relevant variables. The explicit treatment of instrument selection is rarely done for rational expectations models.[20].

[18] This is the measure proposed by Galí and Gertler (1999), Galí, Gertler, and López-Salido (2001) and Sbordone (2002).

[19] Tauchen (1986) finds in a simulation study that the inclusion of additional instruments that are not relevant or are only marginally relevant leads to increasing bias of the parameter estimates.

[20] Much empirical work employs instrument sets that have been used in previous studies without checking whether they are really relevant. An early exception is Pesaran (1987), who emphasizes

Figure 1.1: Data series for Germany



Note: Inflation is measured as the annualized percentage change of the GDP deflator. The marginal cost measure corresponds to the labor income share defined as total compensation of employees divided by nominal GDP (measured as percentage deviations around the mean).

Source: Statistisches Bundesamt; own calculations

As a starting point, we consider a wide range of possible instruments that include important macroeconomic indicators. This potential instrument list contains Galí and Gertler's (1999) instrument set with inflation, real marginal cost, real-time detrended GDP, wage inflation, commodity price inflation, and the long-short interest rate spread. Further, we include as an additional potential instrument the short-term interest rate (defined as the three month bill). For the variables $\widehat{\pi}_t$ and $\widehat{s}_t$ we allow a potential lag length of 5 quarters; for the remaining candidate variables we use a maximal lag length of 2.[21] The first step of instrument selection contains a preselection of possible instruments within a VAR, the two endogenous

pre-checking the conditions for identification in models with rational expectations. Other examples are Fuhrer and Olivei (2005) who use an instrument set that is derived from theory and Nason and Smith (2008), who check how many lags of the labor share can be used as instruments. More recently, there are some studies that employ factor analysis to construct an instrument set for rational expectation models in an environment with many potential instruments (e.g. see Bai and Ng, 2010; Beyer, Farmer, Henry, and Marcellino, 2008)

[21] The choice of the potential lag length is orientated according to previous experience. Galí, Gertler, and López-Salido (2001) use a lag length for inflation of 5 quarters. The same lag length is also considered for the labor share as the driving variable. For the remaining variable we consider only the most recent lags, as these are the variables that ought to be mostly correlated with the variables they instrument.

variables $\widehat{\pi}_{t+1}$ and $\widehat{s}_t$ are therefore regressed on all potential instruments. This specification can be formalized as

$$\begin{bmatrix} \widehat{\pi}_{t+1} \\ \widehat{s}_t \end{bmatrix} = \nu + \sum_{i=1}^{L_1} A_i y_{t-i} + \sum_{j=1}^{L_2} B_j x_{t-j} + u_t, \qquad (1.17)$$

with ν a deterministic term, $y_{t-i} = [\widehat{\pi}_{t-i}\ \widehat{s}_{t-i}]'$ and x_{t-j} the vector of all other predetermined variables with lag j. The maximal lag length is $L_1 = 5$ and $L_2 = 2$.

After estimating the full model, we apply a model reduction procedure, which works by means of a sequential elimination of regressors, in order to obtain a model that leads to the lowest value of the particular information criterion. We base the selection procedure on two selection criteria (AIC and SC), which are frequently used in time-series analysis (e.g., Lütkepohl, 2005). Accordingly, we end up with restrictions on A_i and B_i which determine our instrument sets $\mathbf{z_{t-1}}(c^{AIC})$ and $\mathbf{z_{t-1}}(c^{SC})$, where c^j denotes which elements of the candidate set are included in a particular moment condition. Besides the two instrument sets based on the information criteria, we also take Galí, Gertler, and López-Salido's (2001) set as a benchmark.

We thus have three candidate instrument sets with the following sizes:

- AIC-based instrument set: that includes 14 of 21 potential instruments (see Table 1.5),
- SC-based instrument set: that includes 11 instruments (see Table 1.5),
- Galí, Gertler, and López-Salido's (2001) instrument set: that includes inflation with lags $t-1$ to $t-5$, labor share, wage inflation and output gap from $t-1$ to $t-2$ (all together 11 instruments),

The sensitivity of our results vis-á-vis different instrument sets may also indicate whether there are any problems with redundant or weak instruments.

As a second step we also apply a moment selection check after we have carried out the GMM estimation to evaluate our preselection based on model reduction techniques. This strategy is based on the relevance condition. We therefore use a moment selection criterion proposed by Hall, Inoue, Jana, and Shin (2007). This criterion is defined as

$$RMSC(c) = ln\left[\left|\hat{V}_{\theta,T}(c)\right|\right] + (|c| - p)\, ln(T^{1/3})/T^{1/3} \tag{1.18}$$

where $\hat{V}_{\theta,T}(c)$ is the covariance matrix of the model parameters conditional on the instrument set c. The second term is a BIC-type penalty term with T the sample size and p the number of parameters to be estimated. The idea is to select the instrument vector that minimizes this criterion. Since the relevance condition can be interpreted as statement about the asymptotic variance of the estimator, the sample analog is the natural basis on which to construct an information criterion. Hall, Inoue, Jana, and Shin (2007) show that the natural logarithm of the determinant of the variance–covariance matrix can serve for this purpose. Note that this procedure works only when there are no weak instrument problems present. This means it is necessary to have at least some variables that are correlated to an appreciable extent with the endogenous variables they have to instrument (otherwise the selection criterion may produce strange results).

1.5 Estimation results

In this section we present the results of the structural model and their robustness to several empirical considerations. First, we check for sensitivity relating to different instrument sets and orthogonality conditions. As pointed out above, the instrument relevance is essential to the reliability of GMM point estimates and confidence intervals. For this reason we report estimation results with the instrument set used by Galí, Gertler, and López-Salido (2001) and compare them with those based on a pre-selection, as discussed in section (**??**). In addition, we consider different degrees of real rigidity and the effects on the estimated Calvo parameter. We then show how results change when we augment the PC model by additional lags of inflation. Finally, we present results based on the AR statistic and we compare them with the baseline GMM estimates.

1.5.1 GMM results

We begin by presenting our baseline GMM results for the partial indexation model as well as for the model that includes rule-of-thumb firms. These estimates take as given the degree of real rigidity with calibrated values for α and ϵ (see Section **??**). Further, the SC based instrument set serves as our benchmark instrument set, since this set is associated with the smallest RMSC criterion (see Table 1.6 and 1.7). Table 1.2 shows the results based on the partial indexation model. Point estimates for θ vary from 0.61 to 0.69. These are different from zero, as well as from one (the latter is necessary for the model to hold at least from an economic perspective). The estimates display reasonable values for θ which implies that firms re-optimize prices about every 3 quarters. In addition, the J test of over-identification does not indicate any problems for this specification. The point estimates of the discount factor β are somewhere in the vicity of one, which is also plausible from an economic point of view. We find little evidence for the full indexation scheme ($\xi = 1$) as proposed by Christiano, Eichenbaum, and Evans (2005), as the coefficient tests reject this hypothesis. Furthermore, we do not find much evidence for partial indexation in general, which implies that ξ is close or equal to zero. This finding favors a pure forward-looking specification without a lagged inflation term.

The evidence is more mixed when we consider the model with rule-of-thumb firms (Table 1.2). Here the results differ considerably with respect to the way in which the orthogonality condition is formulated. This is particularly true for point estimates of θ where the first orthogonality condition produces results similar to those of the model with partial indexation. But with orthogonality condition (2), the estimated values for θ are far lower. Additionally, the J test is significant for that specification. This casts doubt on the estimation results based on condition (2), as well as on the model in general. This sensitivity to the normalization of the orthogonality condition may indicate some form of model mis-specification. The estimates for the remaining parameters do not differ considerably from the ones obtained with the partial indexation scheme. Again, the discount factor is close to one and the backward-looking inflation term (ω in this specification) seems to be unimportant.

1.5.2 Sensitivity analysis

The general results are relatively robust in relation to the particular instrument set used (see Table 1.6 and 1.7). However, differences between the two orthogonal-

Table 1.1: Partial indexation model (unrestricted)

	β	θ	ξ	J	Freq.
(1)	1.030	0.611	0.248	8.454	2.6
	(0.058)	(0.181)	(0.156)		
	[0.000]	[0.001]	[0.112]	[0.390]	
(2)	1.036	0.690	-0.182	10.820	3.2
	(0.038)	(0.270)	(0.097)		
	[0.000]	[0.011]	[0.059]	[0.212]	

Notes: Standard errors in round brackets and p-values in square brackets. Rows (1) and (2) correspond to the two specifications of the orthogonality conditions eqs. (1.9) and (1.10) in the text, respectively. A 5-lag Newey-West HAC estimate was used. Sample period: 1973:1-2004:4. SC based instrument set.

Table 1.2: Rule-of-thumb model (unrestricted)

	β	θ	ω	J	Freq.
(1)	1.026	0.590	0.146	8.454	2.4
	(0.052)	(0.184)	(0.092)		
	[0.000]	[0.001]	[0.111]	[0.390]	
(2)	0.908	0.178	-0.019	16.945	1.2
	(0.104)	(0.035)	(0.020)		
	[0.000]	[0.000]	[0.340]	[0.031]	

Notes: Standard errors in round brackets and p-values in square brackets. Rows (1) and (2) correspond to the two specifications of the orthogonality conditions eqs. (1.13) and (1.14) in the text, respectively. A 5-lag Newey-West HAC estimate was used. Sample period: 1973:1-2004:4. SC based instrument set.

ity conditions get more pronounced for the instrument set from GGL. We check the sensitivity of our results further for different assumptions about firm-specific marginal cost. We first show how the estimates of θ change when we assume a markup of 25% ($\mu = 1.25$) instead of 10%, as assumed in our baseline specification (Table 1.8, Appendix). When the degree of real rigidity is degreased, the point estimates for θ rise slightly, whereas the remaining parameters in principal remain uneffected. However, the estimates of θ remain in a meaningful economic range and cannot be rejected on empirical grounds. Next, we relinquish the assumption of firm specific marginal costs and assume equal marginal cost across firms

($A = 1$) as in the baseline model of Galí and Gertler (1999). This leads to a further rise in the estimated parameter θ to about 0.8 in the partial indexation model and to 0.6 and 0.8 in the model with rule-of-thumb firms (this implies an average frequency of price-re-optimization between 3 and 9 quarters). This specification continues to coincide with a sticky price framework, which manifests in a higher degree of nominal price rigidity. From an empirical point of view we cannot favor one specification over the other, when they differ only in the way in which firm-specific marginal cost deviates from average marginal cost. Since the model is compatible with different assumptions about firm-specific marginal cost, it also introduces an additional source of uncertainty in estimating θ and the frequency of re-optimization.

The basic findings also hold if we restrict the different model specifications to the pure forward-looking specification and a discount factor of $\beta = 0.99$ (Table 1.3). We therefore employ a likelihood ratio-type test in which we check whether the imposed restrictions can be rejected (Table 1.9 and 1.10). The tests indicate that the restrictions cannot be rejected, so they thus set. With these restrictions both model specifications (the partial indexation model and the model with rule-of-thumb firms) are the same. This specification is purely forward-looking (it does not include a lagged inflation term), the coefficients being non-linear functions of the parameter θ. Once again, we can construct two different orthogonality conditions that differ in respect to the particular normalization. As with the rule-of-thumb specifications, the estimation results for θ differ considerably. But when we impose less real rigidity ($A \rightarrow 1$), the values for θ converge slightly, but the frequency of the re-optimization for price changes of orthogonality condition (1) is always twice as high in comparison with condition (2).

We also consider the sensitivity of inflation to our marginal cost variable. We denote the reduced form coefficient in front of the marginal cost variable with λ (which is defined as $\lambda = ((1-0.99\theta)(1-\theta)/\theta)A)$. To evaluate whether λ is significant we use the point estimates for θ and its variance to construct standard errors for λ with the delta method. The results are displayed in Table 1.11 and are quite heterogeneous with respect to parameter values as well as for their significance level. For the first specification we find small values of λ that are not significant at conventional levels. The opposite is true for the second orthogonality condi-

Table 1.3: Frequency of re-optimization (restrictions: $\beta = 0.99$, $\xi = 0$, $\omega = 0$)

	$A = 0.1750$		$A = 0.3182$		$A = 1$	
	θ	$\frac{1}{1-\theta}$	θ	$\frac{1}{1-\theta}$	θ	$\frac{1}{1-\theta}$
(1)	0.577	2.36	0.664	2.98	0.795	4.88
	[0.34,0.81]		[0.46,0.87]		[0.65,0.94]	
(2)	0.179	1.22	0.326	1.48	0.607	2.54
	[0.11,0.24]		[0.24,0.41]		[0.52,0.69]	

Notes: Confidence intervals in square brackets. Rows (1) and (2) correspond to the two specifications of the orthogonality conditions eqs. (1.13) and (1.14) in the text given the imposed restrictions, respectively. A 5-lag Newey-West HAC estimate was used. Sample period: 1973:1-2004:4. SC based instrument set. J test never rejects any model.

tion. There we find larger values for λ that are always significant. These results cast doubt on whether marginal cost is indeed the driving variable for inflation or whether the labor share is the correct measure of marginal cost.[22]

Finally, we consider additional variables in our structural model of the Phillips curve. Since it has sometimes been argued that the New Keynesian Phillips curve omits further inflation lags (e.g., Jondeau and Le Bihan, 2006), we check whether our basic results hold when we put three more lags of inflation into our Phillips curve specifications. When the former specification is correct, additional lags should not be a determinant of actual inflation (they should be solely a predictor of future inflation).

Tables 1.12 and 1.13 show the results of these augmented specifications. Although the general interpretation still holds, we find that in either case the estimate of θ is higher than that based on our baseline specification. The other parameters do not change significantly and remain within a plausible range. Another important feature, the differences between orthogonality conditions (1) and (2) in the rule-of-thumb model, is still present and is not overcome by the inclusion of the additional variables. Some of these lags indeed turn out to be significant determinates of inflation (specifically the fourth lag). Like Galí and Gertler (1999), we also test whether the sum of these coefficients is different from zero. We use a Wald test but find no evidence that the sum of additional lags is important.

[22] Mavroeidis (2006) also shows that for the U.S. the marginal cost variable does not turn out to be significant.

Overall, the inclusion of additional lags does not lead to a complete rejection of our original specification. But it further demonstrates how sensitive estimates of θ are to small changes of the model.

1.5.3 Results based on the identification robust procedure

So far our analysis rests on the assumption that our instrument set is sufficiently correlated with the endogenous variables under consideration. Whenever this assumption is violated and weak instrument problems occur, standard GMM estimates are unreliable. To examine whether those problems are present in this analysis we confront the baseline GMM estimates with the identification alternative as outlined in Section 1.3.2.

Table 1.4: AR type test of the estimated parameters

Null Hypothesis	Test Statistic	p-value
$H_0 : \beta_0 = 0.99, \theta_0 = \hat{\theta}_{GMM1} = 0.58$	19.28	0.056
$H_0 : \beta_0 = 0.99, \theta_0 = \hat{\theta}_{GMM2} = 0.18$	33.10	0.001

Notes: The test is evaluated with the CUE objective function. The SC based instrument set is used. A Newey-West HAC estimate with 5 lags was used. Sample period: 1973:1-2004:4.

First, we check whether our GMM point estimates also included the S-sets. This should be the case when the model is correctly specified and there are no weak instrument problems present. We start with the purely forward-looking specification, testing the null hypothesis of whether β and θ are $(0.99, 0.58)$ or $(0.99, 0.18)$ which corresponds to the GMM estimates of Table 1.3 with $A = 0.175$. According to Stock and Wright (2000), $S(\beta_0, \theta_0) \xrightarrow{D} \chi^2_k$, where $S(\beta_0, \theta_0)$ is the objective as defined above, evaluated at the true parameter values. Table 1.4 reports the results of this test type. The results indicate that problems with the orthogonality conditions may be present, as the test rejects the hypothesis for both GMM point estimates, at least at the 10% level.[23] Now, we ask whether there exists a value of the parameter vector for which the model is not rejected. Given this particular instrument set (based on the SC), we find no parameter

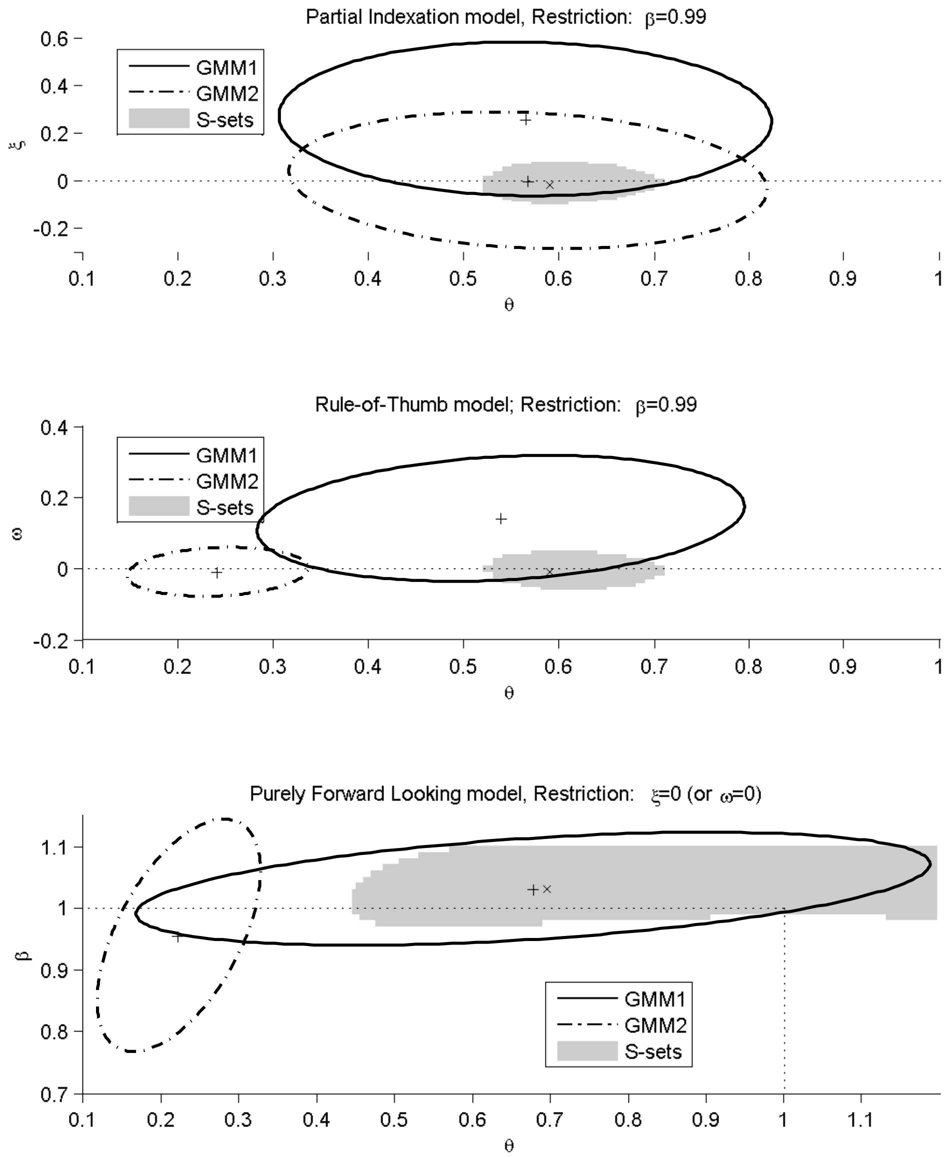
[23] The test is more in favour of the first estimate, denoted by $\hat{\theta}_{GMM1}$. We follow Stock and Wright and take the 90% S-set as the criterion for our final decision.

combination, inside the 90% S-set. That means that the confidence interval is empty and we have to reject the model. As mentioned above, this indicates that the overidentifying conditions are invalid. There may be one or more variables in our instrument set that do not fulfill the orthogonality condition. A natural candidate is a variable that is measured in $t-1$, so agents do not use this kind of information (owing to a possibly larger lag of publication). We exclude some of the instruments from period $t-1$ variable-by-variable and discover that wage inflation is the variable causing the AR type test to reject the model. So we exclude that variable and redo the analysis.

With the adjusted instrument set, the S-sets are non-empty and can be used for inference of our model. Figure 1.2 shows the 90% confidence regions obtained with that method along with the standard GMM results and their 90% confidence ellipsis for different model specifications (both methods use the same instrument set). Generally, we find somewhat small S-sets, irrespective of which particular model is being used or which restrictions are imposed. For the partial indexation model, the computed S-set lies completely inside the two GMM ellipses (for the economic reasonable values). The regions all include the null of parameter ξ, implying that this value is not significantly different from zero. The results based on the S-sets also imply a parameter value for θ of about 0.6, which translates into a frequency of price re-optimization of 2.5 quarters. The GMM results are similar. This estimate is in line with Coenen, Levin, and Christoffel (2007), who find an average frequency of price re-optimization of 2 quarters for the German economy, although with a different estimation strategy and a higher degree of real rigidities. Compared with mirco evidence, the point estimates are more in line with those of Stahl (2006) than with Hoffmann and Kurz-Kim (2006), indicating that prices are less rigid and more comparable with estimates for the U.S..

The results based on the rule-of-thumb model are, in principle, identical, even though the GMM estimates differ quite substantially in respect of the transformation of the orthogonality condition. As mentioned above, S-sets are invariant to the normalization of the orthogonality conditions. From an empirical point of view, we cannot distinguish between the partial indexation model and the rule-of-thumb. But, as shown in the chapter, the GMM estimates are sensitive to transformation of the orthogonality condition. That becomes very obvious in the

Figure 1.2: Joint 90% S-sets and 90% GMM confidence ellipses for different specifications



Notes: SC based instrument set excluding wage inflation. Sample period: 1973:1-2004:4.

rule-of-thumb model and the pure forward-looking model, where the differences between the two specifications are rather substantial.

As the hybrid version of the Phillips curve has been rejected, we concentrate once more on the pure forward-looking specification. While the S-set for this specification is once more quite small, it already includes values for θ between 0.45 and 1. This implies that the uncertainty about θ is quite high when no further restrictions are imposed on β. This also translates into the sensitivity of inflation to marginal cost ($\theta = 1 \longrightarrow \lambda = 0$). When $\theta = 1$ prices are never re-optimized, they do not respond to changes in marginal cost. As long as we cannot rule out the case that θ is equal to one, the model is economically meaningless and can also be seen as rejected.

Taken together, we show how different conclusions can be drawn depending on the particular estimation method. Interestingly, the identification robust procedure provides confidence sets smaller than conventional GMM (a standard finding for the U.S. is the fact that identification robust procedures lead to larger confidence sets compared to standard GMM, see e.g., Ma, 2002; Mavroeidis, 2006). Notably, the GMM results are extremely sensitive to the way in which the orthogonality conditions are formulated, a drawback not shared by our identification robust procedure. Additionally, identification robust inference with the nonlinear Anderson–Rubin Statistic may also help to detect model mis-specifications not indicated by the standard J test. These findings also offers an explanation for the large discrepancies in results of different studies of the German NKPC based on standard GMM. According to our analysis, a broad range of parameter values for θ, the measure of nominal price rigidity, is compatible with our model.

1.6 Conclusion

This chapter evaluates standard New Keynesian Phillips curve specifications for Germany within a limited information framework. Besides the standard GMM estimation and test procedures, we also apply identification robust techniques. The evidence presented clearly favors a purely forward-looking inflation equation for Germany, which is in contrast to that in most other countries. The average frequency of price re-optimization of firms is estimated to be about 2 and 3 quarters, given a plausible degree of real rigidity in the German economy. While these estimates seem plausible from an economic point of view, the uncertainty around these estimates is considerable as well as consistent with perfect nominal price rigidity, where firms never re-optimize their prices. Further, this casts doubt on the labor share as the driving variable for inflation.

In contrast with previous studies, confidence intervals from the identification robust procedure are smaller than results based on conventional GMM procedures. There is also some evidence of model mis-specification that is not detected by the standard J test of over-identifying restrictions. These findings explain why results for the German NKPC differ so considerably from existing studies based on GMM. Obviously, further work is needed to extend the basic framework for Germany. Empirical issues include finding a better proxy for the marginal cost measure (e.g., by including real wage rigidities) and deviating from the assumption of rational expectations through direct measures of inflation expectations.

References

Álvarez, L. J., E. Dhyne, M. Hoeberichts, C. Kwapil, H. L. Bihan, P. Lünnemann, F. Martins, R. Sabbatini, H. Stahl, P. Vermeulen, and J. Vilmunen (2006): "Sticky Prices in the Euro Area: A Summary of New Micro-Evidence," *Journal of the European Economic Association*, 4(2-3), 575–584.

Bai, J., and S. Ng (2010): "IV-Estimation in a data-rich environment," *Econometric Theory*, 26(6), 1577–1606.

Banerjee, R., and N. Batini (2004): "Inflation Dynamics in Seven Industrialised Open Economies," Discussion paper, Research Department, International Monetary Fund, Washington, D.C.

Benigno, P., and J. D. López-Salido (2006): "Inflation Persistence and Optimal Monetary Policy in the Euro Area," *Journal of Money, Credit and Banking*, 38(3), 587–614.

Beyer, A., R. E. A. Farmer, J. Henry, and M. Marcellino (2008): "Factor Analysis in a Model with Rational Expectations," *Econometrics Journal,*, 11(2), 271–286.

Calvo, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12(3), 383–398.

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

Coenen, G., A. T. Levin, and K. Christoffel (2007): "Identifying the influences of nominal and real rigidities in aggregate price-setting behavior," *Journal of Monetary Economics*, 54(8), 2439–2466.

Dhyne, E., L. J. Álvarez, H. L. Bihan, G. Veronese, D. Dias, J. Hoffmann, N. Jonker, P. Lünnemann, F. Rumler, and J. Vilmunen (2006): "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, 20(2), 171–192.

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

Dotsey, M., R. G. King, and A. L. Wolman (1999): “State-dependent pricing and the general equilibrium dynamics of money and output,” *Quarterly Journal of Economics*, 114(2), 655–690.

Dreger, C., and C. Schumacher (2000): “Zur empirischen Evidenz der Cobb-Douglas-Technologie in gesamtdeutschen Zeitreihen,” IWH Discussion Papers 113, Halle Institute for Economic Research.

Dufour, J.-M. (1997): “Some implaussibility theorems in econometrics with applications to structural and dynamic Models,” *Econometrica*, 65(6), 1365–1387.

——— (2003): “Identification, weak instruments, and statistical inference in econometrics,” *Canadian Journal of Economics*, 36(4), 767–808.

Dufour, J.-M., L. Khalaf, and M. Kichian (2006): “Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis,” *Journal of Economic Dynamics & Control*, 30(9-10), 1707–1727.

