



Journal of Business Cycle Measurement and Analysis

Edited by Günter Poser



CIRET



KOF

Volume 1, No. 1

© OECD, 2003.

© Software: 1987-1996, Acrobat is a trademark of ADOBE.

All rights reserved. OECD grants you the right to use one copy of this Program for your personal use only. Unauthorised reproduction, lending, hiring, transmission or distribution of any data or software is prohibited. You must treat the Program and associated materials and any elements thereof like any other copyrighted material.

All requests should be made to:

Head of Publications Service,
OECD Publications Service,
2, rue André-Pascal,
75775 Paris Cedex 16, France.

Editorial

For many years colleagues from the Organisation for Economic Co-operation and Development (OECD) and from the Centre for International Research on Economic Tendency Surveys (CIRET) have been working on business cycle measurement and analysis. Several conferences and workshops on these topics organised by the OECD and CIRET have enabled international experts to share ideas and experiences. However, following the *renaissance* of studies on business cycles, which seems very far from being “dead”, as was believed a few years ago, we felt there was a lack of international journals dealing specifically with statistical and economic research on business cycles. This Journal is intended to close this gap.

Business cycle research and analysis encompass a wide range of topics such as the definition, specification and classification of business cycles, the explication of cyclical fluctuations, the description and construction of long and short range indicators and the use of quantitative and qualitative data for microeconomic and macroeconomic analysis and forecasts. Therefore, the Journal will host both theoretical and applied articles, as well as papers on statistical and econometric methods if they are applied to business cycle problems.

In designing and managing the Journal we have involved quite a large group of worldwide experts in business cycle measurement and analysis. Their contribution in promoting the Journal and in selecting papers to be published is fundamental to the final outcome. A referee process has been established to ensure the publication of high-quality papers.

We believe the new journal will serve as a forum for authors and users all over the world who wish to communicate on the state of the art in the field of business cycle research and analysis. For this purpose, we plan to include book reviews, announcements of relevant conferences, and information on international research teams working in this field. We would therefore be glad to receive any information and feedback you think appropriate.



Prof. Günter Poser
Editor-in-Chief
editor-in-chief@ciret.org



Prof. Enrico Giovannini
Chairman of the Editorial Board
enrico.giovannini@oecd.org

Editors

Editor-in-Chief

Prof. Günter Poser, Technische Universität Darmstadt, Germany

Associate Editors

Richard Curtin, University of Michigan,
Ann Arbor, MI, USA

Marco Malgarini, ISAE, Institute for Studies
and Economic Analyses, Rome, Italy

Gernot Nerb, Ifo Institute for Economic
Research, Munich, Germany

Ronny Nilsson, OECD, Paris, France

Giuseppe Parigi, Bank of Italy, Rome, Italy

Gerd Ronning, University of Tübingen, Germany

Bernd Schips, KOF, Swiss Institute for Business
Cycle Research, Zürich, Switzerland

Editorial Board

Chairman

Enrico Giovannini, Chief Statistician, OECD, Paris, France

Ignazio Angeloni, European Central Bank,
Frankfurt, Germany

Daniel Bloesch, KOF, Swiss Institute for
Business Cycle Research, Zurich, Switzerland

Robert A. Buckle, Victoria University
of Wellington, New Zealand

John A. Carlson, Purdue University,
West Lafayette, IN, USA

Jorgen Elmeskov, OECD, Paris, France

Horst Entorf, Technische Universität Darmstadt,
Germany

Hélène Erkel-Rousse, INSEE, National Institute
for Statistics and Economic Studies, Paris, France

Willy Friedmann, Deutsche Bundesbank,
Frankfurt, Germany

Robert J. Gordon, Northwestern University,
Evanston, IL, USA

Chung-Ying Hu, CEPD, Council for Economic
Planning and Development, Chinese Taipei

Jürgen Kröger, European Commission,
Brussels, Belgium

Kajal Lahiri, State University of New York
at Albany, NY, USA

Marco Lippi, University of Rome "La Sapienza",
Italy

Augustin Maravall, Bank of Spain, Madrid,
Spain

Zbigniew Matkowski, Warsaw School
of Economics, Warszawa, Poland

Gian Luigi Mazzi, Eurostat, Luxembourg

Robert H. McGuckin, The Conference Board,
New York, NY, USA

Lars-Erik Öller, Statistics Sweden and
Stockholm University, Stockholm, Sweden

Karl Heinrich-Oppenländer, Former president
of Ifo Institute, Munich, Germany

Sterrenberg Pretorius, South African Reserve
Bank, Pretoria, South Africa

James H. Stock, Harvard University,
Cambridge, MA, USA

Jean-Jacques Vanhaelen, National Bank
of Belgium, Brussels, Belgium

Pentti Vartia, ETLA, The Research Institute
of the Finnish Economy, Helsinki, Finland

András Vértés, GKI Economic Research Co.,
Budapest, Hungary

Martin Weale, NIESR, National Institute
of Economic and Social Research, London, UK

Contents

Introduction	
An Important Subject in Need of Much New Research	
<i>Victor Zarnowitz</i>	7
Survey Expectations, Rationality and the Dynamics of Euro Area Inflation	
<i>Magnus Forsells and Geoff Kenny</i>	13
Modelling Short-term Interest Rates in the Euro Area Using Business Survey Data	
<i>Renata Grzeda Latocha and Gernot Nerb</i>	43
Quantification of Qualitative Survey Data and Test of Consistent Expectations: A New Likelihood Approach	
<i>Christian M. Dahl and Lin Xia</i>	71
Detection of Turning Points in Business Cycles	
<i>Eva Andersson, David Bock and Marianne Frisé</i>	93
Coincident and Leading Indicators of Manufacturing Industry Sales, Production, Orders and Inventories in Switzerland	
<i>Michael Graff and Richard Etter</i>	109
OECD and CIRET News – 2004/1	133

Introduction

An Important Subject in Need of Much New Research

Victor Zarnowitz^{*}

Abstract

Long records of diverse data reflecting surveys, indexes and other measures of production, employment, incomes and sales lend considerable credibility to the view that growth in market economies has long proceeded through recurrent, though definitely nonperiodic, sequences of expansion and contraction in overall business activity.

Business cycles vary greatly in duration and intensity, less in diffusion, timing and interaction of their many constituent processes. Far from being “all alike,” they are both diverse and evolving. Contractions tend to be more frequent and severe in deflationary times than during inflation, which helps to explain why some moderation of business cycles took place in the past half-century. Expansions lengthened, recessions shortened, depressions faded. Favorable structural and institutional developments, along with policy improvements, have probably also contributed to this result.

Nevertheless, continued moderation of economic fluctuations cannot be taken for granted. Business cycles, even if less of a threat, are far from conquered and still represent the most serious form of macroeconomic instability. Each of the recent long expansions has revived forecasts of their abolition, and in each case such hopes were dashed again by another recession.

A new journal devoted to the studies of national and international economic fluctuations is indeed most timely, needed, and welcome. It should set top standards in dealing with empirics and analytics, shorter movements and longer trends, internal dynamics and external shocks, past and contemporary problems of growth, instability, and policy. It must be both comprehensive and fair, i.e., consider a variety of interesting facts and test a variety of interesting models, avoiding both pseudo-conventional wisdom and overly strong priors.

*

Senior Fellow and Economic Counselor, The Conference Board;
Professor Emeritus, The University of Chicago;
Research Associate, National Bureau of Economic Research

Résumé

Les séries longues de données diverses provenant d'enquêtes ou reflétant l'évolution d'indices et d'autres mesures de la production, de l'emploi, des revenus et des chiffres d'affaires accréditent dans une très large mesure l'idée que la croissance des économies de marché s'opère depuis longtemps en séquences récurrentes, mais absolument pas périodiques, d'expansion et de contraction de l'activité économique globale.

Les cycles conjoncturels varient grandement du point de vue de leur durée et de leur intensité, mais moins du point de vue de leur diffusion, de leur calendrier et de l'interaction entre leurs nombreux éléments constitutifs. Loin d'être identiques, ils sont à la fois différenciés et évolutifs. Les contractions ont tendance à être plus fréquentes et plus prononcées en période de déflation qu'en période d'inflation, ce qui contribue à expliquer pourquoi on a assisté au cours des cinquante dernières années à une certaine modération des cycles conjoncturels. Les périodes d'expansion ont duré plus longtemps, les récessions ont été plus brèves et les dépressions ont perdu de leur vigueur. Des évolutions structurelles et institutionnelles favorables, de même qu'une politique économique de meilleure qualité, ont probablement contribué elles aussi à ce résultat. On ne peut néanmoins considérer comme acquise pour l'avenir cette modération des fluctuations économiques. Les cycles conjoncturels, même s'ils font peser de moins lourdes menaces, sont loin d'être maîtrisés et représentent toujours la forme la plus grave d'instabilité macroéconomique. Lors de tous les épisodes récents d'expansion de longue durée, on a annoncé leur disparition, mais, chaque fois, une nouvelle récession est venue anéantir ces espoirs.

Une nouvelle revue consacrée à l'étude des fluctuations économiques nationales et internationales est à la fois opportune, nécessaire et souhaitée. Elle devra se montrer exigeante dans l'examen des éléments empiriques et analytiques, des mouvements courts et des tendances longues, de la dynamique interne et des chocs externes ainsi que des problèmes passés et actuels qui ont trait à la croissance, à l'instabilité et à l'action économique.

1 On the Historical Record and Variety of Business Cycles

Long records of diverse data reflecting surveys, indexes and other measures of production, employment, incomes and sales lend considerable credibility to the view that growth in market economies has long proceeded through recurrent, though definitely nonperiodic, sequences of expansion and contraction in overall business activity. The historical chronologies of these “classical” business cycles developed by scholars at the National Bureau of Economic Research (NBER) cover more than two centuries for Great Britain and United States, more than hundred years for France and Germany. According to the generally much richer and better data for the post-World War II era, principal market economies continued to show significant cyclical instabilities; moreover, similarly persistent and pervasive fluctuations appear in many less developed countries as well.

Business cycles vary greatly in duration and intensity, less in diffusion, timing and interaction of their many constituent processes. Far from being “all alike,” they differ in many respects. Some downturns originate in external shocks, perhaps the best examples being the large oil price hikes by the OPEC cartel in 1973-74 and 1979-80. Some are related to shifts in policy variables such as the rapid reductions in military expenditures in the demobilizations after World War II in 1945 and after the Korean War in 1953; also the sharp rises in nominal and real interest rates on a few occasions (particularly in 1980 and 1981-2 recessions, to counter high inflation). The occurrence of such major disturbances with contractive effects confirmed to many economists the idea that business cycles result from the way a dynamically stable market system propagates such shocks (which are sufficiently frequent to keep the economy fluctuating).¹

There is no doubt that external disturbances (broadly defined as shifts in preferences, technology, policy variables, weather) always influence the economy, at times critically; a comprehensive explanation of business cycles cannot be purely endogenous. But neither can the heavy reliance on shocks and the mainstream dynamic general equilibrium model that developed in macroeconomics of recent years be accepted as satisfactory. There is no evidence that all business cycles are due to specific adverse shocks with proper timing and intensity. Most identifiable disturbances in recent times had short-lived effects and were not timed to initiate cyclical downturns.² And the cumulative effect of a flow of small random shocks would be fluctuations of much greater regularity than, hence quite different from, the business cycles of actual experience.

¹ The idea goes back, a.o., to E. Slutsky [1927] 1937, “The Summation of Random Causes as the Source of Cyclic Processes” (translation from Russian in *Econometrica* 5: 105-146), but its most influential exposition is R. Frisch [1933] 1965, “Propagation Problems and Impulse Problems in Dynamic Economics” in *AEA Readings in Business Cycle Theory*, 155-185, Homewood, Ill.: Irwin. See V. Zarnowitz, “Business Cycles” in D. Glasner, ed., *Business Cycles and Depressions: An Encyclopedia*, New York and London: Garland, 1997, 62-72.

² For a study of potential shocks in the United States since 1962, see Victor Zarnowitz and Jacinto Torres, Jr. “Economic Context and Consequences of 9/11, a Year Later” in *The Conference Board, Business Cycle Indicators*, vol. 7, no.10, September 2002 (Note that the 2001 terrorist attacks occurred closer to the end than to the beginning of the recession).

2 More on the Theories of Shocks vs. Imbalances

When empirically implemented, the linear dynamic equilibrium models generally perform less well in reproducing and forecasting business cycles than the much less rigorous but also less restrictive econometric and time-series models used in practice, mostly by business but also by some academic economists. The former models are based on strong, and theoretically attractive, assumptions of rational expectations, representative agents, and market clearing by relative price changes with little or no friction (despite allowance for imperfect information). But it is these very assumptions that cause the corresponding models to miss recessions; relaxing them in favor of giving more attention to the behavior of cyclical variables and other “stylized facts” would likely result in considerable analytical and forecasting improvements.

The measurement of “shocks” defined as the effects of all shifts in tastes, technology and productivity, policies, etc. is model-dependent and tenuous. It can be argued that this makes it particularly important to relate specific recessions and recoveries to specific shifts or shock events; but, as already noted, such linkages are often difficult to establish and may not exist.

Instead, long historical records provide much evidence of important common elements in what recurrently happens during business cycles. At the center here are the cumulative and mutually reinforcing movements in business profits, investment, and credit. The prime mover in expansion is the high expected profits from capital investments, particularly in areas of technological innovation. These projects are usually encouraged by the initial conditions of relatively low costs and easy credit. In longer and stronger expansions, overconfidence and overindebtedness tend to develop, leading to imbalances. Plans to invest outrun decisions to save, and there is overinvestment in some capital-goods industries; costs rise faster than prices, and profit margins and expectations fall; in reaction, inventory and fixed investment and later employment and output are cut. The booms prove unsustainable and are followed by busts, often marked by opposite imbalances or excesses (lack of confidence, reluctance to borrow and invest). Some of the recessions are aggravated by bubbles and crises in markets for financial and real assets.

The rich classical literature on business cycles developed roughly between 1890 and 1960 stressed the above and related factors such as interest rates, input and output prices, monetary and general price trends, risks and uncertainties. The bulk of it had the cycle driven mainly by internal dynamics rather than external disturbances. There was no insistence yet on rational expectations, representative agents, and equilibrium modeling; expectations were mostly assumed to be adaptive, agents heterogeneous, and the economy temporarily in disequilibrium or between multiple equilibria. Limitations of “human nature,” i.e., self-interested rational behavior, were recognized, not only in form of errors or misperceptions by private agents but even as occasional mass behavior (manias, panics) in normally efficient markets. Nonlinearities and lagged adjustments appeared prominently in some of the more formal and abstract models; others stressed the wage and price

dynamics, the effects of unexpected price level movements, the interplay of consumption and investment (multiplier-accelerator).³

Only in the last forty-odd years did the balance shift gradually from endogenous to exogenous factors and models. In the monetarist theory, the instability of growth in money supply, due to the failings of monetary authorities, causes the fluctuations in incomes and output. In the new equilibrium models, first limitations of information caused misperceptions of absolute for relative changes and then shifts in technology and productivity caused “real business cycles.”⁴

3 Are Business Cycles Moderating?

Business cycles are not only diverse but also evolving. Contractions tend to be more frequent and severe in deflationary times than during inflation, which helps to explain why some moderation of business cycles took place in the past half-century. Expansions lengthened, recessions shortened, depressions faded. Favorable structural and institutional developments, along with policy improvements, have probably also contributed, including growth of the less cyclical service industries; the insurance of bank deposits; the rise of government taxes and transfer payments acting as fiscal stabilizers; more effective monetary policy; better inventory management; and perhaps also strengthened public confidence that the worst debacles will not occur.

Nevertheless, continued moderation of economic fluctuations cannot be taken for granted. Business cycles, even if less of a threat, are far from conquered and still represent the most serious form of macroeconomic instability. Forecasts of their abolition, widely revived in each of the recent long expansions, have in each case been falsified again by another recession. New trends create new problems, e.g., globalization raises fears of greater international synchronization of business cycles.

All this is highlighted by the surprises and disappointments of the recent past. The increased burden of heavy nominal debt in deflationary times was last experienced by the United States in the deep depression of the 1930s, and the fears that it might return here appear to be premature (despite having been briefly rekindled in mid-2003 by none other than the Fed itself). However, the risk of deflation is being vividly recalled again lately in connection with the long sluggishness of both output and prices in Japan of the 1990s and early 2000s. So the theories of debt-deflation, liquidity trap, and financial fragility that many believed have long lost their current relevance may not be so obsolete after all.⁵ Moreover, the old theories of overinvestment, overconfidence, and overindebtedness, long in a limbo,

³ The eminent economists whose work is considered in the preceding paragraph are too numerous for detailed references in this brief article, but let me list the names of Wicksell, Mitchell, Robertson, Pigou, Irving Fisher, Kalecki, Keynes, Hicks, Hayek, Schumpeter, Samuelson, Kaldor, and Goodwin (more or less in chronological order by publication in the area). For a critical review, see Haberler (1962); for an updated discussion, see Zarnowitz (1992, chap. 2; 1996).

⁴ Here the most prominent names are those of Friedman and Schwartz, Lucas, Kydland and Prescott, Long and Plosser. For reviews, see Glasner, *op. cit.* 557-560, and Zarnowitz (1992, ch.2).

⁵ The relevant writings include Irving Fisher 1933, Keynes 1936, Misky 1977, and Kindleberger 1978. For further discussion and references, see Zarnowitz 1992.

apply well with reference to the U.S. boom-and-bubble burst of 1995-2000 and bust-and-slack of 2001-3.⁶ It is simply undeniable that, as so often in the past, this last U.S. cycle has been triggered by an outburst of spending on newly favored types of capital goods, with correspondingly high and unsustainable rises in credit, expected profits, and stock prices.

Taking this old viewpoint of an essentially endogenous theory, I have seen the bubbles and the recession coming at the time when the U.S. expansion was at its height.⁷ In contrast, any equilibrium theory relying on the advent of some adverse outside event would in its absence fail to anticipate the end of the boom. As already noted, while the earlier students of the subject focused on the self-sustaining elements of business cycles, more recent theorists shifted their emphasis from the internal dynamics to external shocks. But the dismissal of the mainly endogenous theories and adoption of the mainly exogenous ones derived much more from the conventional assumption of the economy's inherent stability than from persuasive evidence and modeling.

The strong boom in real and financial investment in the second half of the 1990s created considerable excess capacity and excess debt, but the short slowdown in 2000 and the mild recession of 2001 did little to heal these imbalances. The recession of average duration was followed by a protracted and disappointing recovery in 2002 and 2003 (through August when this is written). The recoveries are usually phases of high growth, but this one is sluggish in output and absent in employment which simply fails to grow. The overcapacities and overindebtedness – legacies of the boom – are still hurting investment and the labor market. And there is a good deal of weakness in economic activity and prices abroad as well, with on-and-off debt and currency crises and recessions in Asia, Europe, and Latin America since mid-1990s.

4 Conclusion

I believe that the above arguments make it very clear that business cycles are (1) neither dead nor dying, (2) still complex, changing, and misinterpreted, and (3) deserving much further study, better understanding, and more effective counterpolicies. Hence a new journal devoted to the studies of national and international economic fluctuations is indeed most timely, needed, and welcome. It should set top standards in dealing with empirics and analytics, shorter movements and longer trends, internal dynamics and external shocks, past and contemporary problems of growth, instability, and policy. It must be both comprehensive and fair, i.e., consider a variety of interesting facts and test a variety of interesting models, avoiding both pseudo-conventional wisdom and overly strong priors.

⁶ Consider Wicksell [1898] 1936, Robertson [1915] 1948, Mitchell 1913, Pigou 1927, Hayek 1939, Schumpeter 1939, and others. For a critical review and synthesis, see Haberler 1962 and Zarnowitz 1992.

⁷ See V. Zarnowitz, "The Theory and History Behind Business Cycles: Are the 1990s the Onset of a Golden Age?" *Journal of Economic Perspectives*, vol. 13, no. 2, Spring 1999, pp.69-90.

Survey Expectations, Rationality and the Dynamics of Euro Area Inflation

Magnus Forsells and Geoff Kenny*

Abstract

This paper uses survey data in order to analyse and assess the empirical properties of consumers' inflation expectations in the euro area and explores their role in explaining the observed dynamics of inflation. The probability approach is used to derive quantitative estimates of euro area inflation expectations from the qualitative data from the European Commission's Consumer Survey. The paper subsequently analyses the empirical properties of the estimated inflation expectations by considering the extent to which they fulfil some of the necessary conditions for rationality. The results suggest an intermediate form of rationality. In particular, the surveyed expectations are an unbiased predictor of future price developments and they incorporate – though not always completely – the information contained in a broad set of macroeconomic variables. In addition, although persistent deviations between consumers' expectations and the rational outcome have occurred, consumers are shown to rationally adjust their expectations in order to eventually “weed out” any systematic expectational errors. Interestingly, perhaps reflecting changes in the monetary policy regime, there is also evidence of “growing” rationality in the sense that expectations have become more rational over the 1990s compared with the 1980s. Moreover, estimates of a hybrid Phillips curve – which nests both backward and forward-looking inflation dynamics – suggests that consumer expectations played a role in determining the actual dynamics of inflation over the period 1985-2001.

Key Words: Inflation Expectations; Surveys; Rationality; Hybrid Phillips Curves; Euro Area

JEL Classification: D12; D84; E31

*

European Central Bank, Frankfurt am Main, Germany

The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank. This paper has benefited from comments from two anonymous referees and discussions at seminars in the European Central Bank and at the 26th CIRET Conference, Taipei, October 2002.

Résumé

A partir de données d'enquête, on analyse et on évalue les propriétés empiriques des anticipations d'inflation des consommateurs dans la zone euro et on examine le rôle de ces anticipations dans la dynamique observée de l'inflation. L'approche par les probabilités permet d'obtenir des estimations quantitatives des anticipations d'inflation dans la zone euro à partir des données qualitatives de l'enquête de la Commission européenne auprès des consommateurs. On analyse ensuite dans cet article les propriétés empiriques des anticipations inflationnistes estimées en se demandant dans quelle mesure elles remplissent certaines des conditions nécessaires pour pouvoir être considérées comme rationnelles. Les résultats aboutissent à une forme intermédiaire de rationalité. En particulier, les anticipations qui ressortent des enquêtes constituent un prédicteur non biaisé de l'évolution future des prix et elles intègrent – mais pas toujours totalement – les informations contenues dans un large ensemble de variables macroéconomiques. En outre, malgré la présence d'écarts persistants entre les anticipations des consommateurs et le résultat rationnel, il apparaît que les consommateurs ajustent rationnellement leurs anticipations pour éliminer toute erreur systématique d'anticipation. On note également avec intérêt – ce qui est peut-être la conséquence de la modification du régime de politique monétaire – une rationalité de plus en plus grande en ce sens que les anticipations sont devenues plus rationnelles au cours des années 1990 que durant la décennie 1980. De plus, les estimations d'une courbe de Phillips hybride – qui intègre une dynamique de l'inflation à la fois rétrospective et prospective – montrent que les anticipations des consommateurs ont joué un rôle dans la dynamique effective de l'inflation entre 1985 et 2001.

1 Introduction

Central banks have long had an interest in monitoring the behaviour of inflation expectations in the economy and in understanding the nature of the process by which expectations are formed. Two simple examples serve to highlight the crucial importance of access to reliable measures of expected inflation: First, to the extent that they provide a useful or unbiased predictor of future inflation, measures of expected inflation may represent an important information variable in a forward-looking analysis of price developments. Second, higher expectations of inflation may lead employees to demand higher wage settlements, giving rise to cost-push effects on inflation. Moreover in a situation where overall inflation is expected to rise firms may be more willing to pay higher wages because they believe that they can more easily pass on any change in costs in the form of higher selling prices.¹

More fundamentally, however, macroeconomic theory accords a pivotal role to expectations of inflation in understanding the functioning of the economy. From the perspective of aggregate price formation, the Phelps-Friedman expectations augmented Phillips curve posits a relationship whereby the actual change in the price level is driven – in part – by aggregate expectations of its future change. Conversely, from the perspective of business cycle analysis, natural rate theory predicts that the scale of any real stimulus from inflation will crucially depend on the extent to which such inflation is anticipated. Under fully rational expectation, with no systematic errors in forecasting inflation, only unexpected changes in inflation have an impact on real variables such as output and unemployment. More recently, the importance of inflation expectations for aggregate price dynamics has been given rigorous microeconomic foundations by considering the optimal price-setting behaviour of firms in a dynamic framework under imperfect competition and some constraint on the frequency of price adjustment (see, for example, Roberts (1995), Sbordone (2000a) and (2000b), Galí and Gertler (1999) and Galí, Gertler and López-Salido (2001)).² A defining feature of this New Keynesian theory of price dynamics is that inflation is a forward-looking

¹ Christiano and Gust (2000) term this the cost-push version of the *expectations trap hypothesis* because it poses a dilemma to the central bank which must choose between producing the inflation that firms want or risk putting the economy into a recession. There are numerous other examples of the relevance of inflation expectations from a central bank's perspective. Most notably, inflation expectations can be viewed as a determinant of money demand and they are likely to play an important role in the monetary transmission mechanism (e.g. by effecting real interest rates). In the recent literature on monetary policy rules, inflation expectations have also been posited to be among the key arguments in a central banks reaction function (see Taylor, 1999). Inflation expectations, particularly over medium-term horizons, also provide information on the credibility of a central bank's commitment to achieving price stability.

² Wolman (1999) provides a recent review of new Keynesian theories of inflation.

phenomenon, driven by its own expectations.³ In a recent contribution, Taylor (2000) has argued that this “expectations theory of pricing” can help explain the observed lower rates of inflation in many industrialised countries in the 1990s compared with the 1980s and the 1970s.⁴ Similarly in a recent empirical analysis, Galí, Gertler and López-Salido (2001) find that “forward-looking behaviour is dominant in shaping the dynamics of inflation” (p. 17) both in the euro area and in the United States.

All this serves to emphasise the importance attached to measures of expected inflation in macroeconomic analysis. Not surprisingly, a large body of empirical literature exists seeking to both estimate and examine the empirical properties of inflation expectations. One track in this literature has been to focus on direct survey measures of expectations. While much of this literature dates from the 1970s and the 1980s, there has been a relative dearth of empirical research examining data from the 1990s (see, for example, Thomas (1999)).⁵ Given the significant difference in the behaviour of inflation in the 1990s compared with the preceding two decades, and the possible role of inflation expectations in explaining these developments, there is a need for more up-to-date evidence on the properties of survey measures of inflation expectations and particularly – given the adoption of a single currency by twelve member states of the European Union – of expectations in the euro area.

This paper contributes to this renewed interest in survey-based measures of inflation expectations for the specific case of the euro area.⁶ In particular, a measure of expected inflation in the euro area is derived from the European Commission’s Consumer Survey (EC survey) and its predictive performance for future inflation is examined. Following on from this, in order to shed light on the rationality of inflation expectations in the euro area, the properties of the derived series are examined and assessed along a number of dimensions. Firstly, we consider whether or not consumers have correctly anticipated the inflation outcome on average by testing for any *bias* in the survey indicator. Second, we consider the *dynamic*

³ The enthusiasm for this New Keynesian theory of price dynamics is by no means universal. For instance, Fuhrer (1997a) using US data finds that expectations of future price changes are unimportant in explaining price and inflation behaviour. Similarly, Fuhrer (1997b) argues that the fit of such New Phillips curves can be routinely improved by adding lags of inflation. As a result, the inflationary process exhibits a high degree of persistence whereby the effects from a one-time shock will last well beyond the lifetime of the initial shock. More recently, again for the US, Roberts (2001) finds that the New Keynesian model requires additional lags of inflation that suggest that some fraction of the population uses a simple univariate rule when forming their inflation expectations.

⁴ Using a simple staggered pricing model with market power, Taylor (2000) illustrates the crucial role of expectations. In particular, he shows that the extent to which a firm matches an increase in costs or an increase in another firm’s price depends positively on how persistent such changes are *expected* to be. In other words, a decline in persistence of either costs or inflation may be associated with a decline in observed market power.

⁵ Some notable recent exceptions are Berk (1999) and Berk (2000), Roberts (1998), Grant and Thomas (1998). See also Deutsche Bundesbank (2001).

⁶ In this paper, attention is focussed on the properties of an aggregate measure of expectations in the euro area and any possible heterogeneity in the properties of inflation expectations across the individual euro area countries, although potentially important, is overlooked.

adjustment properties of expectations and, in particular, provide estimates of the speed with which consumers revise their expectations to be in line with the fully rational outcome. Following on from this, in a test of the *efficiency* of consumers' inflation expectations in the euro area, the extent to which they incorporate the information contained in a broad set of macroeconomic variables is evaluated. Lastly, we also explore the possible role of surveyed expectations in explaining the observed behaviour of actual inflation in the euro area over the period 1985-2001. This is carried out by estimating various Phillips curves, which nest both backward and forward-looking price dynamics.