——— (2007a): “Structural Multi-Equation Macroeconomic Models: A System-Based Estimation and Evaluation Approach,” Working paper, McGill University.

Dufour, J.-M., L. Khalaf, and M. Kichian (2007b): “Which NKPC? Identification-Robust Estimates of Price Stickiness and Real Wage Rigidity,” Working paper, McGill University.

——— (2008): “How much do Real Wage Rigidities Matter for Canadian Inflation?,” Working paper, McGill University.

Eichenbaum, M., and J. D. M. Fisher (2007): “Estimating the frequency of price re-optimization in Calvo-style models,” *Journal of Monetary Economics*, 54(7), 2032–2047.

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.

FUHRER, J. C. (1997): "The (un)importance of forward-looking behavior in price specifications," *Journal of Money, Credit, and Banking*, 29(3), 338–350.

FUHRER, J. C., AND G. P. OLIVEI (2005): "Estimating forward-looking Euler equations with GMM estimators: an optimal-instruments approach," *Proceedings, Board of Governors of the Federal Reserve System (U.S.)*, pp. 87–114.

GALÍ, J., AND M. GERTLER (1999): "Inflation dynamics: A structural econometric analysis," *Journal of Monetary Economics*, 44(2), 195–222.

GALÍ, J., M. GERTLER, AND J. D. LÓPEZ-SALIDO (2001): "European inflation dynamics," *European Economic Review*, 45(7), 1237–1270.

HALL, A. R. (2005): *Generalized Method of Moments*. Oxford University Press, Oxford.

HALL, A. R., A. INOUE, K. JANA, AND C. SHIN (2007): "Information in generalized method of moments estimation and entropy-based moment selection," *Journal of Econometrics*, 138(2), 488–512.

HANSEN, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50(4), 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14(3), 262–280.

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–1286.

HOFFMANN, J., AND J.-R. KURZ-KIM (2006): "Consumer price adjustment under the microscope - Germany in a period of low inflation," Working Paper Series 652, European Central Bank.

JONDEAU, E., AND H. LE BIHAN (2006): "Testing for a Forward-Looking Phillips Curve. Additional International Evidence," *Economic Modelling*, 22(3), 521–550.

——— (2008): "Examining bias in estimators of linear rational expectations models under misspecification," *Journal of Econometrics*, 143(2), 375 – 395.

Kleibergen, F. (2005): "Testing Parameters in GMM without assuming that they are identified," *Econometrica*, 73(5), 1103–1124.

——— (2007): "Generalizing weak instrument robust IV statistics toward multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139(1), 181–216.

Kleibergen, F., and S. Mavroeidis (2009): "Weak instrument robust tests in GMM and the new Keynesian Phillips curve," *journal of business & Economic Statistics*, 27(3), 293–311.

Kocherlakota, N. R. (1990): "On tests of representative consumer asset pricing models," *Journal of Monetary Economics*, 26(2), 285–304.

Leith, C., and J. Malley (2007): "Estimated open economy New Keynesian Phillips curves for the G7," *Open Economies Review*, 18(4), 405–426.

Lindé, J. (2005): "Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach," *Journal of Monetary Economics*, 52(6), 1135–1149.

Lütkepohl, H. (2005): *New introduction to multiple time series analysis.* Springer.

Ma, A. (2002): "GMM estimation of the new Phillips curve," *Economics Letters*, 76(3), 411–417.

Martins, L. F., and V. J. Gabriel (2009): "New Keynesian Phillips Curves and potential identification failures: A Generalized Empirical Likelihood analysis," *Jornal of Macroeconomics*, 31(4), 561–571.

Mavroeidis, S. (2005): "Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37(3), 421–449.

——— (2006): "Testing the New Keynesian Phillips Curve without assuming identification," Working paper, Brown University.

Nason, J. M., and G. W. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23(5), 525–551.

Newey, W. K., and K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55(4), 703–708.

Pesaran, M. H. (1987): *The Limits to Rational Expectations.* Basil Blackwell, Oxford.

Rumler, F. (2007): "Estimates of the open economy New Keynesian Phillips Curve for Euro Area countries," *Open Economies Review*, 18(4), 427–451.

Sahuc, J.-G. (2004): "Partial indexation and inflation dynamics: what do the data say?," *Applied Economics Letters*, 11(13), 827–832.

Sbordone, A. M. (2002): "Price and unit labor costs: a new test of price stickiness," *Journal of Monetary Economics*, 49(2), 265–292.

Smets, F., and R. Wouters (2003): "An estimated dynamic stochastic general equilibrium model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Stahl, H. (2006): "Producer price adjustment at the micro level: evidence from individual price recordings underlying the German PPI," Discussion paper, Deutsche Bundesbank, mimeo.

Stock, J. H., and J. H. Wright (2000): "GMM with weak identification," *Econometrica*, 68(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 20(4), 518–29.

TAUCHEN, G. (1986): "Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data," *journal of business & Economic Statistics*, 4(4), 397–416.

TAYLOR, J. B. (1980): "Aggregate dynamics and staggered contracts," *Journal of Political Economy*, 88(1), 1–23.

TILLMANN, P. (2005): "The Calvo Model of Price Setting and Inflation Dynamics in Germany," *Applied Economics Quarterly*, 51(4), 307–321.

WALSH, C. E. (2003): *Monetary theory and policy.* The MIT Press, Cambridge, Massachusetts, 2nd edn.

Appendix

Data

The data are mainly taken from the national accounts database provided by the German Federal Statistical Office (known as Fachserie 18 Series 1.3). Additionally, data before reunification (prior 1991) are also available from the German Federal Statistical Office (Fachserie 18 Series p. 28). In detail the series are defined as:

- Inflation: The inflation measure is constructed as the first difference of the quarterly log GDP deflator. The GDP deflator is defined as the ratio of nominal GDP and real chain-weighted GDP.

- Real marginal costs: The labor income share is computed as the total compensation of employees divided by nominal GDP.

- Instruments: Additional instruments that are considered are wage inflation (Δw) defined as the first difference of the log of compensation of employees; output gap (y^{gap}) constructed recursively as percentage deviation of real GDP from an Hodrick-Prescott filtered trend with $\lambda = 1600$; the three month money market rate (r^s); the long-short interest rate spread ($r^l - r^s$) between the ten year government bond yield and the three month money market rate; and commodity price inflation (Δp^{comm}) constructed as the first difference of the log of the HWWA commodity price index (defined in euro).

Supplementary tables and figures

Table 1.5: Instrument selection based on information criteria

	AIC		SC	
	$\widehat{\pi}_{t+1}$	$\widehat{s}_t$	$\widehat{\pi}_{t+1}$	$\widehat{s}_t$
$\widehat{\pi}_{t-1}$	-0.012	0.060	-0.030	0.042
	[-0.12]	[1.34]	[-0.29]	[0.94]
$\widehat{s}_{t-1}$	-0.008	0.825	-0.012	0.824
	[-0.06]	[14.14]	[-0.09]	[13.71]
$\widehat{\pi}_{t-3}$	0.409	0.012	0.397	0.014
	[5.09]	[0.36]	[4.86]	[0.41]
$\widehat{\pi}_{t-4}$	0.125	0.099	0.109	0.088
	[1.53]	[2.82]	[1.33]	[2.50]
$\widehat{s}_{t-4}$	0.146	0.191	0.151	0.232
	[0.76]	[2.32]	[0.78]	[2.80]
$\widehat{\pi}_{t-5}$	0.003	0.150	-0.006	0.139
	[0.04]	[4.12]	[-0.07]	[3.73]
$\widehat{s}_{t-5}$	-0.120	-0.144	-0.102	-0.177
	[-0.70]	[-1.96]	[-0.59]	[-2.37]
y^{gap}_{t-1}	2.382	1.670	2.588	1.608
	[2.62]	[4.30]	[3.13]	[4.52]
$(r^l - r^s)_{t-1}$	0.277	-0.687	-0.066	-0.163
	[0.54]	[-3.15]	[-0.64]	[-3.72]
Δw_{t-1}	0.124	-0.031	0.137	-0.021
	[2.21]	[-1.30]	[2.46]	[-0.86]
r^s_{t-1}	0.518	-0.479	0.147	-0.076
	[1.16]	[-2.52]	[2.18]	[-2.61]
Δp^{comm}_{t-2}	0.009	0.004		
	[1.85]	[1.97]		
$(r^l - r^s)_{t-2}$	-0.324	0.515		
	[-0.66]	[2.45]		
r^s_{t-2}	-0.397	0.391		
	[-0.90]	[2.09]		
AIC	1.1094		1.1168	
SC	1.7429		1.6146	

Notes: t-statistics in brackets.

Table 1.6: Partial indexation model (unrestricted)

Instruments		β	θ	ξ	J	RMSC
GGL's set	(1)	0.996	0.632	0.309	8.877	-9.76
		(0.084)	(0.216)	(0.153)		
		[0.000]	[0.004]	[0.044]	[0.353]	
	(2)	1.047	0.980	-0.333	11.095	-1.63
		(0.040)	(54.43)	(0.072)		
		[0.000]	[0.986]	[0.000]	[0.196]	
AIC based	(1)	1.035	0.646	0.294	9.787	-9.18
		(0.062)	(0.217)	(0.147)		
		[0.000]	[0.003]	[0.045]	[0.550]	
	(2)	1.039	0.743	-0.178	12.125	-10.14
		(0.038)	(0.367)	(0.094)		
		[0.000]	[0.043]	[0.059]	[0.354]	
SC based	(1)	1.030	0.611	0.248	8.454	-10.51
		(0.058)	(0.181)	(0.156)		
		[0.000]	[0.001]	[0.112]	[0.390]	
	(2)	1.036	0.690	-0.182	10.820	-11.63
		(0.038)	(0.270)	(0.097)		
		[0.000]	[0.011]	[0.059]	[0.212]	

Notes: Standard errors in round brackets and p-values in square brackets. Rows (1) and (2) correspond to the two specifications of the orthogonality conditions eqs. (1.9) and (1.10) in the text, respectively. A 5-lag Newey-West HAC estimate was used. Sample period: 1973:1-2004:4.

Table 1.7: Rule-of-thumb model (unrestricted)

Instruments		β	θ	ω	J	RMSC
GGL's set	(1)	0.997	0.601	0.186	8.877	-11.11
		(0.072)	(0.224)	(0.097)		
		[0.000]	[0.007]	[0.056]	[0.353]	
	(2)	0.836	0.121	-0.019	16.050	-17.05
		(0.160)	(0.027)	(0.015)		
		[0.000]	[0.000]	[0.214]	[0.042]	
AIC based	(1)	1.030	0.621	0.182	9.787	-10.47
		(0.055)	(0.223)	(0.095)		
		[0.000]	[0.005]	[0.057]	[0.550]	
	(2)	0.883	0.181	-0.018	17.971	-16.20
		(0.102)	(0.033)	(0.020)		
		[0.000]	[0.000]	[0.353]	[0.082]	
SC based	(1)	1.026	0.590	0.146	8.454	-11.87
		(0.052)	(0.184)	(0.092)		
		[0.000]	[0.001]	[0.111]	[0.390]	
	(2)	0.908	0.178	-0.019	16.945	-16.83
		(0.104)	(0.035)	(0.020)		
		[0.000]	[0.000]	[0.340]	[0.031]	

Notes: Standard errors in round brackets and p-values in square brackets. Rows (1) and (2) correspond to the two specifications of the orthogonality conditions eqs. (1.13) and (1.14) in the text, respectively. A 5-lag Newey-West HAC estimate was used. Sample period: 1973:1-2004:4.

Table 1.8: Sensitivity to different values of A

	Partial indexation model				Model with Rule-of-Thumb firms			
	β	θ	ξ	J	β	θ	ω	J
$\alpha = 0.3$, $\mu = 1.25 \longrightarrow \epsilon = 5$, $A = 0.3182$								
(1)	1.030	0.690	0.248	8.454	1.027	0.669	0.165	8.454
	(0.058)	(0.151)	(0.156)		(0.053)	(0.158)	(0.101)	
	[0.000]	[0.000]	[0.112]	[0.390]	[0.000]	[0.000]	[0.101]	[0.390]
(2)	1.036	0.755	-0.182	10.820	0.970	0.335	-0.033	15.695
	(0.038)	(0.151)	(0.156)		(0.080)	(0.050)	(0.037)	
	[0.000]	[0.001]	[0.097]	[0.212]	[0.000]	[0.000]	[0.377]	[0.047]
$A = 1$								
(1)	1.030	0.805	0.248	8.454	1.028	0.788	0.195	8.454
	(0.058)	(0.096)	(0.156)		(0.055)	(0.106)	(0.118)	
	[0.000]	[0.000]	[0.112]	[0.390]	[0.000]	[0.000]	[0.112]	[0.390]
(2)	1.036	0.846	-0.182	10.820	1.030	0.635	-0.078	13.410
	(0.038)	(0.135)	(0.156)		(0.053)	(0.051)	(0.069)	
	[0.000]	[0.000]	[0.097]	[0.212]	[0.000]	[0.000]	[0.255]	[0.099]

Notes: see above. SC based instrument set.

Table 1.9: Restrictions in the partial indexation model

	$H_0: \beta = 0.99,\ \xi = 0$	
	LR-Test	p-value
(1)	0.0581	0.9714
(2)	4.3360	0.1144

Notes: SC instrument set.

Table 1.10: Restrictions in the rule-of-thumb model

	$H_0: \beta = 0.99,\ \omega = 0$	
	LR-Test	p-value
(1)	5.9409	0.0513
(2)	2.1815	0.3360

Notes: SC instrument set.

Table 1.11: Sensitivity to marginal cost (restrictions: $\beta = 0.99$, $\xi = 0$, $\omega = 0$)

$$\lambda = \frac{(1-0.99\theta)(1-\theta)}{\theta} A$$

	$A = 0.1750$		$A = 0.3182$		$A = 1$	
	λ	J	λ	J	λ	J
(1)	0.055	11.236	0.055	11.236	0.055	11.236
	(0.042)		(0.042)		(0.042)	
	[0.195]	[0.339]	[0.195]	[0.339]	[0.195]	[0.390]
(2)	0.663	14.873	0.445	14.136	0.259	13.054
	(0.178)		(0.117)		(0.072)	
	[0.000]	[0.137]	[0.000]	[0.137]	[0.001]	[0.221]

Notes: Standard errors are computed with the delta method.

Table 1.12: Partial indexation model with additional lags

	β	θ	ξ	ϕ_2	ϕ_3	ϕ_4	$H_0: \phi_2 + \phi_3 + \phi_4 = 0$	J
(1)	0.793	0.846	0.141	0.0911	-0.196	0.275	2.227	5.609
	(0.146)	(0.428)	(0.109)	(0.093)	(0.081)	(0.069)		
	[0.000]	[0.048]	[0.193]	[0.327]	[0.015]	[0.000]	[0.527]	[0.468]
(2)	0.825	0.868	0.046	0.113	-0.217	0.289	2.568	6.367
	(0.131)	(0.507)	(0.099)	(0.093)	(0.089)	(0.074)		
	[0.000]	[0.087]	[0.641]	[0.226]	[0.015]	[0.000]	[0.463]	[0.383]

Notes: SC based instrument set (plus inflation at the second lag).

Table 1.13: Rule-of-thumb model with additional lags

	β	θ	ξ	ϕ_2	ϕ_3	ϕ_4	$H_0: \phi_2+\phi_3+\phi_4=0$	J
(1)	0.798	0.831	0.118	0.0911	-0.196	0.275	2.227	5.609
	(0.146)	(0.454)	(0.094)	(0.093)	(0.081)	(0.069)		
	[0.000]	[0.067]	[0.211]	[0.327]	[0.015]	[0.000]	[0.527]	[0.468]
(2)	1.081	0.273	0.016	0.0120	-0.105	0.059	0.398	13.180
	(0.186)	(0.082)	(0.039)	(0.031)	(0.033)	(0.044)		
	[0.000]	[0.001]	[0.690]	[0.526]	[0.001]	[0.177]	[0.941]	[0.059]

Notes: SC based instrument set (plus inflation at the second lag).

Chapter 2

Does the cost channel matter for inflation dynamics? An identification robust structural econometric analysis*

2.1 Introduction

This chapter investigates whether the costs of external funds affect firms' marginal costs, thereby influencing the aggregate inflation rate. Recently, many authors – including Christiano, Eichenbaum, and Evans (2005), Chowdhury, Hoffmann, and Schabert (2006), Ravenna and Walsh (2006) and Tillmann (2008) – provide evidence of a cost channel relevant to inflation dynamics. This cost channel is introduced through the cost of working capital into the standard New Keynesian

*All computations used in this Chapter are done in MATLAB programs. Some computations are based on the MATLAB Optimization Toolbox and on replications files from Jim Stock's homepage.

model, which is motivated by cash-in-advance, i.e. factors of production, which have to be paid before the proceeds from the sale of output are received. Empirically, the existence of a cost channel can be tested by augmenting the New Keynesian Phillips curve with an interest rate variable as an additional regressor. So the cost channel implies an extension of the standard measure of marginal cost by interest rate effects.

Chowdhury, Hoffmann, and Schabert (2006) test such an augmented Phillips curve specification for G7 countries with GMM and find empirical support for this model for most of them. Ravenna and Walsh (2006) employ the same method, but instead of relying on the reduced form parameters, they estimate the structural parameters of a pure forward-looking specification for the U.S. and draw similar conclusions. The existence of a cost channel is also supported by methods of indirect inference (e.g., Christiano, Eichenbaum, and Evans, 2005; Huelsewig, Henzel, Wollmershaeuser, and Mayer, 2009). However, other studies cast doubt on the existence of a cost channel (e.g., Rabanal, 2007; Gabriel, Levine, Spence, and Yang, 2008). Their estimation includes both Bayesian Methods (Rabanal, 2007) and GMM (Gabriel, Levine, Spence, and Yang, 2008).

In this chapter, we extend the Phillips curve specification of Ravenna and Walsh (2006) to a model that allows for backward-looking behavior in price setting owing to partial indexation. Then we reexamine the existence of the cost channel by estimating structural form parameters for the U.S. (similar to Ravenna and Walsh, 2006). Instead of relying on a standard two-step GMM estimator, we use a continuous-updating GMM (CUE) estimator as proposed by Hansen, Heaton, and Yaron (1996). This estimator is preferable in terms of small sample properties, and it does not depend on the normalization of the orthogonality conditions. Moreover, we combine several additional tools in order to analyze the empirical model in great detail. We use identification robust econometric techniques, which can be readily based on the CUE objective function (as suggested by Stock and Wright, 2000). So this procedure guards against problems which are induced by weak instruments and which might be present in estimates of the new Phillips curve (see Ma, 2002; Mavroeidis, 2005; Dufour, Khalaf, and Kichian, 2006). Confidence intervals for the individual parameters are then computed by means of the projection technique. We also apply simulation techniques to analyze the complete distribution of the structural parameters. So that the available information of private agents can be used most effectively, factors are used as additional instrumental variables (as proposed by Bai and Ng, 2010; Kapetanios and Marcellino, 2010). To facilitate the comparison of non-nested models, we also make use of Andrews and Lu's (2001) model selection criteria for GMM estimation.

The results of this chapter indicate that empirical evidence of the cost channel is not as clear-cut as previously indicated in literature. Generally, the standard procedure of testing one parameter as significant (which is then interpreted as evidence of a cost channel) is found to be inappropriate owing to substantial identification problems. We find that the structure of the model does not allow for drawing strong conclusions about certain aspects of the model on account of weak identification. However, we are able to compare a standard Phillips curve model with the cost channel augmented version, with interesting findings. In fact, the estimated degree of nominal rigidity in a cost channel augmented Phillips curve model is more in line with economically plausible values, and turns out to be statistically more reliable. Additionally, model selection criteria clearly favor a model in which the cost channel is present and where the bank lending rate is included as an additional variable in the marginal cost measure.

Although distinguishing between forward-looking and backward-looking behavior proved difficult, we also document that backward-looking aspects (in the form of indexation) seem to be negligible in all the versions of the model. For the U.S. therefore, a purely forward-looking model that considers the cost channel is most compatible with the data. The inclusion of a cost channel can indeed improve the reliability of estimates of the New Keynesian Phillips curve by introducing the interest rate affecting real marginal costs. Interestingly, this model incorporates enough inflation persistence even without a lagged inflation term (which would in any case be difficult to motivate according to proper microfoundations).

The rest of the chapter is organized as follows. Section 2.2 introduces the theoretical model setup. The empirical strategy is outlined in Section 2.3. Section 2.4 presents the estimation results of the interest-rate augmented Phillips curve. Section 2.5 concludes.

2.2 The basic model

This section briefly introduces the theoretical model, which consists of a standard New Keynesian framework. More detailed derivations may be found in Walsh (2003) and Woodford (2003). We concentrate on the aspects necessary for characterizing inflation dynamics in the economy. The two basic model features are monopolistically competitive goods markets and sticky prices. In addition, the cost channel is introduced.

More precisely, the economy consists of a continuum of firms (indexed by $i \in [0,1]$), each producing a differentiated good $Y_t(i)$ according to a standard Cobb-Douglas production function

$$Y_t(i) = A_t \overline{K}_t(i)^{\alpha} N_t(i)^{1-\alpha}, \tag{2.1}$$

with A_t a common country-wide technological factor, $\overline{K}_t(i)$ the (fixed) firm-specific capital stock and $N_t(i)$ denoting the labor factor employed by firm i.

Each firm i faces a demand function characterized by the constant elasticity of substitution given by

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} Y_t, \tag{2.2}$$

where Y_t equals the aggregate demand, P_t is the aggregate price level in the economy and $P_t(i)$ is the price of good i charged by firm i. The price elasticity of demand for good i is characterized by the parameter ϵ (with $\epsilon > 1$). This determines the constant mark-up (defined as $\mu = \epsilon/(\epsilon - 1)$) required by firms over nominal marginal costs of inputs.

Next, we introduce a liquidity constraint for firms operating in their factor markets. Input factors like the wage bill have to be paid before revenues for the goods produced have been received. To meet these expenditures, firms have to borrow the outlays from a financial intermediary sector. In each period, the individual firm i is assumed to borrow the amount of $Z_t(i)$ to repay the sum total of the salaries. So the liquidity constraint is given by

$$Z_t(i) \geq W_t N_t(i),$$

with W_t the nominal wage rate and $N_t(i)$ the utilized labor factor of firm i. At the end of the period, when the produced good has been sold, firms have to repay these loans with interest to the amount of $i_t^l Z_t(i)$. With these liquidity constraints, firms' marginal costs are equal to

$$MC_t(i) = \frac{R_t^l W_t / P(i)_t}{(1-\alpha) Y_t(i)/N_t(i)} = \frac{R_t^l S_t(i)}{(1-\alpha)}, \tag{2.3}$$

where $R_t^l = 1 + i_t^l$ and $S_t(i)$ is the firm's specific labor share of production.

Further, we assume that firms face nominal price rigidities which can be characterized by Calvo's (1983) model of staggered price setting. This model implies that firms set prices infrequently owing to the costs of gathering information. The frequency of price re-optimizations is characterized by a stochastic process, with the constant probability that a firm changes its price at one particular point in time. So on the aggregate level at each point in time there is a fraction of firms' $1-\theta$ that optimally adjusts prices. The expected waiting period is then given by $1/(1-\theta)$.

Price re-optimizing firms that set their optimal price $P_t^*(i)$ are faced with the following dynamic maximization problem:

$$E_t \sum_{k=0}^{\infty} (\beta\theta)^k \upsilon_{t,t+k} \left[P_t^*(i) X_{t,t+k} - MC_{t,t+k}(i) \right] \frac{Y_{t+k}(i)}{P_{t+k}}, \tag{2.4}$$

subject to the demand constraints eq. (2.2) and

$$X_{t,t+k} = \begin{cases} \prod_{l=0}^{k-1} \overline{\pi}^{1-\xi} \pi_{t+l}^{\xi} & \text{for } k > 0 \\ 1 & \text{for } k = 0. \end{cases} \tag{2.5}$$

with β a constant discount factor, $\upsilon_{t,t+k} = U'(C_t)/U'(C_{t+k})$ the time-varying portion of the discount factor between t and $t+k$; with $U'(C_t)$ being the marginal utility of consumption. $\overline{\pi}$ denotes the long-run average gross rate of inflation. Whenever a firm does not re-optimize its price, it is reset according to an indexation scheme. $\xi \in [0,1]$ measures the degree of indexation to past inflation rates. Note that this partial indexation scheme nests more specific indexation assumptions as special cases.[1]

[1] This specification is adopted from Smets and Wouters's (2003). With $\xi = 1$ it equals

As shown by Walsh (2003), aggregate inflation $\hat{\pi}$ can be related to average real marginal cost $\widehat{mc}$ according to

$$\hat{\pi}_t = \gamma_f E_t \hat{\pi}_{t+1} + \gamma_b \hat{\pi}_{t-1} + \lambda \widehat{mc}_t, \tag{2.6}$$

where

$$\begin{aligned}
\lambda &= \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta} \frac{1-\alpha}{1+\alpha(\epsilon-1)}, \\
\gamma_f &= \frac{\beta}{1+\beta\xi}, \\
\gamma_b &= \frac{\xi}{1+\beta\xi}.
\end{aligned}$$

This inflation equation is known as the *Hybrid New Keynesian Phillips curve.* Its reduced form coefficients γ_f, γ_b and λ are non-linear functions of the structural parameters β, θ, ξ, α and ϵ.[2] When $\xi = 0$, the equation reduces to the pure forward-looking New Keynesian Phillips curve. When the cost channel is introduced, real marginal costs depend not only on the labor share of output (as derived by Galí and Gertler, 1999) but also on the nominal interest rate:

$$\widehat{mc}_t = \hat{R}^l_t + \hat{s}_t,$$

where $\hat{s}_t = \hat{w}_t + \hat{n}_t - \hat{y}_t$ is the log deviation of the labor share around the steady state and $\hat{R}^l_t$ is the percentage point deviation of the nominal interest rate (defined as the lending rate) around its steady state value.

According to Chowdhury, Hoffmann, and Schabert (2006), it is assumed that the lending rate R^l_t can deviate from the nominal interest rate set by monetary policy which is denoted by R^m_t. This is motivated by financial market imperfections and is, for instance, also motivated by the likelihood of defaults. Adopting the simplified framework of Chowdhury, Hoffmann, and Schabert (2006), where profit

Christiano, Eichenbaum, and Evans's (2005) dynamic indexation scheme, with $\xi = 0$ it simplifies to a pure forward-looking model with an indexation to trend inflation.

[2] Note that there are also other versions of the structural Phillips curve that have a slightly different interpretation (e.g., Galí and Gertler, 1999). As shown by Scheufele (2010), Galí and Gertler's (1999) model leads to conclusions similar to those considered here.

maximization of financial intermediaries leads to a log linear relationship between the risk-free rate $\hat{R}^m$ (which is assumed to be under the control of monetary policy) and the lending rate $\hat{R}^l$. This is given by

$$\hat{R}_t^l = (1 + \psi_R)\hat{R}_t^m, \tag{2.7}$$

where the coefficient $1+\psi_R$ measures the response of the lending rate $\hat{R}_t^l$ to changes in the monetary policy rate $\hat{R}_t^m$. As illustrated by Chowdhury, Hoffmann, and Schabert (2006) for $\psi_R > 0$, indicating the existence of strong financial market imperfections. When the opposite holds ($\psi_R < 0$), then management costs are very high.

Now we can express the Phillips curve as a function of the labor share as well as of the monetary policy rate, which is given by

$$\hat{\pi}_t = \gamma_f E_t \hat{\pi}_{t+1} + \gamma_b \hat{\pi}_{t-1} + \lambda \hat{s}_t + \lambda \phi_m \hat{R}_t^m, \tag{2.8}$$

where $\phi_m = (1 + \psi_R)$. So the idea of the cost channel of monetary transmission follows directly from this equation: whenever the central bank raises its interest rate above its steady state level, this leads to an increase in the current inflation rate over its steady state value. This holds true unless this effect is not over-compensated by the response of the labor share through adjustments of aggregate demand.

2.3 Empirical analysis

Next we introduce the strategy for estimating the interest-rate augmented Phillips curve specification and how it is possible to conduct inference about the parameters of interest. Generally, the choice is between two different econometric methods: full information or limited information methods. The choice between these categories has a long history in econometrics. Full information methods provide the complete range of statistical properties associated with the model under investigation. This is preferable in terms of efficiency, unless the model is correctly specified. Limited information methods do not require a fully specified model, but it is enough to set up certain moment conditions to estimate the parameters of interest. Thus there is the classical trade-off between efficiency and the sensitivity to model misspecifications known from simultaneous equations models. Since we are interested solely in the Phillips curve equation and more specifically in the direct impact of interest rates on inflation, we find it more convenient to use limited information methods, because we do not want to restrict our results to a particular model structure.[3] Additionally, the great advantage of limited information methods is that identification issues for these types of techniques are now well established. On the other hand, there is no widely accepted method in full information for dealing with problems of weak identification.

2.3.1 The basic econometric specification

Limited information methods typically require the application of instrumental variable (IV) estimation methods. To get an empirical traceable specification from the theoretical model eq. (2.8), the unobserved variable $E_t\hat{\pi}_{t+1}$ is replaced by its realization assuming the forecasting error $\eta_{t+1} = [E_t\hat{\pi}_{t+1} - \hat{\pi}_{t+1}]$ to be orthogonal to past information. So we obtain the estimable equation

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1} + \frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\left(\hat{s}_t + \phi_m\hat{R}_t^m\right) + u_t, \qquad (2.9)$$

where $u_t = \nu_t + \gamma_f\eta_{t+1}$. We allow the error term to follow a very general structure - so u_t may be autocorrelated and / or heteroskedastic.[4] The natural setup for estimating potentially non-linear dynamic models is to employ GMM as proposed by Hansen (1982). With the assumption $E_{t-1}u_t = 0$, the moment conditions are given by $E_{t-1}\{f_t(\vartheta)\}$, where $f_t(\vartheta) = u_t(\vartheta)\mathbf{z_{t-1}}$ with $\mathbf{z_{t-1}}$ the vector of instruments including predetermined variables dated $t-1$ or earlier. $\vartheta = (\beta, \theta, \xi, \alpha, \epsilon, \phi^i)$ denotes the parameter vector of interest.

When it comes to parameter estimation, we do not consider the convenient

[3] Since we also stress the importance of identification robust inference, full information methods like ML are not immune from that kind of problem. However, certain LI methods are able to deal with them. For this reason, once weak identification problems appear, ML, with its asymptotic theory, can generally not be relied on, and there are no full information methods that are identification robust.

[4] When we assume ν_t to be white noise and use instruments dated $t-1$ or earlier, then u_t follows an MA(1) process per construction. See Appendix for a discussion.

two-step GMM (2GMM) estimator that is frequently used for estimating NKPC models (see e.g., Galí and Gertler, 1999; Galí, Gertler, and López-Salido, 2001; Eichenbaum and Fisher, 2007). Instead, we adhere to the continuous updating GMM (CUE) estimator as proposed by Hansen, Heaton, and Yaron (1996). This estimator is superior in terms of finite sample properties (Hansen, Heaton, and Yaron, 1996; Stock and Wright, 2000). It is more closely related to LIML than to 2SLS (which is related to the 2GMM estimator).[5] Moreover, it does not share the property of standard GMM that estimation bias increases with the inclusion of irrelevant instruments (Tauchen, 1986; Kocherlakota, 1990, as documented by). For non-linear settings, another favorable property is its insensitivity to the statement of the moment conditions. Since the NKPC in its structural formulation is non-linear in its parameters, it is possible to reformulate the orthogonality conditions, for instance through multiplying by a certain parameter. 2GMM estimates may be sensitive to this kind of transformation (see Hall, 2005, for a general discussion and Scheufele, 2010, for this problem in the context of the NKPC).

The CUE estimates can be obtained by minimizing the objective function

$$\mathbf{S}(\vartheta) = \left[\frac{1}{T}\sum_{t=1}^{T} f_t(\vartheta)\right]' W(\vartheta)^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} f_t(\vartheta)\right], \tag{2.10}$$

where $W(\vartheta)$ is a $k \times k$ positive semi definite weighting matrix. It can be shown that the weighting matrix is given by the inverse of the asymptotic variance matrix S_T^{-1}. This matrix is computed to be heteroscedastic and autocorrelation consistent (HAC), as proposed by Newey and West (1987). The peculiarity of this estimator is that the covariance is estimated together with the parameter vector ϑ. Instead, the 2GMM computes first an initial estimate of ϑ with a pre-specified weighting matrix (e.g., the identity matrix) and then uses this initial estimate to specify the weighting matrix in the second step.

To estimate the structural form parameters, it is necessary to calibrate some parameters, as when there are four variables one can at least identify the same number of parameters. Like Galí, Gertler, and López-Salido (2001), we choose to

[5]For the estimation of a single equation in the linear simultaneous equation model, the two-step GMM estimator is 2SLS, whereas the continuous updating estimator is LIML. The superior characteristics of LIML in comparison with 2SLS in finite samples has been well documented in the literature (see e.g., Judge, Griffiths, Hill, Luetkepohl, and Lee, 1985, Chapter 15).

calibrate α and ϵ (like them, we set $\alpha = 0.270$ and $\epsilon = 11$ for the U.S. economy). Given these values, the point estimates for β, θ, ξ and ϕ_i can be computed.

2.3.2 Coping with potentially weak instruments

Once the empirical model has been set up, it must be remembered that the GMM approach is extremely susceptible to problems brought about by weak instruments (see e.g., Stock, Wright, and Yogo, 2002; Dufour, 2003; Andrews and Stock, 2005, for recent surveys). These may arise whenever the instruments are not sufficiently correlated with the variables they instrument. This pathology results in parameter distributions that may be far from normality. This leads standard test statistics (e.g., Wald and significance tests) to spurious over-rejections and may lead to wrong conclusions.

Broadly speaking, weak instrument problems do result in identification difficulties for relevant parameters. For example, consider the standard linear instrumental variable (IV) model in the extreme case of an exact zero correlation of the instruments with the right-hand side endogenous variable(s). Obviously, this is a violation of the rank condition for identification (implying that the coefficient in the structural equation is unidentified). However, a correlation of exactly zero is never seen. So problems occur even when the correlation between the instruments and the instrumented variables is not strong enough (then the parameters are close to being unidentified).

It is also important to stress that the problem of weak identification is not specific to GMM or IV estimation. Instead, ML-methods or other methods of Matching Moments may also be affected by this pathology. Because of their high-dimensionality and their non-linear structure, the parameters of DSGE models have generally been found difficult to identify (see e.g., Canova and Sala, 2009, for a general discussion for this aspect).

In order to avoid the problems of weak identification, we combine two empirical strategies. First, we carry out a careful selection of relevant instruments $\mathbf{z}$. Like Bai and Ng (2010) and Kapetanios and Marcellino (2010), we apply Factor-GMM estimation, whereby factors are extracted from a large data set of macroeconomic indicators and used as additional instrumental variables. The advantage here is that instruments can be used parsimoniously, but factors still provide a great deal of informational content. Our second strategy is to rely on identification robust inference methods that are robust to the problems associated with weak instruments. These methods have been successfully applied to the standard NKPC by Ma (2002), Dufour, Khalaf, and Kichian (2006), Mavroeidis (2007), Nason and Smith (2008), Martins and Gabriel (2009) and Kleibergen and Mavroeidis (2009b) and have the advantage that inference based on these methods are valid, whether or not identification difficulties arise. In certain situations, particularly when some parameters are well identified and others are not, these methods allow for important insights.

Instrument selection

When setting up the moment conditions for the GMM approach, one has to be specific about the choice of instrument variables $\mathbf{z}$. Rational expectations models such as our Phillips curve specification suggest that all information available at the time when the forecast is made can be used as valid instruments. Because of simultaneity and potential publication lags, we consider only instruments dated time $t-1$ or earlier.

Since the potential instrument set contains the full information set of private agents, the number of candidate instruments is infinite. This really means that any measured past variable can serve as a potential instrument. Additionally, there is no theoretical justification or practical guide for deciding which instruments to use for estimation. However, early Monte Carlo experiments for GMM suggest that using as many instruments as possible is not a good idea. Instead, one should be somewhat selective when choosing $\mathbf{z}$ (see Tauchen, 1986; Kocherlakota, 1990). In empirical macroeconomic applications it has become standard practice for instruments to be chosen relatively unsystematically from among recent macroeconomic variables that are thought to predict the instrumented variables well.

The success of IV methods clearly depends on the quality of instruments, so it is logical to put some effort into careful selection of the instrument set in order to get more reliable results. We therefore adopt the Factor-GMM approach by Bai and Ng (2010) and Kapetanios and Marcellino (2010), who employ principle components obtained from a large data set as potentially relevant instruments. Based on the idea of Stock and Watson (2002) that the information of large datasets can be summarized by a few factors, those factors should reflect most of the private agents' information set. Additionally, Stock and Watson (1999) show that those factors are useful in forecasting inflation, which suggests that factors may be relevant instruments (since they are correlated with future inflation).

More precisely, let us assume that there are N potential instrumental variables x_t available and that these are generated by the factor model:

$$x_t = \Lambda' g_t + \nu_t, \tag{2.11}$$

where it is assumed that the number of static factors r is much smaller than N.[6] Consequently, the factors g_t are natural instrument choices. In addition to the factors, we also allow key macroeconomic indicators and lags of the endogenous variables as part of the instrument set. As potential additional indicators, we consider variables used in other studies (see Galí and Gertler, 1999; Ravenna and Walsh, 2006).

The second step is to apply a general-to-specific modeling strategy to eliminate redundant instruments (those which correlate only marginally with the variables they instrument). This is carried out separately for each instrumented variable.[7] Those variables that remain in one of the equations enter into the final instrument set.[8] If only the most relevant instrumental variables are considered in the estimation process, it promotes identification robust methods on account of power gains. The exclusion of redundant instruments can thus help to avoid the well-documented power loss of the AR statistic when the number of instruments increases (Andrews and Stock, 2005).

Identification robust inference

Conducting inference of the parameters of interest relying on standard Wald-type tests and t-statistics is usually done in the standard GMM framework, which is

[6] As a decision rule, we consider all factors that explain at least 10% of the overall variance.

[7] The general-to-specific approach always deletes the least significant variable from the equation until all remaining p values are below a pre-specified threshold (which we set at 0.1).

[8] We also consider a system-based approach whereby all equations are considered jointly and a blockwise elimination is performed (until a certain information criterion is minimized). However, this approach results in very few instruments (sometimes with fewer instruments as instrumented variables), whereas the equation strategy leaves us with slightly more instrumental variables (7 indicators) which are more convenient to work with.

problematic when identification difficulties occur. Typically, the GMM estimate is treated as if

$$\hat{\vartheta} \approx N(\vartheta_0, \hat{V}/T)$$

where

$$\hat{V} = \left(\hat{D}\ \hat{S}_T^{-1}\hat{D}'\right)^{-1} \quad \text{and} \quad \hat{D} = 1/T\sum_{t=1}^{T}\frac{\partial f_t(\vartheta)}{\partial \vartheta'}|_{\vartheta=\hat{\vartheta}} \tag{2.12}$$

and S_T is the long-run covariance matrix (see Hamilton, 1994, Chapter 14). From that, standard $(1-\alpha)100$ confidence intervals for the individual parameter i can be computed as

$$\hat{\vartheta}_i \pm z_{\alpha/2}\sqrt{\hat{V}_{ii}/T},$$

where $\hat{V}_{ii}$ is the $i-i^{th}$ element of the matrix $\hat{V}$ given in eq. (2.12) and $z_{\alpha/2}$ is the 100(1-$\alpha/2$) percentile of the standard normal distribution. One strategy increasingly applied in IV and GMM estimation is to pre-check whether the instruments are strong enough (see e.g., Hahn and Hausman, 2002; Stock and Yogo, 2005). If this is the case, standard methods of inference are applied as outlined above.

In contrast, this study adopts a different perspective and applies identification robust methods of inference. We concur with the arguments by Dufour (2003), Andrews and Stock (2005) and Kleibergen (2007), who have shown that using conventional inference methods after pretesting for identification is both unreliable and unnecessary. The unreliability resides in the fact that the size of such a two-step procedure cannot be controlled. The sequential procedure is unnecessary because identification robust methods are as powerful as the standard methods when instruments are strong and more powerful than the two-step procedures when instruments are weak (see Kleibergen and Mavroeidis, 2009b). Moreover, identification robust methods can be helpful only when certain parameters are unidentified or close to unidentified, while other parameters are well identified.

As shown by Mavroeidis (2005), identification of the NKPC for economic plausible parameter values is challenging, and weak instrument problems are very likely to occur. This view is supported empirically by numerous studies (Ma, 2002; Dufour, Khalaf, and Kichian, 2006; Mavroeidis, 2007; Martins and Gabriel, 2009; Kleibergen and Mavroeidis, 2009b; Nason and Smith, 2008) comparing weak instrument robust tests with standard Wald-type tests obtained with GMM. We essentially adopt their idea, and do not assume a priori that the parameters are identified. We used **S**-sets, as proposed by Stock and Wright (2000) and applied to the NKPC by Ma (2002) and Mavroeidis (2007), which can be constructed from the CUE objective function (see eq. 2.10). This method has similar important characteristics, described by Dufour (1997) as identification robust. This requires unbounded confidence intervals (thus uninformative) whenever parameters are unidentified. In the situation where parameters are weakly identified, this should translate into fairly large confidence sets. As shown by Dufour (1997), this is not the case with standard Wald-type methods, which hold only when identification is fully guaranteed and when there are no weak instrument problems. Other than this, these methods are unreliable and standard normal approximations provide a very unsatisfactory guide for inference.

The **S**-sets used for constructing confidence sets are very close to the well-known overidentification test of Anderson and Rubin (1949). Several authors (including e.g., Dufour, 1997; Stock, Wright, and Yogo, 2002; Dufour, 2003; Andrews and Stock, 2005; Dufour and Taamouti, 2005, 2007) provide evidence that this statistic is fully robust to weak instrument problems. When it comes to linear simultaneous equation models, Stock and Wright (2000) have shown that **S**-sets are asymptotically equivalent to confidence sets obtained by inverting the Anderson-Rubin (AR) statistic. So **S**-sets can be seen as an extension for the AR test in linear models to GMM as a more general model class. To obtain **S**-sets, that is, a joint confidence set for the parameter vector ϑ, we use Stock and Wright's (2000) result that $S(\vartheta_0) \xrightarrow{D} \chi^2_k$, where $S(\vartheta_0)$ is the CUE objective function eq. (2.10) evaluated at the true parameter values ϑ_0 and k is the number of instruments. The joint confidence interval consists of those parameter values for which the test statistic does not reject. Thus an asymptotically valid $100(1-\alpha)\%$ confidence interval for the parameter vector ϑ is given by

$$\{\theta : T \times \mathbf{S}(\vartheta) < c_k(\alpha)\}, \tag{2.13}$$

where $c_k(\alpha)$ denotes the $100(1-\alpha)\%$ percentile of the χ^2_k distribution.[9]

This procedure can be applied to both the reduced form parameters γ_f, γ_b, λ and ϕ^i and the structural parameters β, θ, ξ and ϕ^i (given the calibrated values for α and ϵ). The resulting **S**-set is four-dimensional for the full model specification.[10]

Confidence intervals for the individual parameters are obtained by using the projection method. The idea is that projection-based tests do not reject the individual hypotheses $H_0 : \beta = \beta_0$ when the joint hypothesis $H^* : \beta = \beta_0, \alpha = \alpha_0$ do not reject for some values of α. This test method is proposed by Dufour (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005) and Dufour and Taamouti (2007). This procedure is fully robust to weak instruments but its drawback is that projection-based tests are conservative.[11]

A further characteristic of identification robust confidence intervals based on the CUE objective function is that they may be empty. This is the case when the test rejects for all possible parameter values. Thus, **S**-sets already include a test of overidentified restrictions comparable to a J test as proposed by Hansen. If no parameter vector is compatible with the specified model, the corresponding confidence sets will be empty. We interpret these results as a rejection of the empirical model.

[9] The construction of joint confidence intervals involves searching for values within an economically plausible range and collecting those values which the test does not reject. This is done by means of a grid search procedure.

[10] Additional methods are now available for dealing with weak instrument problems within the GMM setting (see Kleibergen and Mavroeidis, 2009b, for a comparison of different IV robust methods with application to the Phillips curve). Kleibergen and Mavroeidis (2009b) consider not only **S**-sets, but also a score Lagrange Multiplier (KLM) test, the difference between **S**-sets and the KLM statistics (JKLM) and an extension of the conditional likelihood ration test of Moreira (2003) to GMM (MQLR). Their simulation results indicate that the MQLR is at least as powerful as any of the other tests. However, while MQLR dominates the **S** statistics under some conditions in terms of power, it also imposes additional restrictions on the reduced form models and may consequently be more fragile. This could translate into problems when relevant instruments are missing (this point was raised by Dufour, 2009).

[11] Another available approach can be applied to parameter subsets (see Stock and Wright, 2000; Kleibergen and Mavroeidis, 2009a). In this case, some parameters are assumed to be identified. As long as the assumptions are satisfied, these tests are asymptotically non-conservative and more powerful than projection-based tests. However, when estimating the structural parameters, the parameter space is bounded. This implies that the restricted estimates may fall on the boundary, which violates the conditions for subset tests (Kleibergen and Mavroeidis, 2009b).

An MCMC approach for calculating parameter uncertainty

We also present an alternative to the projection method to approximate the estimation uncertainty of the individual parameters. We therefore make use of simulation techniques as originally proposed in the Bayesian literature. This enables us to systematically characterize the shape of the GMM objective function in situations where non-linearities and dimensionality complicate traditional methods of inference (and where the projection technique is less powerful). We concur with Chernozhukov and Hong (2003) by constructing

$$\varphi_T(\vartheta) \propto \exp\left[-\frac{1}{2}\mathbf{S}(\vartheta)\right],$$

with $\mathbf{S}(\vartheta)$ the CUE objective function. The right-hand side function is scaled so that

$$\int \varphi_T(\vartheta)\mathrm{d}b = 1.$$

Now we make use of MCMC (Markov chain Monte Carlo) methods to summarize $\varphi_T(\vartheta)$ and hence $\mathbf{S}(\vartheta)$. Typically, MCMC methods are used in Bayesian analysis in conjunction with likelihood functions. In this application we use the CUE criterion function instead, and run random parameter searches to evaluate its properties (see Hansen, Heaton, Lee, and Roussanov, 2007, for a similar application based on CUE on the intertemporal elasticity of substitution).[12]

Although we do not give a Bayesian interpretation, we may infer "marginals" for individual components of the parameter vector by averaging out the remaining components. The integration method contrasts with the standard practice of concentration inferior to standard methods of inference, where minimization and computing derivatives at minimized values are required. In applying the simulation technique based on the **S**-sets we hope to gain more insight into the individual parameter uncertainty (since we get the whole distribution) than if we were to rely only on the projection-based confidence intervals.

[12] See Appendix for detailed information.

2.3.3 Selecting among candidate models

So far, we have concentrated on methods of inference for one particular model. However, in comparing different models in relation to one economic phenomenon, it is necessary to choose from among different candidate models. When one model is nested within another, it is possible to test for these parameter restrictions given the methods described above. This is, e.g. the case for deciding whether the Phillips curve model (eq. 2.9) is purely forward-looking. This translates into a test of $H_0: \ \xi = 0$.

However, if this is not the case, i.e. meaning that one model is not a special case of the other, the models are *non-nested* and additional methods have to be employed. Non-nested tests for GMM have been proposed by Singleton (1985) Ghysels and Hall (1990) and Smith (1992). Since none of these approaches is really satisfactory (Hall, 2005), we apply the relatively simple model selection criteria proposed by Andrews and Lu (2001). These information criteria can be seen as GMM counterparts of the likelihood information criteria (BIC, AIC,...) and are based on the J-test statistic for testing over-identifying restrictions (see Hansen, 1982).[13]

Defining the moment and model selection criteria (MMSC) as

$$\text{MMSC}_T = J_n(b) - h(k - |b|)\kappa_T, \tag{2.14}$$

where $J_T(b)$ is the test statistic of the over-identification test, given parameters b. Let $|b|$ denote the number of parameters to be estimated given b and k the number of moment conditions. Comparable to likelihood based criteria $h(k - |b|)\kappa_T$ is a "bonus term" that penalizes the increasing number of estimated parameters and rewards the utilization of more over-identifying restrictions.

Standard examples of MMSC are the BIC and AIC criteria for model selection. Those are defined as

$$\begin{aligned}
&\text{MMSC-BIC:} \quad \kappa_T = \ln T \text{ and } \text{MMSC}_{BIC,T} = J_n(b) - (k - |b|)\ln T,\\
&\text{MMSC-AIC:} \quad \kappa_T = 2 \text{ and } \text{MMSC}_{AIC,T} = J_n(b) - 2(k - |b|).
\end{aligned}$$

[13] Note that the J-test in this setting is based on the continuous updating GMM function, not on the standard two step procedure.

Typically, the MMSC measures describe a trade-off between the magnitude of the J statistic and the number of parameters (and moment conditions) employed. The model with the lowest MMSC value is to be preferred. As shown by Andrews and Lu (2001), the selection procedures can help to specify a model, where the MMSC-BIC in particular is found to work quite well for this purpose.

2.3.4 Data

We use quarterly time series data with a sample period ranging from 1960q1-2005q4 to estimate eq. (2.9) for the U.S. economy. The data are taken from the OECD Quarterly National Accounts database, IMFs International Financial Statistics (IFS) and the indicator database provided on Mark Watson's homepage. Inflation is defined as the quarterly log difference of the GDP deflator. Real marginal cost is proxied by the labor share of output, which is defined as the ratio of total compensation to nominal GDP. As a measure for the short-run nominal interest rate, two definitions are considered: 3-month Treasury bill rates and bank lending rates.[14] Both explanatory variables – labor share and interest rates – are defined as percentage deviations of a steady state value, while inflation rate is expressed as percentage point deviations.

The potential instrument set is composed of lags of inflation, the labor share and short-term interest rates (up to four lags are considered). Additional instruments consist of a yield spread, $(r^l - r^m)_t$, defined as the 10-year government bond yield minus the 3-month Treasury bill rate, wage inflation Δw_t and a quasi-real time detrended output gap $\tilde{y}_t$ (which is computed recursively and contains information only up to period t). These additional instruments contain up to two lags in the instrument selection step. The factors are extracted using the Stock and Watson database covering 108 macroeconomic variables over the full sample period. We take the four principle components $(g^1, ..., g^4)$, which provide the largest explanatory power (at least 10% of the overall variance) as candidate instruments. The step-wise selection approach applied for each instrumented variable results in 17 selected instrument variables (from a candidate set of 24 candidates) plus a constant. This instrument set includes

[14] Bank lending rates for the U.S. are taken from the IMFs International Financial Statistics (IFS).

$$\mathbf{z}_{t-1}^{opt} = [c\ \hat{\pi}_{t-1}\ \hat{\pi}_{t-2}\ \hat{\pi}_{t-3}\ \hat{\pi}_{t-4}\ \hat{s}_{t-1}\ \hat{R}_{t-1}^{i}\ \hat{R}_{t-2}^{i}\ \hat{R}_{t-3}^{i}\ \hat{R}_{t-4}^{i}$$
$$(r^l - r^m)_{t-1}\ \tilde{y}_{t-2}\ g_{t-1}^1\ g_{t-2}^1\ g_{t-1}^2\ g_{t-2}^2\ g_{t-1}^3\ g_{t-2}^3]'. \quad (2.15)$$

As a robustness check we also consider an instrument set similar to Ravenna and Walsh's (2006) set, including

$$\mathbf{z}_{t-1}^{rw} = [c\ \hat{\pi}_{t-1}\ \hat{\pi}_{t-2}\ \hat{\pi}_{t-3}\ \hat{\pi}_{t-4}\ \hat{s}_{t-1}\ \hat{s}_{t-2}\ \hat{R}_{t-1}^{i}\ \hat{R}_{t-2}^{i}\ \hat{R}_{t-3}^{i}\ \hat{R}_{t-4}^{i}$$
$$\Delta w_{t-1}\ \Delta w_{t-2}\ \tilde{y}_{t-1}\ \tilde{y}_{t-2}\ (r^l - r^m)_{t-1}\ (r^l - r^m)_{t-2}]'. \quad (2.16)$$

The main differences between these two sets is the use of factors in the instrument set $\mathbf{z}_{t-1}^{rw}$ from eq. (2.15). The size is similar.