To anticipate our findings, our results suggest that consumer expectations from the EC survey provide important information on actual future developments in euro area inflation. In particular they provide an unbiased predictor of inflation one-year ahead and, most strikingly, correctly anticipated the trend decline in inflation over the 1990s.⁷ In line with this, our analysis of the adjustment of expectations suggests that consumers revise their expectations in the light of new information and that, while both actual and expected inflation may drift apart in the short-run, they ultimately revert toward one another. Furthermore, there is also some evidence to suggest that consumers' expectations are not fully rational with respect to all the available information. In particular, while consumers appear to take full account of past developments in inflation, the empirical evidence suggests that a wider set of macroeconomic information (most notable monetary and financial variables) can explain consumers' errors in forecasting inflation. In this sense, expectations are weak but not strong-form efficient. Nonetheless, there is evidence of "growing" rationality over the sample period: when data from the late 1980s is excluded from the sample, the evidence against strong-form efficiency is much weaker.

Lastly, also consistent with hypothesis of rational inflation expectations, we find a role for consumer expectations in explaining the actual dynamics of inflation in the euro area using a hybrid version of the Phillips curve. In particular, as suggested by the recent results of Galí, Gertler and López-Salido (2001), there is a quantitatively significant forward-looking component to the inflationary process in the euro area. The backward-looking component is also significant but less so. Moreover, this finding is robust with respect to various empirical specifications of the Phillips curve.

2 Estimating Euro Area Inflation Expectations from Survey Data

Economic surveys provide a direct source of information on expected inflation.⁸ For example, the EC Survey asks nearly 20,000 consumers in the euro area for information on

⁷ This is striking because purely backward looking or adaptive expectations would tend to systematically over-predict inflation when it is declining.

⁸ The other main source is to derive inflation expectations from the prices of traded financial assets. A common argument favouring such measures is that they are likely to reflect the true beliefs of economic agents whereas respondents to surveys have little incentive to report their true expectations or the expectations upon which they base economic decisions.

their expectations for consumer price trends over the next twelve months. The survey is conducted at a national level, and the results for the euro area are compiled by aggregating the country data using weights based on each country's share in total euro area private final consumption expenditure (at constant prices). Participants in the survey are asked the following question, which is harmonised across all countries:⁹

"By comparison with what is happening now, do you think that in the next 12 months...

1. ...there will be a more rapid increase in prices (S_1),
2. ...prices will increase at the same rate (S_2),
3. ...prices will increase at a slower rate (S_3),
4. ...prices will stay about the same (S_4), or
5. ...prices will fall slightly (S_5)."

While the results of this survey are usually summarised in the form of a "balance statistic", computed as a difference among the proportion of respondents opting for the different response categories, this provides only qualitative information on the likely direction of change in inflation in the next 12 months.¹⁰ However, recent literature (e.g. Berk (1999) and Berk (2000)), building on the earlier contributions of Carlson and Parkin (1975), Batchelor (1981, 1982 and 1986a) and Batchelor and Orr (1988) and others, has further developed the probabilistic approach to obtaining a quantitative estimate of the expected rate of inflation from qualitative surveys such as those carried out by the European Commission. An important feature of these recent advances in methodology is that, unlike earlier approaches, unbiasedness is not imposed in deriving a measure of expected inflation.

2.1 The probability approach to estimating inflation expectations

The central idea behind the probability approach is to interpret the share of respondents replying to each category as maximum likelihood estimates of areas under the density function of aggregate inflation expectations, i.e. as probabilities (see Figure 1). The distribution is segmented by various response thresholds. For example, as in the original Carlson and Parkin (hereafter CP) methodology, respondents expect prices to rise (fall) if the expected rate of inflation for month t (π_t^e) but surveyed in month $t-12$ (for monthly data) is

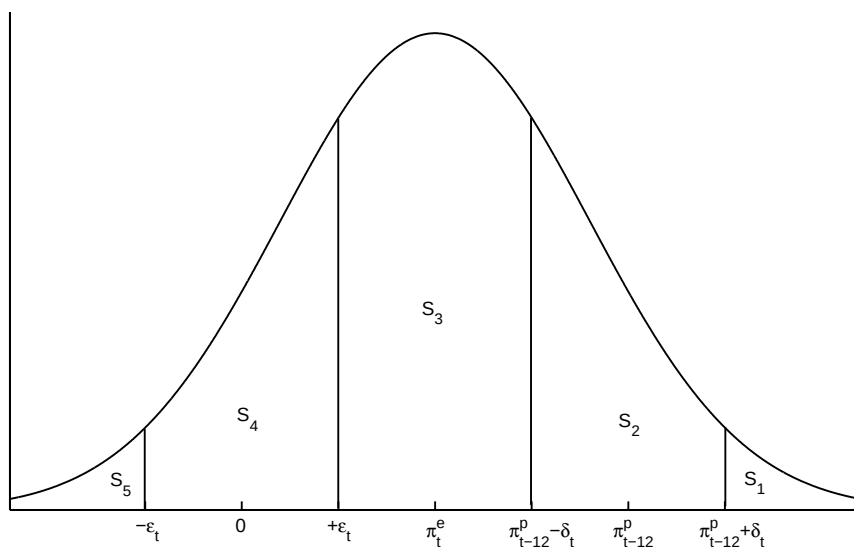
⁹ Respondents are also given the option to reply that they do not have any opinion about the future behaviour of prices.

¹⁰ Denoting S_i (for $i = 1, 2, 3, 4$ and 5) as the sample proportions opting for each of the five response categories, one widely reported balance statistic is calculated as $(S_1 + \frac{1}{2}S_2) - (\frac{1}{2}S_4 + S_5)$.

at least ε_t units above (below) zero.¹¹ In addition, respondents are assumed to condition their replies on their perceptions of past inflation. In particular, respondents report that prices are expected to increase at a more (less) rapid pace if the expected rate of inflation is at least δ_t units above (below) the currently perceived rate of inflation at the time the survey was carried out (π_{t-12}^p).

One option to estimate the perceived inflation rate is to use the currently available rate of inflation, i.e. to assume that respondents correctly perceive the actual rate of inflation at the time they form their expectations. However, it is a quite strong assumption to assume that consumers perceive inflation correctly and, therefore, in this paper π_{t-12}^p is estimated independently using the results from a question on past price developments also contained in the Commission's consumer survey.¹²

Figure 1 The aggregate probability distribution of inflation expectations



¹¹ There is some ambiguity as to whether or not the question in the consumer survey in month $t-12$ refers to the expected year-on-year rate of price increase in month t or the expected *average* rate of inflation in the 12 month period between months $t-12$ and t . Throughout this paper, the consumer survey is assumed to refer to the former.

¹² The question on past price developments asks consumers the following: Compared with what it was 12 months ago, do you think that the cost of living is now 1. Very much higher, 2. Quite a bit higher, 3. A little higher, 4. About the same, 5. Lower or 6. Don't know.

Using the above framework, conditional on an assumed form of the aggregate distribution, it is possible to solve for the level of the expected inflation rate, its standard error (σ_t) as well as for the two response thresholds (ε_t and δ_t). Denoting S_t^i (for $i=1,2,3,4$ and 5) as the sample proportions opting for each of the five response categories in the survey undertaken in month t , the solutions are given by equations (2.1) to (2.4).¹³

$$\pi_t^e = -\pi_{t-12}^p \left[\frac{(Z_{t-12}^3 + Z_{t-12}^4)}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right] \quad (2.1)$$

$$\sigma_t^e = \pi_{t-12}^p \left[\frac{2}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right] \quad (2.2)$$

$$\delta_t = \pi_{t-12}^p \left[\frac{(Z_{t-12}^1 + Z_{t-12}^2)}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right] \quad (2.3)$$

$$\varepsilon_t = \pi_{t-12}^p \left[\frac{(Z_{t-12}^3 - Z_{t-12}^4)}{Z_{t-12}^1 + Z_{t-12}^2 - Z_{t-12}^3 - Z_{t-12}^4} \right] \quad (2.4)$$

Where all variables are defined previously and σ_t^e denotes the standard deviation of the aggregate distribution for inflation expectations. $N^{-1}[\cdot]$, is the inverse of the assumed probability distribution function and $Z_{t-12}^1 = N^{-1}[1 - S_{t-12}^1]$, $Z_{t-12}^2 = N^{-1}[1 - S_{t-12}^1 - S_{t-12}^2]$, $Z_{t-12}^3 = N^{-1}[1 - S_{t-12}^1 - S_{t-12}^2 - S_{t-12}^3]$ and $Z_{t-12}^4 = N^{-1}[S_{t-12}^5]$. The above expressions for the mean and standard error of expected future inflation are quite similar to the original CP results. They express the mean and the uncertainty of expected inflation as a function of the Z_{t-12}^i and the perceived rate of inflation, which has a scaling function. However, in contrast to the original CP approach, where the scaling parameter was estimated by imposing unbiased expectations, an important advantage of equation (2.1) is that it does not impose unbiasedness as an *a priori* property of the measure of expectations.¹⁴ This facilitates subsequent empirical testing of the extent of any bias in the inflation expectations of consumers. However, a disadvantage of the probability approach is that it may give rise to counterintuitive movements in the derived expectations series. For example, one would expect that a *ceteris paribus* increase in the share of respondents expecting prices to rise at a more rapid pace (e.g. as a result of a decline in the share of respondents giving a “don’t know” reply) would be associated with a higher quantitative estimate of expected inflation.

¹³ See the Appendix in Berk (1999) for a full derivation and further discussion of these equations.

¹⁴ Batchelor (1982) suggests a least squares rather than an unbiasedness criterion, i.e. by choosing a scaling factor such that the sum of squared expectation errors is minimised.

However, consideration of the partial derivative of π_t^e in (2.1) with respect to S_{t-12}^1 , shows that this is not necessarily the case.¹⁵

A second important advantage compared with the CP methodology is that the response thresholds are permitted to vary over time. Such time-variation may represent an important feature of the problem faced by consumers in responding to the survey. Batchelor (1986b) provided arguments for time-varying response thresholds based on signal extraction theory, suggesting that the response threshold will tend to be high when the aggregate uncertainty surrounding inflation is also high. Similarly, Seitz (1988) models the response thresholds as a linear function of the actual inflation rate and the dispersion of price changes across industries. In contrast, the above measures permit time-varying response thresholds to be derived directly from the survey replies without the need to make any ad hoc assumptions on their determinants.

2.2 Results for the euro area

Chart 1 plots the derived measure of inflation expectations in the euro area together with the actual year-on-year rate of increase in the Harmonised Index of Consumer Prices (HICP) over the period January 1986 to February 2001 (see Annex 1 for a full description of the data sources).¹⁶ The results suggest a strong relationship between actual and expected inflation. In particular, although there is some evidence to suggest that consumers underpredicted inflation in the second half of the 1980s, they appear to have broadly anticipated the trend decline in inflation over the course of the 1990s (from a high of 5.2% in July 1991 to only 0.8% in February of 1999). Indeed, over the period from the beginning of the 1990s to the start of Economic and Monetary Union in January 1999, the evidence strongly supports a significant degree of forward-looking information in the derived expectations series. In particular, consumers appear to have performed much better than if they had attached a very high weight to past inflationary trends when forming their expectations. For example, in January of 1992, inflation had average 4.3% over the preceding year. However, in that same month consumers were anticipating an inflation rate of 3.2% for the next 12 months and this compared more favourably with the actual outcome for January 1993 (3.4%). This pattern is repeated over much of the 1990s. Hence, it suggests a reasonably rational anticipation of the trend decline in inflation that would not have been possible if consumers simply would have extrapolated in a backward-looking manner past inflation when forming their expectations.

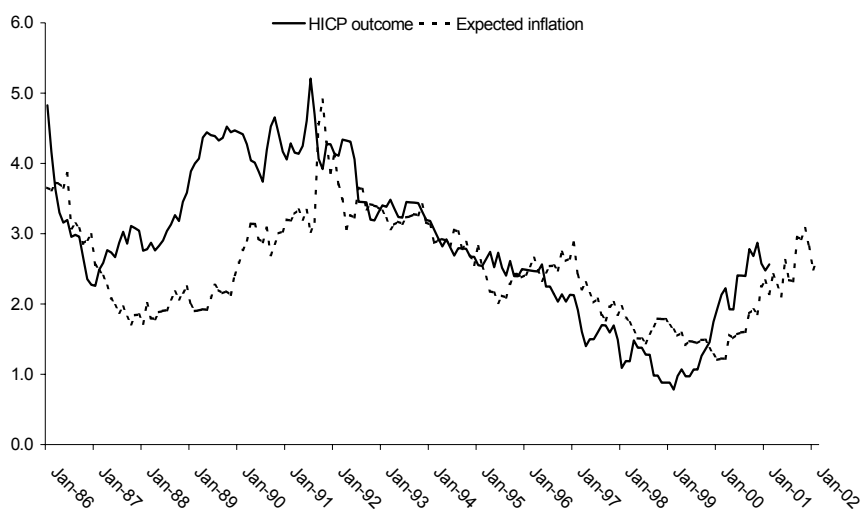
To provide a more quantitative evaluation of the expectations series, Table 1 presents some standard forecast performance statistics. The first statistic is the mean error (ME), which shows the average forecast error over the sample periods. Hence, a large mean error provides evidence of systematic over- or under-prediction. The second measure is the mean absolute error (MAE), which says something about the accuracy of the consumers expectations as a predictor of the future inflation outcome. Large but counterbalancing

¹⁵ See Batchelor (1986a) for a discussion of this problem.

¹⁶ The expectations reported for month t in Chart 1 were formed in month $t-12$. Hence, the difference between the two lines in Chart 1 represents the expectational error.

forecast errors would produce a low mean error but a higher mean absolute error. The third measure is the root mean squared error (RMSE), which also provides a measure of accuracy. An important difference between the MAE and the RMSE is that the former is less sensitive to very large expectational errors or outliers. For comparison reasons the performance measures have also been calculated for two alternative measures of expectations: (i) a naive expectation which simply extrapolates the current year-on-year rate as the expectation for the next twelve months ($\pi_t^e = \pi_{t-12}$) and (ii) a proxy for financial market expectations derived from a Fisher relation for the 1-year ahead money-market interest rate.¹⁷ Consistent with the graphical evidence in Chart 1 that performance has improved in the 1990s compared with the 1980s, the sample is split into two sub-periods. This split is also motivated by the observed decline in inflation over the past decades, which may have changed the inflationary process and, thereby, also the behaviour and forecasting performance of consumers' inflation expectations. The chosen break (1992:12) represents approximately the mid-point of the sample. The performance statistics are then calculated for the two sub-samples as well as the whole sample.

Chart 1 **Expected and actual inflation: Euro area – 1986 - 2001**



17

In order to calculate a measure of expectations from the 1-year interest rate, the real interest rate is estimated as a weighted moving average of the past twelve *ex post* real rates where the weights diminish with the lag length, i.e. $\pi_t^e = i_{t-12} - w_i \sum_{i=0}^{11} (i_{t-12-i} - \pi_{t-12-i})$, where i_{t-12} is the 12-month nominal interest rate and w_i is the weight which is defined as $w_i = (12-i)/78$, where $\sum_{i=0}^{11} w_i = 1$.

The performance statistics in Table 1 indicate quite strongly the change in the performance of the derived expectation series in the 1990s compared with the late 1980s. From the table it is clear that any bias in expectations relates almost entirely to the late 1980s with the mean error falling from 0.94 over the period 1986:1-1992:12 to only -0.06 in the period 1993:1-2000:12. Similarly, there is a dramatic improvement in accuracy statistics in the second sub-sample compared with the first: In the 1990s both the MAE and the RMSE are around 40% of their levels in the late 1980s. Indeed a striking feature of the surveyed expectations for the euro area is their relative precision as a predictor of one-year ahead inflation compared with the two other benchmarks over the 1990s. Most notably, they significantly outperform the expectations based on the naive model and those derived from the Fisher relation. The performance of the measure of consumer expectations also compares reasonably well with those reported for measures of US inflation expectations. For example, Thomas (1999) for the US obtains RMSE for various one-year ahead inflation expectations that range between 1.09 and 1.57.¹⁸ Lastly, in terms of accuracy, the expectations series also compare rather favourably with the out-of-sample forecasts of various indicator models for inflation. For US inflation see, for example, Cecchetti (1995) and Cecchetti, Chu and Steindel (2000). For euro area inflation see Nicoletti Altinari (2001).

Table 1 **Forecast performance statistics: Alternative measures of expected inflation**

	Expected inflation	Fisher	Naïve
Full period: 1986:1 to 2000:12			
ME	0.40	-0.19	-0.22
MAE	0.70	0.76	0.69
RMSE	0.94	1.03	0.95
Subperiod: 1986:1 to 1992:12			
ME	0.94	-0.40	-0.28
MAE	1.07	0.95	0.87
RMSE	1.29	1.29	1.20
Subperiod: 1993:1 to 2000:12			
ME	-0.06	-0.01	-0.16
MAE	0.38	0.59	0.53
RMSE	0.47	0.74	0.65

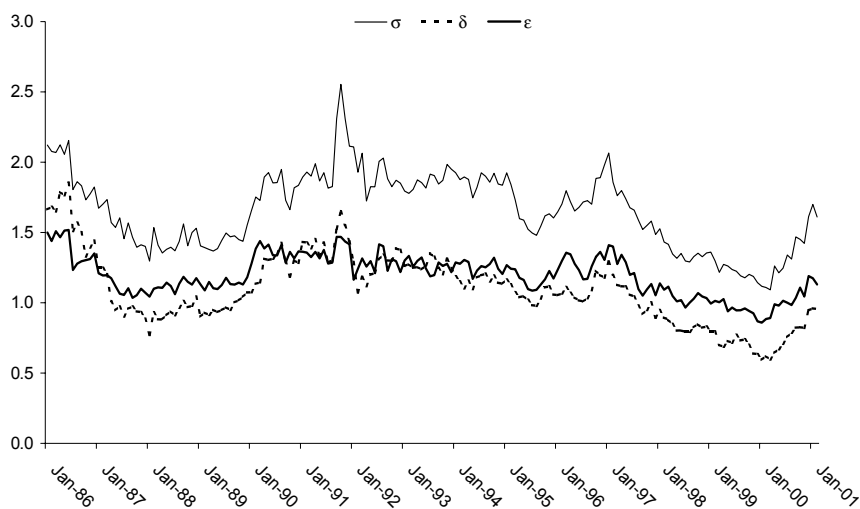
Note: ME = Mean Error, MAE = Mean Absolute Error, RMSE = Root Mean Squared Error.

In order to complete the presentation of the results from applying the probability approach to the euro area consumer survey, Chart 2 plots the estimate of uncertainty (σ_t)

¹⁸ These were calculated over the period 1983:3-1997:4.

along with the estimated response thresholds (ε_t and δ_t) over the period 1985:1 – 2000:12. Inflation uncertainty (σ_t) appears to have trended down somewhat over the 1990s from a high of over 2.5 percentage points in October 1990 to 1.1 percentage points in January 1999. This suggests a high degree of correlation between uncertainty surrounding expected future inflation and the level of actual inflation. Inflation uncertainty did, however, increase abruptly following the oil shock in early 1999. As regards the response thresholds, they also correlate positively both with each other and with the measure of uncertainty.

Chart 2 **Uncertainty of expectations and time-varying response thresholds**



3 Properties of Euro Area Inflation Expectations

The evidence described in section 2 suggests that the measure of expectations derived from the EC survey provides potentially useful information on actual future price developments compared with some simple benchmark alternative measures. In this section, the properties of the derived measure of expectations are examined more formally. We first report the results of regression-based tests for unbiasedness – a necessary condition for rationality. However, tests of bias shed light only on the behaviour of expectations on average. Hence, we also consider the dynamic properties of expectations and, in particular, employ time series techniques in order to investigate the speed with which expectations are revised over time in the light of new information. Following on from this, the weak and strong-form efficiency of expectations is investigated by testing the extent to which the expectations

incorporate the information contained in past price developments as well as that contained in a broader set of macroeconomic variables – available at the time the expectations are formed.

3.1 Long-run properties: Testing for bias

Bias refers to the limiting properties of expectations. In particular, biased expectations imply that consumers, on average, systematically under or overpredict inflation over the long run. A formal test for bias in the expectations series can be carried out using the following equation:

$$\pi_t = \alpha + \beta\pi_t^e + u_t, \quad (3.1)$$

where π_t is the observed inflation rate in month t and π_t^e represents expectations for inflation in month t – but formed in month $t-12$. If the joint null hypothesis $H_0 : (\alpha, \beta) = (0, 1)$ cannot be rejected it can be concluded that the expectations are unbiased in a statistical sense. In line with the previous evidence of an improved performance of the expectations indicator in the 1990s compared with the 1980s, a Chow breakpoint test conducted on equation (3.1) confirms that a structural break has occurred in the relation between consumers' expectations and actual inflation.¹⁹ Hence, the hypothesis of unbiased inflation expectations is conducted for the full sample period and the two sub-sample periods as defined earlier. The results are presented in Table 2. Overall, the results suggest that consumers' inflation expectations have been an unbiased predictor of inflation 12 months ahead. Over the full sample, the null hypothesis cannot be rejected at a reasonable level of significance. However, looking at the results for the first sub-period, the null hypothesis is rejected thus confirming the previous finding of poor performance in the late 1980s. The positive α suggests that expectations have on average underestimated inflation and, in addition, β is well below unity. In contrast, for the second sub-sample the unrestricted parameter estimates accord almost exactly with their theoretical values under the null hypothesis of unbiased expectations. Moreover, if a similar test is conducted but replacing the derived expectation series with expectations based on a naïve model the null hypothesis of unbiased expectations is rejected for all of the three sample periods.²⁰ This finding is consistent with the view that the inflation expectations are more rational compared with a simple model that just extrapolates past outcomes.

¹⁹ The test strongly rejects the null hypothesis of no structural change with a log likely ratio of 91.91, which is to be compared with the critical value of 5.99.

²⁰ A test of unbiasedness of naïve expectations uses the regression $\pi_t = \alpha + \beta\pi_{t-12} + u_t$. The χ^2 statistics for the full sample, the first and second sub-samples are 31.97, 35.84 and 46.38 respectively and these all imply a clear rejection of unbiasedness at standard significance levels.

Table 2 **Test for unbiasedness, sub-periods, ($\pi_t = \alpha + \beta\pi_t^e + u_t$)**

	α	β	R^2	χ^2 for H_0	n
Full period					
1986:1 to 2000:12	0.55 (.40)	0.90 (.14)	.41	5.86 [.053]	180
Subperiods					
1986:1 to 1992:12	2.48 (.48)	0.37 (.15)	.12	32.4** [.000]	84
1993:1 to 2000:12	-0.07 (.39)	1.00 (.14)	.63	0.91 [.635]	96

Notes: ** significant at .01 level. Figures in parentheses are standard errors. Figures in brackets are P values. n is number of observations. Chi-squared statistics pertain to null hypothesis $H_0: (\alpha, \beta) = (0, 1)$. Equations are estimated by OLS using covariance matrix corrections suggested by Newey and West (1987).

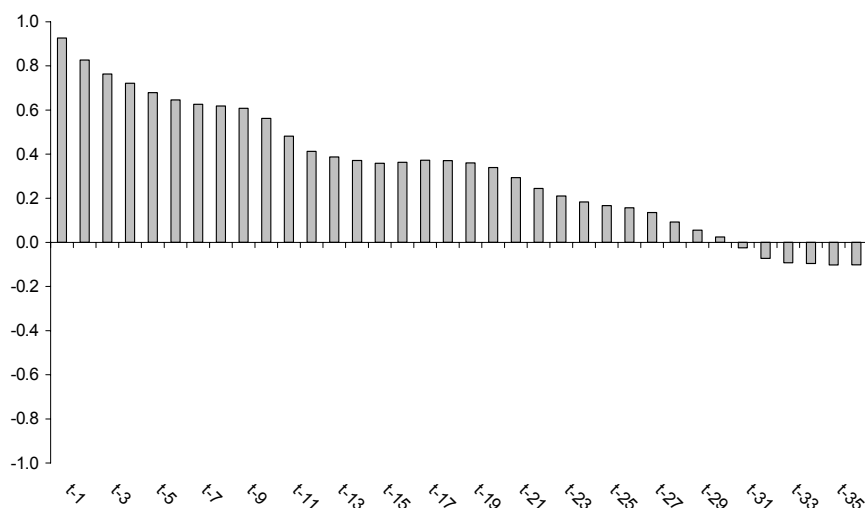
3.2 *Dynamic properties: The adjustment of expectations*

The above simple tests for bias shed light on the question of whether or not – *on average* – the derived measure of consumers' inflation expectations provide an accurate estimate of inflation 12 months ahead. However, tests of bias do not shed any light on the dynamics of the expectations' formation process over time. From a more dynamic perspective, an important issue is the extent to which consumers revise their expectations to reflect the flow of new information, including their knowledge of their own previous forecast errors. A common articulation of the rational expectations hypothesis is that non-overlapping errors in predicting inflation should be uncorrelated – otherwise consumers could have improved their expectations by taking better account of their past errors. In the present context with monthly data on expectations for one-year ahead, the errors are overlapping and this may induce some positive autocorrelation although this should dampen out for lags greater than 12.

The view that the forecast errors of rational agents should be broadly uncorrelated for lags greater than 12 is not undisputed. For example, since the process generating aggregate inflation is uncertain, Cuckierman (1986) argued that even perfectly rational agents may not be able to distinguish permanent from transitory shocks. If this is the case and if permanent shocks are mistakenly perceived to be transitory, agents may make repeated one-sided errors in forming their expectations (see also Thomas, 1999). The evidence from the EC consumer survey is broadly consistent with this view. For example, Chart 3 reports the autocorrelation function for these errors and, while they decay gradually over time, there does appear to be some positive autocorrelation for lags greater than 12. However, even if

uncertainty about the nature of the shocks effecting the inflationary process may give rise to such persistence in expectational errors, they should not persist indefinitely. In particular, as argued in McCallum (1980), because errors are costly (e.g. they can lead to poor decisions) “purposeful agents” have an incentive to acquire sufficient information to weed out systematic expectational error (see also Carlson, 1987).

Chart 3 **Expectation errors: Autocorrelations**



Time series econometrics and, in particular, cointegration analysis suggest an insightful way to investigate the dynamic properties of inflation expectations and, in particular, the speed with which consumers revise their expectations to take account of the flow of new information and weed out any systematic error. If actual and expected inflation are cointegrated with cointegrating vector $[1, -1]$, it is possible to estimate an error correction model whereby the adjustment of expected inflation toward its fully rational value can be examined.²¹ The adjustment parameter from the error correction model reveals important information on the extent and speed with which consumers adjust their expectations toward the fully rational outcome. In addition, within such a framework, it is also insightful to test whether there is any feedback from expected inflation to the actual inflation outcome. Given this potential for feedback in both directions, we first test for the existence of a cointegrating

²¹ Other recent papers applying cointegration techniques to shed light on the rationality of expectations include Grant and Thomas (1998), Dutt and Ghosh (2000), Berk (1999) and (2000). See also Paquet (1992).

relation between actual and expected inflation. In a second step, a bi-variate error correction system of the form given by (3.1) below is estimated and its dynamic properties examined.

$$\Delta\pi_t^e = \alpha_o^e + \sum_{i=1}^p \phi_i^e \Delta\pi_{t-i}^e + \sum_{i=1}^p \psi_i^e \Delta\pi_{t-i} + \alpha^e [\pi_{t-1}^e - \pi_{t-1}] + \varepsilon_t^e \quad (3.1a)$$

$$\Delta\pi_t = \alpha_o + \sum_{i=1}^p \phi_i \Delta\pi_{t-i}^e + \sum_{i=1}^p \psi_i \Delta\pi_{t-i} + \alpha [\pi_{t-1}^e - \pi_{t-1}] + \varepsilon_t \quad (3.1b)$$

Under the hypothesis of cointegration, one of the $\alpha^e, \alpha \neq 0$.²² An interesting feature of this system is that the equation for expectations has a forward – as opposed to a backward-looking dimension. In particular, in the case of equation (3.1a), the test of the hypothesis $\alpha^e \neq 0$ tests whether the change in expectations (formed between periods $t-13$ and $t-12$) is such that the level of inflation expected for period t will be closer to the rational outcome (which is unknown at the time expectations were formed). In the case of equation (3.1b), the test of the hypothesis $\alpha \neq 0$ examines whether the change in actual inflation also adjust in order to maintain a cointegrating relation with expected inflation. Within such a system, it is also possible to test for additional short-run dynamic effects associated with the coefficients $\phi_i^e, \psi_i^e, \phi_i$ and ψ_i and to compute the dynamic impulse response of either variable following an orthogonalised shock to the equations residuals ($\varepsilon_t^e, \varepsilon_t$).