2.4 Estimation results

In the following, we present the baseline results for the econometric model as outlined in section 2.3. We therefore start by considering the most general specification (denoted as I_{Full}), including the cost channel measured by the Treasury bill rate (in line with Chowdhury, Hoffmann, and Schabert, 2006; Ravenna and Walsh, 2006). We assume partial indexation (measured by ξ) as well as the existence of real rigidities (reflected in the calibrated term $\frac{1-\alpha}{1+\alpha(\epsilon-1)}$). Results are displayed only for the structural form, where structural parameters are estimated directly from the non-linear equation. The advantage of this is that economically relevant quantities become directly apparent. In particular, the degree of nominal rigidities, reflected in θ and the corresponding average frequency of price reoptimization (or adjustment), can be easily deduced. However, the reduced form coefficients can be easily computed once the structural parameters have been determined. So one can compare the results directly to Chowdhury, Hoffmann, and Schabert's (2006) findings. They used the same explanatory variables but estimated the linear form.[15]

[15] Owing to the application of the CUE (instead of the 2GMM) the point estimates of the directly estimated reduced form coefficients would be exactly the same as those recalculated from the structural parameters.

Table 2.1 provides the CUE point estimates together with identification robust 90% confidence intervals (**S**-Sets) calculated by using the projection method. For comparison, we also display standard Wald-type confidence intervals with the same level of significance. Further, the average frequency of price reoptimization is calculated from the Calvo parameter θ together with its confidence interval. In addition, p-values associated with the CUE point estimates are provided, which can be interpreted in the same way as the standard test for overidentifying restrictions.[16] To compare different model specifications (which may or may not be nested), we present the GMM information criterion MMSC-BIC for each model. As outlined above, models with lower MMSC-BIC values are to be preferred. All specifications presented in Table 2.1 are based on the same degree of real rigidity; namely $\alpha = 0.27$ (the firm specific capital share) and $\epsilon = 11$ (which implies a steady state mark-up of 10%). Further, the West (1997) HAC estimator with an MA(1) process is used in estimation.

Turning to the results for the full specification I_{Full}, we obtain a CUE parameter vector for $(\hat{\beta}, \hat{\theta}, \hat{\xi}, \hat{\phi}^m)$ of $(0.97, 0.66, 0.15, 1.25)$. The estimates for β, θ and ϕ^m are very close to Ravenna and Walsh's (2006) results, although we allow for partial indexation (which provides a rationale for including a lagged inflation term) and we used a different instrument set. The implied average frequency of price adjustment is roughly three quarters, which is basically in line with the literature. However, when turning to the uncertainty of the estimates reflected by the **S**-sets, it is obvious that the length of the identification robust confidence interval for ϕ^m is infinite.[17] This extreme finding results from the fact that we cannot rule out the case of $\theta = 1$. Economically, this is the case of total price rigidity where prices are never reoptimized (which also results in a Phillips curve slope of zero).

[16]Note that the CUE point estimates can also be interpreted as Hodges-Lehmann estimators (which are the least-rejected values and can be interpreted as point estimates). It may be the case that the CUE objective function evaluated at the least-rejected values (the CUE estimates) exceeds the $100(1-\alpha)$ percentile of the χ^2_k. This case results in an empty confidence set (this happens when the associated p-value is below 0.1) and can be interpreted as a rejection of the model rather like the Hansen test of overidentification. The main difference between the two is that the Hansen test employs a χ^2_{k-s} distribution, with $k-s$ degrees of freedom (where s is the number of estimated parameters), as opposed to k.

[17]Note that a confidence interval of infinity reflects the fact that this parameter is unidentified (given the model structure and the data). As emphasized by Dufour (1997), this feature of allowing for unbounded confidence intervals whenever parameters are unidentified is exactly what characterizes identification robust procedures.

Table 2.1: Estimates of the structural parameters

Model specification:

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1}\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)}\left(\hat{s}_t + \phi_i\hat{R}_t^i\right) + u_t$$

	$\mathbf{I}_{\text{Full}}$	$\mathbf{II}_{\text{NKPC}}$	$\mathbf{III}_{\text{RW}}$	$\mathbf{IV}_{LR}$	$\mathbf{V}_{\text{RW-LR}}$
β	0.9704	1.0110	0.9793	0.9637	0.9645
	(0.92,1.03)	(0.96,1.06)	(0.93,1.03)	(0.92,1.01)	(0.92,1.01)
	[0.84,1.00]	[0.92,1.00]	[0.84,1.00]	[0.88,1.00]	[0.86,1.00]
θ	0.6567	0.6212	0.6369	0.6327	0.6184
	(0.47,0.85)	(0.44,0.80)	(0.53,0.74)	(0.58,0.68)	(0.56 ,0.68)
	[0.45,1.00]	[0.45,1.00]	[0.42,1.00]	[0.55,0.95]	[0.55,1.00]
ξ	0.1528	0.1138	0	0.2067	0
	(0.01,0.30)	(-0.08,0.21)		(0.05,0.36)	
	[0.00, 0.60]	[0.00,0.70]		[0.00,0.65]	
ϕ^m	1.2462	–	0.8828	–	–
	(-0.84,3.33)		(0.86,0.90)		
	$[-\infty,\infty]$		$[-\infty,\infty]$		
ϕ^l	–	0		1	1
Implied Freq.	2.90	2.64	2.75	2.72	2.62
$1/(1-\theta)$	$[1.82,\infty]$	$[1.82,\infty]$	$[1.82,\infty]$	[2.13,20]	$[2.00,\infty]$
p-Value	0.3776	0.2840	0.4013	0.4534	0.3799
MMSC-BIC	-53.39	-56.90	-58.22	-60.38	**-63.80**

Notes: Point estimates are obtained using CUE. 90% projection based confidence intervals in squared brackets and 90% Wald confidence intervals in round brackets. P-values report the test for the joint confidence set evaluated at the CUE point estimates. West's (1997) HAC estimate is used. Specification $\mathbf{I}_{\text{Full}}$ is the most general model. Specifications $\mathbf{II}-\mathbf{V}$ involve different restrictions on the parameters. Sample period: 1960:1-2005:4. Factor-augmented and preselected instrument set eq. (2.15) is used.

Provided that $\theta = 1$, it also follows that $\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)} = 0$, which implies that ϕ^m may take any value. This obviously leads to an unidentifiable parameter ϕ^m. Identification of ϕ^m thus requires $\theta = 1$ not to be part of the confidence set, so that $H_0: \ \theta = 1$ can be rejected. Obviously this is not guaranteed in specification I_{Full}. It also implies that ϕ^m is unidentified in the linearized version of the model, where, if $\lambda = 0$ cannot be rejected, ϕ^m cannot be identified.[18] Thus, it can be also

[18]This finding is generally consistent with those of Gabriel and Martins (2010) who also employ identification robust methods (namely Generalized Empirical Likelihood methods) to a similar specification and test the null hypothesis that $\phi^m = 0$. Since they cannot reject this hypothesis, they interpret this as evidence against the cost channel. However, they fail to recognize that this finding is due to a more deeply rooted identification problem of the model which makes it impossible to identify ϕ^m whenever the joint confidence interval also includes $\theta = 1$

seen that the model structure estimated by Chowdhury, Hoffmann, and Schabert (2006) is confronted by the same identification difficulties as the structural model, at least for the U.S. economy.[19]

Typically, Wald-type confidence intervals turn out to be smaller in comparison with identification robust sets and may not detect the identification problem of ϕ_m, even when the Wald confidence interval for θ would include the case $\theta = 1$. Other studies using identification robust methods for evaluating the standard NKPC (without considering a cost channel) likewise cannot rule out the case of total nominal price rigidity ($\theta = 1$) which also translates into an identification difficulty of ϕ^m (see e.g., Ma, 2002; Kleibergen and Mavroeidis, 2009b). Despite the identification problems for ϕ_m, other parameters can still be analyzed. For the measure of partial indexation ξ, we can conclude that it is basically unimportant and not significantly different from zero. In contrast, the case of full indexation, which implies $\xi = 1$ and has been assumed by Christiano, Eichenbaum, and Evans (2005) and Eichenbaum and Fisher (2007) can be clearly rejected. The confidence interval for β is relatively tight and is in an economically plausible range of something close to 1.

Although we demonstrated that problems exist for parameter ϕ^m identification and that the model structure is more or less useless as far as revealing any conclusion about the existence of a cost channel goes, we proceed with alternative specifications in order to assess the importance of the cost channel effect. We therefore take a standard Phillips curve specification with the restriction $\phi^i = 0$ as a benchmark to see whether different restrictions (with or without the cost channel) are to be preferred relative to a standard NKPC. The results for the standard Phillips curve are given by specification II_{NKPC} and the findings are in line with those of Kleibergen and Mavroeidis (2009b). Similar to specification I_{Full}, we get a rather large confidence interval for ξ, where we cannot reject the case of $\xi = 0$. This means that either indexation seems to play a minor role in inflation dynamics or that the lagged inflation term is relatively unimportant. However, owing to the

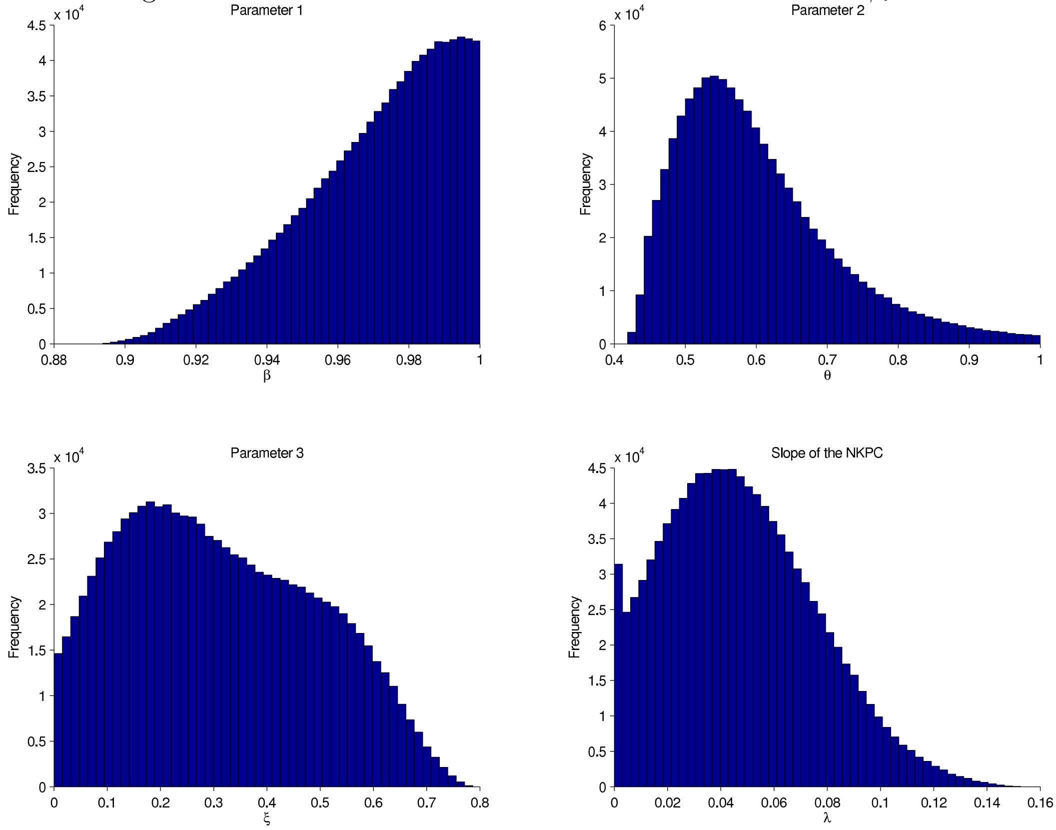
[19] Another problem that becomes evidence when estimating the general model is that the CUE objective function becomes very flat. The resulting ill-behaved shape of the function also impacts on point estimation, obtained by minimizing this function. Thus by using standard MATLAB numerical optimization routines (e.g. fminsearch) one may get stuck in local minima. The results based on identification robust methods including the projection method are unaffected by such an ill-behaved objective function.

large confidence intervals there is substantial uncertainty. We can again reject the null of full indexation. This is in contrast to Kleibergen and Mavroeidis's (2009b) findings, which point to an uninformative confidence interval for the backward-looking component which consists of the entire parameter space. The difference may be attributed to the slightly different model specification (we assume partial indexation, whereas Kleibergen and Mavroeidis assumes rule-of-thumb firms). Even more important may be the use of factors as instruments, which seems to make identification easier. This is in line with Wright's (2009) argument that the choice of instruments should be conducted according to principles of forecasting. He illustrated this by using better predictors for inflation and the labor share (namely inflation expectations from surveys) as instruments and obtaining smaller confidence sets, as do Kleibergen and Mavroeidis (2009b). Since factors have been successfully used to forecast output and inflation, the same principle may apply here. However, for the Calvo parameter θ, we cannot reject total nominal price rigidities because the confidence interval for θ includes 1 and the confidence interval for the average frequency of price adjustment is unbounded from above.

Turning to specification III_{RW}, this is essentially the estimated model of Ravenna and Walsh (2006), which omits the lagged inflation term and assumes static indexation ($\xi = 0$). Given this model, the same identification difficulties for ϕ^m arise as in specification I_{Full}. The restriction of $\xi = 0$ does not help in identifying parameter ϕ^m, and standard test procedure would lead to an entirely incorrect conclusion, namely a significant cost channel parameter ϕ^m, associated with low estimation uncertainty. The results for the remaining coefficients are similar to I_{Full} and II_{NKPC}. However, it is remarkable that the MMSC-BIC criterion favors III_{RW} over II_{NKPC}, providing additional evidence that partial indexation is unimportant in our setting.

If we employ a more direct measure of short-term liabilities, namely the lending rate (as proposed by Tillmann, 2008) instead of the Treasury bill rate, we can make use of the restriction that $\phi^l = 1$. This implies that changes in the lending rate fully translate into firms' marginal costs. This follows naturally from the eqs. (2.3, 2.7 and 2.8). Specifications IV_{LR} and V_{LR-RW} make use of this fact. Although the point estimates remain roughly the same in comparison with previous specifications, it turns out that the latter are to be preferred, according to the model

Figure 2.1: MCMC with CUE criterion function: with $\phi_l = 0$

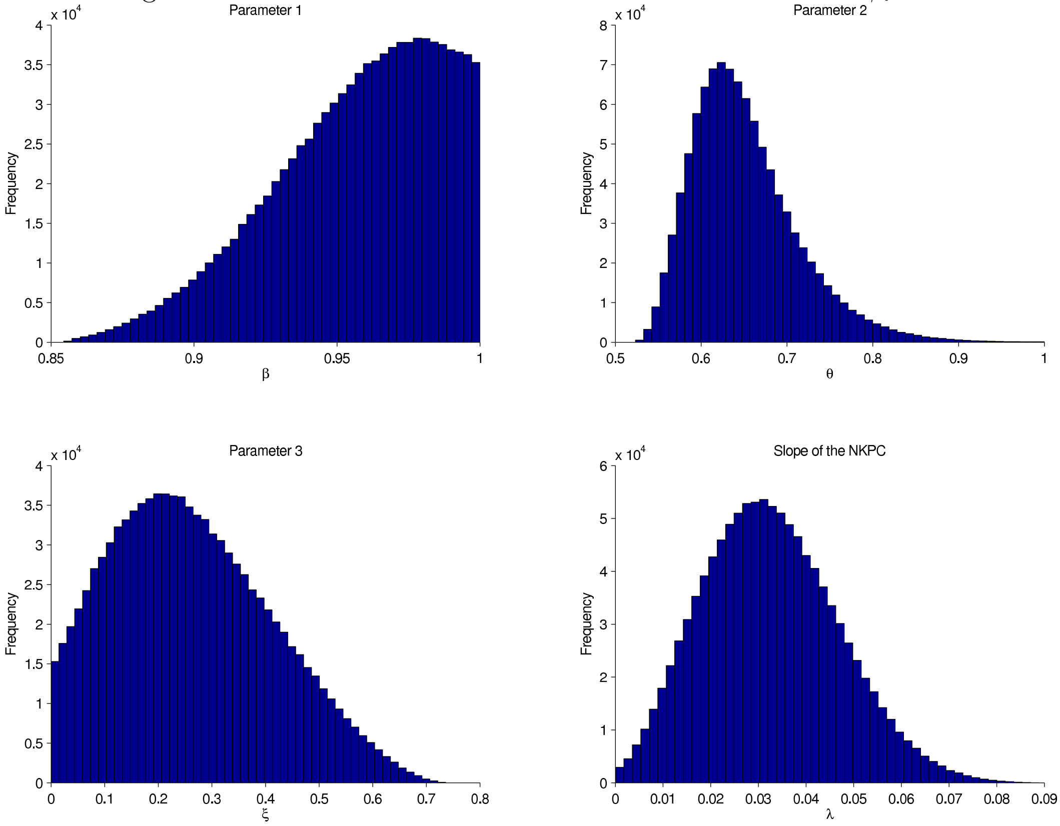


Notes: The histograms contain 1 million valid draws of the MCMC algorithm as described in the Appendix. The parameters β, θ and ξ are restricted to the interval $[0, 1]$ and to be inside the joint **S**-set.

selection criterion. Thus a model including the cost channel is clearly preferable. Moreover, we can now reject the case of full price rigidity $\theta = 1$, which can be also interpreted in favor of the model including the cost channel. This further implies that the confidence interval for the average frequency of price optimization is now bounded from above. In addition, it supports the omission of the lagged inflation term given, that the hypothesis $\xi = 0$ cannot be rejected. This is remarkable, because many studies have criticized pure forward-looking Phillips curve specifications on account of their inability to provide enough inflation persistence.

To strengthen our results, we make use of the MCMC approach (see section 2.3.2 and Appendix) to get a more complete picture of parameter uncertainty, given different Phillips curve specifications, and to find out whether or not those

Figure 2.2: MCMC with CUE criterion function: with $\phi_l = 1$



Notes: The histograms contain 1 million valid draws of the MCMC algorithm as described in the Appendix. The parameters β, θ and ξ are restricted to the interval $[0, 1]$ and to be inside the joint **S**-set.

provide evidence in favor of a cost channel effect. As a first attempt, we simulate the parameter uncertainty associated with specification II_{NKPC} and IV_{LR}, which implies different calibrations for the parameter ϕ^l.[20] The simulation reveals different shapes for the structural parameters β, θ and ξ and the consequences for the Phillips curve slope λ (see Figure 2.1 and 2.2). Overall, the parameter uncertainty seems to be smaller in the case of $\phi^l = 1$. Particularly, the distribution of θ is much more centered, involving less uncertainty, and it is less skewed in case of

[20]In principle, one can also simulate the full model specification I_{Full}. However, owing to the severe identification difficulties of parameter ϕ^m, this is extremely cumbersome. Additionally, the results depend heavily on the tolerated parameter space for ϕ^m. However, when simulating from the full model an ill-behaved distribution is obtained for parameter ϕ^m characterized by multimodality and a large overall dispersion.

$\phi^l = 1$. Even more important is the fact that there is no probability mass close to $\theta = 1$, which implies that the extreme case of perfect nominal price rigidity is quite unlikely. Instead, most of the probability mass lies between 0.55 and 0.85. Further, the simulated density for ξ looks more standard with the restriction $\phi^l = 1$. In both cases, full indexation ($\xi = 1$) can be rejected. For the standard NKPC ($\phi^l = 0$), the density is strongly skewed, still with a substantial probability mass for $\xi > 0.5$. Only by considering the cost channel $\phi^l = 1$ can the uncertainty about the slope of the Phillips curves be characterized by a symmetric distribution located with high confidence in the positive parameter space. In contrast, with $\phi^l = 0$, the dispersion for λ is larger and the distribution is pulled towards zero, which results from the greater uncertainty of θ in the case of the NKPC. Generally, we interpret all these findings as evidence in favor of a cost channel.

2.4.1 Robustness

Having set up the baseline model estimates for evaluating the cost channel along the lines of a Phillips curve model, we now check the robustness of the results for specific modifications of the standard model. We tackle the issues of further parameter restrictions, an alternative HAC estimator, using different instrument sets (one of them similar to the set of Ravenna and Walsh and another which excludes the information in period $t - 1$), a pure forward-looking Phillips curve specification excluding the cost channel (see Table 2.2), and sensitivity with respect to the different calibrations that determine the degree of real rigidity.

First, we naturally employ the restriction $\beta = 0.99$ and check whether this can help in identifying the remaining parameters (see Table 2.2 specification VI and VII). While the additional restriction does not provide much towards estimating ϕ^m, which is still unidentified (see VI), it does give some assistance in identifying θ in specification VII. Together with the restriction on the discount rate, we can reduce the uncertainty surrounding θ and the implied average frequency of price adjustment. The point estimates and the uncertainty associated with parameter ξ are in line with the unrestricted ones (see III_{RW} Table 2.2).

Second, we replace West's (1997) MA HAC estimator with the standard Newey-West estimator with a fixed bandwidth ($q = 6$) of the Bartlett kernel (specification VIII). This hardly affects any of the previous conclusions. Switching between the

Table 2.2: Sensitivity analysis

Model specification:

$$\hat{\pi}_t = \frac{\beta}{1+\beta\xi}\hat{\pi}_{t+1} + \frac{\xi}{1+\beta\xi}\hat{\pi}_{t-1}\frac{(1-\theta\beta)(1-\theta)}{(1+\beta\xi)\theta}\frac{1-\alpha}{1+\alpha(\epsilon-1)}\left(\hat{s}_t + \phi_i\hat{R}_t^i\right) + u_t$$

	VI	**VII**	**VIII**	**IX**	**X**	**XI**
β	**0.99**	**0.99**	0.9594	0.9485	0.7539	1.0063
			(0.90,1.02)	(0.91,0.99)	(0.52,0.98)	(0.95,1.06)
			[0.82,1.00]	[0.78,1.00]	[0.00,1.00]	[0.92,1.00]
θ	0.6533	0.6490	0.6445	0.6605	0.5930	0.6115
	(0.59,0.71)	(0.60, 0.69)	(0.56,0.73)	(0.62,0.70)	(0.47,0.72)	(0.43,0.79)
	[0.50,1.00]	[0.60,0.85]	[0.50,1.00]	[0.60,0.90]	[0.25,1.00]	[0.45,1.00]
ξ	0	0.1989	0.2659	0.3712	0.5361	**0**
		(0.14,0.25)	(0.10,0.43)	(0.23,0.51)	(0.32,0.74)	
		[0.00,0.65]	[0.00,0.65]	[0.00,0.90]	[0.15,0.95]	
ϕ^m	0.8406	–	–	–	–	–
	(0.71,0.96)					
	$[-\infty,\infty]$					
ϕ^l	–	1	1	1	1	**0**
Implied Freq.	2.88	2.85	2.81	2.95	2.46	2.57
$1/(1-\theta)$	$[2.00,\infty]$	[2.50,6.67]	$[2,\infty]$	[2.5,10]	$[1.33,\infty]$	$[1.75,\infty]$
Instr.	opt	opt	opt	**rw**	**lag**	opt
HAC	West	West	**NW**	West	NW	West
p-Value	0.3280	0.4584	0.5451	0.6882	0.8177	0.2564
Wright-Test	$\infty^\dagger$	$5.02^\dagger$	1.25	$2.19^\dagger$	1.25	$1.90^\dagger$
MMSC-BIC	-62.91	-65.04	-61.13	-58.92	-17.09	-61.53

Notes: Point estimates are obtained using CUE. 90% projection based confidence intervals in squared brackets and 90% Wald confidence intervals in round brackets. P-values report the test for the joint confidence set evaluated at the CUE point estimates. The specifications **VI** – **IX** involve different HAC estimates, instrument sets and restrictions on the parameters. West refers to West's (1997) HAC estimate and NW denote a Newey and West (1987) HAC estimator using 6 lags. Sample period: 1960:1-2005:4. Different instrument sets are used: opt (Factor-augmented and preselected instrument set) rw (the instrument set of Ravenna and Walsh) and lag (using only preselected variables and factors dated $t-2$ or ealier). The statistic of Wright is provided where † denotes significance at the 5% level.

two therefore does not affect the result. This may be also be interpreted as evidence that the residuals do not exhibit any substantial additional form of autocorrelation that would drive a wedge between the two estimators.

Third, we vary the instrument set. Ravenna and Walsh's (2006) instrument set B (labeled as $\mathbf{z}_{t-1}^{rw}$) is used as a comparison.[21] This task results in a slightly

[21] We also experimented with instrument set A which is substantially larger than instrument

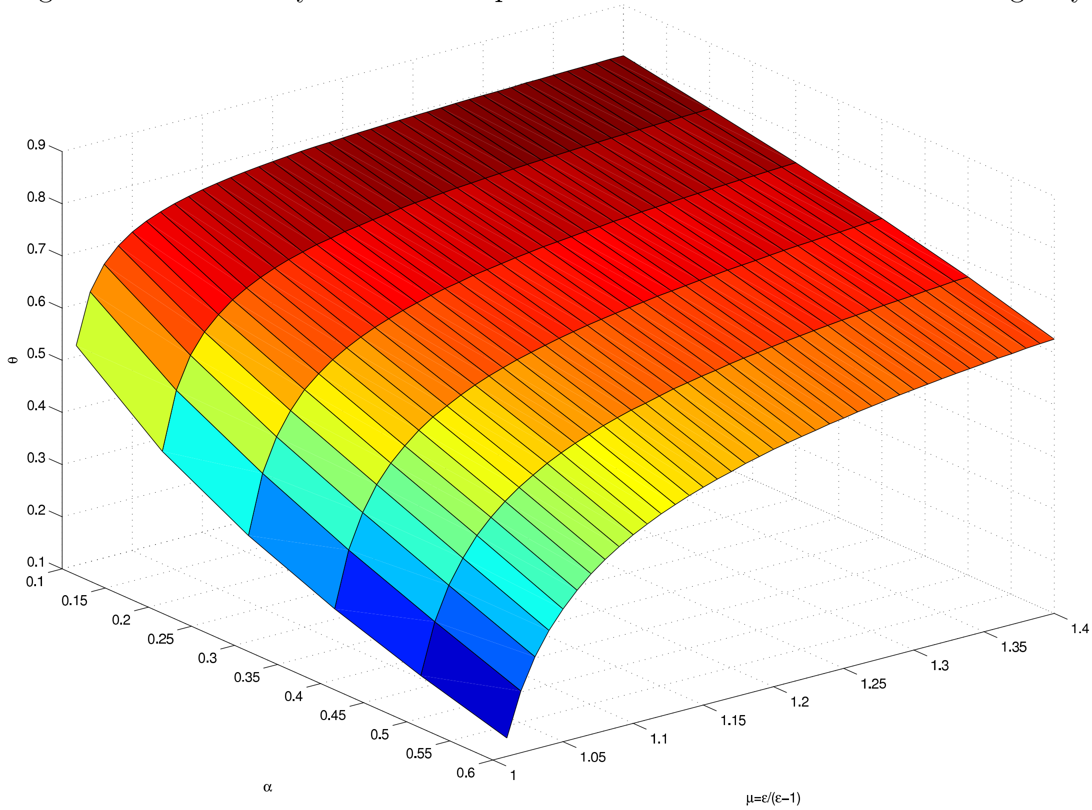
smaller confidence interval for θ (larger values above 0.9 can now be rejected). However, at the same time, the confidence interval for ξ becomes much larger and contains nearly all the possible values – except for cases close to full indexation.[22] In addition, we try to take into account the possibility that the error term of the Phillips curve contains a cost-push shock that might be autocorrelated. Kuester, Müller, and Stölting (2009) show that in this case the validity of moment conditions is violated. Namely, the error term is correlated with the instruments, and θ will be biased upwards. To check whether this could be the case in our setup, we use an instrument set containing only variables dated $t-2$ or earlier (so we exclude all instruments dated in $t-1$). If an autocorrelated cost-push shock is present, this will mitigate the bias effect (since correlation between the error term and the instruments should be smaller). When we observe a large deviation of the point estimates conditioned on the instrument set, we can interpret this as evidence in favor of an autocorrelated cost-push shock (this would imply that all previous results are biased on account of this). From specification X, it follows that omitting the first lag in the instrument set does not lead to a much lower value for $\hat{\theta}$. Thus the estimated degree of nominal rigidity is basically unaffected by the exclusion of instruments dated at time $t-2$ or earlier. However, the omission of relevant instruments results in a higher estimated degree of indexation and the estimate of discount factor is much less precise.