As the preliminary step in testing for the existence of such an error correction representation, Table 3 reports standard unit root tests for both actual and expected inflation over the period 1985-2001. In general, the tests suggest that both actual and expected inflation have unit roots but that they are stationary when differenced once. However, it is possible to reject the null hypothesis of a unit root in the expectation error suggesting that both actual and expected inflation cointegrate with known cointegrating vector $[1, -1]$.²³

Table 3 **Unit root tests**

	π_t	$\Delta\pi_t$	π_t^e	$\Delta\pi_t^e$	$\pi_t^e - \pi_t$
ADF	-1.30	-9.78**	-2.02	-11.47**	-3.01**

Notes: ADF presents the Augmented Dickey-Fuller test-statistic for the null hypothesis of a unit root in the indicated series against the alternative hypothesis of stationarity. Δ denotes the first difference of each series.

A ** indicates that it is possible to reject the null of a unit root at the 5% level of significance.

²² Granger (1988) pointed out that, if two series are cointegrated, then there must be Granger causality in at least one direction.

²³ Additional tests also support the hypothesis of cointegration between actual inflation and the rate expected by consumers. In particular, using Johansen's maximum likelihood technique, both the maximum eigenvalue and trace statistics (see Hansen and Juselius, 1995) support the hypothesis of a single cointegrating vector at the 90% level.

In a second step, the bi-variate error correction model (3.1a and 3.1b) is estimated. Consistent with the monthly frequency of the data, the lag length p is set equal to twelve.²⁴ The equilibrium correction coefficients (α^e, α) and F-tests for Granger-causality associated with the parameters $\phi_i^e, \psi_i^e, \phi_i$ and ψ_i are reported in Table 4 for both the full and two sub-samples identified earlier.

Table 4 Error correction coefficients and F-Statistics for Granger Causality

	Expectations Regressions			Actual Inflation Regressions		
	α^e	F^1 $\phi_i^e \equiv 0$	F^2 $\psi_i^e \equiv 0$	α	F^3 $\phi_i \equiv 0$	F^4 $\psi_i \equiv 0$
Full period						
1986:1 to 2000:12	-0.05 (2.29)	1.92 [0.04]	4.13 [0.00]	0.02 (0.77)	0.86 [0.49]	1.29 [0.23]
Sub periods						
1986:1 to 1992:12	-0.18 (2.89)	0.66 [0.76]	4.03 [0.00]	0.06 (0.81)	0.68 [0.75]	1.04 [0.42]
1993:1 to 2000:12	-0.08 (1.94)	1.83 [0.06]	1.00 [0.45]	0.07 (1.18)	0.46 [0.92]	1.08 [0.39]
Notes: For the coefficients α^e and α absolute t-statistics are given in (.). For the F-statistics, F^i ($i=1,2,3,4$), P-values are reported in [.].						

In the regression for the change in expectations, the adjustment parameter on the expectational error is significant and suggests that consumers revise their expectations and adjust them toward the fully rational outcome. In terms of the speed of this adjustment, given the monthly frequency of the data, the estimated coefficient (-0.05) implies quite low persistence in the deviations of actual from expected inflation.²⁵ Overall, therefore, these results suggest that consumers react quite quickly when revising their expectations in the light of new information and that, while both actual and expected inflation may drift apart in

²⁴ This was the lowest number of lags that was consistent with serially uncorrelated residuals in the bi-variate system.

²⁵ Abstracting from the other short-term dynamics, this coefficient implies a half-life of deviations from long-run equilibrium of about 12 months. This evidence is close to that reported by Roberts (1998) for the US which implies that expectations adjust 50% to 60% toward the fully rational outcome within a year.

the short-run, they ultimately revert toward one another.²⁶ Considering the other adjustment coefficient (α), while positively signed it is not significant, thereby suggesting that the adjustment of actual inflation plays no significant role in maintaining the cointegrating relation. The finding that expectations adjust towards actual inflation, rather than the other way around, suggests that the process generating inflation is not strongly influenced by expectations, which in turn contrasts with New Keynesian theories of price dynamics. In general, all of the above findings are robust across both of the sub samples considered (1986:1 - 1992:12 and 1993:1 - 2000:12).²⁷

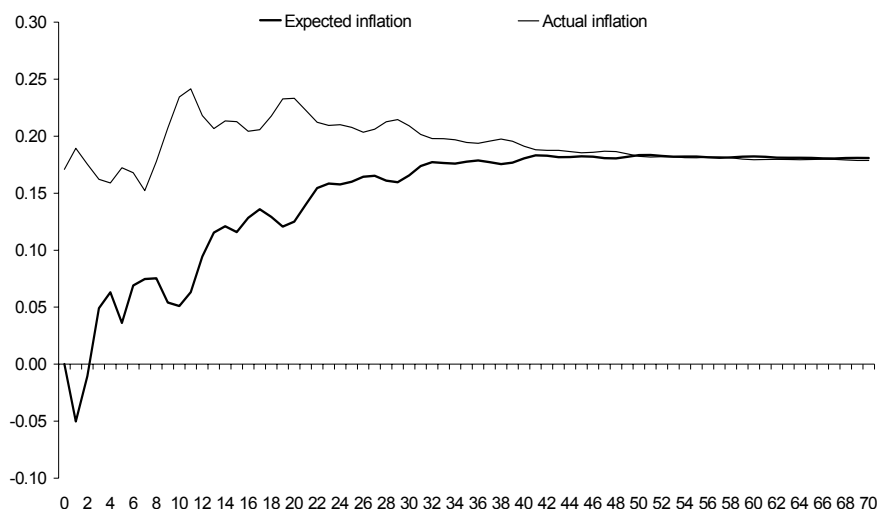
To shed further light on the dynamic interaction between actual and expected inflation, Table 4 also reports the F-statistics for the tests that the other short-run coefficients $\phi_i^e, \psi_i^e, \phi_i$ and ψ_i can be excluded from the bi-variate model. These are denoted F^1, F^2, F^3 and F^4 in Table 4. The results suggest a significant role for actual inflation as determinants of changes in expectations in a Granger-causal sense (F^2). However, the feedback from expectations to actual changes in inflation is quite weak and insignificant at standard levels (F^3). Lastly, to shed further light on the adjustment of expectations, Chart 4 plots the dynamic impulse responses of both actual inflation and expected following a shock to actual inflation. Following such a shock, these responses give the conditional forecasts for the path of both variables that is implicit in the bi-variate model.²⁸ From the chart, and consistent with the finding that it has a unit root, a shock to inflation is highly persistent and has a permanent effect on its level. However, in line with the preceding analysis, expected inflation adjusts gradually to the new permanently higher level of inflation. After approximately 12 months 45% of the deviation between actual and expected inflation has been corrected, while it takes over 30 months before 90% of the gap between expected inflation and its new steady state level has been closed. Overall, this finding of gradual adjustment of expected inflation to its fully rational outcome is in line with the view that, while consumers may face information problems when forming their expectations they eventually adjust them in order to “weed out” any systematic deviations of expected inflation from its fully rational outcome.

²⁶ Some caution is, however, warranted. In particular, under the null hypothesis that actual and expected inflation are not co-integrated, the absolute t-values in Table 4 are not t-distributed. Using the correct critical values in Banerjee et al. (1998), the null hypothesis of non cointegration would not be rejected at standard significance levels (except for the sub-period 1986-1992). This conclusion, however, should be weighed against the alternative evidence from Table 3 on the existence of a known cointegrating relationship between actual and expected inflation. Taking account also of the graphical evidence reported in Chart 1, the evidence appears on balance to support the hypothesis of cointegration.

²⁷ Given the previous evidence of less bias in the measure of expectations in the 1990s compared with the 1980s, one surprising difference is that the adjustment of expectations is estimated to be faster in the first sub-sample compared with the second.

²⁸ The impulse response functions are based on the Choleski decomposition of the covariance matrix of the shocks. As is well known, they are not unique and, in general, depend on which variable is placed first in the bi-variate system. While we report the impulse responses for a bi-variate model in which expected inflation is placed first, the results do not depend significantly on the ordering of the variables.

Chart 4 Impulse response to one standard error shock in actual inflation



3.3 Simple tests of macroeconomic efficiency

A more general formulation of the rational expectations hypothesis is that the expectational errors should be orthogonal with respect to the information set that was known to consumers at the time they formed their expectations, i.e. the *ex post* error cannot be explained by past economic developments. By distinguishing between the breadth of the information set that is assumed to be available to consumers, it is possible to distinguish different degrees of efficiency. In particular, expectations are said to be *weak-form efficient* if the expectational error cannot be explained by an information set that includes only past values of inflation. *Strong-form efficiency*, on the other hand, requires that the expectational error be orthogonal with respect to a much wider information set encompassing many of the macroeconomic variables that are thought to have an influence on price developments.²⁹

The efficiency of consumers' inflation expectations may be tested by estimating the following equation:

²⁹ The distinction between weak- and strong-form efficiency is partly motivated by the idea that information about the past history of inflation is easily accessible to consumers at low cost. In contrast, consumers would have to incur more significant costs to assess the implications for inflation of a broader set of macroeconomic information. For a further discussion of weak and strong-form efficiency see, for example, Thomas (1999) and references cited therein.

$$\pi_t - \pi_t^e = \delta + \phi \Omega_{t-12} + u_t, \quad (3.2)$$

where Ω_{t-12} represents the set of information variables that are relevant for predicting inflation and are available at the time the expectations are formed. In the empirical analysis, this includes past inflation as well as a wide range of other macroeconomic indicators capturing demand and cost pressures as well as monetary and financial conditions. In order to exploit the monthly frequency of the expectations series monthly data are employed for each of the information variables if available. However, for wages, GDP and the nominal effective exchange rate only a quarterly frequency is available so the monthly expectations series is averaged to produce estimates at a quarterly frequency. While in principle it is appropriate to include all variables in a multivariate context this could lead to severe econometric complications (e.g. multicollinearity). Hence, the efficiency of consumers' expectations is evaluated by running a sequence of univariate regressions where the dependent variable is the year-on-year change in the information variable at the time that expectations were formed.³⁰

A statistically significant ϕ suggests that the effect on inflation from past developments in the information variable (Ω_{t-12}) has been incorrectly estimated. Moreover, if the sign on the true correlation between the information variable and inflation is known, it is also possible to judge whether a certain variable's effect on inflation is over or under estimated by consumers. Assuming a positive correlation between the information variable and inflation, a positive (negative) ϕ suggests that the effect from that particular variable on inflation has been under (over) estimated. If a negative correlation is assumed, however, the interpretation is reversed, i.e. a positive (negative) ϕ would suggest that the effect from that particular variable on inflation has been over (under) estimated. The constant (δ) cannot be interpreted in a meaningful way. The results from the macro efficiency tests are presented in Table 5.

Earlier findings based on forecasting performance and tests of unbiasedness indicate that expectations have become more rational over the sample period. This also seems to be the case as regards their efficiency. Over the full sample-period past price developments do not explain consumers' prediction errors thereby providing evidence of weak form efficiency. As regards the other variables all cost indicators, industrial production, the 12-month interest rate and the two euro exchange rate measures also do not explain some of the prediction error. However, consumers' inflation expectations do not appear to be fully efficient with respect to the information contained in the monetary aggregates or some of the other interest rate variables.

³⁰ Hence, this test ignores the effect of publication lags and possible revisions on the information set available to consumers.

Table 5 Test for efficiency, sub-periods, ($\pi_t - \pi_t^e = \delta + \phi \Omega_{t-12} + u_t$)

	Full period		Subperiods			
	1986:1 to 2000:12		1986:1 to 1992:12		1993:1 to 2000:12	
	ϕ	P value	ϕ	P value	ϕ	P value
Demand variables						
GDP	0.28	.028*	0.53	.000**	0.01	.921
Industrial production	0.02	.553	0.00	.986	-0.01	.420
Unemployment	-0.39	.000**	-1.26	.000**	-0.05	.631
Monetary and Financial variables						
M1	0.07	.001**	0.10	.090	0.05	.081
M3	0.20	.000**	0.16	.147	0.07	.016*
3-month interest rate	0.20	.000**	0.24	.013*	-0.01	.721
Long-term interest rate	0.25	.003**	0.27	.019*	0.14	.018*
12-m. real interest rate ¹⁾	0.08	.208	-0.12	.594	-0.02	.636
USD/euro	0.01	.176	-0.00	.678	0.01	.289
Nominal effective exchange rate	0.01	.487	-0.01	.695	0.00	.970
Price & Cost variables						
π	0.02	.842	-0.49	.000**	0.00	.966
Compensation per employee	0.10	.195	-0.44	.032*	0.80	.433
Producer prices	-0.00	.970	-0.02	.679	-0.06	.300
Commodity prices	0.00	.643	0.01	.201	0.01	.215
– Energy	0.00	.893	0.00	.917	0.00	.127
– Non-energy	0.01	.242	0.04	.000**	0.00	.790

1) Derived from the Fisher relation for the 12-months ahead money-market interest rate.

Notes: * denotes significance at .05 level. ** denotes significance at .01 level. For unemployment and interest rates Ω_{t-12} is represented by the year-on-year difference in the variable at the time the survey was conducted. For all other variables Ω_{t-12} represents the annual percentage change.

Looking at the two sub-samples there is, however, clear evidence of a substantial improvement in terms of efficiency in the 1990s compared with the 1980s. In the second sub-sample there are only two variables that significantly explain the errors; M3 and the long-term

interest rate.³¹ For both of them ϕ is positive, suggesting that consumers have underestimated their impact on inflation (assuming a positive relationship between each of these variables and inflation). As regards the former variable this finding is particularly interesting given the monetary policy strategy that has been adopted by the ECB. In particular, since consumers do not take account of the information contained in monetary aggregates when forming their expectations, the results tend to support the independent and incremental information role that is assigned to M3 in the strategy.

Overall, the evidence from the efficiency tests suggests an intermediate rather than a full form of rationality with consumers taking into account a wide ranging – but not complete – set of information in forming their expectations. In addition, consistent with the results relating to bias, there is evidence of “growing” rationality when the performance of expectations in the 1990s is compared with that in the late 1980s. However, the higher degree of efficiency in the second sub-sample may not only be due to greater rationality on the part of consumers. Euro area inflation has declined over the last decades and become less volatile. This, in combination with greater central bank independence and credibility, has made it easier to predict future inflation. Hence, consumers may have become more efficient at predicting inflation but policy makers may also have made the job easier than it was 10 years ago.

4 Expectations and Inflation Dynamics

The preceding sections have shown that the consumer survey measure of inflation expectations broadly fulfils several (though not all) of the necessary conditions for rationality. However, as discussed in the introduction, a key insight of the rational expectations hypothesis is that rational expectations of inflation have a strong influence on actual inflation. This view has been re-articulated in recent new Keynesian models of price dynamics.³² The contribution of the New Keynesian models is to emphasise the role of nominal rigidities in interpreting the Phillips curve. In particular, the reason why expectations matter is because there is some constraint on the frequency of price adjustment and as a result, an optimising firm will partly incorporate expected future changes in costs (or in other firms prices) into current prices (see, for example, Roberts (1995), Roberts (1998), Sbordone (2000a) and (2000b), Galí and Gertler (1999), Galí, Gertler and López-Salido (2001)). As a result, these authors contrast the traditional Phillips curves where the impact of expectations is captured solely by past lags of inflation with the New Phillips Curve (NPC) where inflation is a truly forward-looking phenomenon. Denoting D_t as an excess demand measure, these two divergent views of inflation dynamics are represented by equations (4.1) and (4.2):

³¹ The consistent lack of efficiency with respect to monetary variables that is found for the euro area is at odds with the evidence reported in Ball and Croushore (1998) for US inflation expectations.

³² Wolman (1999) provides a recent survey of these models.

$$\pi_t = \gamma_b \pi_{t-1} + \beta_1 D_t \quad (4.1)$$

$$\pi_t = \gamma_f \pi_{t+1}^e + \beta_2 D_t \quad (4.2)$$

The difference between (4.1) and (4.2), i.e. between whether or not inflation is a backward as opposed to a forward-looking phenomenon, is of more than just academic interest. For example, under (4.2) if a central bank announced a move to reduce inflation and if this announcement was fully credible and thereby led to a reduction in expectations, then actual inflation could fall without the need for any reduction in output. However, under (4.1), the process of disinflation always implies a reduction in output. More generally, the distinction between backward- and forward-looking inflation dynamics also has important implications for a central bank attempting to assess – in real time – the interaction between the business cycle and inflation. For example, if inflation is forward-looking then it will tend to lead (i.e. help predict) future changes in output. This can be seen by recursively solving (4.2) forward to obtain an expression for current period inflation as a (discounted) weighted sum of current and future deviations of output from its steady-state value (see for example Galí, Gertler and López-Salido, 2001). Conversely, if inflation is purely backward looking and primarily driven by past price developments, it will tend to lag (or be driven by) the business cycle.

Given the previous findings on the relatively significant information content of the EC consumer survey data on inflation expectations, such data provide a strong basis on which to test the role of expectations in the inflationary process.³³ To examine this role, appending a cost shock variable (Z_t), a hybrid model which nest (4.1) and (4.2) may be estimated.³⁴ This is given by (4.3) below:

$$\pi_t = \alpha + \gamma_f \pi_{t+4}^e + \gamma_b \pi_{t-4} + \beta D_t + \delta Z_t + u_t \quad (4.3)$$

In applying (4.3) to the data, we use non-overlapping quarterly observations of the year-on-year change in prices: π_t denotes the change in quarter t compared with quarter $t-4$ while π_{t-4} is the change in quarter $t-4$ over the same quarter one year earlier ($t-8$). Similarly π_{t+4}^e is the average of the three monthly expectations surveyed in quarter t for the next 12 months. In order to ensure robustness of the results with respect to the chosen measure of excess demand, two alternative proxies for D_t are considered. First, a standard output gap is considered defined as the log of actual GDP in the euro area minus the log of potential GDP (estimated using a HP filter). Second, we also consider the primitive version of the NPC suggested by Galí *et al.* (2001) according to which the deviations in real marginal

³³ In particular, from the new Keynesian perspective, the one-year ahead horizon for the expectations from the consumer survey should broadly capture the horizon over which costs or prices are sticky.

³⁴ Galí and Gertler (1999) derive a hybrid version of the NPC by assuming an economy populated by two groups of firms. One group sets prices in a forward looking way given the constraints on price adjustment and using all the available information on expected future costs. The remaining set of firms base their prices on past price developments alone.

costs from their steady state value should replace the traditional and more *ad hoc* output gap measures. The deviation of real unit labour costs from their sample mean is used as a proxy for marginal costs. The real price of oil is used to capture the impact of transitory cost shocks (Z_t). The equations are estimated over the period 1985Q1-2000Q4.

Table 6a reports the coefficients from the estimated hybrid Phillips curves. The two equations appear to capture quite adequately the dynamics of inflation over the sample period. A striking feature of the results is that for both specifications, the surveyed measure of inflation expectations enters with a highly significant positive coefficient, which is close to unity. This suggests to a certain degree a forward-looking dimension to the inflationary process in the euro area and supports the findings in Galí *et al.* (2001). However, the backward-looking component is also significant but quantitatively less important. The specification using real unit labour costs appears to have the most satisfactory fit. Surprisingly, the real price of oil does not enter significantly into any of the regressions.

Table 6a **Estimates of hybrid Phillips curve,**
 $(\pi_t = \alpha + \gamma_f \pi_{t+4}^e + \gamma_b \pi_{t-4} + \beta_1 D_t + \delta Z_t + u_t)$

	α	γ_f	γ_b	β	δ	$\bar{R}^2$	n
Output gap	-0.920 [.303]**	1.151 [.205]**	0.310 [.088]**	-0.058 [.085]	0.002 [.002]	.880	64
Real unit labour cost gap	-0.344 [.362]	1.062 [.184]**	0.221 [.106]*	0.080 [.029]**	0.003 [.002]	.899	63

Notes: * denotes significance at .05 level. ** denotes significance at .01 level. Figures in brackets are standard errors.
 n is the number of observations.

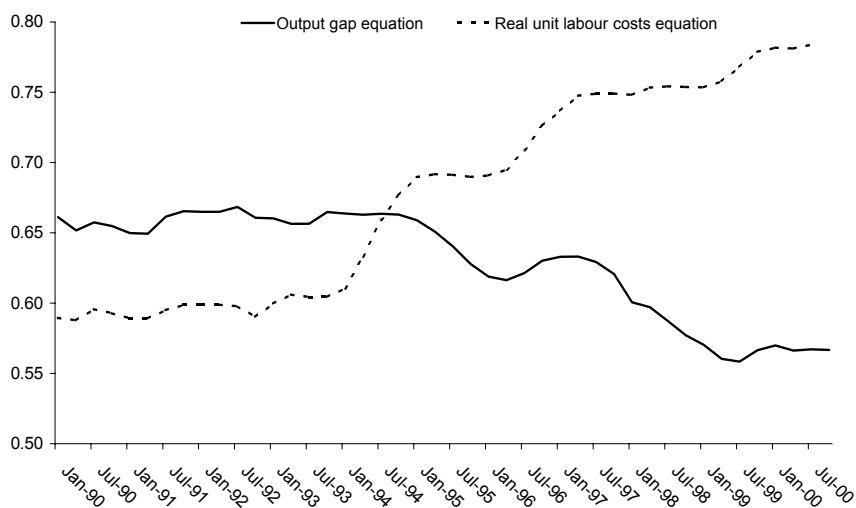
Table 6b reports the results of similar regressions but under the restrictions that the weights on the backward and forward looking components sum to unity. This restriction is accepted only at the 1% level of significance in the equation using the output gap. However, for both specifications the role for expectations in determining actual inflation dynamics is maintained and remains quantitatively more important than the backward-looking component. Most notably, in the specification using real unit labour costs, the expectations term has a weight close to four times that on past changes in prices. Lastly, in order to check for stability, Chart 5 plots recursive estimates of the restricted parameter on consumers' expectations in each of the two models. While there is some evidence that the weight attaching to the forward looking component has risen in the real unit labour costs specifications but fallen for the output gap equation, the estimated parameters range between 0.56 and 0.78.

Table 6b **Restricted estimates of hybrid Phillips curve,**
 $(\pi_t = \alpha + \gamma_f \pi_{t+4}^e + (1 - \gamma_f) \pi_{t-4} + \beta_2 D_t + \delta Z_t + u_t)$

	α	γ_f	$(1 - \gamma_f)$	β	δ	$\bar{R}^2$	χ^2 for H_0	n
Output gap	0.145 [.142]	0.567 [.107]**	0.433 [.107]**	0.171 [.073]*	0.003 [.003]	.829	10.00 (.002)**	64
Real unit labour cost gap	0.388 [.134]**	0.784 [.123]**	0.216 [.123]	0.117 [.030]**	0.005 [.002]*	.874	3.97 (.046)*	63

Notes: * denotes significance at .05 level. ** denotes significance at .01 level. Figures in brackets are standard errors. n is number of observations. Figures in parentheses are P values. Chi-squared statistics pertain to null the hypothesis ($H_0: \gamma_f + \gamma_b = 1$).

Chart 5 **Hybrid Phillips curve - Recursive parameter estimates (γ_f)**



5 Conclusions

Surveys are useful because they provide independent (or relatively non-model dependent) measures of inflation expectations, a key variable that a central bank can use in its design of an optimal monetary policy geared toward the achievement of price stability. This paper analyses and assesses the empirical properties of consumers' inflation expectations in the euro area over the period from the mid 1980s to the end of 2000. Using the approach set out in Berk (1999), quantitative estimates of euro area inflation expectations are derived from the qualitative data from the European Commission Consumer Survey. The paper subsequently analyses the empirical properties of the estimated inflation expectations by considering the extent to which they fulfil some of the necessary conditions for rationality.

The results suggest consumers' inflation expectations in the euro area satisfy an intermediate rather than a complete form of rationality. Firstly, compared with other benchmark indicators of expected inflation, consumers' expectations are shown to provide a reasonably accurate predictor of inflation one-year ahead. Second, the surveyed expectations are found to be an unbiased predictor of future price developments over the long-run. In line with this, the evidence is also generally supportive of the hypothesis that the level of expected inflation and the rational inflation outcome are cointegrated. As a result, consumers are shown to gradually adjust their expectations in order to "weed out" any systematic expectational error. Lastly, consumers' expectations in the euro area appear to incorporate – though not always completely – the information contained in a broad set of macroeconomic variables. In particular, past price developments as well as various cost and activity indicators do not explain consumers' prediction errors. However, consumers' inflation expectations do not appear to be fully efficient with respect to the information contained in the monetary aggregates or some interest rate variables. This suggests that there is an independent and incremental role for these variables in explaining future inflation, which is not taken into account by consumers when forming their expectations.

Overall, the evidence on efficiency of consumers' expectations when combined with the evidence on their unbiasedness and the nature of the process whereby they are adjusted to converge toward the fully rational outcome suggests an intermediate form of rationality (similar to that cited in Roberts (1998) for the US). A further important feature of the empirical analysis has been the investigation of potential changes over time in the properties of the derived expectations series. In this regard, there is some significant evidence of "growing rationality" in the sense that expectations have become a more accurate, less biased and more efficient predictor of inflation over the 1990s compared with the 1980s. This improvement is particularly notable during the period of convergence in preparation for Economic and Monetary Union and, in particular, in the manner in which consumers' anticipated the trend decline in inflation over this period. Part of the explanation for this may indeed be that consumers have become better at predicting inflation. However, euro area inflation has also declined over the last decades and became less volatile. This, in combination with greater central bank credibility, may have made it easier to predict future

inflation. Hence, while consumers may have become more efficient at predicting inflation, monetary policy makers may also have made the job easier than it was ten years ago.

Given these findings on the relatively significant and improving information content of the consumer survey data for expected inflation, such data provide a strong basis on which to test the role of expectations in the inflationary process. As suggested by recent new Keynesian models of inflation we have also examined the role of the survey expectations in explaining actual inflation dynamics. Estimates of a hybrid Phillips curve – which nests backward as well as forward-looking inflation dynamics – suggests that consumer expectations have played a role in determining the observed behaviour of inflation over the period 1985-2001. Overall, therefore, our results underline the important information content of survey data in the analysis of euro area inflation.

References

- Ball, L. and D. Croushore (1998) "Expectations and the effects of monetary policy" *Federal Reserve Bank of Philadelphia*, WP 98-13
- Banerjee, A., J.J. Dolado and R. Mestre (1998) "Error correction mechanism tests for cointegration in a single equation framework", *Journal of Time Series Analysis*, 19(3), pp. 267-283
- Batchelor, R. A. (1981) "Aggregate Expectations under the stable laws" *Journal of Econometrics*, 16, pp. 199-210
- Batchelor, R. A. (1982) "Expectations, output and inflation: The European experience" *European Economic Review*, 17, pp. 1-25
- Batchelor, R. A. (1986a) "Quantitative versus qualitative measures of inflation expectations" *Oxford Bulletin of Economics and Statistics*, 48, pp. 99-120
- Batchelor, R. A. (1986b) "The psychophysics of inflation" *Journal of Economic Psychology*, 7, pp. 263-290
- Batchelor, R. A. and A. B. Orr (1988) "Inflation expectations revisited" *Economica*, 55, pp. 317-31
- Berk, J. M. (1999) "Measuring inflation expectations: a survey data approach" *Applied Economics*, 3, pp. 1467-1480
- Berk, J.M. (2000) "Consumers' inflation expectations and monetary policy in Europe" *Research Memorandum, Vrije Universiteit, Amsterdam*, 2000-20
- Carlson, J. A. and M. Parkin (1975) "Inflation expectations" *Economica*, 42, pp. 123-38
- Carlson J. B. (1987) "Learning, Rationality, the Stability of Equilibrium and Macroeconomics", *Federal Reserve Bank of Cleveland, Economic Review*, Fourth Quarter, pp. 1-12
- Cecchetti, S. G. (1995) "Inflation Indicators and Inflation Policy" *NBER Macroeconomics Annual*.
- Cecchetti, S. G., R. S. Chu and C. Steindel (2000) "The unreliability of inflation indicators" *Current Issues in Economics and Finance, Federal Reserve Bank of New York*, 6 N4
- Christiano, L. J. and C. Gust (2000) "The expectations trap hypothesis" *Federal Reserve Bank of Chicago*.
- Cukierman, A. (1986) "Measuring Inflation Expectations" *Journal of Monetary Economics* 17, pp. 315-324
- Deutsche Bundesbank (2001) "The information content of survey data on expected price developments for monetary policy" *Monthly Report*.
- Dutt, S. D. and D. Ghosh (2000) "Inflation expectations: are they rational? An empirical examination" *Applied Economics Letters*, 7, pp. 103-106
- Fagan, G., J. Henry and R. Mestre (2001) "An area wide model (AWM) for the euro area" *ECB working paper*, 42

- Fuhrer, J. (1997a) "The unimportance of Forward-Looking Behaviour in Price Specifications", *Journal of Money, Credit and Banking*, 29, pp. 338-350
- Fuhrer, J. (1997b) "Inflation Persistence" *The Quarterly Journal of Economics*, 110, pp. 127-159
- Gali, J. and M. Gertler (1999) "Inflation dynamics: A structural econometric analysis" *Journal of Monetary Economics*, 44, pp. 195-222
- Gali, J., M. Gertler and D. López-Salido (2001) "European Inflation Dynamics" *NBER Working Paper*, No. w8218
- Granger, C. (1988) "Recent development in the concept of causality" *Journal of Econometrics*, 39, pp. 199-211
- Grant, A. P. and L. B. Thomas (1998) "Inflationary expectations and rationality revisited" *Economics Letters*, 62, pp. 331-338
- Hansen, H. and K. Juselius (1995), CATS in RATS Version 1.00 – Installation and Users Guide, Estima, IL.
- McCallum, B. T. (1980) "Rational Expectations and Macroeconomic Stabilization Policy: An Overview" *Journal of Money, Credit and Banking*, XII, pp. 697-746
- Newey, W. K. and K. D. West (1987) "A Simple, Positive Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix" *Econometrica*, 55, pp. 703-708
- Nicoletti Altimari, S. (2001) "Does money lead inflation in the euro area" *ECB Working Paper*, 63
- Paquet, A. (1992) "Inflationary expectations and rationality" *Economic Letters*, 40, pp. 303-308
- Roberts, J. M. (2001) "How well does the New Keynesian Sticky-Price Model fit the data", Working Paper, Board of Governors of the Federal Reserve System, February.
- Roberts, J. M. (1998) "Inflation expectations and the Transmission of Monetary Policy" Finance and Economics Discussion Series 1998-43, Federal Reserve Board Washington D.C.
- Roberts, J. M. (1995) "New Keynesian economics and the Phillips Curve" *Journal of Money, Credit and Banking*, 27 N4, pp.939-1506
- Sbordone, A. M. (2000a) "Prices and unit labour costs: a new test of price stickiness" unpublished manuscript, Rutgers University.
- Sbordone, A. M. (2000b) "An optimizing model of US Wage and Price Dynamics" unpublished manuscript, Rutgers University.
- Seitz, H. (1988) "The estimation of inflation forecasts from business survey data" *Applied Economics*, 20, pp. 427-38
- Taylor, J. B. (1999) "Monetary policy rules" *University of Chicago Press*.
- Taylor, J. B. (2000) "Low inflation, pass-through, and the pricing power of firms" *European Economic Review*, (44) 7, June 2000, pp. 1389-1408
- Thomas, L. B. (1999) "Survey Measures of Expected U.S. Inflation" *Journal of Economic Perspectives*, 13, N4-125-144
- Wolman, A. L. (1999) "Sticky Prices, Marginal Cost, and the Behaviour of Inflation" *Federal Reserve Bank of Richmond, Economic Quarterly*, 85/4

Annex 1 The Data Set

Variable	Source	Freq.	Remark
Survey data	European Commission (DG ECFIN)	M	Detailed responses
GDP	Eurostat	Q	
– Potential GDP		Q	Estimated using a HP filter ($\lambda = 1,600$)
– Output gap		Q	Log of actual GDP minus the log of potential GDP
Industrial production	Eurostat	M	Excluding construction
Unemployment rate	Eurostat	M	
M1	ECB	M	
M3	ECB	M	
3-month interest rate	ECB calculations	M	Aggregation based on national data from the BIS databank
Long-term interest rate	ECB	M	
12-month real interest rate	ECB calculations	M	Derived from the Fisher relation for the 12-months ahead money-market interest rate (EURIBOR)
USD/euro	ECB	M	USD/ECU until December 1998
Nominal effective exchange rate	Fagan, Henry and Mestre (2001)	Q	From Q1 1990 onwards linked with EER as published in the ECB's Monthly Bulletin
Inflation			
– CPI	ECB calculation	M	Aggregation based on national data from the BIS databank
– HICP	Eurostat	M	Linked with CPI in January 1990
Compensation per employee	ECB calculation	Q	Aggregation based on national data from the BIS databank
Producer prices	Eurostat	M	Excluding construction
Commodity prices	HWWA - Institut für Wirtschaftsforschung (Hamburg)	M	Converted into euro
Real unit labour costs gap		Q	Log of real unit labour costs minus its mean
– Real unit labour costs		Q	Unit labour costs deflated with HICP
– Unit labour costs	ECB calculation	Q	Aggregation based on national data from the BIS databank

Modelling Short-term Interest Rates in the Euro Area Using Business Survey Data

Renata Grzeda Latocha* and Gernot Nerb**

Abstract

Interest rates play a key role in free market economies. According to the Taylor rule, short-term interest rates depend on deviations of the current inflation rate from a normative value and of the output gap. Given the difficulty of observing the output gap, we postulate that alternative indicators of pressure on capacity should be monitored, especially those obtained from business surveys. Primarily, our work compares those empirically observed capacity utilization figures (CU) with the difference between actual and potential GDP, and tries to assess monetary policy on the basis of these indicators. The analysis of monetary policy is done in the first case by using a simple single equation framework and in a next step by using a more advanced vector autoregressive model. Given the fact that the survey measure of output gap captures only manufacturing and thus omits the important service sector which has a high share in value added we constructed a more comprehensive measure of capacity utilization (CU*). The new indicator is based on CU in manufacturing industry and on the confidence indicator for services. As our analysis shows, this indicator appears to capture better the capacity utilization in the whole euro area economy and should primarily be used as a benchmark for analyzing ECB monetary policy. Whereas the Taylor rule model with CU indicates that the current central bank interest rate should be significantly higher than it actually is, a model with our new indicator CU* rate assesses the current ECB policy as almost fully appropriate. Applying a VAR model which also takes care of additional influencing factors such as exchange rate, oil price and the impact of foreign economies, particularly the US, there appears to be even room for a moderate rate cut of about 25 basis points in autumn 2003, from 2.00 to 1.75.

Key Words: Short-term interest rates; Taylor rule; Output gap; Capacity utilization; Business confidence indicator in the service sector

JEL Classification: E43; E52

* MEAG MUNICH ERGO Asset Management GmbH, Research, RGrzedaLatocha@meag.com
 ** ifo Institute for Economic Research, Business Survey Department, nerb@ifo.de

Note: The paper reflects the views of the authors, not necessarily those of the MEAG and the ifo

Résumé

Les taux d'intérêt sont des rouages essentiels de l'économie de marché. Selon la règle de Taylor, les taux d'intérêt à court terme dépendent de l'écart entre le taux courant de l'inflation et celui de la production. Étant donné la difficulté d'évaluer l'écart de production, on part du postulat qu'il faut suivre l'évolution d'autres indicateurs des tensions sur les capacités, en particulier ceux qui peuvent être dérivés des enquêtes de conjoncture. Tout d'abord, notre travail consiste à comparer les chiffres de la capacité d'utilisation (CU), chiffres empiriquement observés, avec l'écart entre le PIB observé et le PIB potentiel, ainsi qu'à essayer d'évaluer une politique monétaire sur la base de ces indicateurs. L'analyse d'une politique monétaire est effectuée en premier lieu par l'utilisation d'une équation simple et ensuite par l'utilisation d'un modèle VAR. Vu que l'enquête qui mesure l'écart de production prend en compte uniquement le secteur manufacturier et omet donc le secteur des services qui a une part importante dans la valeur ajoutée, nous avons construit une mesure plus complète de la capacité d'utilisation (CU*). Ce nouvel indicateur est basé sur la capacité d'utilisation dans le secteur manufacturier et sur l'indice de confiance du secteur des services. Comme l'a montré notre étude, il semble que cet indicateur capte mieux la capacité d'utilisation dans l'économie de la zone euro tout entière et qu'il devrait être utilisé en premier lieu comme référence pour évaluer la politique monétaire de la Banque centrale européenne (BCE). Bien que le modèle de la règle de Taylor avec CU montre que l'actuel taux d'intérêt de la banque centrale devrait être significativement plus élevé qu'il ne l'est actuellement, un modèle utilisant notre nouvel indicateur CU* évalue la politique de la BCE comme étant presque entièrement appropriée. Appliquer un modèle VAR qui tient également compte d'autres facteurs influents comme le taux de change, le prix du pétrole et l'impact des économies étrangères, et en particulier les États-Unis, il peut être envisagé une réduction modérée de 25 points à l'automne 2003, de 2.00 à 1.75.

1 Introduction

Interest rates play a key role in the free market economy. As prices for the inter-temporal transfer of purchasing power, they are the main influencing factors in a large number of different saving decisions and investment choices in the economy. Depending on the planning horizon either short-term or long-term interest rates are of relevance. In this work, we aim to understand, theoretically and empirically, the behaviour of short-term interest rates.

Economic theory provides several determinants of interest rates. The path of short-term interest rates is determined mainly by central bank decisions. The role of every monetary policy and of course also that of the ECB is to find the best compromise between inflation variability and output gap variability (Svensson 2003). How the central banks should respond to these target variables, inflation and output gap, is reflected in a famous equation known as Taylor rule. However there is a lot of trouble with the output gap estimation and monetary policy makers setting interest rates look mostly at growth and inflation forecasts or business cycle indicators. Given the difficulty of observing the output gap, we postulate that alternative indicators of pressure on capacity should be monitored, especially those obtained from business surveys.

Information generated from business and consumer surveys is primarily designed to deal with judgments, estimates and expectations of economic agents. According to a statement of the European Commission (1997) the usefulness of these survey data is more and more stressed by empirical researchers because this type of data is not only timely but can often be used to construct good leading indicators. The survey data have been collected for more than 30 years by institutions like the Ifo Institute in Munich and are analyzed intensively on a harmonized basis by international institutions like the European Commission (EC), the European Central-Bank (ECB) and the Organization for Economic Co-operation and Development (OECD). Nevertheless, these data have been rarely used until now in the econometric analysis of financial markets. To fill this gap, we decided to use the data on the capacity utilization (CU) rate from business surveys in monetary policy analysis.

The goal of this work is to check the performance of survey data in econometric analysis of interest rates.

2 The Role of the Output Gap in Short-term Interest Rates Modelling

Macroeconomists have long been interested in giving guidance as to how monetary policy should respond to different economic developments. There are several “instrument” rules: Taylor rule, the McCallum rule and Friedman’s constant money growth rule. Officially the ECB does not use a mechanical instrument rule to set monetary policy. Price stability is the primary objective of the ECB. In October 1998 the Governing Council of the ECB agreed that price stability is defined as a year-on-year increase in the Harmonized Index of

Consumer Prices (HICP) for the euro area¹ of below 2 per cent. The Governing Council explicitly announced that price stability is to be maintained over the medium term, thereby acknowledging that price levels may be temporarily distorted by short-term factors (Eijffinger and de Haan 2000). According to the revision of monetary strategy in May 2003 the ECB slightly modified this target by stressing that HICP inflation should in the medium term be below 2% but at the same time relatively close to this figure. This modification should signal that the ECB is both taking into consideration the risk of inflation as well as that of deflation when designing monetary policy.

According to Taylor (1993), author of a relatively simple monetary policy rule,² the main central-bank interest rate should be set primarily with regard to the level of inflation and the trend of real output. An interest rate based on these factors constitutes a fully acceptable monetary policy, according to Taylor.³ Such a rule is easy to explain and monitor. We do not postulate that the ECB should use such a rule when setting its policy because it is not necessarily an optimal rule. However to assess the stance of monetary policy of the ECB a measure is needed. Therefore the Taylor rule is used as a benchmark for comparison with the ECB prevailing policy rate.

The output gap in the Taylor rule plays an important role in interest rates modelling and forecasting. It can be interpreted not only as a measure of the condition of the business cycle but also as an indicator of future inflation in the context of a Phillips curve model⁴ (Judd and Rudebusch 1998). Experts like Gerlach and Svensson (2001) find little empirical support for the prominent role of money-growth indicators in predicting future inflation. The output gap in their investigation seems to have less predictive power for future inflation as well. However, the researchers point out that their empirical results may be sensitive to the exact choice of

¹ The statement “for the euro area” highlights that area-wide developments, instead of specific national or regional factors are relevant for the ECB’s monetary policy decisions.

² The Taylor rule is a conventional decision rule. The “opportunistic” decision rule is a new approach to monetary policy. According to this approach in case of inflation above the long-run target, the opportunistic monetary policy authorities would not take deliberate anti-inflation measures but wait for external circumstances such as a fall in oil prices or an unforeseen recession. When inflation is high the opportunistic policymaker fights inflation, when inflation is low he stabilizes output (see Orphanides and Wilcox 1996).

³ Taylor’s Rule leads to some practical difficulties if inserted into a simple practical model. Researchers blame it for: simplicity (Clarida and Gertler 1996); non-simultaneity among variables in monetary transmission mechanism (Christiano, Eichenbaum and Evans 1999); contemporaneity in reaction of the central bank to key economic variables which are not known in the moment of setting interest rates (McCallum 1998; Orphanides 1997; McCallum and Nelson 1998); backward looking character of the formula because of not including future inflation (Bernanke and Woodford 1997, Clarida, Gali, Gertler 1997); examination problems because of practical definitions of variables (Isard, Laxton 2000).

⁴ The curve was introduced by Phillips (1958) and discussed then by Lipsey (1960) and several other authors. It focuses on the relationship between the rate of inflation and the level of unemployment. Instead of the unemployment gap, the output gap is often included as a measure of excess demand in the Phillips Curve analysis. The connection between unemployment and GDP growth is often formally summarized by the statistical relationship known as Okun’s Law (Okun 1962).

data. It is possible that other measures of the output gap, particularly those which are more up-to-date, like business survey results, give better information on future inflation trends than the ones they have used. Therefore we focus on the capacity utilization rate and confidence indicators obtained from surveys as a proxy for the output gap in monetary policy analysis.

3 Capacity Utilization Rate from Surveys as a Measure of the Output Gap

One of the critical points in Taylor's rule is how to test the hypothesis empirically. The estimation of the Taylor rule can vary a great deal depending on the applied output gap which can be measured as:⁵

- the gap between the unemployment rate and the so-called NAIRU,⁶
- the gap between actual real GDP and potential GDP,⁷
- the industrial capacity utilization (CU), and a more comprehensive measure of capacity utilization based on CU and business survey results in the service sector (CU*).

There are some conceptual differences between these measures of output gap. The industrial capacity utilization rate measures the operating rate of the nation's industrial capacity. It is *the greatest level of output that a plant can maintain within the framework of a realistic work pattern, taking account of normal downtime and assuming the availability of inputs to operate the machinery in place* (Kenessey 1990).

Given an amount of capacity in place initially as well as actual and future demand, the firm may change output without changing capacity, change output and change capacity using existing technology or change output and change capacity, using new technology. The chosen strategy depends on expected profits from covering the expected demand (Bansak et al. 2003). Thus higher output gap does not indicate necessarily higher capacity utilization.

⁵ For surveys, see Chagny and Döpke (2001).

⁶ Natural rate of unemployment or Non-Accelerating Inflation Rate of Unemployment was introduced by Edmund Phelps (1967) and Milton Friedman (1968) and is defined as the level of unemployment that is consistent with constant inflation. In the structural approach the NAIRU is based on an equation that explains wage inflation, which is then used to determine the equilibrium value of another variable, i.e. unemployment. The direct method of NAIRU estimation is based on time series analysis of unemployment (univariate analysis) and other relevant variables (multivariate analysis) (see Fabiani and Mestre 2000).

⁷ The potential GDP is the GDP which the economy has the capacity to produce. The techniques for estimation of potential output can be divided into three classes: structural (the production function of Cobb-Douglas is mostly used), non-structural (statistical) methods (drawing a straight line through the peaks or fitting a linear or polynomial trend through actual output or Hodrick and Prescott (1997) (HP) filter) and finally semi-structural methods (mixture of both previous mentioned methods). The advantages and disadvantages of these methods are laid out very well in the EPC report on Potential Output and Output Gap (Economic Policy Committee of the EU, 2001).

The industrial capacity utilization rate in the euro area is provided by business surveys. It is the most direct and obvious means of obtaining numerical output gap ratios by asking firms for their own assessment of how intensively production capacity was used. Almost in all industrialized countries conducting business surveys, this question is now a topic in the questionnaires. For the euro area the data are published in quarterly intervals by the European Commission (Directorate General for Economic and Financial Affairs, DG ECFIN). The firms of the manufacturing sector taking part in the quarterly survey are asked the following question at the end of each quarter: *The utilization of our plant for the production of XY (usual full utilization=100%) amounts currently to: up to 30/40/50/60/70/80/85/90/95/100/ more than 100%, being....% (Question for Germany in Ifo Business Survey).*⁸

However, there are some problems associated with such a question on capacity utilization. One problem relates to possible sampling errors, which do not cause big concern as long as the business survey is conducted on sound statistical grounding which can be assumed in the case of the harmonized EU business surveys. A problem may be a measurement error, which may occur, when incorrect answers are given by respondents in the companies. However these measurement errors are reduced by the fact that for the analysis mainly the survey-to-survey changes are of concern. Here the panel character of the EU business surveys, i.e. asking each time with only small deviations the same companies, has a very positive effect, especially when the person answering the question is always the same.

Another and more specific problem is how individual respondents define the term "capacity" when answering this question. Firms may interpret it in the narrow sense of installed technical capacity or in a broader sense, which includes all input factors (labour and material). Similarly, respondents may confuse the concept of economic capacity with their own subjective or preferred level of capacity utilization (Kenny 1996). Lindlbauer (2000) mentions reasons why the definition of "capacity" rightly rests largely on the shoulders of the survey participant in a firm. The main reason is that this person should not be burdened with complicated calculation procedures, which possibly do not coincide with those used in a specific firm. Generally speaking, it is not possible to formulate the question in such a precise way that it is fully applicable to all companies given specific particularities in different industries or company size classes. Another problem is the interpretation of the survey results. Does an increasing rate of capacity utilization, for example, reflect a temporary cyclical movement in demand or a more permanent shift in the installed technical capacity (Kenny 1996). The increasing capacity due to the demand will change the output level and thus output gap.

Despite those potential problems there are some important advantages of survey-based capacity utilization (CU) data. Ehrmann and Ellison (2001) argue that these data are better

⁸ The results are expressed as a weighted average of the individual company answers. The open space at the end of the scale in which the respondent can fill in the exact number in case the degree of utilization exceeds 100% shows that the variable "capacity" cannot be defined as an absolute technical limit of the installed machine capacity. Instead the basis for replies is up to the firm's assessment of the "usual full capacity".

than alternative measures, which are mainly based on industrial production figures because they better capture changes in productivity and other adjustments firms may have introduced to improve the production process and thus potential production. Another advantage of survey-based capacity utilization lies in the fact that no de-trending of these data is necessary. From a practical point of view, CU data – and generally business survey results – are currently the most accessible and timely economic indicators in the euro area. Therefore, there is good reason to further test their usefulness in explaining the development of short-term interest rates.

In our analysis of monetary policy in the euro area, we formulate the following hypothesis: The CU data from surveys are an appropriate measure of the output gap and the central banks should therefore look at this qualitative indicator when making their decisions.

4 Relative Power of Survey Data ...

4.1 ... as measure of the output gap

A time series plot of the surveyed capacity utilization rate⁹ in the euro area is given in Figure 1b and 2b together with the following other measures of the output gap:¹⁰

- logarithmic difference between real GDP¹¹ and potential GDP obtained from HP filter (called then GDP gap) (see Figure 1a).
- logarithmic difference between the actual rate of unemployment and NAIRU obtained from HP filter (called then unemployment gap) (see Figure 2a).

Though the study focuses on the euro area the period covered in our research sets in already in 1991 and thus long before the start of the ECB at the beginning of 1999. This procedure is mainly due to statistical reasons (to have long enough time series to apply the econometric methods described in the next sections). However, preliminary statistical tests for different periods, particularly the periods before and after 1999, have shown that there are no structural changes in the relationships presented in this paper.

A positive GDP gap, i.e. actual production above potential growth, can be expected to place upward pressure on prices due to the growing tightness in industrial and labor markets (here negative unemployment gap).¹²

⁹ Source: DG ECFIN.

¹⁰ Our analysis is based on the period 1991:Q1-2003:Q1 for GDP, 1991:Q1-2003:Q1, for CU, 1993:Q1-2003:Q1 for unemployment in analysis.

¹¹ Gross Domestic Product in billions of chained 1995 euro (ECU until the end of 1998), seasonally adjusted. Source: Eurostat.

The cyclical profiles of the survey-based CU figures and the GDP gap data are similar. However, the extent of excess CU appears to be substantially larger than that implied by the GDP gap.¹³ The same observations have been made by Christiano (1981). He found that survey-based measures tend to indicate a higher level of output gap than those based on quantitative statistics.

There is also some evidence that the level of capacity utilization in the euro area is rising; at least it did so in the 1990s. Peaks in capacity utilization occurred in 1995/Q2, 1998/Q3, 2000/Q4 compared with 1995/Q2, 1998/Q1, 2000/Q2 in the case of the GDP gap. The CU in the euro area reached its lowest level of 77% in 1993/Q3 compared with a trough in the GDP gap of 1.23% in 1993/Q2. Figure 1b also illustrates that the CU ratio provides a reasonable coincident indicator of the euro area business cycle (changes in CU are, as a rule, in good accordance with those identified by GDP gap measures). However, the correlation between CU and the GDP gap was significantly closer in the period before 1995 than afterwards. A main reason for this phenomenon might be the fact that the service sector showed a strong variability since 1995: From 1995 until the mid 2001 we observe vigorous growth in gross value added in financial and business services (in 2000 rate of growth 5%). On the other hand, from mid 2001 till mid 2003 business services growth slowed down dramatically, much more pronounced than in the case of manufacturing (see Figure 3a). This is also demonstrated by a comparison of the EU confidence indicators in manufacturing and for business services (see Figure 3b).

Though it is too early to present a statistical proof it appears that the increased maturity of many service branches has increased the cyclical volatility of business services. Thus, the argument of the European Commission (1997) stressing the usefulness of a harmonized survey in the services sector is obvious becomes even more important, considering the size of the sector.

Many studies confirm the leading character of business survey data (see Nerb et al. 1999). Although there is no statistical evidence on the leading character of CU in relation to the GDP gap for the euro area, there is good reason to assume that the business survey data have at least an informational lead as the official GDP data are published, as a rule, with a rather long time lag and are subsequently often revised substantially. Thus, the critique of Orphanides (1997) that there are no real time data for the output gap available when setting interest rates by the central banks can be rejected.

¹² Instead of the output gap, the unemployment gap is often included as a measure of excess demand in the Phillips curve analysis. The connection between unemployment and GDP growth can be formally summarised by the statistical relationship known as "Okun's law" and relates decreases in the unemployment rate to increases in output growth. Therefore a positive output gap (actual output above potential output) indicates a negative unemployment gap (actual unemployment below the "natural rate" of unemployment).

¹³ The long-term average of the CU rate in the euro area is 82% and corresponds to 100% "usual" capacity utilization. Thus, a CU rate of 86%, for example, implies a positive output gap of 4% percentage points.

Figure 1a Actual and potential GDP

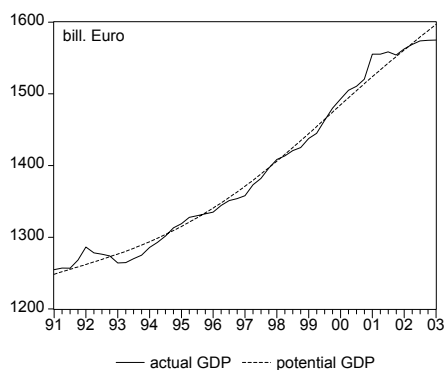


Figure 1b Capacity utilization and GDP gap

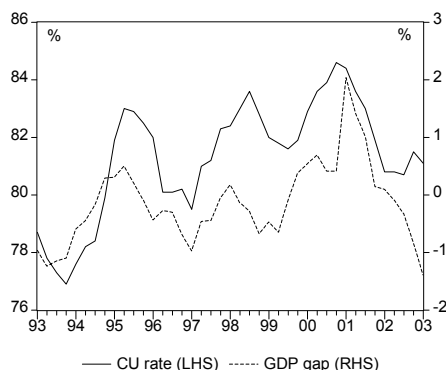


Figure 2a Actual unemployment and NAIRU

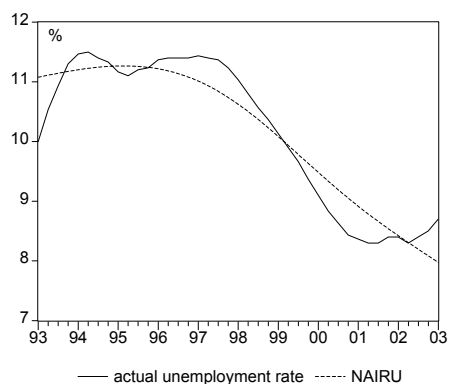
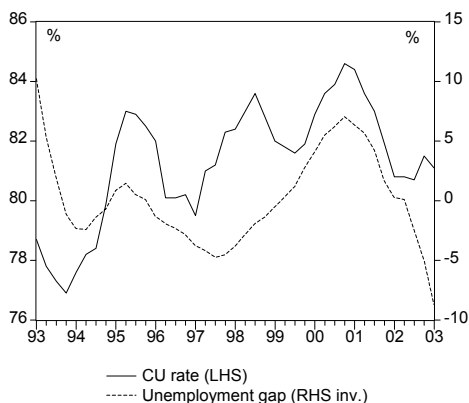
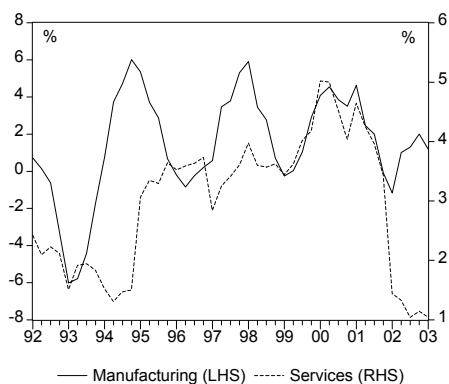
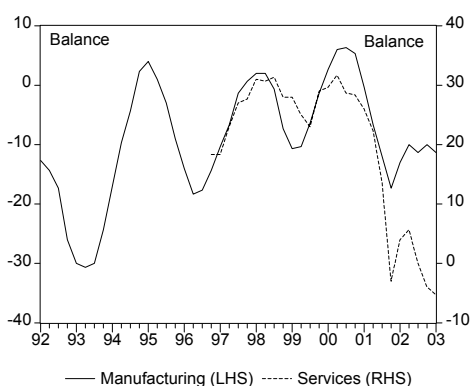


Figure 2b Capacity utilization and unemployment gap



After this analysis the question is if it would be desirable for policy to be guided by a variable like CU rate. Because of their character as real time data, yes. However this variable covers only manufacturing industry while GDP data cover all industries and additionally agriculture and services. Thus, it is a more complex measure of national economic activity. In order to make up for this deficiency we shall construct a more comprehensive measure of the capacity utilization rate (CU*) including also the service sector.

In the following section the use of CU and CU* as well as a measure for the GDP gap in Taylor rule estimation models shall be demonstrated.

Figure 3a **Gross value added (GVA) growth rates**Figure 3b **Confidence indicators**

4.2 ... in Taylor rule estimation

We estimate a reaction function of the type:

$$i_t = \alpha + \beta y_t + \gamma \pi_t + \delta(i_{t-1}) + \varepsilon_t \quad (1)$$

where the variables are:

i_t short-term nominal interest rate, defined as 3-month EURIBOR (LIBOR until the end of 1998); average of period. The interbank short-term interest rates EURIBOR and LIBOR are closely related to the monetary policy rates (and are as a rule about 10-20 basis points higher),

y_t percentage deviation of real output from trend,

π_t inflation rate (percent change in price level).

Because interest rates demonstrate serial correlation, interest rate smoothing parameter is included here.

For the measurement of output gap we use first the well-known GDP gap and CU gap, calculated as CU rate minus 82%.¹⁴ Estimating the Taylor rule, we include three measures of inflation: headline inflation (HICP), core inflation and inflation derived from deflator on consumer

¹⁴ It is the equilibrium rate of capacity, called NAICU (Non-Accelerating Inflation Rate of Capacity Utilization). Studies on the topic agreed that the NAICU was somewhere around 82% (see McElhattan (1978, 1985), Bauer (1990), Franz and Gordon (1993), and Garner (1994).

expenditure. The first measure is the official measure of inflation in the euro area. However the ECB signalled in November 2000 that in the future it would attach a larger weight than before to the core rate of inflation when specifying its interest rate decisions. Therefore the core inflation is included in the analysis as well. Inflation derived from the deflator for consumer expenditure is a standard measure of inflation proposed by Taylor. The three-month money market rate is used as measure of monetary policy. All data are extracted from the ECB database.

The following table presents the results for the traditional and "smoothed" Taylor rule, with different measures of inflation and output gap.