Fourth, when we choose a pure forward-looking Phillips curve specification without a cost channel given by XI (and the restriction $\xi = 0$ and $\phi^l = 0$) then the MMSC-BIC gives a higher value than specification V_{RW-LR}. Once again, this can be interpreted as evidence in favor of the cost channel. For all the additional results, we provide a test of adequacy of conventional asymptotics recently proposed by Wright (2010). The idea is thus simply to compare the volume of the robust confidence interval (**S**-set) with those of Wald confidence intervals obtained from the 2GMM estimator (see Appendix for details). If this test is significant, standard asymptotics will be inappropriate (critical values are provided by the author). The results for the different specifications are displayed in Table 2.2. It transpires that

set B. With instrument set A we get much larger confidence intervals compared to B.

[22]Figure 2.5 in the Appendix show bivariate plots for θ and ξ (given $\beta = 0.99$ and $\phi^l = 1$), depending on the two instrument sets.

Figure 2.3: Sensitivity of θ with respect to different calibrations of real rigidity



Notes: Different combinations of the calibrated values α and $\mu = \frac{\epsilon}{\epsilon - 1}$ are used for estimating θ, given the values $\xi = 0$ and $\beta = 0.99$.

the test rejects in four out of six cases.[23] This points to the necessity for using identification robust methods.

Another important issue is the calibration of the degree of real rigidities in the U.S. economy. Since real rigidities in this model are empirically indistinguishable from nominal rigidities, we experimented with different calibrations to robustify our findings. From the production function and the assumption that the firms' specific capital stock is fixed in the short run, it follows that the degree of real rigidities depends on the capital share α and the constant demand elasticity ϵ. We investigate different calibrations and their effect on the estimated degree of nominal price rigidity (parameter θ). Since the steady-state markup rate is equal

[23]Note that in this setting we have bounded the confidence set for the **S**-sets for the structural parameters in an economically plausible range between 0 and 1, while we leave the Wald confidence set unbounded. This may result in an underrejection of the test.

to $\mu = \frac{\epsilon}{\epsilon-1}$, we consider economically plausible mark-up rates of between 1 and 40 percent. Although there is generally more consensus on the capital share (around 1/3), we compute calibrations of α between 0.1 and 0.6.[24] Figure 2.3 plots the estimated Calvo parameter θ, depending on different calibrations. It turns out that only with a high capital share and a high elasticity of demand ($\epsilon \rightarrow \infty$), which is equivalent to a very low mark-up, the estimated θ will be considerably smaller in comparison with the baseline estimates. On the other hand, with a smaller α and a larger mark-up, θ will be pulled upwards, but rarely above 0.8. Generally, we find that it needs large (and economically implausible) deviation from the baseline calibrations, particularly for the capital share, to affect the estimated Calvo-parameter to any great extent.

2.4.2 Fit

For the sake of completeness, we also evaluate the fit of the Phillips curve specifications to see how the models track the data on actual inflation. We therefore concentrate on forward-looking specifications and compare the models employing different marginal cost definitions – the standard version with the labor share and the cost channel version (labor share plus lending rate).

Operationally, for comparing the model prediction with the actual inflation rate, we need a measure for the expected inflation rate. As in the 2SLS approach, we run a first-stage regression where realized future inflation rates are regressed on the instruments:

$$\hat{\pi}_{t+1} = \mathbf{z}_{t-1}^{'}\gamma + \nu_{t+1},$$

where the vector of coefficients γ is simply estimated by OLS. The predicted inflation rates are given by

$$\hat{\pi}_{t+1}^{p} = \mathbf{z}_{t-1}^{'}\hat{\gamma}$$

and $\hat{\pi}_{t+1}^{p}$ is the direct measure for inflation expectations which can be plugged in for $E_t\hat{\pi}_{t+1}$ in the structural model, given the estimated coefficients β and θ

[24]e.g., Coenen, Levin, and Christoffel (2007) note that α may be higher than normally expected to match the estimated degree of real rigidities.

Table 2.3: In-sample fit (error measures)

	IV	**XI**	**RW**	**AR(1)**
RMSE	**0.9596**	0.9729	1.2104	1.1714
MAE	**0.7576**	0.7655	0.9207	0.8827

Notes: **IV** and **XI** refer to the specification used in Table 2.1 and 2.2. **RW** and **AR**(1) denote the random walk model and autoregressive model with one lag, respectively. Root mean square errors (RMSE) and mean absolute errors (MAE) are displayed.

obtained from the GMM estimate.

It is now possible to compare the specifications employing different measures of marginal cost ($\hat{s}$ vs. $\hat{s}+\hat{R}^l$). Figure 2.4 (Appendix) shows the actual inflation rates against those predicted, assuming the existence of a cost channel (point estimates from specification V are used). Since graphically the two Phillips curve versions are difficult to distinguish from each other, we compute in-sample measures of fit, namely the root mean squared error (RMSE) and the mean absolute error (MAE) for both versions. In addition, we calculate the error terms for two simple benchmark models: a random walk (RW) model $\hat{\pi}_t = \hat{\pi}_{t-1} + \upsilon_t$ and an AR(1) model $\hat{\pi}_t = \delta\hat{\pi}_{t-1} + \eta_t$. As expected from the results based on MMSC-BIC, the model including the cost channel dominates the pure labor share-based Phillips curve version as well as the simple univariate time series models.

2.5 Conclusion

We find that the frequently applied test for the existence of the cost channel, i.e. $\phi^m = 0$, is misleading, because this parameter is basically unidentifiable, given the standard model structure. For this reason, there is not a great deal to be learned by relying on this practice, since robust confidence intervals should be uninformative in this case. This follows automatically with the application of identification robust econometric methods. This chapter therefore compares different model specifications (which are non-nested) and assesses whether the marginal cost variable composed of the labor share and the bank lending rate (the model with cost channel) is more compatible with the data in comparison with the standard labor share version of the Phillips curve.

We use various techniques to gain a broader perspective on the empirical model. First, we use factors as additional instruments to efficiently characterize the agent's information set, as implied by rational expectations in a parsimonious way. Then we apply identification robust techniques to evaluate the structural parameters of the model. Therefore, Stock and Wright's (2000) **S**-sets are used in combination with the projection method to get individual confidence sets that guarantee a high degree of robustness. We also apply an MCMC to characterize the uncertainty of the estimated parameters more efficiently. In addition, we use an information criterion (MMSC-BIC), which allows us to compare different model specifications (nested or non-nested) and which can be used to assess the relative model quality.

Our results reveal that a cost channel-based Phillips curve version is empirically more compatible with the data and with the theoretical arguments than the standard Phillips curve version. Moreover, in nearly all specifications, we cannot reject the fact that the degree of indexation (given by $\xi = 0$) is zero. This implies that the backward-looking element of inflation seems rather unimportant. Generally, a pure forward looking cost channel version of the Phillips curve characterizes inflation dynamics well. The direct effect of bank lending rates on firms' marginal costs may also account for the observed persistence of inflation rates beyond what can be explained by the standard NKPC.

Thus, it can be concluded that a pure forward-looking interest rate augmented Phillips is most compatible with the data. This suggests that considering the cost channel is an important aspect for monetary policy which should be also included in theoretical models.

References

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20(1), 46–63.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59(3), 817–859.

ANDREWS, D. W. K., AND B. LU (2001): "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models," *Journal of Econometrics*, 101(1), 123–164.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60(4), 953–966.

ANDREWS, D. W. K., AND J. H. STOCK (2005): "Inference with weak instruments," NBER Working Paper T0313, NBER.

BAI, J., AND S. NG (2010): "IV-Estimation in a data-rich environment," *Econometric Theory*, 26(6), 1577–1606.

BROCKWELL, P. J., AND R. A. DAVIS (1990): *Time Series: Theory and Methods.* Springer, 2nd edn.

CALVO, G. A. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12(3), 383–398.

CANOVA, F., AND L. SALA (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56(4), 431–449.

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

CHOWDHURY, I., M. HOFFMANN, AND A. SCHABERT (2006): "Inflation dynamics and the cost channel of monetary transmission," *European Economic Review*, 50(4), 995–1016.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. EVANS (2005): “Nominal rigidities and the dynamic effects of a shock to monetary policy,” *Journal of Political Economy*, 113(1), 1–45.

COENEN, G., A. T. LEVIN, AND K. CHRISTOFFEL (2007): “Identifying the influences of nominal and real rigidities in aggregate price-setting behavior,” *Journal of Monetary Economics*, 54(8), 2439–2466.

DEN HAAN, W. J., AND A. LEVIN (1997): “A practitioner’s guide to robust covariance matrix estimation,” in *in Handbook of Statistics: Robust Inference (Vol 15)*, pp. 291–341. North-Holland.

DUFOUR, J.-M. (1997): “Some implausibility theorems in econometrics with applications to structural and dynamic Models,” *Econometrica*, 65(6), 1365–1387.

——— (2003): “Identification, weak instruments, and statistical inference in econometrics,” *Canadian Journal of Economics*, 36(4), 767–808.

——— (2009): “Comment,” *Journal of Business & Economic Statistics*, 27(3), 318–321.

DUFOUR, J.-M., AND J. JASIAK (2001): “Finite sample limited information inference in econometrics,” *International Economic Review*, 42(3), 815–843.

DUFOUR, J.-M., L. KHALAF, AND M. KICHIAN (2006): “Inflation dynamics and the New Keynesian Phillips curve: An identification robust econometric analysis,” *Journal of Economic Dynamics & Control*, 30(9-10), 1707–1727.

DUFOUR, J.-M., AND M. TAAMOUTI (2005): “Projection-based statistical inference in linear structural models with possibly weak instruments,” *Econometrica*, 73(4), 1351–1365.

——— (2007): “Further results on projection-based inference in IV regressions with weak, collinear or missing Instruments,” *Journal of Econometrics*, 139(1), 133–153.

EICHENBAUM, M., AND J. D. M. FISHER (2007): “Estimating the frequency of price re-optimization in Calvo-style models,” *Journal of Monetary Economics*, 54(7), 2032–2047.

Gabriel, V., P. Levine, C. Spence, and B. Yang (2008): “On the (ir)relevance of direct supply-side effects on monetary policy,” Discussion Papers in Economics 04/08, University of Surrey.

Gabriel, V. J., and L. F. Martins (2010): “The cost channel reconsidered: A comment using an identification-robust approach,” *Journal of Money, Credit and Banking*, 42(8), 1702–1712.

Galí, J., and M. Gertler (1999): “Inflation dynamics: A structural econometric analysis,” *Journal of Monetary Economics*, 44(2), 195–222.

Galí, J., M. Gertler, and J. D. López-Salido (2001): “European inflation dynamics,” *European Economic Review*, 45(7), 1237–1270.

Ghysels, E., and A. Hall (1990): “Testing nonnested Euler conditions with quadrature-based methods of approximation,” *Journal of Econometrics*, 46(3), 273–308.

Hahn, J., and J. Hausman (2002): “A new specification test for the validity of instrumental variables,” *Econometrica*, 70(1), 163–189.

Hall, A. R. (2005): *Generalized Method of Moments*. Oxford University Press, Oxford.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Hansen, L. P. (1982): “Large sample properties of generalized method of moments estimators,” *Econometrica*, 50(4), 1029–1054.

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov (2007): “Intertemporal substitution and risk aversion,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 61. Elsevier.

Hansen, L. P., J. Heaton, and A. Yaron (1996): “Finite sample properties of some alternative GMM estimators,” *Journal of Business & Economic Statistics*, 14(3), 262–280.

Huelsewig, O., S. Henzel, T. Wollmershaeuser, and E. Mayer (2009): "Price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock?," *Journal of Macroeconomics*, 31(2), 268–289.

Judge, G. G., W. Griffiths, R. C. Hill, H. Luetkepohl, and T.-C. Lee (1985): *The Theory and Practice of Econometrics*. John Wiley and Sons, second edn.

Kapetanios, G., and M. Marcellino (2010): "Factor-GMM estimation with large sets of possibly weak instruments," *Computational Statistics & Data Analysis*, 54(11), 2655–2675.

Kleibergen, F. (2007): "Generalizing weak instrument robust IV statistics toward multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139(1), 181–216.

Kleibergen, F., and S. Mavroeidis (2009a): "Inference on subsets of parameters in GMM without assuming identification," Discussion paper, Brown University.

——— (2009b): "Weak instrument robust tests in GMM and the New Keynesian Phillips curve," *journal of business & Economic Statistics*, 27(3), 293–339.

Kocherlakota, N. R. (1990): "On tests of representative consumer asset pricing models," *Journal of Monetary Economics*, 26(2), 285–304.

Kuester, K., G. Müller, and S. Stölting (2009): "Is the New Keynesian Phillips curve flat?," *Economics Letters*, 103(1), 39–41.

Ma, A. (2002): "GMM estimation of the new Phillips curve," *Economics Letters*, 76(3), 411–417.

Martins, L. F., and V. J. Gabriel (2009): "New Keynesian Phillips curves and potential identification failures: A Generalized Empirical Likelihood analysis," *Journal of Macroeconomics*, 31(4), 561–571.

Mavroeidis, S. (2005): "Identification issues in forward-looking models estimated by GMM with an application to the Phillips curve," *Journal of Money, Credit and Banking*, 37(3), 421–449.

——— (2007): "Testing the New Keynesian Phillips curve without assuming identification," Working paper, Brown University.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71(4), 1027–1048.

Nason, J. M., and G. W. Smith (2008): "Identifying the New Keynesian Phillips curve," *Journal of Applied Econometrics*, 23(5), 525–551.

Newey, W. K., and K. D. West (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55(4), 703–708.

Newey, W. K., and K. D. West (1994): "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61(4), 631–653.

Rabanal, P. (2007): "Does inflation increase after monetary policy tightening? Answers based on an estimated DSGE model," *Journal of Economic Dynamics & Control*, 31(3), 906–937.

Ravenna, F., and C. E. Walsh (2006): "Optimal monetary policy with the cost channel," *Journal of Monetary Economics*, 53(2), 199–216.

Scheufele, R. (2010): "Evaluating the German (New Keynesian) Phillips curve," *The North American Journal of Economics and Finance*, 21(2), 145–164.

Singleton, K. J. (1985): "Testing specifications of economic agents' intertemporal optimum problems in the presence of alternative models," *Journal of Econometrics*, 30(1-2), 391–413.

Smets, F., and R. Wouters (2003): "An estimated dynamic stochastic general equilibrium model of the Euro area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Smith, R. J. (1992): "Non-nested tests for competing models estimated by Generalized Methods of Moments," *Econometrica*, 60(4), 973–980.

Stock, J., and M. Yogo (2005): "Testing for weak instruments in linear IV regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge: Cambridge University Press.

Stock, J. H., and M. W. Watson (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44(2), 293–335.

——— (2002): "Macroeconomic forecasting using diffusion indexes," *journal of business & Economic Statistics*, 20(2), 147–162.

Stock, J. H., and J. H. Wright (2000): "GMM with weak identification," *Econometrica*, 68(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 20(4), 518–29.

Tauchen, G. (1986): "Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data," *journal of business & Economic Statistics*, 4(4), 397–416.

Tillmann, P. (2008): "Do interest rates drive inflation dynamics? An analysis of the cost channel of monetary transmission," *Journal of Economic Dynamics and Control*, 32(9), 2723–2744.

Walsh, C. E. (2003): *Monetary theory and policy.* The MIT Press, Cambridge, Massachusetts, 2nd edn.

West, K. D. (1997): "Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator," *Journal of Econometrics*, 76(1-2), 171–191.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press, Princeton.

Wright, J. H. (2009): "Comment," *Journal of Business & Economic Statistics*, 27(3), 323–326.

——— (2010): “Testing the adequacy of conventional asymptotics in GMM,” *Econometrics Journal*, 13(2), 205–217.

Appendix

Determining the appropriate HAC estimator

Determining the estimated long-run covariance matrix $\hat{S}_T$ is also crucial for inference, particularly when heteroskedasticity and/or autocorrelation are present. The class of HAC estimators has therefore been proposed. These estimators may be classified into two broad categories: non-parametric kernel-based procedures and parametric procedures.

The most frequently applied method is Newey and West (1987), a kernel-based procedure. In this case, $\hat{S}_T$ would be

$$\hat{S}_T^{NW} = \hat{\Gamma}_{0,T} + \sum_{v=1}^{q} (1 - [v/(q+1)]) \left(\hat{\Gamma}_{v,T} + \hat{\Gamma}'_{v,T}\right)$$

where

$$\hat{\Gamma}'_{v,T} = (1/T) \sum_{t=v+1}^{T} \left[f_t(\hat{\vartheta})\right] \left[f_t(\hat{\vartheta})\right]' .$$

With $f_t(\hat{\vartheta})$ denoting the k orthogonality conditions for this model which depend on the parameters and the data. Note that $\hat{\vartheta}$ and $\hat{S}_T$ are estimated jointly owing to the application of the CUE.

A critical aspect of this approach is the determination of the optimal bandwidth q. Andrews (1991) and Newey and West (1994) have therefore proposed selection rules for determining the optimal bandwidth. In addition Andrews and Monahan (1992) and Newey and West (1994) promoted the construction of pre-whitened and re-colored HAC estimators to provide a better finite sample performance. However, neither of these more advanced procedures works without problems in our application, and endogenous selection leads to an implausibly high value of q for each of the selected methods (regardless of whether pre-whitening is employed).

As an additional method of HAC estimation parametric approaches exists (see Den Haan and Levin, 1997, for an overview). Given the model structure employed here with our instrument set, we can assume that the residuals are generated by a specific parametric model. Since we use only instruments dated $t-1$ or earlier, we can use the procedure of West (1997), using the MA(1) structure. More specifically, we can define

$$\hat{u}_t = h_t(\hat{\vartheta}),$$

where $\hat{u}_t$ is given by the model structure (see eq. 2.9). Then let $\hat{\theta}_1$ be a consistent estimator of θ_1 based on $\hat{u}_t = \hat{\epsilon}_t + \hat{\theta}_1 \hat{\epsilon}_{t-1}$. We determine $\hat{\epsilon}$'s and $\hat{\theta}_1$ by applying a weighted nonlinear least squares procedure (see e.g., Brockwell and Davis, 1990, chapter 8.7). Then we define the vector $\hat{d}_{t+1}$ as

$$\hat{d}_{t+1} = \left(\mathbf{z}_{t-1} + \mathbf{z}_t \hat{\theta}_1 \right) \hat{\epsilon}_t$$

and can estimate S_T as

$$\hat{S}_T^{MA} = (T-1)^{-1} \sum_{t=1}^{T-1} \hat{d}_{t+1} \hat{d}_{t+1}^{\prime}.$$

$\hat{S}_T^{MA}$ is positive semidefinite per construction.

West (1997) shows that this method works well in situations where cross-products of instruments and disturbances are sharply negatively correlated. This is also the situation where "truncated" (non-parametric) estimators are likely to fail. This seems to be the case, because in this application θ is estimated to be around -0.6. We also calculated the residuals of the estimated model and ran some diagnostic tests for autocorrelation, so we can confirm the MA(1) process. The implication is that we cannot confirm the existence of a cost push shock as stressed by many DSGE models estimated under full information, which are assumed to be positively correlated (mostly following an AR(1) process). To allow for more flexible specifications of the error term, we also consider Newey and West's (1987) standard procedure, whereby the bandwidth is set up by the rule of thumb $q = int\left\{T^{1/3}\right\}$.

MCMC

The MCMC simulations follow a version of the standard Metropolis-Hastings algorithm (see Chernozhukov and Hong, 2003; Hansen, Heaton, Lee, and Roussanov, 2007). Let the parameter combination corresponding to the ith draw be $b^{(i)} = \left[\beta^{(i)}, \theta^{(i)}, \xi^{(i)}\right]$, given the calibrated parameters α, ϵ and ϕ. Then

1. Take $b^{(0)}$ as the CUE point estimator $b^{(0)} = b^{(CUE)}$.
2. Draw ς from the conditional distribution $q(\varsigma|b^{(i)})$.
3. With probability $\inf\left(\frac{\exp(-\mathbf{S}(b^{(i+1)}))q(b^{(i)}|\varsigma)}{\exp(-\mathbf{S}(b^{(i)}))q(\varsigma|b^{(i)})}, 1\right)$ update $b^{(i+1)} = \varsigma$; otherwise keep $b^{(i+1)} = b^{(i)}$.

We take a Gaussian transition density, which results in a Markov chain, which is a random walk. We also constrain the parameter space to match the economically plausible range between 0 and 1 (which guarantees a compact set). Further, we allow only parameter combinations to include that pass the χ^2 test; i.e. $T \times \mathbf{S}(b) < c_k(\alpha)$ (see eq. 2.13). Let ϕ be the trivariate normal density centered around zero with cdf Φ. Then

$$q(x|y) = \frac{\phi(x-y)}{\Pr(x \in A)}, \quad \text{where } x = y + z, \; z \sim \Phi,$$

where A is the event that b falls outside its boundaries or the χ^2 test rejects. In the simulation, the truncation is accomplished by discarding those values that do not fulfill these requirements. A choice has to be made concerning the dispersion of ϕ (the different variances). We take different values for each parameter in order to obtain an acception rate of the algorithm that matches a value slightly above 25%. The reported results are based on simulations with 1,000,000 accepted draws.

Test of adequacy of convential asymptotics in GMM

Wright (2010) proposed a test for the null hypothesis that a GMM model is sufficiently well identified for conventional asymptotics to be reliable. This test can be applied for various Phillips curve specifications in order to decide whether it is necessary to rely on more robust procedures, which could also be associated with a power loss.

Setting up the test statistic as proposed by Wright (2010), we define W_1 as the maximum distance between any two values of ϑ in the $\mathbf{S}$-set for ϑ. If the $\mathbf{S}$-set is unbounded, W_1 will be infinite. When ϑ is a vector, ϑ_1 and ϑ_2 are the lowest and largest values, obtained respectively from the projection method and W_1 will be the value that maximizes these distances.

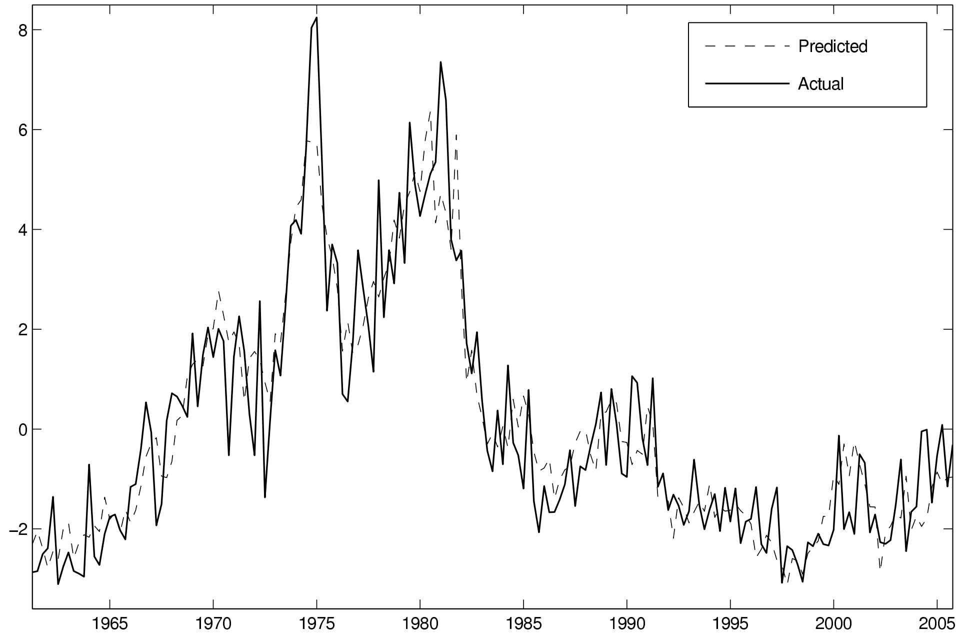
Likewise, W_2 can be defined as the maximum between any two points in the usual 2GMM Wald confidence set for ϑ, which is given by $T(\hat{\vartheta}_{TS}-\hat{\vartheta}_i)'\hat{J}(\hat{\vartheta}_{TS}-\hat{\vartheta}_i) \leq c_n(\alpha)$ for $i = 1, 2$, where $c_n(\alpha)$ is the $100(1-\alpha)\%$ percentile of the χ^2 distribution. Numerically, W_2 can be computed as $W_2 = \frac{2}{\sqrt{T}}\sqrt{\frac{c_n(\alpha)}{\hat{\lambda}}}$, where $\hat{\lambda}$ is the smallest eigenvalue of $\hat{J}$. The test then simply compares the two volumes of the confidence set equal to

$$L = \frac{W_1}{W_2}.$$

The author provides critical values for this test statistic (see Wright, 2010, Table 1).