Table 1 Taylor rule for euro area with different measures of inflation and output gap

	Model	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
	deflator	2.10*** (0.18)	0.01 (0.13)					2.18*** (0.20)	0.22* (0.13)				
Inflation	core			2.24*** (0.15)	(-0.09) (0.13)					2.42*** (0.16)	-0.01 (0.15)		
	headline					2.56*** (0.23)	-0.01 (0.12)					2.51*** (0.22)	0.13 (0.12)
	GDP	0.00 (0.25)	0.32*** (0.08)	0.60*** (0.22)	0.32*** (4.28)	-0.18 (0.28)	0.33*** (0.08)						
Output gap	CU							0.09 (0.10)	0.12*** (0.03)	0.27*** (0.08)	0.10*** (0.03)	-0.05 (0.10)	0.10*** (0.03)
Constant		-0.12 (-0.24)	0.94 (0.16)	0.33 (0.41)	0.14 (0.14)	-0.44 (0.60)	0.12 (0.73)	-0.13 (-0.24)	-0.15 (-0.85)	0.01 (0.40)	0.00 (0.15)	-0.33 (0.57)	-0.10 (0.16)
Interest rate (t-1)			0.94*** (0.05)		0.99*** (0.05)		0.95*** (0.04)		0.91*** (0.05)		0.98*** (0.06)		0.94*** (0.04)
R-squared		0.77	0.98	0.83	0.98	0.74	0.98	0.77	0.97	0.84	0.98	0.74	0.98
Adjusted R-squared		0.76	0.97	0.82	0.98	0.72	0.98	0.76	0.97	0.93	0.98	0.72	0.97
S.E. of regression		1.19	0.38	1.12	0.37	1.40	0.38	1.17	0.39	1.10	0.41	1.40	0.39
Akaike		3.24	1.00	3.12	0.95	3.56	0.96	3.22	1.06	3.08	1.11	3.56	1.08
DW		0.47	1.00	0.24	1.46	0.34	1.39	0.47	1.23	0.20	1.23	0.32	1.18

Note: */**/** denote significance at the 10%/5%/1% level respectively, sample: 1991/Q2-2003/Q1.

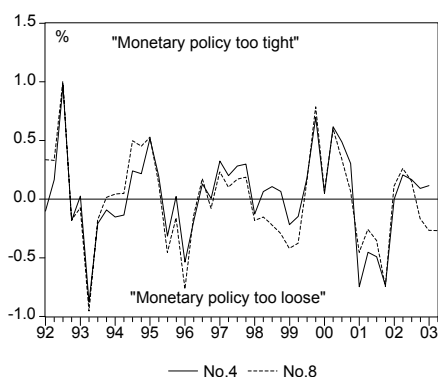
The results shown in Table 1 indicate that the inflation rate is highly significant. However, the estimated coefficients differ from the ones proposed by Taylor. The ECB appears to respond systematically in a pronounced way to inflationary developments in the economy. A counter-cyclical response to the business cycle is significantly less pronounced and appears only significant when the output gap is estimated by using core inflation or inflation measured as deflator of consumer expenditures. This is due to the fact that core inflation or GDP deflator are stronger correlated with the output gap than the headline inflation. The elasticity of interest rate with respect to the output gap is 0.32 to 0.33 for GDP

gap and 0.10 to 0.12 for CU gap. Such a difference in the magnitude of elasticity is not surprising. In the previous section we showed that the extent of excess CU appears to be substantially larger than that implied by the GDP gap.

The specification with interest rate smoothing gives a better fit than the restricted model. Such results indicate that the central bank adjusting interest rates to the current economic situation is looking at the past levels of interest rate and its dynamics. However, there appears to be still some serial correlation left, which is probably mainly due to missing variables like exchange rate, oil price and the impact of the foreign economies like the US, and some additional lags of euro area interest rates. This is the reason why we present in the next section a vector autoregressive model which takes care of those influences.

To select the best subset of interest rate predictors to assess ECB monetary policy an Akaike information criterion is adopted. Two models are selected to describe the behavior of ECB during past years (No. 4 and No. 8). The episodes when the market rate deviated from the one according to the policy rule can be called “policy mistakes” according to Taylor (1999) and the gap between these two rates would be a measure of a policy mistake.¹⁵

Figure 4 **The ECB policy mistakes (1)**



The set of results (Figure 4) suggests the ECB's policy should have been more restrictive during the crisis of 1993 and also during the 2001 crisis. However, such statements should be interpreted with some care. The models are limited to the concrete sample period and the researcher may have obtained other results, having chosen a different sample period and different variables. Because of the weakening growth prospects across most countries of the euro zone, the model with GDP gap suggests that the current monetary policy (2002/2003) is too restrictive. This view is contrary to the consensus view according to which the ECB policy in this period appears to be appropriate. Other results are given by the model

¹⁵ Defined as the difference between the actual interest rate and the one estimated in models.

with the CU gap: According to this model the current ECB monetary policy is too accommodative and higher interest rates are suggested (+30bp,+40bp higher). Such a difference in assessment of the monetary policy at the end of sample is due to the fact that the negative GDP gap is significantly more pronounced than indicated by the CU rate. This is probably mainly due to the omission of the business service sector which proved to be very sensitive to cyclical fluctuations in the recent past. For that reason we constructed a more comprehensive measure of capacity utilization (CU*).

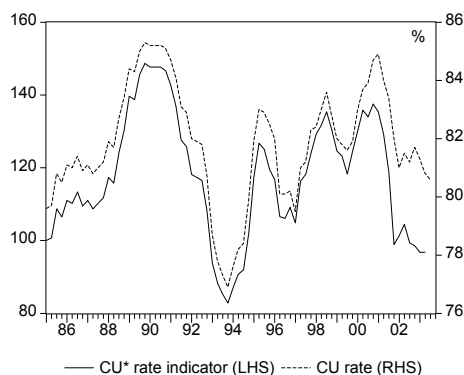
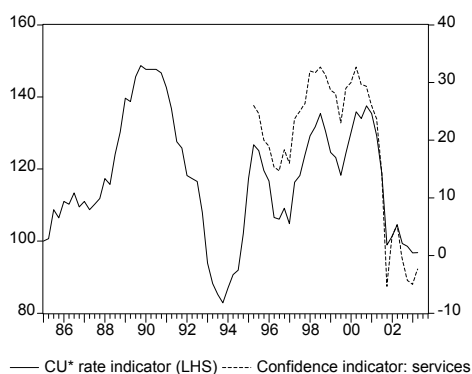
This new indicator we propose is based on CU in manufacturing industry and on the confidence indicator for services. For construction of this indicator we applied a method developed by the National Bureau of Economic Research (NBER) and adopted also by DG ECFIN in construction of the economic sentiment indicator (see Box 1).

Box 1 Steps in construction of CU* rate index

1. We calculate quarter-to-quarter changes for CU rate in manufacturing sector and quarter-to-quarter changes for confidence indicator in services.
2. Then the new series are multiplied by their weight. Because manufacturing sector has a 25% share in valued added in the euro area, for the CU transformed series a weight of 25% is used, for the transformed confidence indicator series a weight of 75% is used.
3. These standardized rates of changes are then averaged (arithmetically) so as to provide a series of composite rates of change for these both series entering into CU* rate index.
4. The above calculated series is in turn normalized by dividing it by the average and multiplying it by the average rate of change in the industrial production index (average captures 40 quarters from 1985/Q2 to 1995/Q1).
5. Finally, the index of CU* rate is established by applying the standardized rates of change to a fictitious starting value of 100.

Source: DG ECFIN, EU Harmonised Business and Consumer Surveys: User guide including new questionnaires

The next Figures 5a and 5b show our new indicator CU* rate index in comparison with both row series used for its construction.

Figure 5a **CU rate and CU* rate index**Figure 5b **Confidence indicator for services and CU* rate index**

The output gap captured by this new indicator is greater than indicated only by the CU rate. This is obvious because the movement in services has been included. Given the high share of services in valued added this new indicator appears to capture better the capacity utilization in the whole economy and should primarily be used as a benchmark for analyzing ECB monetary policy. In the next step we estimate the Taylor rule, using different measures of inflation and the new indicator CU*.

The results shown in the Table 2 indicate that the constructed indicator performs very well in the estimation being highly significant (in a similar way as the different inflation measures). As expected the specification with interest rate smoothing gives a better fit than the restricted models what is indicated also by the Akaike information criterion and higher DW. Two models are selected to describe the behaviour of the ECB during past years, the model No. 14 and No. 18.

As we see in the Figure 6 these models justify the consensus view assessing the ECB policy in the period 2002/2003 as almost appropriate. The policy mistake is very low, near to zero.

The analysis above provides evidence that different output gap indicators may lead to different conclusions about the stance of monetary policy. To analyze more deeply monetary policy in the euro area a multivariate model is needed. Thus, in the next section we create a VAR¹⁶ model for short-term interest rates in the euro area.

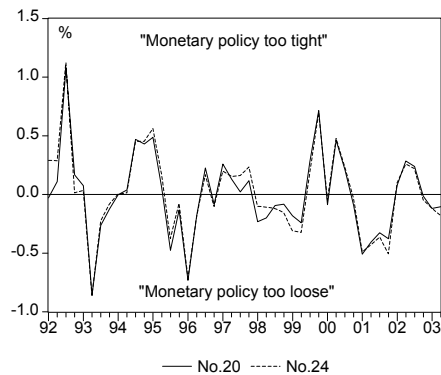
¹⁶ The VAR methodology is described detailed in Hamilton (1994), Johnston and DiNardo (1997), Chatfield (2001).

Table 2 Taylor rule for euro area with CU* rate index

Model	No.13	No.14	No.15	No.16	No.17	No.18
deflator	2.37*** (0.19)	0.31*** (0.12)				
Inflation						
core			2.57*** (0.13)	0.19 (0.16)		
headline					2.56*** (0.22)	0.20*** (0.11)
Output gap						
CU* rate index	0.03*** (0.01)	0.02*** (0.00)	0.06*** (0.00)	0.02*** (0.00)	0.01*** (0.01)	0.01*** (0.00)
Constant	-4.42*** (1.68)	-2.48*** (0.54)	-6.87*** (1.15)	-2.09 *** (0.60)	-2.17*** (1.62)	-1.87*** (0.45)
Interest rate (t-1)		0.88*** (0.04)		0.91*** (0.06)		0.92*** (0.03)
R-squared	0.79	0.98	0.88	0.97	0.73	0.98
Adjusted R-squared	0.77	0.97	0.88	0.97	0.72	0.97
S.E. of regression	1.14	0.36	0.91	0.38	1.41	0.37
Akaike	3.17	0.89	2.72	1.00	3.52	0.97
DW	0.53	1.30	0.27	1.14	0.29	1.19

Note: *** denote significance at the 1% level, sample: 1991/Q2-2003/Q2.

Figure 6 The ECB policy mistakes (2)



5 Structural Vector Autoregressive Model (SVAR)

5.1 Specification and restrictions

Our benchmark VAR will include a vector of euro area variables¹⁷ and a vector of foreign variables¹⁸ including variables representing the world and the United States¹⁹ to capture the international transmission of U.S. monetary policy shocks.²⁰ Table 3 outlines the variables used in the SVAR estimation along with the abbreviation for each variable.

Firstly, the user needs to determine whether the variables to be included are stationary $I(0)$ or non-stationary $I(1)$. Since not all variables seem to be stationary we estimate first differences.²¹ The next step involves estimating the reduced form VAR using OLS, for the period 1991/Q1–2003/Q1, ensuring that enough lags are incorporated to avoid serial correlation of the residuals. We estimate three VARs: The first one with the industrial capacity utilization rate in the euro area as a proxy for the output gap, the second one with the GDP gap as a proxy for the output gap. The last one includes our new indicator, CU^* rate. In selecting the order of the VAR we rely on a criterion widely used by other researchers, namely a likelihood ratio (LR) test. For the system we chose four lags, the maximum number of lags we can handle (degrees of freedom). Because matrices of estimated VAR coefficients

¹⁷ Definitions and sources of the data as in previous section.

¹⁸ Peersman (2001) created a model for the euro area, including foreign and domestic variables which will be the basis of our analysis.

¹⁹ Given its sheer size and weight in international trade, the US economy is at the centre of global trade flows. If the U.S. were to experience a "hard landing" or severe slowdown in its domestic economy, a flow-on effect would spread to other stakeholding countries through reduced demand for overseas exports. Because of strong trade-connections with the U.S. the euro area economy can be significantly affected by developments there.

²⁰ The analytical tools for studying the international transmission of monetary shocks were developed in the 1960s by Robert Mundell (1962) and Marcus Flemming (1962). The two-country version of the Mundell-Flemming (FM) model predicts that a monetary expansion in one country will produce an increase in home output and a negative output response in the other country. The transmission mechanism is the trade balance, thus the trade links are very important: the home country will lower its interest rate, depreciate the exchange rate and gain competitiveness, thus producing an expenditure switching effect from foreign goods to home goods, resulting in an improvement of the home country trade balance, an increase in its output and a decrease in foreign output, the so-called beggar-thy-neighbour effect. On the financial side of the model, perfect capital mobility requires interest rates to equal (in the version with static expectations), hence foreign interest rate will decrease and this should make foreign output grow (see Obstfeld and Rogoff 1996).

²¹ Estimating a VAR in differences has some advantages. The system is robust and better forecasts will result if there are structural breaks. If forecasts of changes are wanted, VAR in differences will be more accurate than VAR in levels, even when the true model is an Error Correction Model (ECM) (Clements and Hendry, (1995), Monte Carlo evidence). If the variables have characteristic roots close to one, imposing unit roots will improve forecasts (Lin and Tsay 1996, Diebold and Kilian 2000). There are also disadvantages: If the variables are already stationary, differencing induces a moving average term into the equation. In real data, when there is evidence of one or more co-integrating vectors, ECM gives better forecasts.

are difficult to interpret directly and not important, they are not reported here. Estimated VARs are commonly summarized by certain functions of these matrices like impulse response functions and forecast error decompositions (Stock and Watson 2001). To interpret the impulse response function we have to impose some restrictions (ordering, Cholesky decomposition) to the VAR model because many different structural models have the same reduced form. We impose several restrictions according to economic theory (see Table 4).

- According to the Taylor rule, policymakers have contemporaneous information about non-policy variables (implying that the policy variables should be ordered last in the VAR) (see Bernanke and Blinder 1992, Christiano and Eichenbaum 1992).²²
- Oil prices are self determined depending only on an AR process in this model and should capture the future course of inflation (see Bernanke and Michov 1998, Christiano et al. 1996).
- The exchange rate is modeled to include all available information in the system, so the euro rate equation includes all US and EMU variables plus oil prices. This reflects the fact that the exchange rate is a financial variable and reacts quickly to all information (Cushman and Zha 1997).
- The U.S. literature often models inflation using a closed economy approach (Gordon 1998). Here U.S. inflation and EMU inflation are modelled as dependent on the contemporaneous and lagged impact of the oil prices and domestic output and self-lagged values.
- The euro area is modelled as an open economy. The output equations in each economy contain lagged values of domestic output and the contemporaneous and lagged commodity price data. The output equation in the EMU also includes the influences of U.S. output.

The imposed restrictions imply a system of dynamic responses²³ presented in the next section.

²² Sims (1980) argued that the policy instrument could be ordered first ahead of output and prices since contemporaneous measures of output and prices were not available to the monetary authorities when making decisions regarding the appropriate rate for the monetary policy variable.

²³ The quantitative results are available; here we provide only the interpretation and graphical illustration of results.

Table 3 **Key variables in the two-economies QVAR system**

Variables	World	US	EMU
oil prices	C_t ^{a)}		
output gap		Y_t^{US} ^{b)}	Y_t^{EMU}
inflation		Π_t^{US} ^{c)}	Π_t^{EMU} ^{d)}
interest rate		R_t^{US} ^{e)}	R_t^{EMU} ^{f)}
exchange rate			E_t ^{g)}

a) Cushing, OK WTI Spot Price FOB (\$/bbl), average of period. Source: U.S. Department of Energy, Energy Information Administration. Time series transformation: natural logarithm.

b) Gross Domestic Product in billions of chained 1996 USD, seasonally adjusted. Source: BEA. Time series transformation: first logarithmic difference of raw series and HP filter.

c) Consumer Price Index, year-on-year changes in %, Source: Eurostat.

d) Harmonized Consumer Price Index, year-on-year changes in %, Source: Eurostat.

e) Federal Funds Rate, average of period. Source: Federal Reserve Board of Governors.

f) Defined as 3month EURIBOR rate (LIBOR until the end of 1998), average of period, percent. Source: ECB.

g) Exchange rate euro per 1\$, average of period. Source: Eurostat. Time series transformation: natural logarithm.

Table 4 **Contemporaneous structure of the system**

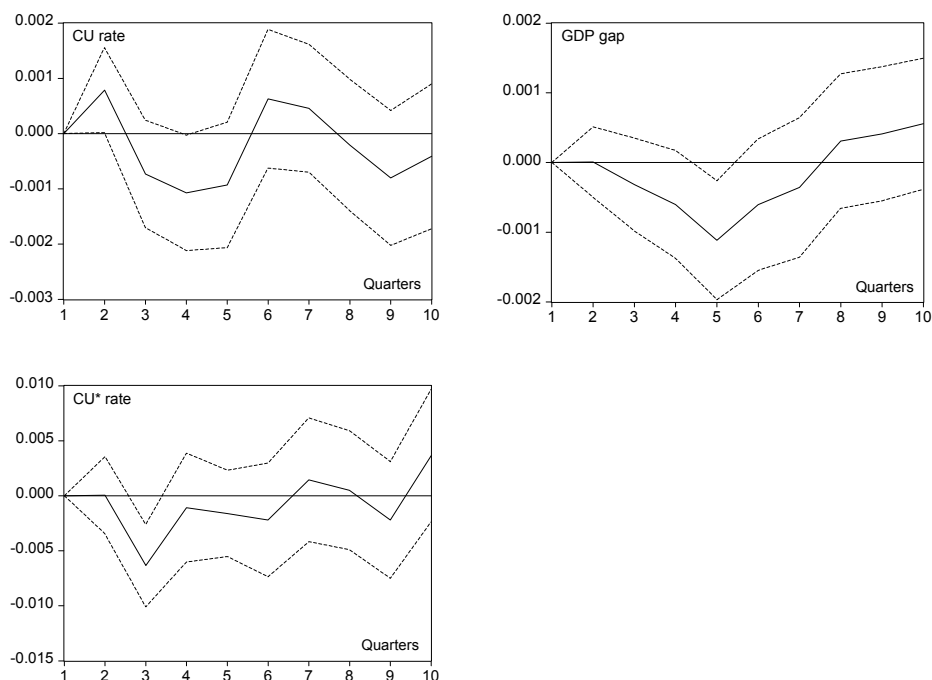
	World		US		EMU			
	C_t	Y_t^{US}	Π_t^{US}	R_t^{US}	Y_t^{EMU}	Π_t^{EMU}	R_t^{EMU}	E_t
C_t	*							
Y_t^{US}	*	*						
Π_t^{US}	*	*	*					
R_t^{US}		*	*	*				
Y_t^{EMU}	*	*			*			
Π_t^{EMU}	*				*	*		
R_t^{EMU}					*	*	*	
E_t	*	*	*	*	*	*	*	*

Note: The rows show the dependent variable in each equation, while the columns show the variables appearing in it. For instance, inflation in the U.S. is dependent on the current and lagged output gap in the U.S., current and lagged oil prices and lagged own value. The ordering of the variables is in accordance with the Cholesky decomposition.

5.2 The dynamic responses of the system

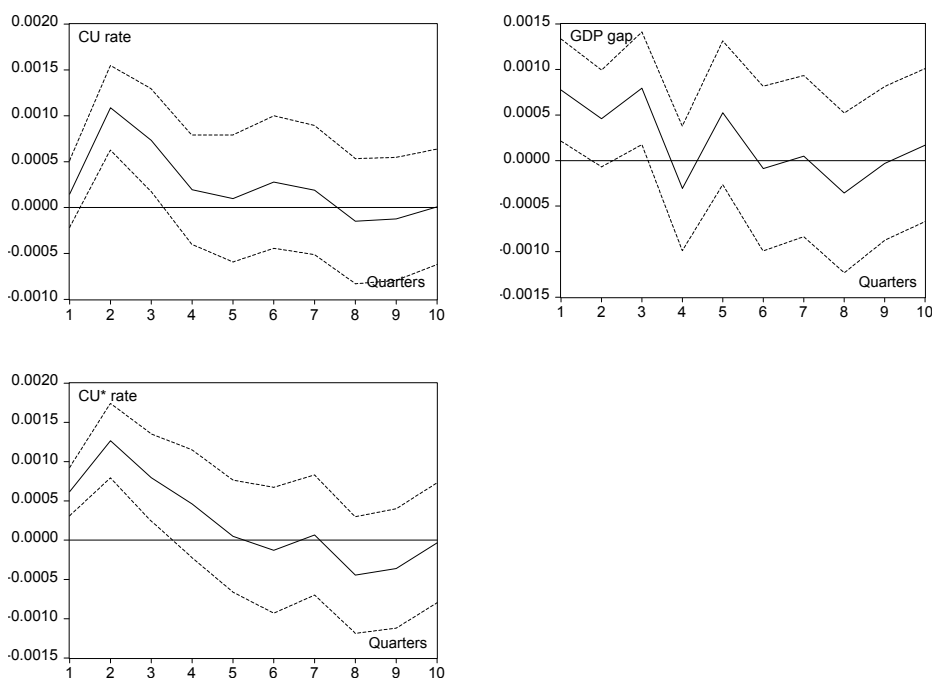
The ECB reacts in a different way to the shocks in inflation changes and output gap growth as the FED: shocks in inflation changes seem to not influence monetary policy in the U.S. whereas they do in the euro area in the time period studied. Moreover, the U.S. national transmission mechanism differs from the one in the euro area with regard to responses to the magnitude of the output gap. Regarding the reaction of an output gap (measured by GDP gap and the industrial capacity utilization rate) in the euro area to monetary policy shocks the following conclusions can be drawn: The adjustment to monetary policy shocks is quicker for an output gap measured by survey data than for the one based on traditional statistics. However, the reaction of the capacity utilization rate at the beginning of an interest rate shock is different, as expected: Instead of a negative impact a positive one is observed. Thus, it appears that survey respondents do not view a first change of central bank interest rates as the beginning of a new trend of monetary policy. After one quarter, when respondents are sure that the policy action is compatible with other fundamental factors determining the outlook for interest rates, they reduce production activity.

Figure 7 Monetary policy shocks in the euro area and reaction of the output gap



This shows that firms in the euro area adjust their production quicker and the reaction is stronger than indicated by the GDP gap based on quantitative data. Firms tend to cut production in order to run down inventories and to be so prepared for the expected lower demand in the future. The reaction of CU* rate is like the reaction of the GDP gap and CU rate on average. In case of CU* the negative impact of monetary policy appears to be immediate. This may be due to the fact that the business and financial services are most sensitive to monetary policy changes and are predominant in the euro area service confidence indicator (Figure 7).²⁴

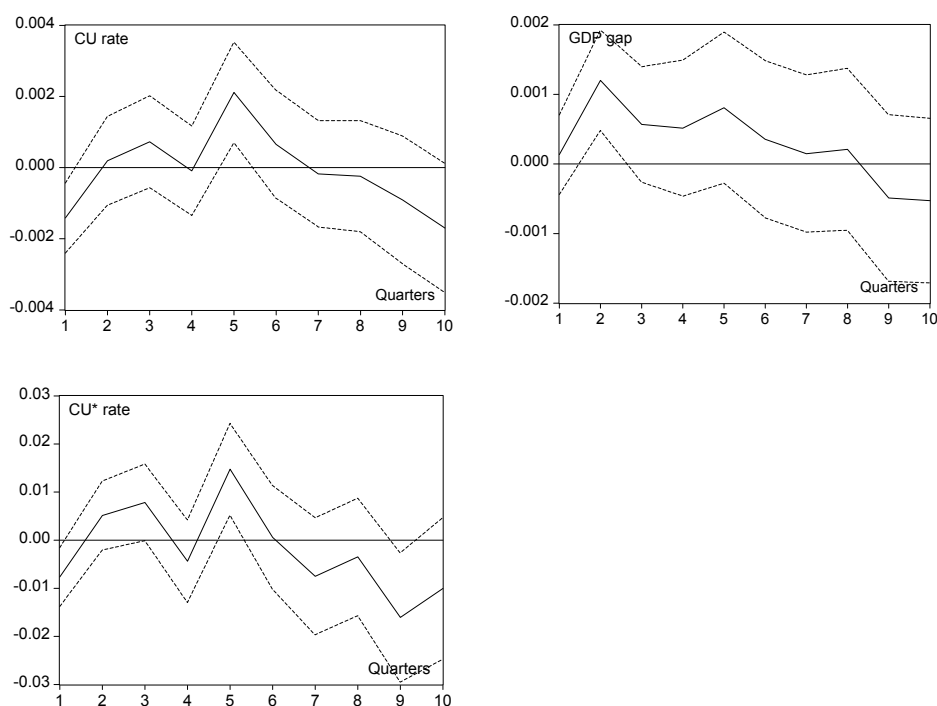
Figure 8 Shocks to different measures of the output gap in the euro area and the reaction of monetary policy



²⁴ The impulse-response functions represent dynamic responses of each variable included in the VAR model when one standard deviation structural shock occurs. The response is graphed over 10 quarters. In addition, the asymptotic standard errors are computed and the impulse responses are graphed with 95% confidence bounds. The upper and lower bounds contain impulse-response values with a probability of 95%. For example, in case of a 1%-point increase of the interest rate the output gap measured by CU* falls with a probability of 95% at most by 1% in the 3rd quarter

Our studies in Section 4.1 indicated that the CU rate fluctuates more strongly than the GDP output gap. Thus, the survey-based indicators of the output gap show greater reactions of monetary policy to this measure of output gap than to the output gap measured by traditional statistics. The more comprehensive measure of the output gap CU* shows a significantly higher impact (twice as high) on interest rates than CU which is restricted to manufacturing industry. This is understandable given the high share of services in total value added in the euro area (Figure 8).

Figure 9 **The U.S. real economy shocks and reaction of the output gap in the euro area**

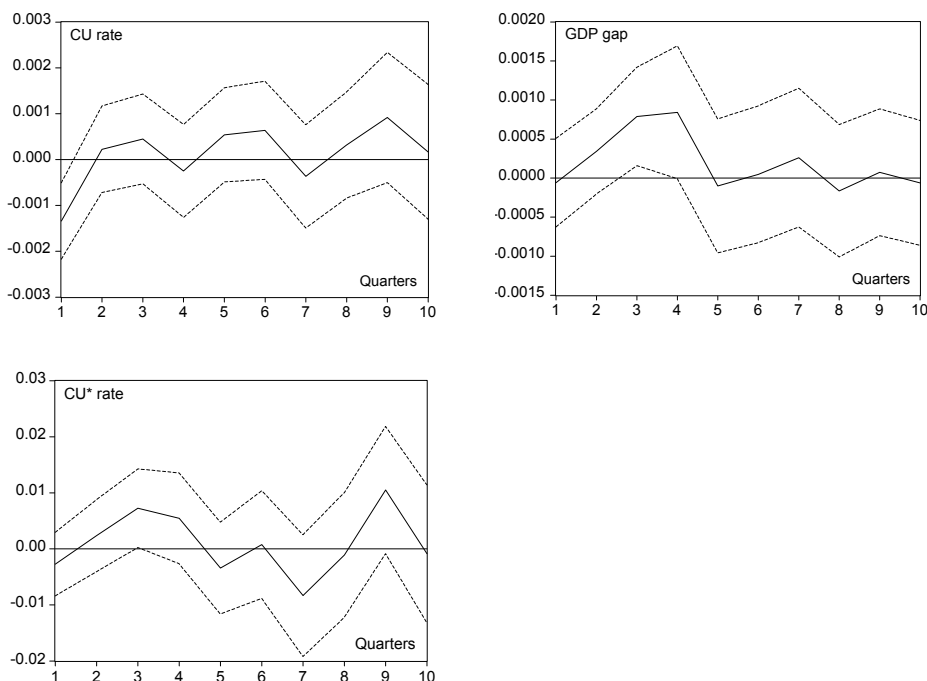


The model provides evidence on the so-called beggar-thy-neighbour effect, which is reflected in negative response of the euro area output gap growth to positive shocks in the U.S. real economy. This result gives evidence for international transmission mechanism through the trade balance. A domestic monetary expansion leads to a worsening of the foreign trade balance through the expenditure switching effect and a decrease in foreign output (beggar-thy-neighbour-policy). The model also comprises the expenditure switching

effect, so that the foreign output may decrease. These findings are consistent with our assumptions and support our specifications of the VAR model. However, the reaction of the euro area output gap is stronger when measured by survey data than by traditional statistics. There is also a difference in the response of the euro area figures for CU on the one and CU* on the other side to real economy shocks in the US. The CU* rate including manufacturing and services is more sensible to international business cycle fluctuations than the CU rate alone (see Figure 9).

The reaction function in Figure 10 reflects the typical anticipations of survey respondents in the euro area with regard to changes in U.S. monetary policy. If the FED raises interest rates, the respondents anticipate appreciation of the U.S. dollar through inflow of foreign capital to the U.S. Thus, they see chances to increase their own production because their exports to the United States become more competitive and the opposite holds true for U.S. exports. The Europeans can increase the capacity utilization of their plants; in the case of service sector this is even more easy than in the case of manufacturing. Thus, the reaction in the CU* is more pronounced than in the case of CU.

Figure 10 **US monetary policy shocks and reaction of the output gap in the euro area**



5.3 Variance decomposition

Variance decomposition of interest rates in the euro area is also of interest of this work, which looks into the role of capacity utilization for explaining and forecasting ECB interest rate decisions. The variance decomposition of the change in interest rates is the percentage of the within-sample forecast error that is accounted for by the explanatory variables at various time horizons. If the forecast horizon is large and the time series vector is stationary, then the forecast error variance approximates the actual variance of the series. If a series is truly exogenous, then all of its forecast variance will be explained by its own shocks. If a large fraction of the forecast variance of the change in interest rates is explained by the shocks to the capacity utilization rate, then this variable is important in explaining volatility of interest rates. The variance decomposition of the change in interest rates can therefore add important information about the relative influence of the output gap.

The decomposition is shown in Table 5 for a forecast horizon of 20 quarters for three models with GDP gap, CU rate and CU* rate and the results can be summarized as follows: the CU rate and CU* rate account for more variance in interest rates, about 20%, while GDP gap accounts only for 4%. This can be attributed to the greater variance of CU rate and CU* rate than output gap that was stated in previous sections.

Table 5 **Variance decomposition of the interest rate in the euro area**

Measure of output gap in the euro area	Oil price	Output gap	Inflation	Interest rate	Output gap	Inflation	Interest rate	Exchange rate
			US				EMU	
GDP Gap	21.36	31.69	11.37	12.48	4.23	2.64	14.05	2.18
CU Rate	14.90	10.55	15.10	8.61	18.49	15.45	7.59	9.29
CU* Rate	26.18	21.25	8.51	7.10	17.04	5.11	4.93	9.88

Ordering: like variables in columns.

The table shows the contribution of each variable (first row) to the variation of the short-term interest rate in the model with GDP gap (third row), respectively in the model with CU rate (fourth row) and the one with CU* rate (fifth row). The total variation is assumed to be 100 percent.

The results give evidence that the output gap measured by survey data is an important factor in interest rates forecasting and thus should be regarded carefully. If the survey data are a good approximation of the output gap, it means the forecast of the ECB interest rate can be improved.

5.4 Forecast

The main interest for the policy makers is not in the historic behavior of interest rate but in the forecast. We generate basic VAR forecast for the short- term interest rate in the euro area without prior assumptions about the likely future paths of one or more of the variables. Therefore our forecast is unconditional. Forecasts are useful both in their own right and to cross-check forecasts produced by traditional macroeconomic models. However, as with other techniques, VAR forecast have several weaknesses. Firstly, they are susceptible to the Lucas critique and more generally are likely to perform poorly following structural breaks (Clements and Hendry 1995). Secondly, the inclusion of too many lags or too many variables can lead to VAR- derived forecasts performing poorly, even though the model fits the data well in the estimation period. This occurs if the lags of variables in the system pick up non-systematic relationship between the variables (Doan et al. 1984).

Despite the fact that the forecast from VAR model may be inconsistent and may tend to random values (Philips 1998) we would like to present the dynamic forecast from our models with different measures of output gap.

The results obtained are interesting and each one of them reflects the nature of the chosen output gap. Because the CU* rate and GDP gap show a more pronounced deterioration in the latest cyclical downturn than the CU rate the first two models suggest a lower interest rate at the end of 2003 than the model based on the CU rate in manufacturing industry (Table 6).

Table 6 **Dynamic forecast for the interest rate in the euro area from SVAR models**

	Actual	Model		
		with CU rate	with GDP gap	with CU* rate
2003/Q1	2.68			
2003/Q2	2.23			
2003/Q3	*	3.1	2.0	1.8
2003/Q4	*	2.5	1.5	1.8

6 Summary

In our paper we propose a survey based measure of output gap for interest rate modelling. Our analysis provides evidence that the capacity utilization in manufacturing (CU) rate may be used as a proxy for the output gap. The cyclical profiles of the survey-based CU figures and the GDP gap data are rather similar. However, the extent of excess CU appears to be substantially larger than implied by the GDP gap. A main advantage of this approach is

the fact that business survey data have at least an informational lead as the official GDP data are published as a rule with a rather long time lag and are subsequently often revised substantially.

However, as experience in the last years has shown that the cyclical variability of the service sector has increased, it is proposed to use in addition to CU a confidence indicator for business services as a proxy for the not existing measure of capacity utilization in this sector. Given the high share of services in valued added this new indicator appears to capture better the capacity utilization in the whole economy and should primarily be used as a benchmark for analyzing ECB monetary policy.

According to the Taylor rule model with CU the interest rate should have been significantly higher at the end of 2002 than implied by a model with a GDP gap. A model with our new indicator CU* rate assesses the current ECB policy as almost fully appropriate.

Given the model with the new indicator CU*, which appears to be superior to the one with the industrial CU rate alone, we expect the ECB policy rate to continue to decline slightly in late autumn 2003 (to about 1.75) and stay at that low level till early next year before starting to pick up in mid 2004.

References

- Bansak, C., Morin, N. and Starr, M. (2003): Technology, Capital Spending and Capacity Utilization, Department of Economics, Center for Public Economics San Diego State University Discussion Paper 03-06.
- Bauer, P.W. (1990): A Reexamination of the Relationship Between Capacity Utilization and Inflation, *Economic Review*, 26, 2, 2-12.
- Bernanke, B.S. and Blinder, A. (1992): The Federal Funds Rate and the Channels of Monetary Transmission, *American Economic Review*, 82, 4, 901-921.
- Bernanke, B.S. and Mihov, I. (1998): The Liquidity Effect and Long- Run Neutrality, NBER Working Paper 6608.
- Bernanke, B.S. and Woodford, M. (1997): Inflation Forecasts and Monetary Policy, *Journal of Money, Credit and Banking*, 29, 4, 653-684.
- Chagny, O. and Döpke, J. (2001): Measures of the Output Gap in the Euro-Zone: An Empirical Assessment of Selected Methods, *Vierteljahrshefte zur Wirtschaftsforschung*, 70, 3, 310-330.
- Chatfield, Ch. (2001): Time-Series Forecasting, London.
- Christiano, L. (1981): A Survey of Measures of Capacity Utilization, IMF Staff Papers, 28, 144-198.
- Christiano, L. and Eichenbaum, M. (1992): Identification and the Liquidity Effect of a Monetary Policy Shock, [in:] Political Economy, Growth and Business Cycles, eds. A. Cukierman, L. Hercowitz and L. Leiderman, Cambridge, 335-370.
- Christiano, L.J., Eichenbaum, M., and Evans, C.L. (1996): The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds, *Review of Economics and Statistics*, 78, 1, 16-34.
- Christiano, L.J., Eichenbaum, M. and Evans, C.L. (1998): Monetary Policy Shocks and their Consequences: Theory and Evidence, NBER Working Paper 6400.

- Clarida, R. and Gertler, M. (1996): How the Bundesbank Conducts Monetary Policy, [in:] *Reducing Inflation: Motivation and Strategy*, Chicago, 363-406.
- Clarida, R., Gali, J. and Gertler, M. (1997): Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Economic Research Reports*, 86, 14.
- Clements, M. and Hendry, D. (1995): *A Theory of Economic Forecasting*, Cambridge.
- Cushman, D.O. and Zaha, T.A. (1997): Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates, *Journal of Monetary Economics*, 39, 3, 433-448.
- Diebold, F.X. and Kilian, L. (2000): Unit Root Tests Are Useful for Selecting Forecasting Models, *Journal of Business & Economic Statistics*, 18, 3, 265-273.
- Doan, T., Litterman, R. and Sims, C.A. (1984): Forecasting and Conditional Projection Using Realistic Prior Distributions, *Econometric Reviews*, 3, 1-100.
- Economic Policy Committee (EPC) (2001): Report on Potential Output and the Output Gap, Brussels.
- Eijffinger, S. and de Haan, J. (2000): *European Monetary and Fiscal Policy*, Oxford.
- Ehrmann, M. and Ellison, M. (2001): The Changing Response of US Industry to Monetary Policy, mimeo.
- European Commission (1997): The Joint Harmonised EU Programme of Business and Consumer Surveys, *European Economy* No. 6.
- Fabiani, S. and Mestre, R. (2000): Alternative Measures of the NAIRU in the Euro Area: Estimates and Assessment, *European Central Bank Working Paper Series* 17.
- Flemming, J.M. (1962): Domestic Financial Policies under Fixed and Floating Exchange Rates, *IMF Staff Papers*, 9, 369-379.
- Franz, W. and Gordon, R.J. (1993): German and American Wage and Price Dynamics, *European Economic Review*, 37, 4, 719-62.
- Friedman, M. (1968): The Role of Monetary Policy, *American Economic Review*, 58, 1-21.
- Garner, C.A. (1994): Capacity Utilization and U.S. Inflation, *Economic Review*, 79, 4, 1-21.
- Gerlach, S. and Svensson, L.E.O (2001): Money and Inflation in the Euro Area: A Case for Monetary Indicators? BIS Working Paper 98.
- Gordon, R.J (1998): Foundations of the Goldlocks Economy: Supply Shocks and the Time Varying NAIRU, *Brookings Papers on Economic Activity*, 2, 297-333.
- Hamilton, J.D. (1994): *Time Series Analysis*, Princeton NJ.
- Hodrick, R. J. and Prescott, E.C. (1997): Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, 29, 1, 1-16.
- Isard, P. and Laxton, D. (2000): Inflation-Forecast Targeting and the Role of Macroeconomic Models, [in:] IMF and Czech National Bank, *Inflation Targeting in Transition Economies: The Case of the Czech Republic*, eds. Warren Coats Monetary and Exchange Affairs Department of the International Monetary Fund, Washington (DC) and Czech National Bank (Prague).
- Johnston, J. and DiNardo, J. (1997): *Econometric Methods*, New York.
- Judd, J.P. and Rudebusch, G.D. (1998): Taylor's Rule and The Fed, 1970-1997, *Economic Review*, 3, 3-16.
- Kenessey, Z.E. (1990): Industrial Production and Capacity Utilization, [in:] *Macroeconomic Data: A User's Guide*, Federal Reserve Bank of Richmond, 14-16.
- Kenny, G. (1996): Capacity Utilization in Irish Manufacturing, Technical Paper 96/3, Central Bank of Ireland.
- Lin, J-L. and Tsay, R.S. (1996): Co-integration Constraint and Forecasting: an Empirical Examination, *Journal of Applied Econometrics*, 11, 519-538.
- Lindlbauer, J.D. (2000): Capacities and Capacity Utilisation from the Ifo Business Survey - Knowledge and Background, [in:] *Use of Survey Data for Industry, Research and Economic Policy*, eds. K.H. Oppenländer, G. Poser, B. Schips, Aldershot.
- Lipsey, R.G. (1960): The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom 1862-1957: A Further Analysis, *Economica*, 27, 1-31.
- McCallum, B.T. (1998): Issues in the Design of Monetary Policy Rules, [in:] *Handbook of Macroeconomics*, eds. J. B. Taylor and M. Woodford, Amsterdam.

- McCallum, B.T. and Nelson, E. (1998): Performance of Operational Policy Rules in an Estimated Semi-Classical Structural Model, National Bureau of Economic Research Working Paper 6599.
- McElhattan, R. (1978): Estimating a Stable-Inflation Capacity-Utilization Rate, Federal Reserve Bank of San Francisco, Economic Review, 20-30.
- McElhattan, R. (1985): Inflation, Supply Shocks and the Stable-Inflation Rate of Capacity Utilization, Federal Reserve Bank of San Francisco, Economic Review, 45-63.
- Mundell, R. (1962): The Appropriate Use of Monetary and Fiscal Policy for Internal and External Stability, IMF Staff Papers, 9, 70-79.
- Nerb, G., Englmeier, K., Goldrian, G., Lehne, B., Lindlbauer, J., Nierhaus, W. and Stock, K. (1999): Creating a Leading Indicators and Financial Warning System for the Turkish Economy, Republic of Turkey, Prime Ministry, Undersecretariat of Treasury.
- Obstfeld, M. and Rogoff, K. (1996): Exchange Rate Dynamics Redux, *Journal of Political Economy*, 103, 3, 624-660.
- Okun, A. (1962): Potential GNP: Its Measurement and Significance, Proceedings of the Business and Economics Section, American Statistical Association, 98-104.
- Orphanides, A. (1997): Monetary Policy Rules based on Real-Time Data, Federal Reserve Board Working Paper 1998-3.
- Orphanides, A. and Wilcox, D.W. (1996): The Opportunistic Approach to Disinflation, Board of Governors of the Federal Reserve System, Finance and economics discussion series; 1997-36.
- Peersman, G. (2001): The Transmission of Monetary Policy in the Euro Area: Implications for the European Central Bank, diss., Universitaet Gent.
- Phelps, E. (1967): Money Wage Dynamics and Labour Market Equilibrium, *Journal of Political Economy*, 75, 678-711.
- Phillips, A.W. (1958): The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom 1861-1957, *Economica*, 25, 283-99.
- Phillips, P.C.B. (1998): Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VAR's, *Journal of Econometrics*, 83, 21-56.
- Sims, C.A. (1980): Macroeconomics and Reality, *Econometrica*, 48, 1-49.
- Stock, J.H. and Watson M.W. (2001): Vector Autoregressions, *Journal of Economic Perspectives*, 15, 4, 101-115.
- Svensson, L. (2003): Monetary Policy and Real Stabilization, NBER Working Paper 9486.
- Taylor, J.B. (1993): Discretion Versus Policy Rules in Practice, *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.
- Taylor, J.B. (1999): A Historical Analysis of Monetary Policy Rules, [in:] Monetary Policy Rules, eds. J.B. Taylor, Chicago.

Quantification of Qualitative Survey Data and Tests of Consistent Expectations: A New Likelihood Approach

Christian M. Dahl and Lin Xia*

Abstract

In this paper, we develop a likelihood approach for quantification of qualitative survey data on expectations and perceptions and we propose a new test for expectation consistency (unbiasedness). Our quantification scheme differs from existing methods primarily by using prior information (perhaps derived from economic theory or well established empirical relations) on the underlying process driving the variable of interest. To investigate the properties of our novel quantification scheme and to analyze the size and power properties of the new expectation consistency test, we perform Monte Carlo simulation studies. Overall, the simulation results are very encouraging and show that efficiency gains from including prior information can be substantial relative to existing quantification schemes. Finally, we provide an empirical illustration. We argue that a nonlinear regime switching model, historically, provides a more adequate characterization of price changes in the British manufacturing industry relative to a wide range of alternative linear representations. We then show how to use this information to quantify the qualitative survey data on price change expectations and perceptions from the CBI survey in the U.K. manufacturing industry. Our findings support the existence of so-called ex post bias, i.e., that rationally formed forecasts might appear biased when compared to ex post realizations. As a result, tests of expectation consistency by comparing forecasts with realizations become invalid. The new test we propose, however, is not affected by the existence of ex post bias and by applying this statistic we cannot reject the expectation consistency hypothesis.

Key Words: Qualitative Survey Data; Quantification Schemes; Test of Expectation Formation Hypothesis

JEL Classification: C33; C35; C42; C51

* Department of Economics, Krannert School of Management, Purdue University,
dahlc@mgmt.purdue.edu; xia@mgmt.purdue.edu

Comments and suggestions from John A. Carlson and an anonymous referee are gratefully acknowledged. The notation used follows the "new" standard proposed by Abadir and Magnus (2002). The software developed for this paper is available from the authors upon request.

Résumé

Nous exposons dans cet article une approche par les probabilités pour la quantification des données d'enquête qualitatives concernant les anticipations et les perceptions, et nous proposons un nouveau test de cohérence des anticipations (absence de biais). Notre méthode de quantification diffère des méthodes actuelles essentiellement en ce qu'elle utilise les informations préalables (issues éventuellement de la théorie économique ou des relations empiriques bien établies) sur le processus régissant la variable qui nous intéresse. Pour étudier les propriétés de notre nouvelle méthode de quantification et pour analyser les propriétés de taille et de puissance du nouveau test de cohérence des anticipations, nous procédons à des simulations de type Monte Carlo. Au total, les résultats des simulations sont très encourageants et les gains d'efficacité obtenus en prenant en compte les informations préalables peuvent être importants par rapport aux méthodes actuelles de quantification. Enfin, nous donnons une illustration empirique. Nous faisons valoir qu'un modèle non linéaire à changements de régime caractérise mieux historiquement l'évolution des prix du secteur manufacturier du Royaume-Uni qu'un large éventail d'autres modèles linéaires. Nous montrons ensuite comment utiliser ces informations pour quantifier les données qualitatives sur les anticipations et perceptions d'évolution des prix recueillies via l'enquête CBI, qui porte sur le secteur manufacturier du Royaume-Uni. Nos travaux confirment l'existence d'un biais dit « ex post », qui consiste en ce que les prévisions formulées rationnellement peuvent apparaître biaisées lorsqu'on les compare aux réalisations ex post. En conséquence, les tests de cohérence des anticipations reposant sur une comparaison des prévisions avec les réalisations perdent toute validité. Le nouveau test que nous proposons n'est pas entaché d'un biais ex post et son application ne permet pas de rejeter l'hypothèse de cohérence des anticipations.

1 Introduction

The analysis on how economic agents form their expectations on economic variables has been treated as one crucial issue in explaining many important economic phenomena. Although there seems to be no consensus among economists as to whether the true expectation formations are rational, adaptive or mixed, the rational expectations hypothesis has been more and more accepted by macroeconomists and is playing a central role in contemporary macroeconomic models. Both new classical economists and later new Keynesian proponents take the rational expectations hypothesis as one of the key assumptions in their respective modeling frameworks. While the rational expectations hypothesis has been applied to numerous economic models, its validity is nevertheless very difficult to verify empirically because tests on the hypothesis and further analysis on the expectation formation process require direct observations on such expectations, which are rarely available.

Many empirical researchers in the past have tried to construct proxy measures for such subjective expectations. Such proxies are usually susceptible to the underlying economic theories, and the empirical tests of any expectations formation theories are in fact joint tests of the economic theory under consideration as well as the validity of the proxies constructed. Subsequently, rejection of the tests might result from the inappropriate choices of proxies, rather than from the falsehood of the hypotheses. Obviously, this potentially possibility of Type I errors is the weakness of the methodology. In addition, procedures using proxies typically suffer from generated regressors problems and measurement errors, which can have unfortunate effects if not properly accounted for.¹ In short, testing the hypothesis of expectation formation using proxies is not without drawbacks.

Alternative approaches adopted by many researchers to circumvent these problems have been to obtain direct measures of expectations from expectation survey data. With the availability of survey evidence on economic agents' expectations, it is possible to conduct direct tests on the rational expectations hypothesis without being concerned with the problems mentioned above. However, in most situations information on expectations obtained from surveys is qualitative instead of quantitative. Survey respondents are only questioned about the expected directions of, say, changes of the variable in question, e.g., whether the variable of interest is expected to go up, stay the same, or go down. This is mainly due to the fact that surveys that require respondents to give point forecasts for the variables in question are more likely to be susceptible to sampling and measurement errors than the surveys that only question about the expected directions of changes, see, e.g., Katona (1975) and Pesaran (1987). Often qualitative variables on expectations are of limited interest since they are difficult to interpret and methods of quantification are therefore needed. While there is a vast literature on this topic, no consensus has emerged among researchers on how to quantify the expectation survey data. The two prevalent pioneering

¹ On this argument, see Pagan (1984).

quantification approaches are the subjective probability approach by Theil (1952) and Carlson and Parkin (1975) and the regression model approach by Pesaran (1984, 1987).

Being purely statistical methods, the above mentioned quantification schemes work well in converting qualitative expectations survey data into quantitative time series of expectations. However, these quantification approaches do not attempt to include any prior information about the underlying process driving the variable of interest. Such prior information is typically available from economic theory or from empirical (stylized) facts. We argue that using prior information, say by including a behavioral model of the variable of interest, could lead to significant efficiency gains and, if so, such information should be included in the quantification scheme. One of the aims of this paper is therefore to develop a quantification approach that may take such information into account. In particular, we develop a general likelihood based modeling framework for quantification of qualitative survey data on expectations and perceptions, and propose a likelihood ratio test for the weak form of the rational expectations hypothesis – the consistency of expectations.

The main differences between our approach and previous ones lie in at least three aspects: First, our approach is likelihood based and estimation of unknown parameters or other population magnitudes of interest therefore not restricted to any of the distributional assumptions previous approaches rely critically upon. In particular, our approach facilitates the use of a very general class of mixture distributions that seems very useful in many situations as we will illustrate later. Because of its generality, the likelihood approach can easily be applied to quantify a wide range of economic variables like inflation rates, unemployment rates, stock prices and foreign exchange rates, etc. Secondly, unlike previous quantification approaches, the new approach is structural, in the sense that the underlying process of the variable of interest is modelled explicitly. The imposed economic structure is a crucial feature in the new test of consistency of expectations. We argue that the (weak) rational expectations hypothesis can only be tested under an explicitly specified economic modeling framework. Without a proper modeling framework to describe the process driving the variable of interest, the validity of inferences on expectations from purely statistical methods might not always be guaranteed, an implication that has been largely ignored in the literature. Thirdly, previous approaches test the expectation consistency hypothesis by comparing expectations of the variable of interest with its ex post realizations. As pointed out by Evans and Wachtel (1993) and Dahl and Hansen (2001), this will, due to ex post bias, lead to too many false rejections of the expectation consistency hypothesis in situations where a structural break or a recurrent regime shift has occurred. By contrast, we model the variable of interest taking into account the possibility of regime shifts, and the expectation consistency hypothesis is tested by comparing quantified expectations of the variable of interest from the survey data set with the mathematical expectation derived from the postulated behavioral/-structural economic model. Therefore, the new test will not suffer from ex post bias.

Monte Carlo simulation results show that the likelihood ratio test of expectation consistency delivers encouraging size and power properties. Moreover, our approach is more

efficient in terms of minimizing the mean squared error (MSE) relative to the quantification scheme proposed by Pesaran (1984, 1987).

Finally, an empirical illustration of the likelihood approach is provided by demonstrating how to quantify qualitative survey data on CBI industrial Trends Survey data on the British manufacturing industry. This data set was originally used by Pesaran (1984, 1987). Statistical evidence is presented, showing that the actual price change rates can be characterized well by a two-state Markov regime-switching model. The empirical results support the expectation consistency hypothesis and also help explain the phenomenon of ex post bias in inflation forecasts, see, e.g., Evans and Wachtel (1993) and Dahl and Hansen (2001).

The paper is organized as follows: In Section 2 the general setup of our quantification scheme is introduced and discussed under alternative survey sampling designs. In Section 3 the size and power properties of the proposed likelihood ratio test of expectation consistency are examined and the relative performance of the likelihood based quantification scheme is analyzed. In Section 4 an empirical illustration is provided using the CBI industrial Trends Survey data on the British manufacturing industry. Finally, Section 5 contains concluding remarks.

2 A Likelihood Approach for Quantification of Qualitative Survey Data

Let $y_{it+1|t} \in \mathbb{N}$, for $t=1,2,\dots,T$, $i=1,2,\dots,N$ be a panel survey data set (with T time periods and N respondents) on expectations with respect to $y_{t+1} \in \mathbb{R}$. Assume the survey consists of $j=1,2,\dots,J$ discrete response categories. If respondent i anticipates y_{t+1} to fall within an interval $[\beta_{j-1}; \beta_j)$, and we let $G(y_{t+1} | \Gamma_{it})$ define respondent i 's subjective probability function, the survey data sampling scheme can be described as $y_{it+1|t} = \sum_{j=1}^J j_{(\beta_{j-1} \leq y_{it+1|t} < \beta_j)}$ where $y_{it+1|t}^* = E(y_{t+1} | \Gamma_{it}) = \int y_{t+1} dG(y_{t+1} | \Gamma_{it}) \in \mathbb{R}$ is respondent i 's latent subjective expectation on the random variable y_{t+1} .² Notice that expectations are formed conditional on the information set Γ_{it} , which includes all the relevant information that respondent i has collected up to time t . When respondents form expectations by correctly using all the information they possess we can write $y_{it+1|t}^* = E(y_{t+1} | \Gamma_t) + \sigma_\eta \eta_{it+1}$, where $\Gamma_t = \bigcup_{i=1}^N \Gamma_{it}$ and η_{it+1} is a zero mean idiosyncratic error term with probability function $\Psi(x) = \Pr(\eta_{it+1} \leq x)$. In order to compute $E(y_{t+1} | \Gamma_t)$ we need to define $G(y_{t+1} | \Gamma_t)$. We suggest a simple regression model approach for this purpose and think of the y_{t+1} as being generated according to $y_t = g(\Gamma_{t-1}; \theta) + \sigma_u u_t$, where $g(\cdot)$ is a general nonlinear measurable function. Let u_t have a stationary distribution with probability function $U(x) = \Pr(u_t \leq x)$. Under the hypothesis that $(\delta_1, \delta_2)' = (0, 1)'$ the model can be written as

² Following standard notation, $j_{(\cdot)}$ is an indicator function taking the value j or zero.

$$y_t = g(\Gamma_{t-1}; \theta) + \sigma_{ut} u_t, \quad (1)$$

$$y_{it+|t}^* = \delta_1 + \delta_2 E(y_{t+1} | \Gamma_t) + \sigma_{\eta} \eta_{it+1}, \quad (2)$$

$$y_{it+|t} = \sum_{j=1}^J j_{(\beta_{j-1} \leq y_{it+|t}^* < \beta_j)}, \quad (3)$$

where $\psi = (\theta', \sigma_{ut}, \delta_1, \delta_2, \sigma_{\eta}, \beta_1, \dots, \beta_{J-1})'$ is the collection of unknown parameters.³

We suggest to choose $g(\cdot)$ and Γ_{t-1} based on prior information obtained from economic theory. To give a simple example, consider a situation where qualitative data on exchange rate expectations is available. Classical economic theory suggests that the actual exchange rate (measured as the domestic price of foreign currency) would obey the purchasing power property (PPP) and accordingly a natural choice would be $\Gamma_{t-1} = (p_{t-1}, p_{t-1}^*)'$ and $g(\Gamma_{t-1}; \theta) = \theta_1(p_{t-1} - p_{t-1}^*)$ where p_{t-1} and p_{t-1}^* denote the domestic and foreign prices in logarithms respectively. In more "restricted" situations, where such information is unavailable we recommend to employ simple dynamic empirical representations. In the simplest AR(1) case one would set $\Gamma_{t-1} = y_{t-1}$ and $g(\Gamma_{t-1}; \theta) = \theta_1 y_{t-1}$. In Section 3 and 4 we consider examples of more evolved empirical models where the dynamics governing the variable of interest are allowed to change between regimes. This can be modelled by assuming that y_t has a mixture of, say, normal distributions. This "flexible" distributional assumption seems relevant when modelling a wide range of "fat tailed" nominal or financial variables such as prices, interest rates, exchange rates, etc. In all situations, however, it is crucial to evaluate the adequacy of the choices made, which we also emphasize in the empirical illustration in Section 4. Misspecification of $g(\cdot)$ and Γ_{t-1} will deteriorate the quality of the estimated quantities and lead to too many rejections of the expectation consistency hypothesis. Finally, it should be noted that economic theory rarely gives any guidance on the short run dynamics which also is the case in the PPP example above. To obtain a well specified representation of equation (1) one would therefore typically be forced to augment Γ_{t-1} and $g(\Gamma_{t-1}; \theta)$ with additional lags of domestic and foreign price levels.

To motivate a test for rationality of expectations based on the model given by (1)–(3) we use Muth's (1961) central idea of the rational expectations hypothesis which states that individuals form expectations of a variable rationally only if they are making efficient use of all information pertinent to the determination of that variable. This definition implies that if agents form expectations rationally then the difference between the individuals subjective expectations and the objective/mathematical expectation should be random and equal to zero on average, hence be systematically unbiased. A test for unbiasedness or consistent expectations is often considered as the weak test of rationality, see, e.g., Rich (1990).

³ It is implicitly assumed that $(\beta_0, \beta_J)' = (-\infty, \infty)'$.