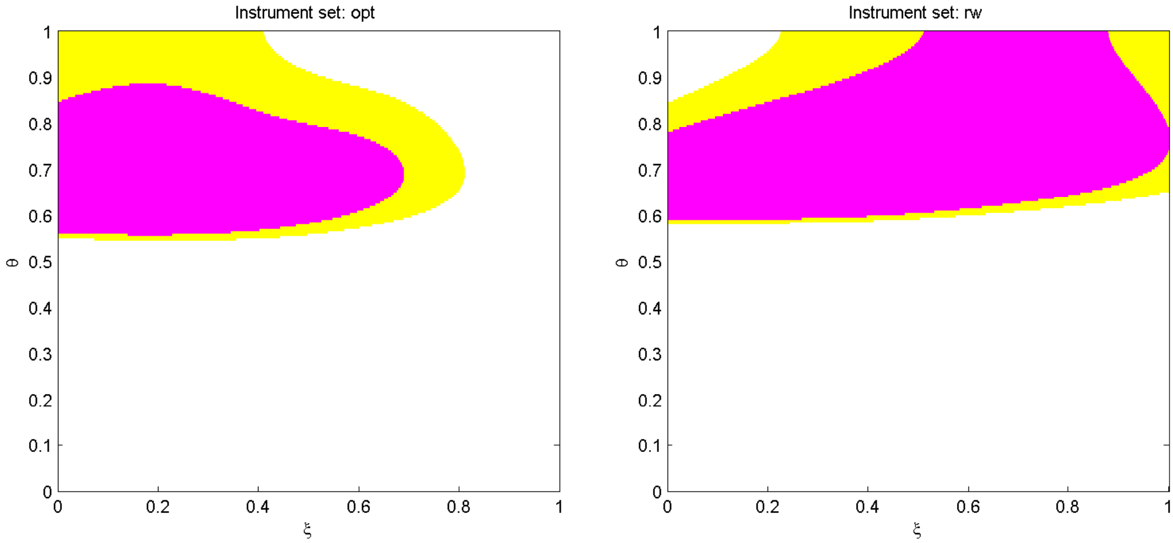
Figures

Figure 2.4: Fit of the pure forward-looking model with the cost channel (Actual vs. predicted Inflation)



Notes: Actual and predicted annualized demeaned inflation rates based on the GDP deflator.

Figure 2.5: **S**-sets for the structural parameters



Notes: Joint confidence sets on the structural parameters (θ,ξ) in the NKPC model with a cost channel given the calibrated values $\beta = 0.99$ and $\phi^l = 1$. Different instrument sets used. The shaded areas contain joint 95% (yellow) and 90% (magenta) confidence sets.

Chapter 3

Are qualitative inflation expectations useful to predict inflation?*

3.1 Introduction

Nowadays, inflation expectations play a central role in conducting monetary policy. Since many central banks have explicitly or implicitly adopted an inflation targeting regime, stabilizing inflation expectations has become the primary policy objective. Because there is a lag between policy actions and their impact on inflation, monetary authorities are guided by medium term forecasts. This makes inflation forecasting essential for effective monetary policy. Although monetary authorities seek to stabilize long-term inflation expectations, monitoring short- and medium-term inflation is also important. Whenever inflation exhibits some inertia, good short-term inflation forecasts translate into more-accurate longer-term

*All computations used in this Chapter are done using EViews 6.1.

projections.

In this chapter we use survey data collected from economic experts and characterize its properties as indicators for German inflation expectations. Although Germany is part of the EMU, and the ECB has to control inflation in the aggregate Euro area, it is still important to monitor inflation in the major member countries for the purpose of aggregate inflation projections. However, inflation expectations are not only important for conducting monetary policy, but are also essential for decision-making of private households and firms. Examples include price-setting of firms and wage negotiations.

The inflation expectation data we use in this chapter has the advantage that it is available at monthly frequency and has a fixed time horizon (six months). We use several methods to compute quantitative measures from qualitative responses. First, we consider different variants of the well known method of Carlson and Parkin (1975). Second, we use the regression approach proposed by Pesaran (1984) (in which the balance statistic is also a special case). In a first step, we compare all the different quantification procedures relating to their in-sample ability to match realized inflation rates, test for Granger causality between inflation expectations and realizations and then test for asymmetric thresholds.

In addition, we test whether our inflation expectation measure matches the orthogonality condition which is often associated with rational expectations. The orthogonality condition requires forecast errors (defined as the difference between expected and realized inflation) to be uncorrelated with past information. Moreover, we investigate the measure of inflation expectations concerning its information content to predict future inflation rates (over different forecasting horizons). In an out-of-sample experiment, we compare models that employ these survey measures with other inflation models. These other models consist of univariate time series models, Phillips curve specifications and term structure models. With tests for predictive accuracy we are then able to assess whether these models display significantly different relative predictive accuracy.

The first contribution of this chapter is that it rigorously compares different methods for the quantification of qualitative survey data on inflation expectations. The second contribution is the comparison among a broad range of popular forecasting models of inflation for Germany on a monthly basis. As Ang, Bekaert, and Wei (2007) document the superiority of survey-based methods over many alternative inflation models, we follow this line of research and investigate whether these results also hold for qualitative survey data for the German economy. Our third contribution is the provision of insights into the expectation formation by economic experts. Here, the intension is to find out which information and forecasting models are used or not used by participants in the survey.

Our findings can be briefly summarized as follows. Inflation expectations obtained from the ZEW Financial Market Survey are not fully consistent with rational expectations, as they do not contain all the available (costless) information and thus violate the orthogonality assumption. However, a Granger causality test reveals that this series contains information about future inflation. In an out-of-sample experiment we find that the pure survey measure performs poorly in comparison with other inflation models. When an augmented model is used that includes not only the expected inflation series but also additional lags of actual inflation, it outperforms most of the alternative specifications in terms of root mean squared forecast error (RMSFE) and mean absolute forecast error (MAFE). It performs even better for longer time horizons (i.e. 12 months). It performs significantly better than standard Phillips curve models and some univariate time series models. Encompassing tests indicate that the model employing inflation expectations already contains the information of univariate time series models and Phillips curve specifications but disregards some of the information included in financial variables, such as the interest rate or term spreads.

The remainder of this chapter is as follows. Section 3.2 describes the data set and the conversion method for getting quantitative inflation expectations. Section 3.3 presents the characteristics and some test results. In Section 3.4 the out-of-sample set-up is explained and its main results are presented. Section 3.5 concludes.

3.2 Measurement of expectations via survey data

The use of survey data on inflation expectations has a long tradition in economic literature (e.g. Anderson, 1952; Theil, 1952). Direct measures of expectations allow for analysis of the expectation formation process without reliance on a particular behavioral model, which is typically found in rational expectation models.

In principle, it is possible to distinguish between two types of survey data on inflation expectations: "quantitative" and "qualitative". Quantitative data means that respondents are asked for the exact magnitude of change or level. For instance, the question could be asked: "What inflation rate do you expect next year?". In contrast to exact measures, surveys may also ask for a general tendency. In this case, respondents would give a qualitative statement in response to a question such as: "Do you expect inflation to go up (or down) during the next year?". Although it always seems preferable to obtain point forecasts of expectations relating to future inflation rates, there may be certain drawbacks to using quantitative responses, because, in comparison with tendency statements, direct measures could be affected by sampling and measurement errors (e.g. Pesaran, 1987, Ch. 8.2). Using qualitative survey data as a measure of inflation expectations always necessitates transforming such data into quantitative expressions requiring certain assumptions (which are not necessarily testable).

3.2.1 Expert versus household expectations

Another distinction between different types of surveys can be made according to the population of the survey. The population may compromise households, firms or professional economists. However, the literature distinguishes mainly between professional forecasters and households. Well-known examples are the consumer survey distributed by the European Commission (EC) (known also as the EC Consumer Survey) and the Consensus Forecasts.[1] The latter survey aims to collect individual forecasts of large enterprises (mainly banks and financial institutions) and economic research institutes. Both surveys contain information on both the Euro area as a single area, and some of the member states, including Germany (see Mestre, 2007, Annex A, for a detailed description and some characteristics for both surveys).[2]

It is generally accepted that professional forecasters process information earlier

[1] Consensus Forecasts are provided by Consensus Economics, a London-based macroeconomic survey firm.

[2] Alternative surveys comprise industry surveys from the ifo-institute or the Survey of Professional Forecasters of the ECB. Whereas the last one is only available for the aggregate Euro area.

and more efficiently than do households. This view is rationalized by, for example, Carroll (2003) who provides evidence that household expectations are derived from those of professional forecasters. Empirically, this manifests in terms of improved forecasting accuracy. For example, Carroll (2003) documents for the USA that the Mean Square Error (MSE) of the Michigan Survey (a household survey) on Inflation Expectation is nearly twice as large as that for the Survey of Professional Forecasters (SPF). More important, though, is the finding that there is statistical evidence of Granger causality running from professional forecasts to household forecasts, but not vice versa. This finding is also supported by Doepke, Dovern, Fritsche, and Slacalek (2008) for three major European economies including Germany. Inoue, Kilian, and Kiraz (2009) put forward another argument in this context, finding evidence that consumers with low levels of education find it difficult to articulate their expectations when responding to surveys. Only better educated households seem able to articulate their inflation expectations adequately. These expectations are closer to what professional forecasters expect, which corroborates Carroll's (2003) argument.

We therefore expect that surveys carried out by professional forecasters are more effective than consumer surveys when it comes to forecasting inflation. This is confirmed by Ang, Bekaert, and Wei (2007) in their comprehensive study for the US, in which they compare the forecasting performance of a large set of different models, using information based on survey inflation expectations. While they find that consumer expectations can be useful for inflation forecasting (information from this survey already do better than some regression-based forecasting methods), surveys of professional forecasters, such as the SPF or the Livingston survey, generally lead to better results. Using in-sample tests (Granger causality) for the EC Consumer Survey Forsells and Kenny (2004) find no evidence that inflation expectations are a useful predictor for future inflation rates.

3.2.2 Data set

In the chapter, we use a monthly survey carried out by the Center for European Economic Research (ZEW) to construct a direct measure for inflation expectations. This type of survey is more country-specific than those mentioned above because only German financial analysts are consulted. This may be the primary reason for this type of survey playing only a marginal role in the literature on inflation expectations.[3] The ZEW Financial Market Survey covers approximately 300 experts from banks, insurances and large industrial firms. Each month, the experts are asked whether they expect "a rise", "a decline" or "no change" in the annual inflation rate in the medium term (over the next six months).Obviously this survey, very like the consumer survey of the European Commission, asks about qualitative inflation expectations.

The basic advantage of this data set is that it can be used to construct a monthly measure of inflation expectations with a fixed forecasting interval for each point in time. This constitutes its main advantage over Consensus Economics.[4] A further advantage of the ZEW survey is that it is more representative, the number of participants is approximately 10 times greater than that of the consensus forecast. However, a possible disadvantage of the ZEW survey is its qualitative nature, which makes it necessary to make additional assumptions when constructing a quantitative expected inflation rate. However, the aim of this chapter is precisely to investigate the usefulness of these qualitative data.

3.2.3 Estimating German inflation expectations

The quantification of qualitative response data has a long history in economics. Its general thrust is that aggregate fractions of survey answers reveal something about the magnitude of inflation changes. The first results were obtained using so called balance statistics; this work goes back to Anderson (1952) and Theil (1952). The procedure employs the difference between the percentage of respondents who report an increase and the percentage who report a decrease as a quantitative measure. Another solution to the problem of quantification was offered by Carlson and Parkin (1975), theirs remaining the foremost technique for quantifying qualitative inflation expectations. The Carlson-Parkin (C-P) approach assumes, respondents answer "no change" if the perceived change is below a certain threshold δ.

[3]Notable exceptions are Franz (2005), Heinemann and Ullrich (2006) and Breitung (2008).

[4]Consensus Economics asks about inflation expectations for the current and subsequent year (i.e. flexible horizon). This implies that for converting these data into inflation expectations with a fixed horizon, additional assumptions are necessary (e.g. some authors use moving averages as a conversion technique). Only at quarterly frequency does Consensus Economics provide inflation expectations with a fixed horizon.

The quantification method requires that certain assumptions hold. According to Pesaran (1987), these include:

- There is an interval $[a_{it}, b_{it}]$ of inflation changes in the region of zero which respondents cannot distinguish from zero.
- The subjective probability distributions have certain properties that make it possible to obtain an aggregate probability distribution with first and second order moments where the subjective information set is the union of the individual sets, and where the aggregate expected change in the inflation rate is the average of the subjective expected change of the inflation rate.
- The subjective probability distributions are independent of each other and of the same known form across respondents.
- The thresholds a_{it} and b_{it} are the same across individuals, constant over time and symmetric around zero.

The majority of these assumption cannot be tested. Statistical tests can be constructed only for the last assumption.

While the original Carlson and Parkin approach is employed for price levels, not for inflation rates, we have to modify their approach slightly so that it fits this particular data set. Following Carlson and Parkin (1975), we have to make a distributional assumption about perceived changes in the inflation rate. Carlson and Parkin (1975) choose the normal distribution. Like Dasgupta and Lahiri (1992), we also consider the logistic distribution and a scaled t-distribution as alternative distributional assumptions.

For the derivation of the expected changes of inflation, we need the fraction that report "inflation goes up" and the fraction that report "inflation goes down", which we denote with A_t and B_t, respectively. For the normal distribution we define $a_t = \Phi^{-1}(1 - A_t)$ and $b_t = \Phi^{-1}(B_t)$, where $\Phi^{-1}(\cdot)$ is the inverse of the probability function of the standard normal distribution. We can proceed with other distributional assumptions in the same way. For the logistic distribution a_t and b_t are given by $a_t = \log[A_t/(1 - A_t)]$ and $b_t = \log[(1 - B_t)/B_t]$. For the third alternative – the scaled t-distribution – the standard normal probability distribution by a student t-distribution can be replaced with η degrees of freedom. Note that the distributions differ only in their tails.

The expected change in the inflation rate during the next 6 months, $E\left(\Delta_6\pi^{12}_{t+6}|\Omega_t\right)$, can now be calculated with Ω_t the information set at time t, and π^{12}_t defined as $\pi^{12}_t = 100\ln\left(P_t/P_{t-12}\right)$, the year-on-year inflation rate and $\Delta_6\pi^{12}_{t+6} = \pi^{12}_{t+6} - \pi^{12}_t$ denoting 6 month changes of year-on-year inflation rates.[5] This expression is a function of the variables a_t, b_t and δ_t given by

$$E\left(\Delta_6\pi^{12}_{t+6}|\Omega_t\right) = E_t\left(\Delta_6\pi^{12}_{t+6}\right) = -\delta_t\left(\frac{a_t+b_t}{a_t-b_t}\right). \tag{3.1}$$

Obviously, we need to make further assumptions about the parameter δ_t in order to identify expected inflation changes. Carlson and Parkin assume that this parameter is constant in time $\delta_t = \delta$ and symmetric for price increases and decreases. In our problem, this translates into symmetry in increases and decreases in inflation. They further assume long-term unbiasedness, which implies, in our setting that δ can be calculated as

$$\widehat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\left|\Delta_6\pi^{12}_{t+6}\right| / \frac{1}{T}\sum_{t=1}^{T}\left|\frac{a_t+b_t}{a_t-b_t}\right|. \tag{3.2}$$

This implies that $\left|\frac{a_t+b_t}{a_t-b_t}\right|$ is scaled to match the average absolute inflation change. This is what we refer to as long-term unbiasedness.[6] This scaling technique is critical, since it generally does not correspond to what statisticians mean by unbiasedness. For this reason we choose an alternative method to estimate δ. Batchelor (1982) proposes estimating δ by regressing inflation changes on unscaled estimates $\frac{a_t+b_t}{a_t-b_t}$ (without considering a constant). This procedure is referred to as statistical unbiasedness.[7] This regression is used for the model with the scaled t-distribution to calibrate η in such a way that we maximize the fit (the R^2) be-

[5] All the subsequent analyses employ seasonally adjusted CPI data provided by the Bundesbank. In a previous version of the chapter we analyzed seasonally unadjusted data which is the more prominent inflation measure in the public debate. However, these measures makes it necessary to explicitly model the seasonal pattern which complicates the empirical analysis. In general, main results remain unchanged. Results based on seasonally unadjusted inflation data are available on request.

[6] In the original C-P method, $\frac{a_t+b_t}{a_t-b_t}$ is scaled to match average price changes.

[7] Breitung (2008) discusses this later method more generally in a quasi GMM framework which also allows for the inclusion of instruments in estimating δ.

tween $\Delta_6\pi^{12}_{t+6}$ and the unscaled measures. The result is that the t-distribution with $\eta = 18$ degrees of freedom performs best.

As an alternative to the C-P method, Pesaran (1984) proposes a so-called regression approach. The basic premise of this procedure is that the relationship between official actual time series and the respondents' past perceptions can be used as a yardstick for quantification of their expectations relating to the future. However, the ZEW survey does not possess information on respondents' previous perceptions of inflation rates. Instead we follow others (e.g. Dasgupta and Lahiri, 1992; Breitung, 2008) and use the regression approach for the relation of inflation expectations and future realizations. Under rational expectations, this implies estimating the regression

$$\Delta_6\pi^{12}_{t+6} = \alpha A_t + \beta B_t + u_{t+6}. \tag{3.3}$$

Whenever $\alpha = -\beta$, the method equals the balance statistic mentioned above.

A further extension of the standard probability method and the regression method is to consider asymmetric thresholds in the imperceptibility parameter which implies that the indifference interval is no longer symmetric around 0. In this case, statistical unbiasedness is no longer imposed. For the probability approach, expected inflation changes can be calculated as

$$E_t\left(\Delta_6\pi^{12}_{t+6}\right) = \delta^L\left(\frac{a_t}{a_t - b_t}\right) + \delta^U\left(\frac{-b_t}{a_t - b_t}\right). \tag{3.4}$$

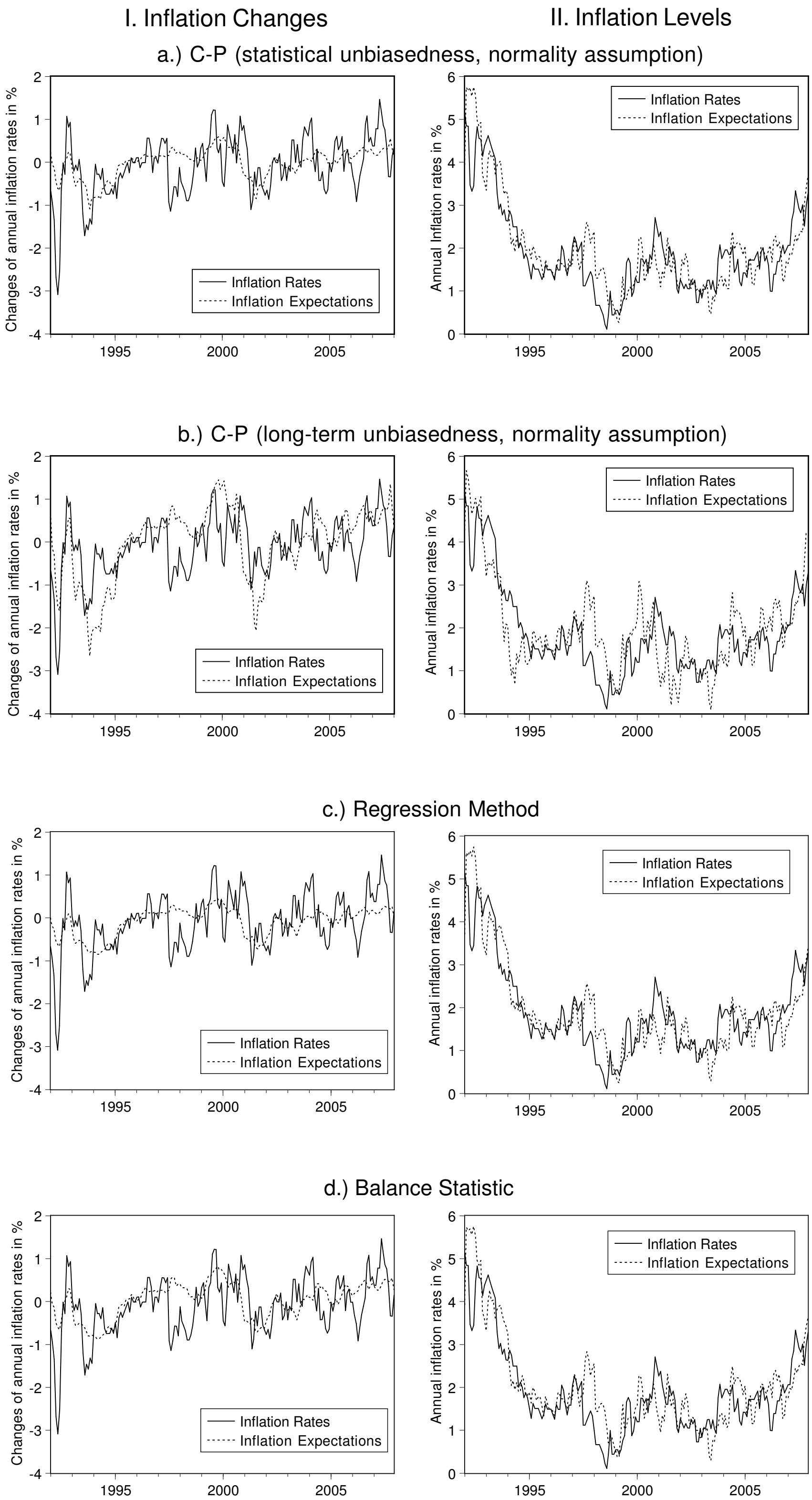
Again estimation of $\widehat{\delta}^L$ and $\widehat{\delta}^U$ can be done by means of a regression when statistical unbiasedness is assumed (using eq. 3.4 by replacing expected changes with those realized) or, in the case of long-term unbiasedness, by scaling to the mean of changes in inflation expectations. In this case the mean differs for positive and negative inflation changes, respectively.

Additionally, we can augment the regression approach by allowing for an asymmetric threshold according to

$$\Delta_6\pi^{12}_{t+6} = \frac{\alpha A_t + \beta B_t}{1 - \lambda A_t} + e_{t+6}, \tag{3.5}$$

which can be estimated using nonlinear least squares.

Figure 3.1: Expected and realized inflation rates



After the calculation of expected inflation changes it is also possible to compute expected inflation levels for the annual inflation rate. Since the actual inflation rate of the ongoing month is unknown at the point when the ZEW survey is conducted, experts base the expected change of the inflation rate on the most recent inflation release, referring to the inflation rate for the previous month. So we follow Heinemann and Ullrich (2006) and calculate the expected annual year-on-year inflation rate as

$$E\left(\pi_{t+6}^{12}|\Omega_t\right) = E_t\pi_{t+6}^{12} = \pi_{t-1}^{12} + E_t\left(\Delta_6\pi_{t+6}^{12}\right) \tag{3.6}$$

where $E_t\pi_{t+6}^{12}$ is the expected annual inflation rate for time $t+6$ and π_{t-1}^{12} is the inflation rate on which experts base their opinion about changes in the inflation rate (as known to the survey participants). Since it is, in principle, unclear as to which inflation rate respondents are referring in their report on expected changes, our subsequent analysis focuses on inflation changes rather than on inflation levels.

3.3 Some basic properties and tests

Because we want to explore whether the ZEW survey by financial experts provides useful information about future inflation rates we first describe the properties of expected inflation changes obtained by means of various quantification methods as outlined in section 3.2. Our sample therefore includes the monthly series from 1992.5 to 2008.7 concerning six month changes of annual inflation rates and the corresponding expected inflation changes. Basic stationarity tests reveal that both year-on-year inflation rates and expected inflation levels can be characterized by stationary behavior (see Appendix, Table 3.8 for the results of two standard unit root tests: the ADF- and PP test). As illustration, we plot how inflation and expected inflation evolve over time (see Figure 3.1). The two versions of the C-P approach (based on the normality assumption) are shown to differ with respect to the computation of $\widehat{\delta}$. In addition, the series based on the regression approach and balance statistic are also displayed. It can be seen that estimated changes in inflation expectations based on the long-term unbiasedness assumption are more volatile than the series obtained with other conversion techniques (see also Appendix, Table 3.7 for a descriptive summary). The remaining series display similar characteristics and can hardly be differentiated when viewed.

Table 3.1 gives an overview of all the methods under consideration. First, we show how different versions of the probability approach yields to different estimates of the imperceptibility parameter $\widehat{\delta}$. Since our proposal in a subsequent section concerns forecasting inflation, we provide some measures of how successfully survey expectations match the realized variables. Second, we consider asymmetric thresholds for all possible quantification methods. Using a standard coefficient test, it is possible to decide whether statistical differences exist.

In addition, we present evidence on the direction of causality within a standard Granger test (see columns 6-8). This causality test helps to determine whether the expected inflation series contains additional information about future inflation beyond what is already contained in the past history of actual inflation. If inflation expectations do not Granger-cause inflation, it might be a bad indicator for inflation.

Our results indicate several interesting facts. The estimated imperceptibility parameter $\widehat{\delta}$ varies considerably between the two estimation methods (statistical unbiasedness vs long-term unbiasedness). The statistical unbiasedness assumption leads to an estimated $\widehat{\delta}$ of about 0.27, which is remarkably close to what Heinemann and Ullrich (2006) report based on additional survey information.[8] The estimates based on long-term unbiasedness assumption are more then twice as large. The regression approach displays a similar fit of the data (MAEs and RMSEs) as compared to the C-P method with statistical unbiasedness.[9] The balance statistic coincides approximately with the fit of the regression approach, although the statistical hypothesis underlying the balance statistic can be rejected.[10] Generally, different distributional assumptions for the C-P approach change the in-sample fit only marginally. In all our specifications, we find no significant evidence for an asymmetric threshold.

[8]The ZEW undertook two polls to get more information on the threshold. In December 1993 this information was used to calculate a symmetric threshold, which was 0.23. Later they allowed for asymmetry and calculated the threshold to be -0.18 and +0.21.

[9]This result is not particularly surprising, since the two methods can be related to each other. Breitung (2008) show under which condition the two approaches produce the same output. One of these conditions is to assume a uniform distribution for the subjective probability distribution.

[10]This hypothesis can be tested with a joint significance test in eq. (3.3) equal to $\alpha = -\beta$. For this reason a Wald test with HAC standard errors is used.