Focusing on the weak form of rationality, the null hypothesis of rational expectations can be formalized as

$$H_0 : \delta_1 = 0 \wedge \delta_2 = 1. \quad (4)$$

in the model (1)-(3). That is, under the null hypothesis, the conditional mean of respondents' aggregate subjective probability function equals the conditional mathematical mean of the objective probability function.

2.1 Estimation based on individual responses on expectations

In order to quantify the survey expectations and to test the hypothesis on expectation formation, estimates of the unknown population parameters in ψ must be obtained. Once $\hat{\psi}$ is obtained the qualitative expectations from the survey data set can be quantified simply as

$$\hat{y}_{it+1|t}^* = \hat{\delta}_1 + \hat{\delta}_2 g(\Gamma_t; \hat{\theta}).$$

In what follows we propose a likelihood based approach for estimating $\hat{\psi}$ and to test H_0 given by (4) using the classical likelihood ratio test. Let the joint conditional probability density of the observed sequence $\{(y_t, y_{it+1|t})'\}_{t=1}^T$ for respondent i be given by

$$\prod_{t=1}^T f(y_t, y_{it+1|t} | \Gamma_t; \psi),$$

where $\{\Gamma_t\}_{t=1}^T$ typically is a vector sequence of weakly exogenous variables. Consequently, the log likelihood function defined over all N respondents is simply

$$\ell(\psi) = \sum_{i=1}^N \sum_{t=1}^T \log f(y_t, y_{it+1|t} | \Gamma_t; \psi). \quad (5)$$

Using Bayes rule we can write (5) more conveniently as

$$\ell(\psi) = \sum_{i=1}^N \sum_{t=1}^T \left(\log f_1(y_{it+1|t} | \Gamma_t; \psi_1) + \log f_2(y_t | \Gamma_{t-1}; \psi_2) \right), \quad (6)$$

where $f_1(\cdot | \cdot)$ is the conditional density function of $y_{it+1|t}$ given Γ_t , $f_2(\cdot | \cdot)$ is the conditional density function of y_t given Γ_{t-1} , and $\psi = (\psi_1', \psi_2')'$. Using (1)-(3) and by defining

$$\Phi_{j,t+1} = \Pr \left(\eta_{it+1} < \frac{\beta_j - \delta_1 - \delta_2 E(y_{t+1} | \Gamma_t)}{\sigma_\eta} \right), \text{ we can write}$$

$$\log f_1(y_{it+1} | \Gamma_t; \psi_1) = \sum_{j=1}^J 1_{(y_{it+1}=j)} \log(\Phi_{j,t+1} - \Phi_{j-1,t+1}).$$

The shape of $\log f_2(y_t | \Gamma_{t-1}; \psi_2)$ will depend on the probability function $U(x)$. In the case of normality

$$\log f_2(y_t | \Gamma_{t-1}; \psi_2) = -\frac{1}{2} \log(2\pi) - \log(\sigma_{ut}) - \frac{1}{2} \frac{(y_t - g(\Gamma_{t-1}; \theta))^2}{\sigma_{ut}^2}.$$

For estimation and testing purposes, however, one cannot use (6) directly since ψ is not uniquely identified.⁴ In order to proceed we therefore need to impose a set of identifying assumptions. One possibility is the following scheme:

Identification Scheme 1 *If β_j for all $j=1,2,\dots,J-1$, are known to the respondents, all the unknown parameters of the model given by equation (1)-(3) are uniquely identified.*⁵

The identification scheme is reasonable to impose whenever the categories in which respondents place their answers are pre-specified numerically by the survey organizers. Survey data sets with such pre-specified threshold parameters include the well-known Survey of Consumers Data on price expectations collected by the Survey Research Center (SRC) at the University of Michigan and Livingston Survey collected for the Philadelphia Inquirer.

Often, however, β_j for $j=1,2,\dots,J-1$ are not pre-specified by the surveyors. In this case, it is not reasonable to assume that respondents know the values of threshold parameters, and consequently identification scheme 1 cannot be used. In the next section we will introduce a second identification scheme that does not rely on this assumption.

2.2 Estimation based on individual responses on expectations and perceptions

Surveys are often organized such that qualitative data on expectations as well as perceptions are collected by the surveyors. That is, in addition to being questioned about future expectations on the variable of interest, respondents are asked about their perceptions on the variable of interest. Observations on perceptions can help to identify the parameters of

⁴ The lack of identification can be seen by noticing that it is only the ratio β_j / σ_η that enters the log likelihood function. Hence β_j and σ_η can be varied proportionally without affecting the log likelihood function.

⁵ This result is well known from the discrete choice literature, see, e.g., Amemiya (1985) Ch. 9.

the model (1)-(3) even when the threshold values are not pre-specified by the surveyors. This idea was originally coined by Pesaran (1984, 1987), who in a different setting derives the parameters used to quantify survey data expectations from the historical relationship between perceptions and actual realizations.

In particular, if $y_{it|t} \in \mathbb{N}$ denotes the categorical observations on respondent i 's perceptions, then by the same line of arguments as used above, our model (1)-(3) can be augmented with the following two equations

$$y_{it|t}^* = \mu_1 + \mu_2 E(y_t | \mathbf{F}_t) + \sigma_\omega \omega_{it}, \quad (7)$$

$$y_{it|t} = \sum_{j=1}^J j_{(\alpha_{j-1} \leq y_{it|t}^* < \alpha_j)}, \quad (8)$$

where $y_{it|t}^* \in \mathbb{R}$ is respondent i 's latent subjective perception with respect to the random variable y_t , and ω_{it} is an idiosyncratic error term with probability function $\Psi(x) = \Pr(\omega_{it} < x)$. We will use that $E(y_t | \mathbf{F}_t) = y_t$, since it must be reasonable to assume that $y_t \in \mathbf{F}_t$. The threshold parameters for the perceptions are given by α_j , for $j=1, 2, \dots, J-1$. Augmented by observations on perceptions and given the assumptions associated with (7) and (8) the log likelihood becomes

$$\ell(\psi) = \sum_{i=1}^N \sum_{t=1}^T \left(\log f_1(y_{it+1|t} | \mathbf{F}_t; \psi_1) + \log f_2(y_t | \mathbf{F}_{t-1}; \psi_2) + \log f_3(y_{it|t} | \mathbf{F}_t; \psi_3) \right), \quad (9)$$

where the additional term is given as

$$\log f_3(y_{it|t} | \mathbf{F}_t; \psi_3) = \sum_{j=1}^J 1_{(y_{it|t}=j)} \log(\Psi_{jt} - \Psi_{j-1t}),$$

and $\Psi_{jt} = \Psi((\alpha_j - \mu_1 - \mu_2 y_t) / \sigma_\omega)$. It is easy to verify that the unrestricted perception augmented model cannot be identified. We propose to obtain identification by imposing the following scheme:

Identification Scheme 2 *All the unknown parameters of the model given by (1), (2), (3), (7) and (8) will be uniquely identified when the following restrictions are imposed:*

$$(\mu_1, \mu_2, \alpha_1, \alpha_2, \dots, \alpha_{J-1})' = (0, 1, \beta_1, \beta_2, \dots, \beta_{J-1})'$$

Accordingly, identification can be obtained, assuming that respondents are at least capable of forming unbiased perceptions and that respondents use the same set of threshold values when placing their answers in the J expectation and perception categories.

2.3 Estimation based on proportion categories data

The log likelihood functions given above are derived under the assumption that the outcome of each of the individual decisions is observed for every time period. Typically, however, observations on individual expectations are unavailable and only the proportions of respondents replying in each of the categories are reported. Consequently, the log likelihood function have to be modified accordingly by re-defining it over the observed proportions, given as $\hat{p}_{jt}^e = (1/N) \sum_{i=1}^N 1_{(y_{it+1|t}=j)}$ for expectations and $\hat{p}_{jt}^p = (1/N) \sum_{i=1}^N 1_{(y_{it|t}=j)}$ for perceptions. In that case (6) and (9) becomes

$$\ell_I(\psi) = N \sum_{t=1}^T \sum_{j=1}^J \hat{p}_{jt}^e \log(\Phi_{jt+1} - \Phi_{j-lt+1}) + N \sum_{t=1}^T \log f_2(y_t | \Gamma_{t-1}; \psi_2),$$

and

$$\begin{aligned} \ell_{II}(\psi) = N \sum_{t=1}^T \sum_{j=1}^J & \left(\hat{p}_{jt}^e \log(\Phi_{jt+1} - \Phi_{j-lt+1}) + \hat{p}_{jt}^p \log(\Psi_{jt} - \Psi_{j-lt}) \right) \\ & + N \sum_{t=1}^T \log f_2(y_t | \Gamma_{t-1}; \psi_2), \end{aligned}$$

respectively. It is important to notice that the i.i.d. assumptions on the error terms ω_{it} and η_{it+1} can be relaxed when observations on individual data are available. When only observations on proportion data are available, such degree of generality cannot be obtained, because information on the possible correlation between ω_{it} and η_{it+1} will not be available in this situation. However, under the weak rationality hypothesis it does not seem unreasonable to assume that perception and expectation errors on average are uncorrelated.

3 Monte Carlo Experiments

The purpose of this section is two-fold. First, a likelihood ratio test statistic for testing expectation consistency hypothesis given by (4) is derived and some limited evidence on the size and power properties of the statistic is provided using Monte Carlo simulations. Secondly, we compare the efficiency of the likelihood based quantification scheme with Pesaran's (1984, 1987) quantification approach under alternative data generating mechanisms. This section will focus entirely on the situation where only proportion categories data are available from the survey.

3.1 Size and power properties of the expectation consistency test

Let

$$\hat{\psi}_I = \arg \max_{\psi} \ell_I(\psi) \text{ and } \hat{\psi}_{II} = \arg \max_{\psi} \ell_{II}(\psi)$$

be the unconstrained maximum likelihood estimates obtained under identification scheme 1 and 2 respectively. Similarly, let $\hat{\psi}_I^C$ and $\hat{\psi}_{II}^C$ be the constrained maximum likelihood estimates when the restrictions given by (4) are imposed. The likelihood ratio test for consistent expectations under the two identification schemes are then given as

$$LR_s = 2(\ell_s(\hat{\psi}_s) - \ell_s(\hat{\psi}_s^C)) \stackrel{a}{\sim} \chi^2(2) \text{ for } s = I, II.$$

To investigate the finite sample properties of LR_s we simulate proportion categories data from the model given by (1), (2), (3), (7) and (8) for alternative values of δ_2 ranging from 0.1 to 1.8.⁶ It is assumed that $N=1000$, $J=4$, $(\delta_1, \mu_1, \mu_2)' = (0, 0, 1)'$, $(\sigma_{\omega}, \sigma_{\eta}, \sigma_{\omega})' = (1, 1, 1)$ and the threshold values chosen are $(\alpha_1, \alpha_2, \alpha_3) = (\beta_1, \beta_2, \beta_3) = (-0.5, 0.5, 1.5)$. The choice of numerical values for the threshold parameters is based on the hypothesis stated by Carlson and Parkin (1975) that there is a range about zero which the respondents cannot distinguish from zero. The magnitudes of the threshold parameters are similar to those used in the SRC survey. For simplicity, the variable of interest y_t is simulated using a Gaussian first order autoregressive process, i.e. $g(\Gamma_{t-1}; \theta) = 0.5y_{t-1}$ and $u_t \sim N(0, 1)$. The results on the size and power of LR_s on a 5% nominal significance level based on 1000 Monte Carlo replications are summarized in Figure 1 and 2.

When $\delta_2 = 1$, the simulated size of LR_I and LR_{II} are between 5%–10% for all the sample sizes considered, implying that strictly speaking the size of the tests seems a little too high. We do not, however, consider the size distortion as being significant. In small sample sizes one could use a simple bootstrap version of the tests, suggested by Dahl (2003), to eliminate the risk of possible size distortion. When δ_2 starts deviating from one the power starts approaching one rapidly as expected. There is no strong evidence, that LR_I is more powerful than LR_{II} , though there are fewer estimated parameters involved. In conclusion, both test statistics seem to perform satisfactorily.

⁶ The same type of analysis was performed with respect to δ_1 , but the results did not differ and therefore are not reported. The results are available from the authors upon request.

Figure 1 **Power of LR_I , at a 5% nominal significance level for alternative sample sizes. The number of Monte Carlo replications equals 1000.**

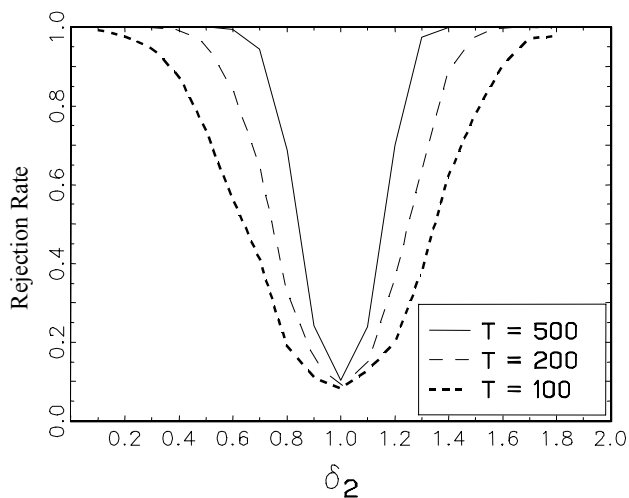
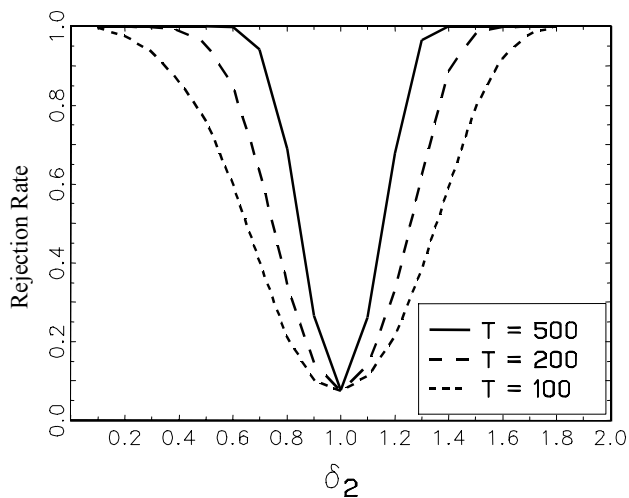


Figure 2 **Power of LR_{II} , at a 5% nominal significance level for alternative sample sizes. The number of Monte Carlo replications equals 1000.**



3.2 Relative efficiency of likelihood based quantification

In order to analyze the relative performance of the likelihood based quantification scheme a numerical comparison to the quantification scheme suggested by Pesaran (1984,1987) is carried out. Pesaran's regression model approach uses observations on survey expectations as well as perceptions, so in this comparison focus is on the likelihood approach under Identification Scheme 2.

We consider two data generating mechanisms (DGP's) that only differ with respect to the dynamic process driving the variable of interest. DGP 1 is as defined in the previous section with the exceptions that $J = 3$, $(\alpha_1, \alpha_2) = (\beta_1, \beta_2) = (-0.5, 0.5)$, and that δ_2 is fixed at $\delta_2 = 1$. Under DGP 2 we let y_t be a mixture of normals, by characterizing it as following a two-state Markov regime-switching process. In particular, we assume

$$g(\Gamma_{t-1}; \theta) = (1 - s_t^*)(c_0 + \phi_0 y_{t-1}) + s_t^*(c_1 + \phi_1 y_{t-1}),$$

$$\sigma_{ut} u_t = (1 - s_t^*)\sigma_0 v_{0t} + s_t^*\sigma_1 v_{1t},$$

where $(c_0, c_1, \phi_0, \phi_1)' = (-1, 2, 0.7, 0.3)'$, $(v_{0t}, v_{1t})' \sim N(0, I_{2 \times 2})$ and s_t^* is an unobservable random state variable ($s_t^* \notin \Gamma_{t-1}$) that can take the values zero (regime 1) and one (regime 2) only. We assume that the transition between the two regimes are governed by a Markov chain with transition probability matrix

$$P = \begin{bmatrix} \Pr(s_t^* = 0 | s_{t-1}^* = 0) & \Pr(s_t^* = 0 | s_{t-1}^* = 1) \\ \Pr(s_t^* = 1 | s_{t-1}^* = 0) & \Pr(s_t^* = 1 | s_{t-1}^* = 1) \end{bmatrix} = \begin{bmatrix} \pi_0 & 1 - \pi_1 \\ 1 - \pi_0 & \pi_1 \end{bmatrix}.$$

In the simulations $(\pi_0, \pi_1)' = (0.9, 0.8)'$ is used.⁷ We measure the accuracy of the two alternative quantification schemes using the mean squared difference between the quantified/estimated expectations and the true expectations. The results based on 1000 Monte Carlo replications are reported in Table 1.

When comparing the results based on Pesaran's approach with those from the likelihood approach the improvement in accuracy by using information on the DGP explicitly seem rather striking. This seems to be particular true when possibly recurrent regime shifts are present in the process driving the variable. Since, as already mentioned, a number of studies have found such evidence in macroeconomic time series on inflation our limited simulation study strongly suggests that the likelihood approach seems to be an important new tool for analysts who wish to quantify qualitative survey data on price change expectations.

⁷ For specification of $\log f_2(y_t | \Gamma_{t-1}; \psi_2)$ in the Markov regime switching case, see, e.g., Hamilton (1994), Ch. 22.

Table 1 Comparison of accuracy of Pesaran's regression model approach with likelihood based quantification. Monte Carlo replications used equals 1000.

Sample Size	DGP 1 with AR(1)		DGP 2 with MS-AR(1)	
	100	500	200	500
MSE: True	1.00	1.00	1.00	1.00
MSE: Pesaran's approach	1.26	1.38	3.69	3.66
MSE: Likelihood approach	1.00	0.99	1.00	1.00

4 Empirical Illustration

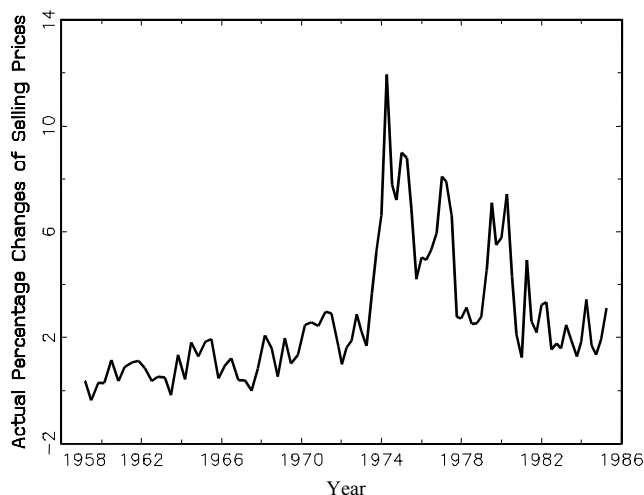
In this section, the CBI industrial Trends Survey data on the British manufacturing industry are used to re-examine the performance of the likelihood approach for quantification and testing the expectation consistency hypothesis. The CBI survey consists of data on qualitative trends for both expectations and perceptions over the past and next four months on average change rates of selling prices of British manufacturing and dates back to June 1958. This data as well as the time series data of the actual rates of price changes in the British manufacturing industry are available from Pesaran (1987) for the period 1959q1 to 1985q2. The data set enables quantification of expected price change rates in British manufacturing industry based on identification scheme 2 of the likelihood based approach and facilitates easy comparison to Pesaran's regression model approach as discussed in Pesaran (1987).⁸ As shown in Figure 3, the actual price change rates in the British manufacturing industry may have experienced a structural break with the mean and variance of price change rates increasing substantially after 1974. The simulation results reported in the previous section suggest that the quality of the quantified expectation can be improved if we account for the possibility of such regime shifts.

For the likelihood based quantification scheme, we begin by modeling the actual price change rates based on the well-known Box-Jenkins modeling philosophy that simpler models provide more robust forecasts. As mentioned in Hamilton (1994), although complicated models can track the data very well over the historical period for which parameters are estimated, they often perform poorly when used for out-of-sample forecasting. Accordingly, to characterize the actual price change rates, we concentrate on low order ARMA models, rather than complicated models with many exogenous variables. In terms of model selection, we first attempt to characterize the actual price change rates with linear representation by selecting the best linear model to fit the actual price change rates data among a pool of candidate ARMA models. The model selection criteria used are the AICC, which is a bias-

⁸ The detailed results of Pesaran's regression method applied to this data set are reported in Pesaran (1987).

corrected version of the AIC suggested by Hurvich and Tsai (1989), BIC and the FPE (Final Prediction Error criterion) developed by Akaike (1969). The results reported in Table 2 indicate that the best linear univariate model for the price change rates seems to be an AR(1) representation.

Figure 3 **Actual rates of selling prices changes in the British manufacturing industry from 1959q1 - 1985q2.**



The graphs of sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of AR(1) model shown in Figure 4 further justify the choice of AR(1) model. In particular, the fact that all of the PACF values beyond one lag fall within the bounds, defined as $\pm 1.96/\sqrt{T}$, supports the choice of the AR(1) model for describing the actual price change rates data. In addition, the AR(1) model passes the diagnosis tests of randomness on residuals such as Portmanteau test, McLeod-Li test, implying white noise residuals.

With rather strong indications of possible regime shifts by inspection of Figure 3, a careful study of the stability of the linear models is required. Several tests are conducted to test the stability of the AR(1) model as well as the other candidate ARMA models. Based on the one-step-up Chow test for structural break within the sample period, depicted in Figure 5, the null hypothesis of stability of the AR(1) model is rejected at one percent level. Moreover, none of the other candidate linear univariate models passes the Chow test at this level. We also test structural stability using CUSUM test proposed by Brown, Durbin, and Evans (1975). Figure 5 (lower panel) shows that most CUSUM statistics stay outside the upper confidence bound at 5 percent level, implying the strong instability of the AR(1) model.

Table 2 Determining the best linear univariate representation of the price change rate in the British manufacturing industry for the period 1959q1 - 1985q2.
Model selection based on alternative information criteria.

Models	Model Selection Criteria		
	AICC	BIC	FPE
AR(1)	323.565	324.635	1.832
AR(2)	325.671	329.587	1.871
AR(3)	325.586	331.814	1.864
AR(4)	327.734	335.880	1.903
MA(1)	370.862	371.634	.
MA(2)	350.815	354.247	.
MA(3)	345.599	351.233	.
MA(4)	335.827	343.471	.
ARMA(1,1)	325.636	329.568	.
ARMA(1,2)	325.993	332.241	.
ARMA(2,1)	327.747	333.967	.
ARMA(2,2)	328.232	336.390	.

Figure 4 Sample ACF (left) and PACF associated with selling price changes in the British manufacturing industry from 1959q1 - 1985q2.

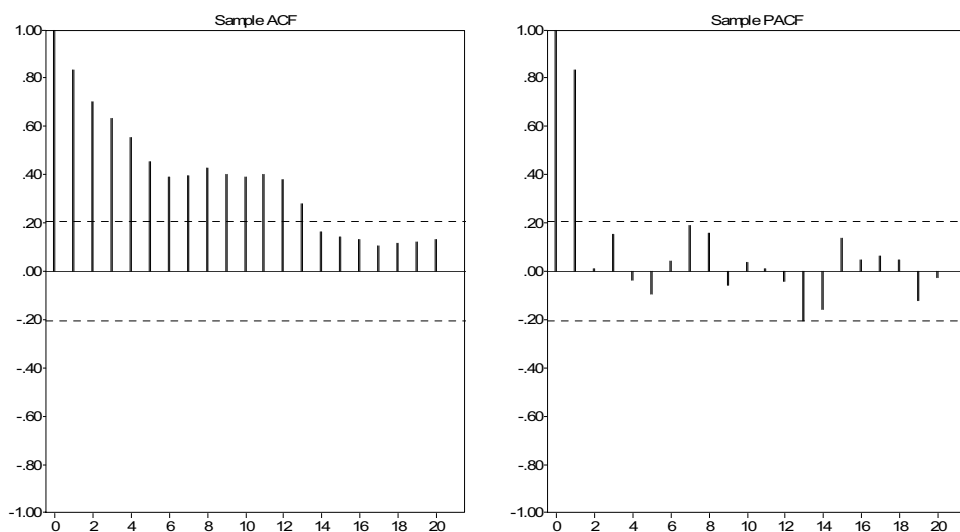
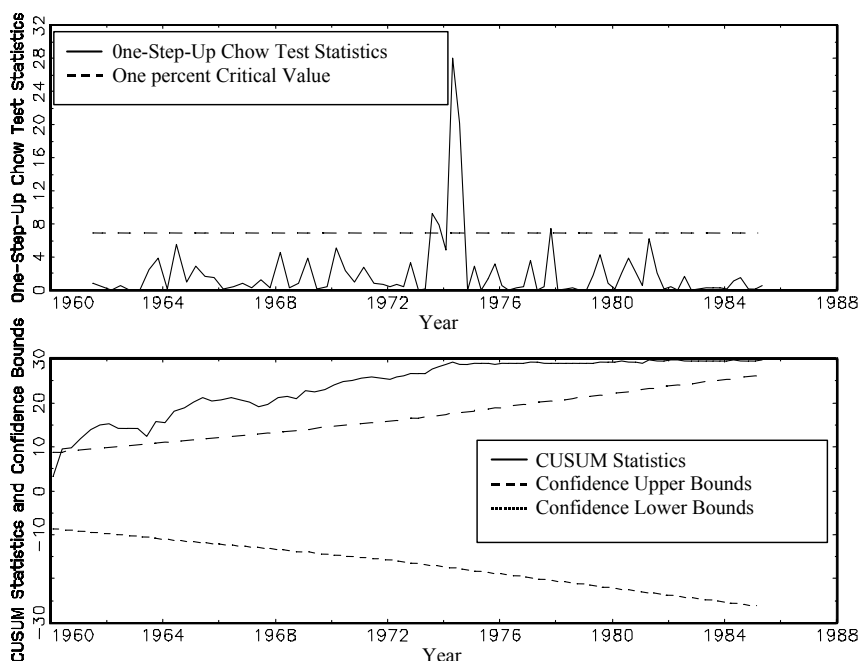


Figure 5 **One-step-up Chow test (top) and CUSUM test for stability of the AR(1) model of price changes in the British manufacturing industry from 1959q1 - 1985q2.**



Andrews, Lee, and Ploberger (1996) suggest three optimal change point tests for detecting the presence of non-stable coefficients in a normal linear-regression model with unknown breakpoints. The three test statistics, denoted as Sup LM, Exp LM and Ave LM, are reported in Table 3. All three tests reject the hypothesis of stability of all parameters in the AR(1) model. In conclusion, all testing evidence rejects the stability of AR(1) modeling of actual price change rates and it seems fairly safe to reject the adequacy of the linear AR(1) model. To model the apparent shift in the parameters in the linear model a non-linear Markov regime-switching model seems a natural choice. It is, however, crucial to test whether the Markov regime-switching specification is preferred over the linear model in a statistical sense. For that purpose the specification test proposed by Hansen (1992, 1996) is employed, and the results are summarized in Table 4. In Table 4, LR^* denotes the standardized likelihood ratio test statistics and M denotes the chosen bandwidth, see, Hansen (1996) for more details. Independent of the choice of bandwidth, the associated p-values are all very small, suggesting a rejection of the null of the AR(1) model as an adequate description of the actual price change rates in favor of the two-state Markov regime-switching model.

Table 3 Andrews-Lee-Ploberger test for instability of the AR(1) model of price change rates in the British manufacturing industry for the period 1959q1 - 1985q2.

	Statistic	Critical values			p-value
		10%	5%	1%	
Sup LM	11.89	10.01	11.79	15.51	0.0436
Exp LM	3.07	2.59	3.22	4.76	0.0574
Ave LM	4.18	3.75	4.61	6.73	0.0699

Table 4 Testing the MS-AR(1) specification of price change rates in the British manufacturing industry against the AR(1) alternative using Hansen's (1992,1996) standardized likelihood ratio test. The sample period is 1959q1 - 1985q2.

	Statistics	p-value				
		M=0	M=1	M=2	M=3	M=4
LR*	5.615	0.0000	0.0000	0.0000	0.0000	0.0000

To determine the order of the MS-AR(p) representation, AIC, HQ (Hannan-Quinn Criterion) and BIC are used. All three model selection criteria give support to the MS-AR(1) process, which also passes the diagnosis test of randomness of residuals, see Table 5 and 6. Consequently, we will take this representation as the underlying process driving price change rates in the likelihood based quantification scheme.

Table 5 Determining the best MS-AR(p) representation of price change rates in the British manufacturing industry for the period 1959q1 - 1985q2.
Model selection based on alternative information criteria.

Model	Model selection criteria			
	AIC	HQ	BIC	logL
MS-AR(1)	321.61	330.46	343.54	-139.94
MS-AR(2)	322.50	333.63	350.09	-136.74
MS-AR(3)	323.58	337.03	356.92	-133.61
MS-AR(4)	328.41	344.19	367.56	-132.14

Table 6 Diagnostic tests of the residuals of MS-AR(1) model of inflation in the British manufacturing industry for the period 1959q1 - 1985q2.