Table 3.1: Estimated imperceptibility parameters, in-sample fit and Granger causality from alternative procedures

Procedure	$\hat{\delta}$		MAE	RMSE	Granger-Causality $\pi^e \rightarrow \pi$ (p-values)	$\pi \rightarrow \pi^e$	lags	
Symmetric Indifference Interval								
i. C.-P. with Statistical unbiasedness								
- Normal	0.27		0.47	0.39	0.08	0.00	13	
- Logistic	0.29		0.47	0.39	0.05	0.00	13	
- Scaled-t	0.28		0.47	0.39	0.06	0.00	13	
ii. C.-P. with Long-term unbiasedness								
- Normal	0.67		0.62	0.63	0.08	0.00	13	
- Logistic	0.70		0.61	0.61	0.05	0.00	13	
- Scaled-t	0.68		0.62	0.62	0.06	0.00	13	
Regression	0.54 / -0.97†		0.47	0.39	0.04	0.13	13	
Balance			0.49	0.41	0.03	0.06	13	
	$\hat{\delta}^L$	$\hat{\delta}^U$	MAE	RMSE	Granger-Causality $\pi^e \rightarrow \pi$	$\pi \rightarrow \pi^e$	lags	Asy p-value
Asymmetric Indifference Interval								
i. C.-P. with Statistical unbiasedness								
- Normal	-0.35	0.20	0.47	0.38	0.08	0.00	13	0.31
- Logistic	-0.38	0.21	0.47	0.38	0.05	0.00	13	0.26
- Scaled-t	-0.36	0.20	0.47	0.38	0.06	0.00	13	0.29
ii. C.-P. with Long-term unbiasedness								
- Normal	-0.89	0.36	0.62	0.62	0.08	0.00	13	0.50
- Logistic	-0.85	0.33	0.58	0.53	0.05	0.00	13	0.55
- Scaled-t	-0.87	0.35	0.60	0.58	0.06	0.00	13	0.52
Regression	0.57 / -0.99 / -0.14†		0.47	0.39	0.04	0.14	13	0.93

Notes: The second and third columns report the estimated parameters for δ. For the regression approach (†) the numbers correspond to the estimated parameters for α and β (as well as λ when an asymmetric indifference interval is considered). The MAE and RMSE are calculated from in-sample deviations of observed and expected inflation changes. The optimal lag length of the Granger causality tests is selected according to AIC. $\pi^e \rightarrow \pi$ corresponds to the null hypothesis that inflation expectations do not Granger-cause actual inflation rates. $\pi \rightarrow \pi^e$ corresponds to the null hypothesis that inflation does not Granger-cause inflation expectations. The last column reports p-values for the tests for asymmetric indifference intervals (denoted by Asy p-value). The corresponding null hypothesis is $\delta^L = -\delta^U$ (and $\lambda = 0$ for the regression approach).

Granger causality tests indicate that a feedback effect between inflation and inflation expectations is present (at least at a 10% confidence level; the only exception is the regression approach where causality runs exclusively from expectations to realizations).[11] These results are in contrast with those of other studies that identify a unidirectional causality running from inflation to inflation expectations (Berk, 1999; Forsells and Kenny, 2004). Their results are obtained by using consumer price expectations, and conclude that these indicators may be not very helpful in forecasting inflation (since most information is already incorporated in past inflation rates). Our results indicate that the series of inflation expectation from the ZEW financial market survey may be suitable for predicting future inflation.

The literature on measured inflation expectations applies various tests to investigate whether expectations are formed rationally. These tests can be characterized by four hypotheses associated with the rational expectation assumption. The test hypotheses may be based on unbiasedness, lack of serial correlation, efficiency and orthogonality (Pesaran, 1987). Since our approach for the quantification of inflation expectations allows only for testing unbiasedness via an asymmetric imperceptibility parameter which cannot be confirmed by our test results, we cannot go any further in this direction to test for this specific assumption. Moreover, we also have to account for measurement errors in our series of inflation expectations. This is an additional argument why tests of unbiasedness and lack of serial correlation are not appropriate in our setting. For instance, the frequently applied test of unbiasedness by regressing the inflation series on expected inflation is very difficult to interpret when we assume measurement errors in the expected series. In this case the test is based on the null hypothesis that the constant is zero and the coefficient of the expected series is equal to one. It is clear that when we allow for measurement errors the coefficient of the expected series is biased toward zero. So when the test for unbiasedness rejects, it is not clear whether this is due to measurement errors or a lack of unbiasedness.

Our emphasis is on testing the orthogonality assumption which is also associated with rationality. The basic aim is to check whether the forecast error of inflation changes $\Delta_6\pi_t^{12} - \Delta_6 E_{t-6}\pi_t^{12}$ is orthogonal to the (costless) information set available at the time when expectations are formed. If this hypothesis cannot be rejected, it would indicate that expectations had been formed in a rational way.

[11] Note that the Granger-causality test for the way in which $\widehat{\delta}$ is estimated does not matter for the specific test result as long as the parameter is constant in time.

Table 3.2: Orthogonality tests for alternative quantification procedures

	C-P, statistical unbiasedness						C-P, long-term unbiasedness						Regression		Balance	
	Normal		Logistic		Scaled-t		Normal		Logistic		Scaled-t					
c	-0.004	-0.010	-0.012	-0.010	-0.007	-0.010	0.180	0.200	0.160	0.170	0.170	0.190	0.082	0.080	-0.025	-0.020
	(0.08)	(0.13)	(0.08)	(0.11)	(0.08)	(0.11)	(0.10)	(0.13)	(0.09)	(0.13)	(0.10)	(0.13)	(0.08)	(0.11)	(0.08)	(0.11)
$\Delta_6\pi^{12}_{t-3}$	-0.208	–	-0.207	–	-0.208	–	-0.200	–	-0.200	–	-0.200	–	-0.218	–	-0.226	–
	(0.09)	–	(0.09)	–	(0.09)	–	(0.08)	–	(0.08)	–	(0.08)	–	(0.09)	–	(0.08)	–
$\Delta_{12}r^s_{-4}$	0.132	0.110	0.140	0.120	0.135	0.120	-0.160	-0.380	-0.120	-0.400	-0.140	-0.390	0.151	0.140	0.099	0.090
	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)	(0.11)	(0.18)	(0.10)	(0.17)	(0.11)	(0.18)	(0.08)	(0.09)	(0.08)	(0.09)
$(r^l - r^s)_{-6}$	-0.286	-0.270	-0.298	-0.280	-0.291	-0.270	-0.588	-0.490	-0.590	-0.510	-0.592	-0.500	-0.304	-0.280	-0.299	-0.270
	(0.14)	(0.09)	(0.14)	(0.13)	(0.14)	(0.13)	(0.28)	(0.34)	(0.25)	(0.29)	(0.27)	(0.32)	(0.14)	(0.12)	(0.13)	(0.12)
$(r^l - r^s)_{-2}$	0.262	0.250	0.276	0.270	0.268	0.260	0.180	0.120	0.210	0.160	0.190	0.140	0.275	0.270	0.247	0.230
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.18)	(0.21)	(0.17)	(0.20)	(0.18)	(0.21)	(0.15)	(0.16)	(0.14)	(0.16)
$\Delta_{12}p^{raw}_{-1}$	-0.817	–	-0.816	–	-0.819	–	-2.540	–	-2.450	–	-2.510	–	-0.650	–	-1.092	–
	(0.47)	–	(0.47)	–	(0.47)	–	(0.61)	–	(0.60)	–	(0.61)	–	(0.48)	–	(0.50)	–
$\Delta_{12}p^{raw}_{-4}$	0.909	–	0.894	–	0.904	–	2.620	–	2.490	–	2.570	–	0.930	–	1.261	–
	(0.51)	–	(0.51)	–	(0.51)	–	(0.64)	–	(0.63)	–	(0.64)	–	(0.51)	–	(0.53)	–
R^2adj.	0.11	0.03	0.11	0.03	0.11	0.03	0.36	0.21	0.34	0.20	0.35	0.21	0.12	0.04	0.13	0.03
χ^2-Stat	12.19	5.72	12.61	6.34	12.38	5.98	62.64	22.39	66.87	23.14	64.75	22.84	12.77	7.22	17.46	7.31
p-value	**0.06**	**0.13**	**0.05**	**0.10**	**0.05**	**0.11**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.07**	**0.01**	**0.06**

Notes: Variable selection is done in a step-wise procedure minimizing SIC. The χ^2 test is the test of joint significance of all coefficients with the corresponding p-value are given. Newey and West standard errors are given in parenthesis. Sample period: 1992.5-2008.2.

However, it is important to point out that, because of the qualitative nature of the data, the interpretation of the orthogonality test is somehow limited. Nevertheless, the orthogonality test can give further insights into the expectation for the formation process, even when the data is of qualitative nature. In practice, it should be clear that it is possible to use only a very small subset of this general information set. In accordance with the the existing literature (e.g. Pesaran, 1987), we consider macroeconomic variables to proxy, at least in part, the available information set. Basically, we can categorize the variables under consideration into three classes: variables that reflect real economic activity (industrial production, the unemployment rate); financial variables (short- and long-term interest rates as well as the spread between the two); and a price variable reflecting price changes for foreign raw materials and energy. We consider all the variables with lags up to one year after the time when expectations were formed.[12] For industrial production and unemployment rates, we also construct detrended measures using a quasi real-time HP-filter (that uses only past information). For other variables we take year-on-year differences (see also Appendix for the exact variable definitions).

The orthogonality test is conducted in the following way. We run regression

$$\Delta_6 \pi^{12}_{t+6} - \Delta_6 E_t \pi^{12}_{t+6} = c + x_t' \beta + e_{t+6}, \tag{3.7}$$

where x_t contains all relevant explanatory variables including lags of inflation changes, the variables mentioned above and c as a constant. We exclude all insignificant variables in a stepwise procedure. Then we test the joint null hypothesis that $c = 0$ and $\beta = 0$. Since autocorrelation and heteroskedasticity may be an issue in this model we use Newey-West standard errors to calculate the test statistic. Table 3.2 shows the results of two different specifications using different conversion methods. The first test considers all explanatory variables together with lagged inflation changes. In the second specification we omitted the lagged values of the endogenous variable.

In the first specification six explanatory variables turn out to be significant and can explain at least some variance of the forecast error. Again results differ a great deal between the two estimation methods for $\widehat{\delta}$. The tests are highly

[12] Owing to publication lags we use only variables that are available at time t when expectations are formed.

significant for long-term unbiasedness and there is clear evidence of correlation between the forecast errors and past information (irrespective of whether or not a lagged inflation change is added).

However, using the statistical unbiasedness assumption makes for some lack of clarity. In this case, it seems that there are still some variables that have some explanatory power, mainly financial variables and prices of raw materials. When lag terms of inflation changes are omitted it becomes difficult to find past variables that are correlated with the forecast errors. These are mainly financial variables like the short-term interest rate and the spread between long-term and short-term interest rates. Further, the changes in commodity prices seem to matter, at least for the first specification. Interestingly, no real economic variable shows up in the regressions (all turned out to be insignificant). Altogether, the orthogonality tests indicate that inflation expectations from the ZEW are not far from rational. However, the findings of the orthogonality test depend strongly on the underlying conversion method. If at all, then financial variables (interest rates and yield spreads) may improve inflation forecasts by using this information more efficiently.

3.4 Survey expectations as indicator for future inflation

In this section we conduct an out-of-sample forecasting experiment with different competing models for inflation forecasting. The main focus is on the relative forecasting performance by models that employ the survey-based measures introduced in the previous sections. Three alternative model classes for inflation forecasting in Germany are examined: univariate time series models (autoregressive and unit root models), regressions with real activity variables (motivated from the Phillips curve) and term structure models. Further, methods for combining these three elements with the information of survey measures are discussed.

3.4.1 Forecasting models and setup

The ability of inflation forecasting plays a role crucial in conducting optimal monetary policy. Further, it is important for private agents for price setting, optimal investment or negotiation of wage contracts. Here we ask whether qualitative inflation expectations contribute to standard inflation models in terms of forecasting performance. By standard inflation models we mean easy applicable statistical models found in the literature to be good for predicting inflation. The intension is therefore that inflation expectations from surveys (particularly those from economic experts) should contain much of the information that exists also in single equation models with a small number of explanatory variables. Ang, Bekaert, and Wei (2007) document the forecasting performance of different surveys for inflation expectations in the US along with their relative forecast accuracy. Another of their findings is that survey measures do indeed provide information important for forecasting future inflation. As far as Germany is concerned, there is no comprehensive available study that assesses survey measures for predicting inflation.

Like Stock and Watson (1999) and Ang, Bekaert, and Wei (2007) we conduct an out-of-sample experiment to assess the relative predictive power of the ZEW survey measure for inflation. This is the dominant model evaluation method because it is not subject to overfitting, in contrast to in-sample measures of model accuracy. Natural benchmark models are univariate time-series models and simple indicator models (including some measure of real economic activity or some financial market indicator). We model German inflation rates as a stationary process. Our test results (see Appendix, Table 3.8) are in line with this assumption. Moreover, they are also in line with New Keynesian models in which the monetary authority follows a stable inflation objective (which seems plausible, seeing that the Bundesbank provided a credible monetary policy regime for many years before being replaced by the ECB, which now have a similar objective). The out-of-sample experiment is conducted for different forecasting horizons. First, we consider the forecasting horizon six month ahead, which coincides with the forecasting horizon of the ZEW survey.[13] Then we compare models of a shorter horizon (three months), after which we consider models of a longer time span, 12 and 18 months ahead forecasts.

Most of the forecasting models employed here have the following structure

[13]Note that the ZEW survey asks for inflation changes (annual inflation rate) within the next 6 months which is generally different form the annualized change of inflation over the next six months. However, each variable can be transformed into the other definition, e.g. $\Delta_6\pi_t^{12} = \pi_t^6 - \pi_{t-12}^6$.

$$\pi_{t+h}^{h} = \alpha + \beta(L)\,\pi_t^h + \gamma(L)\,x_t + u_{t+h}, \tag{3.8}$$

whereas π_t^h is defined as $\pi_t^h = 1200/(100h)\ln(P_t/P_{t-h})$ with P_t the seasonally adjusted CPI-Index in period t and h the forecasting horizon (defined in months). In this specification x_t stands for one additional explanatory variable with its corresponding lags; e.g. real activity measures, financial indicators, or expected inflation rates. To keep this setting more realistic, we consider only the indicators for x_t that are available when forecasts are conducted. $\beta(L)$ and $\gamma(L)$ denote lag polynomials which are defined as $\left(1 - L - L^2 - \ldots - L^k\right)$. However, k is allowed to be different for $\beta(L)$ and $\gamma(L)$. This step-up implies a direct approach for multistep forecasting which is comparable with the approach by Stock and Watson (1999; 2003). Lags are selected according to the AIC criteria in the first in-sample period. To avoid overfitting, we allow for zero restrictions of the coefficients β and γ.[14]

Following this, we describe the forecasting models used in the following out-of-sample forecasting experiment. First, we consider a simple univariate autoregressive model that serves as a first benchmark for all other models under consideration. This specification is a special case of eq. (3.8) where the x_t regressor and its lags are absent. Again, the lag length of the AR model is selected according to AIC with further restrictions.

Besides the AR model we also consider other univariate time series models: an ARMA model (*ARMA*) and a random walk model (*RW*). The ARMA model has a parsimonious specification and includes one AR and one MA term (see Hamilton, 1985, for a motivation). The RW model is inspired by Atkeson and Ohanian (2001). These authors find out that many inflation models for the US (including Phillips curve models) cannot outperform a simple unit root model. This result is confirmed by Fisher, Liu, and Zhou (2002) who documented the superiority of random walk forecasts for inflation in the 1990s over other models of inflation. The unit root model is constructed as a yearly average of past inflation rates.

The second model class consists of Phillips curve models. By Phillips curve

[14] This is carried out with a step-wise approach, where the least significant variable is always excluded. This procedure is repeated until the model with the lowest AIC is obtained. Overall, this procedure leads to smaller forecast errors than those of models where no further coefficient restrictions are imposed.

models is meant specifications that link the inflation rate with some measure of real activity such as the unemployment rate or an output gap variable. The Phillips curve plays a prominent role in the theoretical monetary models known as New Keynesian models (see Galí and Gertler (1999) for a theoretical and empirical analysis as well as Scheufele (2010) for an empirical evaluation for Germany). We have not estimated these structural models, here; instead we have used a reduced form model that can be seen as a rough approximation of a more sophisticated Phillips curve model.[15] Different real activity measures are considered (and included for x_t in the general specification of eq. 3.8): the unemployment rate (PC1, which is the real activity measure in early Phillips curve specifications when a constant NAIRU is assumed), the deviation of the unemployment rate from a quasi real-time Hodrick-Prescott-Trend (PC2, motivated from a time-varying NAIRU assumption) and detrended industrial production (PC3, also quasi real-time detrended using the HP-filter which corresponds to an output gap measure).[16]

Next, financial variables are used to predict inflation. Motivated by Fama (1975) and Mishkin (1990), we consider both the short-term interest rate and a term spread as inflation indicators. Several theoretical hypotheses rest on the idea that the yield curve is forward looking and may thus provide information about future inflation (e.g. see Kozicki, 1997, for theoretical arguments). We employ different models that include either the short interest rate (TS1), the term spread (TS2) or both (TS3) as regressors for x_t in eq. (3.8).[17]

Finally, we investigate the usefulness of survey expectations in forecasting inflation. Note that our inflation definition in the forecasting experiment is somewhat different from those in section 3.2 where annual inflation rates and their changes are considered. We choose the specification in the forecasting experiment to follow as closely as possible the literature on inflation forecasting (e.g. Stock and Watson, 1999) and to avoid having to explain movements in the inflation rate that occurred in the past (which would be the case when investigating π_{t+6}^{12} which originates di-

[15] This corresponds with the approach by Stock and Watson (1999) and Ang, Bekaert, and Wei (2007), who use similar Phillips curve specifications in their forecasting experiment.

[16] See Table 3.9 for details concerning the model specifications.

[17] We also considered monetary aggregates as potential regressors of inflation. But non of them display acceptable forecasting results for all the time horizons under consideration. This may be owing to the changes in response to the start of the European Monetary Union (although continuous series are constructed). So we skip these variables in the subsequent analysis.

rectly from the ZEW inflation expectation). However, it should be borne in mind that it is possible to convert annual inflation expectations $E_t\pi^{12}_{t+6}$ into semi-annual inflation expectations according to $E_t\pi^{6}_{t+6} = E_t\pi^{12}_{t+6} - \pi^{6}_{t}$. Now we can directly compare the expectations with forecasts by other models. This approach is thus directly applied to the raw data for semi-annual inflation rates. Another possibility is that of using inflation expectations as a regressor for x_t in eq. 3.8 (together with possible lags ang lags of realized inflation rate). In doing so, we need make no further assumptions about how to estimate the threshold $\hat{\delta}$ for the probability approach. Instead, we could use the unscaled data $\left(\frac{a_t+b_t}{a_t-b_t}\right)$ directly. In the forecasting setup, the coeffient(s) of x_t could then be seen as the scale factor, which may be different for each forecasting round. Owing to the rolling window, some time variation of the scaling factors is implicitly allowed for. The combination with lagged inflation rates additionally allows for a flexible dynamic structure. Because the orthogonality tests reveal that autocorrelation is an issue, this information can be used for forecasting. All three distributional assumptions are considered: standard normal distribution (NORM), scaled t-distribution (T) and the logistic distribution (LOG). Additionally, the regression approach (REG) and the balance statistic (BAL) are also used in the forecasting experiment. For comparison, we also estimate a model using the balance statistic from the EU Consumer Survey (CBAL) for Germany as a potential regressor.[18]

All models are estimated with the specific selection of regressors for the sample period 1992.5 to 2000.5. Then we obtain out-of-sample forecasts with a rolling window of fixed length (2000.5+h to 2008.7). Parameters are sequentially updated for the remaining time period and we obtain forecasts for inflation π^{h}_{t+h} with a constant in-sample period. So we always have an in-sample estimation period using 97 observations ($R = 97$), and we obtain $99 - h$ out-of-sample estimates for inflation ($P = 99 - h$). We use the rolling scheme because this may guard against moment or parameter drifts which are very difficult to model explicitly (see, for

[18]Note that the balance statistic of the consumer survey is not directly comparable with the ZEW survey. For example, the consumer survey asks about price changes (not inflation changes) and it has a different time horizon (12 months instead of six). Although it would be preferable to use conversion techniques other than the simple balance statistic as well, disaggregate responses are unavailable for Germany on a monthly basis (the European Commission does not report this information for Germany, although it does so for the Euro area in general and for most other member countries).

instance, West, 2006, Section 4, for a discussion).

3.4.2 Forecast evaluation

When it comes to examining the predictive accuracy of different models, we assume a symmetric loss function given by mean square forecast error loss as well as mean absolute forecast error loss. Given these loss functions, we can now evaluate and compare the outcomes of different models. In particular, we would like to assess how well models that employ a survey measure of inflation expectations perform in comparison with other standard models for inflation. First, models are evaluated according to their root mean squared forecast error (RMSFE) as well as their mean absolute forecast error (MAFE). Then we ask whether these differences are statistically significant (which can be done using tests of equal predictive accuracy). Next we investigate the issue of forecast encompassing. The intension is to examine whether forecasts from other models contain additional information not included in the reference forecast (irrespectively of its own forecasting performance in terms of RMSFE or MAFE).

Throughout the evaluation step, we consider parameter uncertainty due to estimation which is relevant for the specific tests (see West, 1996). However, in their seminal work Giacomini and White (2006) employ an alternative asymptotic framework, in which the differentiation between nested and non-nested models becomes irrelevant. Consequently, no further test adjustments are necessary. This framework is valid whenever a rolling (or fixed) estimation window is used. The writers' test rests on the null hypothesis

$$E\left[\Delta L_{t+h}|\Psi_t\right] = 0, \tag{3.9}$$

where ΔL_{t+h} is the loss difference for the $t+h$ variable pair (h denotes the forecasting step). This differs from West's (1996) hypothesis in two ways. First, the losses depend on estimated parameters rather than on their probability limits. Second, the expectation is conditional on the information set Ψ_t. This approach accounts for bias-variance trade-offs, whereas West (1996) and others eliminate the effects of in-sample estimation uncertainty.

In the following analysis we consider two test variants: an unconditional and a conditional test. The unconditional test assumes $\Psi_t = \{\emptyset, \Omega\}$. This test is thus equivalent to that of Diebold and Mariano (1995) on equal predictive ability. The conditional test requires $\Psi_t = F_t$, where we use the test function $g_t = [1\ \Delta L_t]$. Let

$$\Delta L_{t+h}^{(m,n)} = \left(\hat{e}_{m,t+h}^{h}\right)^2 - \left(\hat{e}_{n,t+h}^{h}\right)^2$$

for mean squared forecast error loss.[19] $\hat{e}_{m,t+h}^{h}$ is the forecast error of model m at time $t+h$. The test statistic equals

$$GW_h^{(m,n)} = N\left(\frac{1}{N}\sum_{t=T_1}^{(T_2-h)} g_t \Delta L_{t+h}^{(m,n)}\right)' \hat{\Omega}^{-1}\left(\frac{1}{N}\sum_{t=T_1}^{(T_2-h)} g_t \Delta L_{t+h}^{(m,n)}\right), \qquad (3.10)$$

where $N = T_2 - T_1 - h + 1$ is the sample size (for the out-of-sample period). g_t is a $q \times 1$ measurable test function which we set equal to $g_t = [1\ \Delta L_t]$ and a HAC covariance matrix $\hat{\Omega}$. Under some regularity conditions $GW_h^{(m,n)} \overset{a}{\sim} \chi_q^2$.

In testing for forecasting encompassing we rely on Harvey, Leybourne, and Newbold's (1998) encompassing test, which can be formulated as

$$\hat{e}_{t+h}^{m} = \lambda\left(\hat{e}_{t+h}^{m} - \hat{e}_{t+h}^{n}\right) + u_{t+h}, \qquad (3.11)$$

with $\hat{e}_t^m$ the forecasting error of the benchmark model m and $\hat{e}_t^n$ the error of the competing model n. This test is based upon the null hypothesis that $\lambda = 0$, whereas $\lambda > 0$ is the alternative. Therefore eq. (3.11) is estimated with OLS using HAC standard errors. Under the asymptotics of Giacomini and White (2006) again no further adjustments are needed and the limiting distribution is standard normal. The corresponding loss function equals $\Delta L_{t+h}^{(m,n)} = \hat{e}_{t+h}^{m}(\hat{e}_{t+h}^{m} - \hat{e}_{t+h}^{n})$.

[19] Similarly, we can define the loss differences in mean absolute forecast error loss as $\Delta L_{t+h}^{(m,n)} = \left|\hat{e}_{m,t+h}^{h}\right| - \left|\hat{e}_{n,t+h}^{h}\right|$

Table 3.3: Error measures using raw data of inflation expectations

Step=6	RMSE	MAE
i. C.-P. with Statistical unbiasedness		
- Normal	1.03	0.92
- Logistic	1.03	0.92
- Scaled-t	1.03	0.92
ii. C.-P. with Long-term unbiasedness		
- Normal	2.12	1.28
- Logistic	1.98	1.25
- Scaled-t	2.07	1.27
Regression	1.07	0.93
Balance	1.07	0.93

3.4.3 Results

Univariate and single indicator models

Tables 3.3, 3.4 and 3.5 summarize the main findings of our forecasting experiment. Table 3.3 reports the forecast errors based on raw data (without estimation). These values can be directly compared with the results in Tables 3.4 and 3.5. However, since inflation expectations are constructed for a six-month horizon, only forecast step six can be compared. Obviously raw data display more inferior forecast error characteristics than do most forecasting models. Even simple autoregressive time series models perform much better than raw data on inflation expectations. This holds for all convertibility methods. This result is in line with that of Breitung (2008), who states that ZEW inflation expectations are less accurate than simple time series models.

When inflation expectations are used as indicator models in a dynamic equation, the results change considerably. For a forecast horizon of six months (which is the horizon of the survey), the relative forecasting performance is much better. Both RMSFEs and MAFEs are lower than most competitive models (even if most differences are not significant). The only exceptions are term spread models that display even smaller forecast errors as well as a significantly better predictive ability for the six-months ahead forecast horizon. Interestingly, although there is no great difference in forecast accuracy between conversion methods, the C-P method

Table 3.4: Forecasting performance of alternative models (mean square loss)

	Step=3			Step=6			Step=12			Step=18		
	RMSFE	I	II	RMSFE	I	II	RMSFE	I	II	RMSFE	I	II
Univariate Models												
AR	1.26	–	–	0.93	(0.26)	(0.79)	0.77	(0.08)	(0.28)	0.69	(0.19)	(0.06)
ARMA	1.34	(0.35)	(0.45)	1.06	(0.16)	(0.55)	0.77	(0.40)	(0.16)	0.64	(0.51)	(0.11)
UT	1.34	(0.12)	(0.28)	0.95	(0.48)	(0.85)	0.79	(0.29)	(0.44)	0.66	(0.90)	(0.20)
Single Indicator Models												
PC1	1.30	(0.20)	(0.05)	0.98	(0.08)	(0.10)	0.85	(0.01)	(0.01)	0.84	(0.01)	(0.03)
PC2	1.30	(0.19)	(0.27)	0.94	(0.43)	(0.33)	0.87	(0.00)	(0.01)	0.80	(0.00)	(0.07)
PC3	1.26	–	–	0.92	(0.76)	(0.36)	0.84	(0.05)	(0.44)	0.77	(0.06)	(0.09)
TS1	1.26	(0.92)	(0.85)	0.92	(0.92)	(0.19)	0.73	(0.95)	(0.26)	0.76	(0.06)	(0.28)
TS2	1.26	–	–	0.83	(0.03)	(0.10)	0.75	(0.34)	(0.34)	0.69	(0.19)	(0.06)
CBAL	1.36	(0.16)	(0.26)	0.98	(0.13)	(0.25)	0.84	(0.02)	(0.10)	0.78	(0.01)	(0.00)
NORM	**1.26**	**–**	**–**	**0.91**	**–**	**–**	**0.72**	**–**	**–**	**0.67**	**–**	**–**
T	1.26	–	–	0.91	(0.31)	(0.34)	0.72	(0.25)	(0.30)	0.67	(0.21)	(0.22)
LOG	1.26	–	–	0.91	(0.34)	(0.37)	0.73	(0.15)	(0.26)	0.67	(0.22)	(0.26)
PES	1.26	–	–	0.93	(0.00)	(0.05)	0.76	(0.01)	(0.17)	0.70	(0.04)	(0.06)
BAL	1.26	–	–	0.92	(0.01)	(0.12)	0.75	(0.02)	(0.27)	0.69	(0.09)	(0.06)
Indicator models with 2 Indicators												
TS1 - NORM	1.26	(0.92)	(0.85)	0.90	(0.88)	(0.30)	0.73	(0.90)	(0.41)	0.79	(0.03)	(0.09)
TS2 - NORM	1.26	–	–	0.83	(0.07)	(0.19)	0.72	–	–	0.67	–	–
Model averaging schemes												
MSFE-weights	1.26	(0.48)	(0.74)	0.88	(0.19)	(0.49)	0.72	(0.65)	(0.35)	0.67	(0.66)	(0.22)
AIC-weights	1.27	(0.95)	(1.00)	0.90	(0.40)	(0.69)	0.74	(0.31)	(0.11)	0.68	(0.25)	(0.02)
Equal-weights	1.27	(0.95)	(0.98)	0.89	(0.25)	(0.54)	0.72	(0.49)	(0.29)	0.67	(0.44)	(0.61)

Notes: All results refer to simulated out-of-sample forecasts for different forecasting horizons (3, 6, 12 and 18 months ahead) assuming mean squared loss. The different model specifications can be found in Table 3.9. Columns 2, 5, 8 and 11 report RMSFE for annualized inflation rates. In brackets the corresponding p-values of equal predictive ability are shown. I and II are the unconditional and conditional test of predictive ability, respectively. Reference category is model NORM. Out-of-sample period: 2000.5+h to 2008.7.

(NORM, which assumes normality) performs significantly better than the regression method and the balance statistic (at least for mean squared loss using the unconditional test). It is also interesting that the balance statistic always works better than the regression approach (despite the fact that the latter method is less restrictive).

Turning to the shorter horizon (three months), neither ZEW expectation measure turns out to be significant in the in-sample period so neither enters into the equation (the same is true for the term spread). For short horizons, the simple univariate model (AR) clearly stands out from all other models. At the 12 month horizon, survey expectations (from the ZEW) do even better (in relative terms) compared to the six-month horizon. They display the lowest RMSFE of all models and significantly outperform other model types like Phillips curve models (PC1-PC3) and the AR model (unconditional test at the 10% level). For MAFE, only the model with inflation rates is marginally better. Again, the distributional assumption of the C-P method has a neglectable effect on the forecasting results. At an even longer horizon (18 months) the forecastability of inflation seems to be rather limited by simple indicator models. Although inflation expectations still do

Table 3.5: Forecasting performance of alternative models (mean absolute loss)

	Step=3			Step=6			Step=12			Step=18		
	MAFE	I	II	MAFE	I	II	MAFE	I	II	MAFE	I	II
Univariate Models												
AR	1.02	–	–	0.87	(0.39)	(0.82)	0.79	(0.16)	(0.37)	0.74	(0.07)	(0.02)
ARMA	1.04	(0.51)	(0.48)	0.91	(0.30)	(0.80)	0.76	(0.94)	(0.09)	0.72	(0.90)	(0.20)
UT	1.05	(0.08)	(0.21)	0.89	(0.42)	(0.64)	0.79	(0.50)	(0.69)	0.74	(0.63)	(0.47)
Single Indicator Models												
PC1	1.04	(0.20)	(0.20)	0.91	(0.06)	(0.16)	0.85	(0.02)	(0.01)	0.83	(0.01)	(0.00)
PC2	1.04	(0.12)	(0.30)	0.89	(0.24)	(0.72)	0.84	(0.01)	(0.01)	0.79	(0.00)	(0.00)
PC3	1.02	–	–	0.87	(0.77)	(0.76)	0.84	(0.05)	(0.27)	0.79	(0.05)	(0.00)
TS1	1.01	(0.68)	(0.70)	0.86	(0.93)	(0.28)	0.75	(0.68)	(0.70)	0.80	(0.05)	(0.14)
TS2	1.02	–	–	0.82	(0.01)	(0.09)	0.78	(0.62)	(0.38)	0.74	(0.07)	(0.02)
CBAL	1.04	(0.52)	(0.56)	0.89	(0.19)	(0.14)	0.82	(0.13)	(0.15)	0.79	(0.02)	(0.00)
NORM	**1.02**	**–**	**–**	**0.86**	**–**	**–**	**0.76**	**–**	**–**	**0.71**	**–**	**–**
T	1.02	–	–	0.86	(0.16)	(0.24)	0.76	(0.87)	(0.32)	0.71	(0.88)	(0.69)
LOG	1.02	–	–	0.86	(0.18)	(0.23)	0.76	(0.94)	(0.34)	0.71	(0.80)	(0.83)
PES	1.02	–	–	0.87	(0.06)	(0.43)	0.78	(0.03)	(0.23)	0.72	(0.29)	(0.07)
BAL	1.02	–	–	0.87	(0.15)	(0.55)	0.77	(0.07)	(0.41)	0.72	(0.38)	(0.09)
Indicator models with 2 Indicators												
TS1 - NORM	1.01	(0.68)	(0.70)	0.85	(0.73)	(0.36)	0.75	(0.63)	(0.78)	0.81	(0.06)	(0.08)
TS2 - NORM	1.02	–	–	0.83	(0.08)	(0.27)	0.76	–	–	0.71	–	–
Model averaging schemes												
MSFE-weights	1.01	(0.51)	(0.60)	0.84	(0.14)	(0.35)	0.76	(0.84)	(0.59)	0.74	(0.08)	(0.05)
AIC-weights	1.02	(0.67)	(0.91)	0.85	(0.34)	(0.54)	0.77	(0.36)	(0.05)	0.73	(0.07)	(0.00)
Equal-weights	1.01	(0.90)	(0.97)	0.85	(0.22)	(0.52)	0.77	(0.55)	(0.36)	0.73	(0.06)	(0.05)

Notes: All results refer to simulated out-of-sample forecasts for different forecasting horizons (3, 6, 12 and 18 months ahead) assuming mean absolute loss. The different model specifications can be found in Table 3.9. Columns 2, 5, 8 and 11 report MAFE for annualized inflation rates. In brackets the corresponding p-values of equal predictive ability are shown. I and II are the unconditional and conditional test of predictive ability, respectively. Reference category is model NORM. Out-of-sample period: 2000.5+h to 2008.7.

significantly better than those of other indicator models, the ARMA model and the random walk (RW) are the best models for this horizon. Generally, we find that consumer expectations (CBAL) display both inferior forecasting results and larger forecast errors than do simple benchmarks (note that we consider only the balance statistic and the fact that results may be different for other conversion methods).

Turning to the issue of encompassing, Table 3.6 presents the results for the different horizons. Since expectation measures seem to be unimportant in the short run, we can turn directly to longer horizons. For the semi-annual step, term spreads (TS2) clearly contain information not included in inflation expectations. This is not surprising, as spreads also perform better in terms of average forecasting performance. For the 12-month horizon interest rates (TS1) can add information to expectation measures. For longer horizons the random walk model contains useful information. The encompassing tests also indicate that Phillips curve models and the AR model do not provide any additional information. These findings suggest that economic experts who participate in the ZEW survey use methods similar to those of autoregressive models and Phillips curve models to form their inflation expectations. But they do not fully take into account information from interest rates or term spreads.

Table 3.6: Encompassing tests

Reference Model: NORM	Step=3	Step=6	Step=12	Step=18
AR	–	0.79	0.80	0.54
ARMA	0.40	0.65	0.31	0.07
UT	0.54	0.25	0.05	0.02
PC1	0.65	0.57	0.71	0.64
PC2	0.72	0.51	0.92	0.93
PC3	–	0.36	0.87	0.67
TS1	0.24	0.13	0.05	0.28
TS2	–	0.00	0.37	0.54
CBAL	0.64	0.74	0.84	0.63
NORM	–	–	–	–
T	–	0.20	0.86	0.83
LOG	–	0.21	0.92	0.83
PES	–	1.00	0.97	0.98
BAL	–	1.00	0.94	0.96
TS1 - NORM	0.24	0.07	0.08	0.45
TS2 - NORM	–	0.00	–	–
MSFE-weights	0.17	0.00	0.28	0.28
AIC-weights	0.40	0.09	0.59	0.45
Equal-weights	0.40	0.03	0.40	0.28

Notes: P-values are reported. Out-of-sample period: 2000.11 to 2008.7.

Augmenting the information set

Because encompassing tests suggest that forecasts can be improved by incorporating additional information into single indicator models, we briefly discuss two possible ways of doing so. First, we combine two different indicators and estimate eq. (3.8) by including one additional explanatory variable with potential lags. This is done in the same way as for the single indicator model (using lags selected by AIC with potential zero restrictions on the coefficients). Since financial variables – short-term interest rate and interest rate spreads – are useful in forecasting inflation, we combine each of these measures with the ZEW survey measure. This is done by using the C-P approach assuming normality.

Estimating a composite model may often be subject to overfitting due to estimation uncertainty. In practice, it is difficult to pinpoint the "correct" model when there are so many alternative specifications. For this reason, we consider, as an alternative, the pooling of individual forecasts. There is now growing consensus that combining individual forecasts generally leads to forecast improvements relative to single forecasts (Bates and Granger, 1969; Clemen, 1989; Timmermann, 2006, see e.g.). Pooling provides a way of combining information from many different sources. We consider three simple averaging methods. The first and simplest is to use the mean forecast which is to apply equal weights to all forecasts. Second, we use an in-sample measure for constructing forecast weights. The AIC is thus used to construct the forecast weights. Kapetanios, Labhard, and Price (2008) apply this method successfully to UK inflation forecasts.[20] Finally, we choose a variant of Bates and Granger's (1969) proposed weighting scheme according to past forecast errors (denoted as MSFE-weighting). This scheme gives more weight to those forecasts that have been performed well in the recent past. In contrast to AIC-weights, this method is computed with the out-of-sample performance of each model. For this reason, only past forecast errors are used to construct a mean square forecast error as the discounted sum of squared errors, a monthly discount factor of 0.6, is used (which gives relatively high weight to most recent forecast errors). The weight received by each individual forecast is inversely proportional to its discounted mean square forecast error (see e.g. Stock and Watson, 2004, for details). This implies that indicator models that have recently been performing best receive the greatest weight.

Indicator models using two different candidate indicators may improve the performance of the expectation measure (e.g. for six months). However, it does not lead to lower forecast errors for both single indicator models. This implies that including additional regressors into single indicator models is not a good alternative once forecasting accuracy is the main objective. The pooling of single indicator models displays a robust performance for each horizon, although this does not dominate all single indicator models. It is also evident that the three

[20]Individual weights can be calculated according to $w(m) = \frac{\exp(-1/2\text{AIC}(m))}{\sum_{j=1}^{M} \exp(-1/2\text{AIC}(j))}$, where $w(m)$ is the weight of model m and M is the number of all models under consideration. Since the AIC-weights are computed for each forecast step, we allow $w(m)$ to be time varying.

forecast combination schemes are very similar when it comes to their forecasting characteristics. At the six-month ahead horizon, inflation expectations do not encompass forecast combination schemes. However, at the 12- and 18-months ahead horizon they do. Generally, we can confirm that the inflation expectations of the ZEW do not encompass all the other forecasts, suggesting that professionals could principally do better. Financial variables, in particular interest rates and interest rate spreads are natural candidates for further improving forecast accuracy.

3.5 Conclusion

This study shows how a monthly indicator for German inflation expectations with a fixed horizon may be obtained. Several conversion methods for the quantification of qualitative survey data are presented and compared with each other. The popular Carlson-Parkin method (with the statistical unbiasedness assumption) leads reasonable results and to some extent supersedes other conversion methods. However, the regression approach and even the simple balance statistic show quite similar characteristics.

Basic properties of inflation expectations are then tested by applying rationality and Granger causality tests. While there is some evidence that the concept of rationality can be empirically rejected, the results depend on the exact conversion method and, so when it comes to the orthogonality condition the expectations can be seen as nearly rational. Granger causality tests indicate that the indicators may be useful for forecasting inflation. An out-of-sample experiment is conducted to compare forecasts based on survey measures with other standard inflation models. Forecasts based on raw survey expectations perform poorly in comparison with other models, but once these survey measures are used as regressor in a dynamic specification this improves predictive ability compared to other standard inflation models. In particular, at longer horizons (12 months), these indicators beat nearly all other specifications in terms of RMSFE and MAFE. Statistically, we can differentiate among only some of the models. Further tests of forecasting encompassing reveal that survey measures already contain the information in most of the models, e.g. Phillips curve specifications. But there is some indication that survey data do not include full information on financial variables. This suggests that the forecasting performance could be further improved by considering the information contained in financial variables, such as interest rates and term spreads.

However, the validity of this study is limited in certain respects. First, the analysis is based only CPI inflation. This means that we do not provide evidence as to whether our survey indicator reveals information about other inflation measures, such as the GDP deflator or the deflator for private consumption (since these indicators are not available at quarterly frequency). Nevertheless, other choices are available at monthly frequency such as the HICP definition or some subindicators like "core inflation". Second, we restrict the analysis to simple single equation models and do not consider factor models that constitute an alternative to model combination in dealing with a great deal of information. Further, multivariate models, such as VARs, are also excluded because it is often found that they provide no improvement in comparison with univariate time series models. Fourth, our results apply only to the specific period following reunification and the out-of-sample period 2000-2008. This means that there is no guarantee of the future stability of these models.

In spite of all limitations, this study indicates that quantitative inflation expectations might be a useful indicator for future inflation. Whether the results hold for other surveys as well (e.g. Forecasts from Consensus Economics) must be the subject of future research.

References

ANDERSON, O. (1952): "The business test of the IFO-Institute for economic research, Munich, and its theoretical model," *Revue de l'Institut International de Statistique*, 20(1), 1–17.

ANG, A., G. BEKAERT, AND M. WEI (2007): "Do macro variables, asset markets, or surveys forecast inflation better?," *Journal of Monetary Economics*, 54(4), 1163–1212.

ATKESON, A., AND L. E. OHANIAN (2001): "Are Phillips Curves Useful for Forecasting Inflation?," *Ferderal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

BATCHELOR, R. A. (1982): "Expectations, Output and Inflation: The European Experience," *European Economic Review*, 17(1), 1–25.

BATES, J. M., AND C. W. J. GRANGER (1969): "The Combination of Forecasts," *Operations Research Quarterly*, 20(4), 451–468.

BERK, J. M. (1999): "Measuring Inflation Expectations: A Survey Data Approach," *Applied Economics*, 31(11), 1467–1480.

BREITUNG, J. (2008): "Assessing the Rationality of Survey Expectations: The Probability Approach," *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 228(5+6), 630–643.

CARLSON, J. A., AND M. J. PARKIN (1975): "Inflation expectations," *Economica*, 42(166), 123–138.

CARROLL, C. D. (2003): "Macroeconomic Expectations Of Households And Professional Forecasters," *The Quarterly Journal of Economics*, 118(1), 269–298.

CLEMEN, R. T. (1989): "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5(4), 559–583.

DASGUPTA, S., AND K. LAHIRI (1992): "A Comerative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data," *Journal of Business & Economic Statistics*, 10(10), 391–400.

Diebold, F. X., and R. S. Mariano (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–263.

Doepke, J., J. Dovern, U. Fritsche, and J. Slacalek (2008): “The Dynamics of European Inflation Expectations,” *The B.E. Journal of Macroeconomics*, 8(1), Article 12, (Topics).

Fama, E. F. (1975): “Short-term interest rates as predictors of inflation,” *American Economic Review*, 65, 269–282.

Fisher, J. D. M., C. T. Liu, and R. Zhou (2002): “When Can We Forecast Inflation?,” *Ferderal Reserve Bank Chicago Economic Perspectives*, (Q1), 30–42.

Forsells, M., and G. Kenny (2004): “Survey Expectations, Rationality and the Dynamics of Euro Area Inflation,” *Journal of Business Cycle Measurement and Analysis*, 1, 13–41.

Franz, W. (2005): “Will the (German) NAIRU Please Stand Up?,” *German Economic Review*, 6, 131–153.

Galí, J., and M. Gertler (1999): “Inflation dynamics: A structural econometric analysis,” *Journal of Monetary Economics*, 44(2), 195–222.

Giacomini, R., and H. White (2006): “Tests of conditional predictive ability,” *Econometrica*, 74(6), 1545–1578.

Hamilton, J. D. (1985): “Uncovering financial market expectations of inflation,” *Journal of Political Economy*, 93, 1224–1241.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998): “Tests for Forecast Encompassing,” *journal of business & Economic Statistics*, 16, 254–259.

Heinemann, F., and K. Ullrich (2006): “The Impact of EMU on Inflation Expectations,” *Open Economies Review*, 17(2), 175–195.

Inoue, A., L. Kilian, and F. B. Kiraz (2009): “Do Actions Speak Louder Than Words? Household Expectations of Inflation Based on Micro Consumption Data,” *Journal of Money, Credit and Banking*, 41(7), 1331–1363.

Kapetanios, G., V. Labhard, and S. Price (2008): "Forecasting Using Bayesian and Information-Theoretic Model Averaging: An Application to U.K. Inflation," *Journal of Business & Economic Statistics*, 26(1), 33–41.

Kozicki, S. (1997): "Predicting real growth and inflation with the yield spread," *Federal Bank of Kansas City Economic Review*, 82, 39–57.

Mestre, R. (2007): "Are Survey-Based Inflation Expectations in the Euro Area Informative?," Working Paper 163, European Central Bank.

Mishkin, F. S. (1990): "What does the term structure tell us about future inflation," *Journal of Monetary Economics*, 25(1), 77–95.

Pesaran, H. M. (1984): "Expectations Formations and Macroeconomic Modelling," in *Contemporary Macroeconomic Modelling*, Oxford: Basil Blackwell. Pierre Malgrange and Pierre-Alain Muet.

Pesaran, M. H. (1987): *The Limits to Rational Expectations*. Basil Blackwell.

Scheufele, R. (2010): "Evaluating the German (New Keynesian) Phillips curve," *North American Journal of Economics and Finance*, 21(2), 145–164.

Stock, J. H., and M. W. Watson (1999): "Forecasting inflation," *Journal of Monetary Economics*, 44(2), 293–335.

——— (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 47(3), 788–829.

——— (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23(6), 405–430.

Theil, H. (1952): "On the time shape of economic microvariables and the Munich business test," *Revue de l'Institut International de Statistique*, 20(2), 105–120.

Timmermann, A. (2006): "Forecast Combinations," in *Handbook of Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, vol. 1, chap. 4, pp. 135–196. Elsevier.

West, K. D. (1996): "Inference about Predictive Ability," *Econometrica*, 64(5), 1067–1084.

——— (2006): "Forecast Evaluation," in *Handbook of Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, vol. 1, chap. 3, pp. 99–134. Elsevier, Amsterdam.

Appendix

Descriptive statistics and unit root tests

Table 3.7: Summary statistics of actual and estimated inflation rates and inflation changes

	π^{12}				$\Delta_6\pi^{12}$			
	Mean	Sd	Max	Min	Mean	Sd	Max	Min
Actual Inflation	1.94	1.11	6.41	0.11	-0.07	0.73	2.56	-3.08
C-P (statistical unbiasedness)								
Normal	1.95	1.00	5.75	0.26	-0.01	0.34	0.59	-1.09
Scaled-t	1.95	1.00	5.75	0.27	0.00	0.34	0.59	-1.06
Logistic	1.95	1.00	5.76	0.28	0.00	0.34	0.59	-1.02
C-P (long-term unbiasedness)								
Normal	1.94	0.98	5.05	0.09	-0.02	0.83	1.45	-2.66
Scaled-t	1.94	0.97	5.06	0.08	-0.01	0.82	1.41	-2.56
Logistic	1.96	0.97	5.06	0.06	0.00	0.80	1.39	-2.40
Regression	1.86	0.99	5.74	0.24	-0.09	0.32	0.45	-0.85
Balance	1.98	0.99	5.75	0.30	0.03	0.41	0.83	-0.87

Data description

Most series are taken from the Deutsche Bundesbank database. The following abbreviations are used in the data description: SA = seasonally adjusted; NSA = not seasonally adjusted; HWWI = Hamburgisches WeltWirtschaftsInstitut (additional data source); ZEW = Center for European Economic Research (additional data source)

Definitions

P	consumer price index: total index (2005=100, SA)
U	unemployment rate (SA)
$Ugap$	HP(14400)-filtered unemployment rate using only past information (SA)
$Ogap$	HP(14400)-filtered industrial production using only past information (SA)
r^l	long term government bond yield: 9-10 years maturity (NSA)
r^s	money market rates reported by Frankfurt banks: Three-month funds (NSA)
P^{raw}	HWWI commodity price index for Euro area (euro basis, NSA)
A	fraction of respondes reporting "inflation goes up" (ZEW, NSA)
B	fraction of respondes reporting "inflation goes down" (ZEW, NSA)

Table 3.8: Unit root tests for inflation and estimated inflation expectations

	π^{12}		$\Delta_6\pi^{12}$	
	ADF	PP	ADF	PP
Actual Inflation	-3.598***	-3.608***	-3.517***	-6.070***
C-P (statistical unbiasedness)				
Normal	-3.631***	-3.591***	-3.391***	-2.240**
Scaled-t	-3.621***	-3.578***	-3.405***	-2.249**
Logistic	-3.604***	-3.560***	-3.429***	-2.564***
C-P (long-term unbiasedness)				
Normal	-3.523***	-3.386***	-3.391***	-2.240**
Scaled-t	-3.510***	-3.368***	-3.405***	-2.249**
Logistic	-3.488***	-3.352***	-3.429***	-2.564***
Regression	-3.644***	-3.622***	-3.048***	-2.405**
Balance	-3.611***	-3.552***	-3.135***	-2.360**

Notes: Each test in levels does include a constant term; for inflation changes the constant is omitted (since it is insignificant). ADF and PP correspond to the augmented Dickey-Fuller test and the Phillips-Perron test, respectively. The lag length of the ADF-test is choosen based on SIC. The PP test is calculated with Newey and West standard errors using a Bartlett kernel. ***: 1% and **: 5% significance level

Table 3.9: Model specifications

Abbr.	Specification (included regressors)			
	Step=3	Step=6	Step=12	Step=18
Univariate Time Series Models				
AR	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}	π^{12}_{t-4}	π^{18}_t
ARMA	π^3_t, MA(t)	π^6_t, MA(t)	π^{12}_t, MA(t)	π^{18}_t, MA(t)
RW	π^{12}_t	π^{12}_t	π^{12}_t	π^{18}_t
Phillips Curve Models				
PC1	π^3_{t-8}, U_{t-1}, U_{t-3}	π^6_t, π^6_{t-6}, U_{t-1}, U_{t-3}	π^{12}_t, π^{12}_{t-4}, U_t, U_{t-3}	π^{18}_t , U_{t-3}
PC2	π^3_{t-8}, $Ugap_{t-1}$, $Ugap_{t-3}$	π^6_{t-6}, $Ugap_{t-1}$, $Ugap_t$	π^{12}_{t-4}, $Ugap_t$, $Ugap_{t-3}$	π^{18}_t , $Ugap_t$, $Ugap_{t-3}$
PC3	π^3_{t-4}, π^3_{t-8}	π^6_{t-6}, $Ogap_{t-1}$	π^{12}_{t-4}, $Ogap_{t-2}$	π^{18}_t , $Ogap_{t-1}$
Term Structure Models				
TS1	π^3_{t-8}, r^s_t, r^s_{t-1}	π^6_t, r^s_t	π^{12}_{t-4}, $(r^l - r^s)_{t-1}$	π^{18}_t, r^s_{t-3}
TS2	π^3_{t-8}, $(r^l - r^s)_{t-4}$	π^6_{t-6}, $(r^l - r^s)_{t-1}$	π^{12}_{t-4}, $(r^l - r^s)_{t-1}$	$(r^l - r^s)_t$
Survey Expectation Models				
CBAL	π^3_{t-1}, π^3_{t-8}, E_tP_{t+12}	π^6_{t-6}, E_tP_{t+12}	π^{12}_{t-4}, E_tP_{t+12}	π^{18}_t, E_tP_{t+12}
NORM	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$
T	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$
LOG	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$
PES	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$
BAL	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$
Indicator models with 2 Indicators				
TS1 - NORM	π^3_{t-8}, r^s_t, r^s_{t-1}	π^6_t, $E_t\pi_{t+6}$, r^s_t	$E_t\pi_{t+6}$, r^s_{t-1}	π^{18}_t, $E_t\pi_{t+6}$, r^s_{t-3}
TS2 - NORM	π^3_{t-4}, π^3_{t-8}	π^6_{t-1}, π^6_{t-6}, $E_t\pi_{t+6}$, $(r^l - r^s)_t$, $(r^l - r^s)_{t-1}$	π^{12}_t, π^{12}_{t-4}, $E_t\pi_{t+6}$	π^{18}_t, $E_t\pi_{t+6}$

Erklärung

Ich versichere, dass ich die vorliegende Dissertation selbständig verfasst habe. Andere als die angegebenen Hilfsmittel und Quellen habe ich nicht benutzt. Die Arbeit hat keiner anderen Prüfungsbehörde vorgelegen.

Rolf Scheufele

Halle, den 14. Dezember 2010