Diagnostic Test	Statistics	p-value
Portmanteau (lag=7)	2.9378	0.8166
Normality	2.7014	0.2591
Heteroskedasticity	0.4477	0.7994

The main estimation results of our quantification approach with actual price change rates described as a two-state Markov regime-switching model are reported in Table 7. Figure 6 depicts the actual and the quantified expected price change rates. The probabilities for being in the high inflation regime ($S_t^* = 1$) are plotted in the lower part of Figure 6. It is not surprising that the mean and variance differ substantially in two regimes. In the low-inflation regime, the unconditional mean of the price change rates is equal to 1.5, whereas the unconditional mean equals 6.8 in the high-inflation regime. The estimates of the variance in the two states confirm the conjecture of a positive relationship between the level of inflation and its volatility. That is, higher inflation is associated with higher inflation variance. More importantly, the LR test statistic fails to reject the null hypothesis of expectations consistency, implying that agents make unbiased expectations. This empirical result supports the rational expectations hypothesis. An interesting finding is that the estimates of threshold parameters reveal asymmetry in respondents' decisions on price changes. While a 4.5 percent increase in price makes respondents believe prices are going up, respondents will not believe prices are going down until they observe more than 5.5 percent drop in actual prices. This evidence suggests that economic agents are more sensitive to price increases relative to price decreases.

From Figure 6 (top panel) it appears that respondents tend to over-predict price change rates when they are relatively low, and under-predict when the actual price change rates are relatively high. These seemingly biased inflation expectations, however, may not serve as the evidence against the rational expectations hypothesis. As discussed earlier, the LR test cannot reject the expectations consistency and since it is based on comparing two sets of forecasts, ex post bias does not affect the outcome of the test. If one, however, were tempted to test (4) based on a regression of variable of interest on the quantified expectations derived from Pesaran's approach or the likelihood approach, the null hypothesis of weak rationality would be rejected.⁹ Our empirical results support the conclusion of Evans and Wachtel (1993) who claim that forecasters are acting rationally, but face a complicated forecasting problem that makes systematic forecast errors unavoidable. If the price change rates are in one

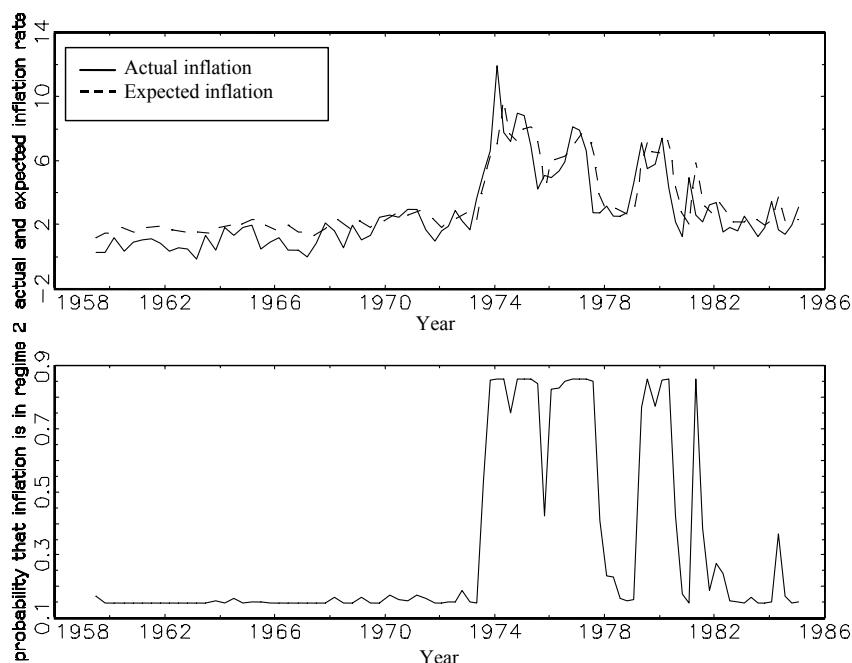
⁹ Testing (4) based on the regression $y_{t+1} = \delta_1 + \delta_2 \hat{q}_{t+1} + \varepsilon_t$, using a F-test, where y_{t+1} is actual price change rates and $\hat{q}_{t+1}$ is the quantified expectations gave $F(2, 89) = 6.3256$ [p-value = 0.0027] for the likelihood approach and $F(2, 89) = 4.6285$ [p-value = 0.0122] when $\hat{q}_{t+1}$ was obtained by Pesaran's approach. In this simple illustrations it is ignored that $\hat{q}_{t+1}$ depends on estimated parameters which probably would have lead to a (very moderate) increase of the p-values.

regime but the forecasters perceive that there is a possibility that price change rates switch to the other regime, then their expectations will be systematically biased if the switch does not occur.

Table 7 Quantification by the likelihood approach. Estimation results based on a MS-AR(1) representation of price change rates in the British manufacturing industry for the period 1959q1 - 1985q2.

Parameters	Regime 1 ($s_t^* = 0$)	
	Estimates	Std. Errors
c_0	0.761	0.157
ϕ_0	0.493	0.074
σ_0	0.739	0.066
	Regime 2 ($s_t^* = 1$)	
	Estimates	Std. Errors
c_1	3.396	1.248
ϕ_1	0.500	0.174
σ_1	1.633	0.292
$\pi_0 = \Pr(s_t^* = 0 \mid s_{t-1}^* = 0)$	0.953	0.028
$\pi_1 = \Pr(s_t^* = 1 \mid s_{t-1}^* = 1)$	0.858	0.078
	Quantification	
	Estimates	Std. Errors
β_1	-5.533	2.555
β_2	4.499	0.910
σ_2	5.564	1.691
δ_2	0.384	1.460
δ_3	1.043	0.464
σ_3	5.069	1.668
Log-likelihood		-205.566
Constrained Log-likelihood		-205.709
LR test statistic		0.286

Figure 6 **Actual (solid line) and quantified (dashed line) price change expectations in the British manufacturing industry from 1959q1 - 1985q2 (top) and the estimated probability of observing regime 2 ($s_t^* = 1$).**



5 Conclusion

In this paper, a new likelihood approach for quantification of qualitative survey data is introduced and as a by-product a likelihood ratio test for consistent expectation is proposed. Evidences from limited simulation studies and from empirical observation analysis demonstrate that the likelihood approach promises to deliver a more efficient quantification scheme whenever prior information on the underlying process driving the variable of interest is available. When actual price change rates are described by a two-state Markov regime-switching model, the likelihood ratio test conducted on CBI survey data supports the hypothesis of consistent expectations. In fact, our empirical conclusion indicates that economic agents behave rationally in making their price change rate forecasts, though they make unavoidable systematic forecast errors due to the possibility of switches of price change regimes.

References

- Abadir, K. M. and Magnus, J. R. (2002), 'Notation in econometrics: a proposal for a standard', *The Econometrics Journal*, 5, 76–90.
- Akaike, H. (1969), 'Fitting autoregressive models for prediction', *Annals of the Institute of Statistical Mathematics*, 21, 243–247.
- Amemyia, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.
- Andrews, D. W. K., Lee, I. P. and Ploberger, W. (1996), 'Optimal change point tests for normal linear regression', *Journal of Econometrics*, 70, 9–38.
- Brown, R., Durbin, J. and Evans, J. (1975), 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society, Series B* 37, 149–172.
- Carlson, J. A. and Parkin, M. (1975), 'Inflation expectation', *Economica*, 42, 123–138.
- Dahl, C. M. (2003), 'A test for unbiased expectations based on qualitative survey data'. Unpublished manuscript, Department of Economics, Purdue University.
- Dahl, C. M. and Hansen, N. L. (2001), 'The formation of inflation expectations under changing inflation regimes', *Studies in Nonlinear Dynamics and Econometrics*, 4, 183–212.
- Evans, M. and Wachtel, P. (1993), 'Inflation regimes and the sources of inflation uncertainty', *Journal of Money, Credit and Banking*, 25, 475–511.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Hansen, B. E. (1992), 'The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GDP', *Journal of Applied Econometrics*, 7, 61–82.
- Hansen, B. E. (1996), 'Erratum: The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GDP', *Journal of Applied Econometrics*, 11, 195–198.
- Hurvich, C. M. and Tsai, C. L. (1989), 'Regression and time series model selection in small samples', *Biometrika*, 76, 297–307.
- Katona, G. (1975), *Psychological Economics*, Elsevier, New York.
- Muth, J. F. (1961), 'Rational expectations and the theory of price movements', *Econometrica*, 29, 315–335.
- Pagan, A. R. (1984), 'Econometric issue in the analysis of regressions with generated regressors', *International Economic Review*, 70, 221–247.
- Pesaran, M. H. (1984), 'Expectations formations and macro-econometric modelling', in P. Malgrange and P. A. Muet, eds, 'Contemporary Macroeconomic Modelling', Basil Blackwell, Oxford, England, 27–55.
- Pesaran, M. H. (1987), *The Limits to Rational Expectations*, Basil Blackwell Ltd.
- Rich, R. W. (1990), 'Another look at the rationality of the Livingston price expectations data', *Applied Economics*, 22, 477–485.
- Theil, H. (1952), 'On the time shape of economic microvariables and the Munich business test', *Revue de l'Institut International de Statistique*, 20, 105–120.

Detection of Turning Points in Business Cycles

Eva Andersson, David Bock and Marianne Frisén*

Abstract

Methods for continuously monitoring business cycles are compared. A turn in a leading index can be used to predict a turn in the business cycle. Three likelihood based methods for turning point detection are compared in detail by using the theory of statistical surveillance and by simulations. One of the methods is a parametric likelihood ratio method. Another includes a non-parametric estimation procedure. The third is based on a Hidden Markov Model. Evaluations are made of several features such as knowledge of shape and parameters of the curve, types and probabilities of transitions and smoothing. Results on the expected delay time [of](to) a correct alarm and the predictive value of an alarm are discussed. The three methods are also used to analyze an actual data set (of) [for] a period of (the) Swedish industrial production. The relative merits of evaluation of methods by one real data set or by simulations are discussed.

Key Words: Business cycle; Monitoring; Optimal; Likelihood ratio; HMM; Non-parametric

JEL Classification: C14; C15; C44; E32

*

Department of Statistics, Göteborg University, Gothenburg, Sweden

Résumé

Les méthodes de suivi continu des cycles conjoncturels sont comparées. Un retournement d'un indice précurseur peut donner un signal sur un retournement du cycle conjoncturel. Trois méthodes fondées sur la probabilité pour la détection des points de retournement font l'objet d'une comparaison détaillée à l'aide de la théorie de la surveillance statistique et au moyen de simulations. Une des méthodes met en œuvre un ratio de probabilité paramétrique. Une autre implique une procédure d'estimation non paramétrique. La troisième repose sur un modèle de Markov caché. Les évaluations comportent plusieurs aspects tels que la connaissance de la forme et des paramètres de la courbe, les types et probabilités de transitions et le lissage. L'étude analyse les résultats concernant le délai prévu pour une alerte correcte et la valeur prédictive d'une alerte. Les trois méthodes sont également utilisées pour analyser une série de données réelles sur une période de la production industrielle suédoise. Sont ensuite examinés les mérites relatifs de l'évaluation des méthodes par une seule série de données réelles ou par des simulations.

1 Introduction

Leading economic indicators can be used to predict the turns of the business cycles. The turning point time of the general business cycle can be predicted by timely detection of the turns in a leading indicator. In this paper, predictions of the actual value of the business cycle are not discussed. Instead the methods discussed here are concerned with detecting the turning point time, i.e. a change from a recession phase to an expansion phase (or vice versa). Predictions of the turning point times can be made by using information from one or several time series, which are leading in relation to the actual business cycle. By applying a system for detecting turning points in a leading indicator we can receive early signals about the future behavior of the business cycle. For reviews and general discussions see e.g. Neftci (1982), Zarnowitz and Moore (1982), Westlund and Zackrisson (1986), Hackl and Westlund (1989), Zellner et al. (1991), Jun and Joo (1993), Lahiri and Wang (1994), Li and Dorfman (1996) and Birchenhall et al. (1999).

Methods based on likelihood or posterior distributions have been in focus in recent years. In the general theory on statistical surveillance there are proofs for their optimality properties (see e.g. Shiryayev (1963) and Frisén and de Maré (1991)). Many of the suggested methods are based on a hidden Markov model (HMM).

In a surveillance situation repeated decisions are made, which is important to consider when the inference is made. Continual observation is made of a time series with the goal of detecting the turning point in the underlying process as soon as possible. For general reviews on statistical surveillance, see Shiryayev (1963), Frisén and de Maré (1991), Wetherhill and Brown (1991), Srivastava and Wu (1993), Lai (1995), Frisén and Wessman (1999) and Frisén (2003).

In this paper some differences and similarities between suggested likelihood based methods for turning point detection of business cycles are shortly described. Furthermore, there is a short description of how the methods perform when used to analyze an actual data set of a period of the Swedish industrial production. For details see Andersson et al. (2002a) and Andersson et al. (2002b).

The paper is organized as follows. Section 2 contains a description of different likelihood based approaches. In Section 3 the results of the Monte Carlo study conducted in Andersson et al. (2002b) in order to compare the performance of three different likelihood based methods are briefly summarized. In Section 4 the three methods are used to analyze a period of the Swedish industrial production and the pros and cons of this way to evaluate methods are discussed. Section 5 contains a summarizing discussion.

2 Likelihood Based Surveillance for Detection of Turning Points

2.1 Model for each state

The variable under surveillance, X , is a leading economic indicator. By monitoring X we want to detect a regime shift (a turning point) as soon as possible. The model for X discussed here is:

$$X(t) = \mu(t) + \varepsilon(t), \quad (1)$$

where $\varepsilon(t) \sim \text{iid } N[0; \sigma^2]$ and $\mu(t)$ is a stochastic process.

The aim is to detect the change in μ , from expansion state to recession state (or vice versa). The definition of a turning point is that the regression of μ on time is monotonic within each regime:

$$E[X_t] = \mu_t : \begin{cases} \mu(1) \leq \mu(2) \leq \dots \leq \mu(t), & t < \tau \\ \mu(1) \leq \dots \leq \mu(\tau-1) \text{ and } \mu(\tau-1) \geq \dots \geq \mu(t), & t \geq \tau \end{cases} \quad (2)$$

where $\mu_t = \{\mu(1), \mu(2), \dots, \mu(t)\}$, $t < \tau$ is in a period of expansion and τ is the turning point from the expansion to a recession. The turning point time, τ , is stochastic and hence μ is stochastic.

Often a parametric assumption is used, for example that the regression is linear within each phase

$$E[X(t)] = \mu(t) = \begin{cases} \beta_0 + \beta_1 t, & t < \tau \\ \beta_0 + \beta_1(\tau-1) - \beta_2(t-\tau+1), & t \geq \tau \end{cases} \quad (3)$$

where $t = \{1, 2, \dots\}$. A consequence is that the expected value of the differentiated process is assumed to be constant, conditional on the state:

$$E[X(t) - X(t-1)] = \begin{cases} \beta_1, & t < \tau \\ \beta_2, & t \geq \tau \end{cases} \quad (4)$$

where β_1 and β_2 often are assumed known.

The standard deviation σ can be assumed to be different for recession and expansion. Here, the variance is assumed equal, as also in Andersson (2002).

2.2 Specifications of the events to be detected

The unknown time when the underlying process μ changes from expansion to recession (a peak), or vice versa is denoted τ . At decision time s an alarm statistic is used to discriminate between $D(s) = \{\tau > s\}$ and $C(s) = \{\tau \leq s\}$. Knowledge of whether the next turn will be a trough or a peak is assumed. For a method that relies only on the assumption of monotonicity restrictions, the aim is to discriminate between the following two events:

$$D(s) : \mu(1) \leq \dots \leq \mu(s) \quad (5)$$

$$C(s) : \mu(1) \leq \dots \leq \mu(\tau-1) \text{ and } \mu(\tau-1) \geq \mu(\tau) \geq \dots \geq \mu(s)$$

where $\tau = \{1, 2, \dots, s\}$ and at least one inequality is strict in the second part.

Often it is assumed that the regression consists of linear functions. This assumption (which is the base for two of the examined methods, but not for the third) is not very realistic and the risks in connection with it are discussed in Section 3.2. The main purpose here is not to model the economic process as accurately as possible, but to study the effect of different assumptions for likelihood based methods. Thus the events D and C , under the often made assumption of the functions being linear within each phase and the additional assumption of the slopes being symmetrical for the two phases ($|\beta_1| = |\beta_2|$), are such that

$$D(s) : \mu(s) = \beta_0 + \beta_1 s \quad (6)$$

$$C(s) = \{\cup \cdot C(\tau)\},$$

where $C(\tau) : \mu(s) = \beta_0 + \beta_1(\tau-1) - \beta_1(s-\tau+1)$ and where $\tau = \{1, 2, \dots, s\}$ and β_0 and β_1 are known constants.

If an HMM is assumed, the situation is such that at decision time s an alarm statistic is used to discriminate between

$$D(s) : \mu(s-1) \leq \mu(s), \quad (7)$$

$$C(s) : \mu(s-1) > \mu(s).$$

When comparing (5) and (6) the difference between D and C is only the assumptions on $\mu(t)$. However, when a HMM is used, as in (7), the events are different also in another aspect. The apparently simpler event in the HMM approach is combined with a more complicated situation for the information of previous states. No knowledge of previous states is utilized in a HMM approach. Thus the events in (7) correspond to families of histories of states.

Because of Markov dependence, the earlier observations carry information of the history of states.

2.3 Assumptions regarding transition probabilities

The common assumption of time invariant transition probabilities between the states is made for all three methods investigated here.

The information about the time and type (peak or trough) of the last turning point can be utilized, as in (5) and (6) and Neftci (1982). Then it is not necessary to involve both types of transition probabilities (both ν_{12} and ν_{21}) in the alarm system. It is sufficient with the transition probability

$$\nu = P(C(t)|D(t-1)) = P(\tau = t | \tau \geq t), \quad (8)$$

where the value of ν can differ depending on the type of transition.

When the information about the time and type of the last turning point cannot be used, it is necessary to make a probability statement regarding the type of the next turning point as well as inference about whether the turning point has occurred. For this purpose, it is necessary to consider all previous possible turns. Therefore, two intensities are needed. The intensity parameters can be given as transition probabilities from expansion state to recession state and vice versa.

One approach which will be described below is to use a non-informative prior for the turning point time.

2.4 Alarm statistics and alarm limits

The full likelihood ratio method (LR) has several optimality properties, see Frisén and de Maré (1991). The likelihood ratio method yields a minimal expected delay of an alarm for a fixed probability of false alarm. The full likelihood ratio method signals an alarm for the first time s for which

$$LR(s) = \frac{f(x_s|C)}{f(x_s|D)} > k_s, \quad (9)$$

where f is the likelihood function and $k_s = (k/(1-k)) \cdot (P(D(s))/P(C(s)))$.

For the situation where C is the complement of D , Frisén and de Maré (1991) demonstrated that the posterior probability approach is equivalent to the likelihood ratio approach. All methods considered here use a likelihood ratio based alarm statistic.

The full likelihood ratio requires an assumption about τ . When no reliable information is available a non-informative prior can be used. This is sometimes named the Shiryaev-Roberts (SR) approach.

For the situation where $\mu(t)$ is modeled using linear functions with a symmetric turning point (see (6)) and where the SR approach is used, the alarm rule for the likelihood ratio at time s is written as

$$\sum_{t=1}^s \frac{f(x_s | \mu = \mu^{Ct})}{f(x_s | \mu = \mu^D)} > k_1 \quad (10)$$

where k_1 is a constant alarm limit. This method is hereafter referred to as the SRlin method.

Frisén (1994) suggested a method where no knowledge of μ is assumed, other than the monotonicity restrictions in (5), and where the SR approach regarding the intensity is used. This method was evaluated by Andersson (2002) and Andersson et al. (2002b). The alarm rule is

$$\sum_{t=1}^s \frac{f(x_s | \mu = \hat{\mu}^{Ct})}{f(x_s | \mu = \hat{\mu}^D)} > k_2, \quad (11)$$

where k_2 is a constant alarm limit, $\hat{\mu}^D$ is the maximum likelihood estimator of the vector μ under the monotonicity restriction D (see e.g. Robertson et al. (1988)) and $\hat{\mu}^{Ct}$ is the maximum likelihood estimator of the vector μ under the monotonicity restriction Ct (see Frisé (1986)). This method is hereafter referred to as the MSR method.

In Hidden Markov Modeling (HMM) approaches, e.g. Neftci (1982), Hamilton (1989), Layton (1996), Koskinen and Öller (2003) and Kim and Nelson (1998) the posterior probability is used as the alarm statistic. Thus, the method signals an alarm as soon as

$$P(C(s)|x_s) > c, \quad (12)$$

where c often is chosen to 0.5. The method where this alarm rule is used and where the expected value of the differentiated series is assumed to be constant, conditional on the regime, is hereafter referred to as the PHM method.

In the literature on surveillance the constants are often set so that the false alarms are under control. The most common way is to control the ARL^0 (the Average Run Length to the first false alarm). Another way, which is used here, is to control the MRL^0 , which is the median run length.

In much of the theoretical work, e.g. Shiryayev (1963) and Friséen and de Maré (1991) and especially where optimality theorems are available, the false alarms are measured by the total probability of false alarm

$$P(t_A < \tau) = \sum_{i=1}^{\infty} P(\tau = i) \cdot P(t_A < i | D). \quad (13)$$

3 Evaluation of Likelihood Based Methods by a Monte Carlo Study

Andersson et al. (2002b) report a Monte Carlo study of the performance of the SRlin, MSR and PHM methods. The comparison was made with respect to the delay of motivated alarms and the predictive value of an alarm. The methods were made comparable by adjusting the alarm limit so that a false alarm property (MRL^0) is equal. Two situations were investigated: i) when the correct parameter values were used and ii) when the regression is miss-specified. In this section a brief summary of the results from the study is made.

3.1 Correct parametric specifications

If the parametric model and the parameter values used by the SRlin and PHM methods are completely known then it is an advantage to use this information. As expected the SRlin method has a shorter expected delay than the MSR method.

The PHM method does not use information on earlier turns and thus has worse properties than the SRlin method (and in some respects also than the MSR method). The conditional (with respect to τ) expected delay of a motivated alarm is highest for MSR, for very small values of τ as is seen in Table 1. The MSR method needs more observations to have enough evidence of a turn. After that, the delay is slightly shorter for MSR, compared to PHM. For PHM, the effect of τ is very small. The SRlin method has the shortest delay for every τ .

Table 1 **Conditional Expected Delay, for different times τ of the turn**

Turning point occurs after	Conditional Expected Delay		
	SRlin	MSR	PHM
1 quarter	1.5	3.5	1.7
8 quarters	1.2	1.7	1.8

The PHM method has more frequent false alarms at early time points, but low alarm probability later compared with that of the SRlin and especially MSR. The total probability of a false alarm, $P(t_A < \tau)$, is investigated for a geometric distribution with intensity ν . When

comparing the methods, MSR has the smallest false alarm probability, whereas PHM has the largest for every value of the intensity, ν . As a result of the assumption of a geometric distribution for τ , the alarms at the beginning have a large influence on $P(t_A < \tau)$. The large false alarm probability for PHM for all ν is a result of the error-spending curve of PHM, with many early alarms.

The predictive value of an alarm at time t , $PV(t) = P(\tau \leq t | t_A = t)$ reflects the trust you should have in an alarm. Under the assumption of a geometric distribution with intensity $\nu = 0.1$, it is shown in Table 2 that the price for the high alarm probability in the first point for the PHM method is that those alarms are of little value. $PV(1)$ is high for MSR, as a result of the very low false alarm rate at $t=1$. The MSR method places no parametric restrictions on the turning point curve. All information about the curve comes from data. For small values of t the number of observed data is very small and thus the data have to be very extreme in order to call an alarm. However, as t increases (and the number of observations increases) the information about the curve is improved.

Table 2 **Predictive value, depending on the time of the alarm**

Alarm at time	Predictive value		
	SRlin	MSR	PHM
1 quarter	0.5	0.5	0.2
8 quarters	0.8	0.8	0.8

3.2 *Parameters miss-specified*

In practice neither the regression model nor the values of the parameters are known. If the slopes (expansion- and recession-) are estimated from a short period, then the parameters might be very wrongly specified. Both the SRlin method and the PHM method use information from previous expansion and recession phases in the alarm statistic and for parameter estimation. We have determined (by simulations) the 95% confidence limits for the parameters when estimated by a procedure used in practice. We have then examined the effect of errors in parameters when using the confidence limit instead of the true value.

If only the slope after the turning point (the post-peak slope) is wrongly specified, then there are only minor effects on the false alarms, the delay of a motivated alarm and the predictive value of an alarm. The predictive value is very similar between the correctly specified post-peak slope and the miss-specified post-peak slope, except at $t=1$. The difference is due to the difference in the error-spending curve. When the post-peak slope is specified as being too flat, then the alarm statistic is optimized to detect a smaller change (flatter post-peak slope) and therefore the alarms are located later on. The result is few

alarms at early time points, which results in a high predictive value for $t=1$. The opposite holds for a too steep post-peak slope.

However, wrong specification of also the pre-peak slope has a major influence. For incorrect specification with both slopes being too steep, the false alarm probability is very low for small t . This is due to the increasing difference between the true and specified states. For small t this difference is small. However, as t increases, so does the difference. Thus, the likelihood for the specified D -state decreases and therefore, the alarm probability increases. The low false alarm probability is a result of the error-spending curve with few early alarms, in contrast to using correctly specified slopes that result in many early alarms.

For small values of τ , the delay is longer if both slopes are incorrectly specified so that both slopes are too steep. This is seen in Table 3. As τ increases, the conditional delay decreases towards an asymptote, zero. Thus for the situation where both slopes are misspecified, the resulting delay is large unless the turning point occurs very late. In contrast, the non-parametric approach, MSR, has a long delay time for extremely early turns, but quickly reaches a reasonable asymptotic value.

Table 3 Conditional expected delay for SRlin: parameters correctly specified and parameters miss-specified in both expansion- and recession-parameters (too steep)

Turning point occurs after	CED	
	SRlin	SRlin, miss-specified
1 quarter	1.5	2.5
8 quarters	1.2	1.6
20 quarters	1.2	0.5

The conclusion is that wrong assumptions about slopes do have a major effect on the properties and that the MSR method gives a safe way to avoid this.

4 Using Methods for Turning Point Detection on Real Data

When suggesting a new statistical method, one often wants to evaluate the method and establish its strong and weak sides. If the evaluation is made by a Monte Carlo study, it is possible to make statements about, for example, the standard errors of the evaluation measures. However, ultimately the method will be used on real data, which presents some additional problems, such as seasonal variation, presence of trend, autocorrelation and noise, how to handle the situation with several cyclical processes. In the paper by Andersson et al. (2002a), these problems are also discussed and here a brief description is given.

4.1 Special data problems

4.1.1 Smoothing

The smoothing of observations reduces the variance and hence reduces the false alarm probability. The alarm limit 0.5 is often used for the posterior probability. This will not correspond to a desired false alarm rate. Therefore one suggestion for reducing the false alarm rate is to smooth the observations. Öller (1986) and Koskinen and Öller (2003) use exponential weights. Hall et al. (1995) use smoothing by kernel estimators. However, there are also disadvantages. The smoothing will introduce an auto-correlation and reduce the distinctness of the turning point. The surveillance method EWMA (see e.g. Crowder (1987), Domangue and Patch (1991) and Frisén and Sonesson (2001)), which is designed for a controlled false alarm rate at the same time as the variability is reduced by smoothing, might be an alternative.

In a Monte Carlo study (Andersson et al. (2002a)), the effect of smoothing is investigated for the PHM method. The skewness of the density of the time of the alarm is different depending on whether the data is smoothed or not. It was shown that the reduced distinctness of the turning point, due to smoothing, decreases the probability of successful detection. Thus, we do not recommend smoothing as a separate step before an analytical method is applied.

4.1.2 Seasonality

Leading economic indicators are often measured monthly or quarterly and thus often contain seasonal variation, which could complicate the monitoring. The seasonal variation can be considerable. Seasonality cannot be neglected in the monitoring without the risk of seriously wrong conclusions. However, when adjusting for seasonality, it is important that the structure of the original series is not disturbed and that the time of the turning point is preserved (see e.g. Ghysel and Perron (1996) and Franses and Paap (1999)).

It is demonstrated, in Andersson and Bock (2001), that in a surveillance situation the detection is delayed when a data transformation such as differentiating or moving average is used. Thus, the method for seasonal adjustment should be chosen with care.

4.1.3 Autocorrelation

Lahiri and Wang (1994) evaluate the performance of models with autocorrelation of order $r = \{0, 1, 2, 3, 4\}$. They find that the introduction of autocorrelation in the errors increases the risk of wrong inference concerning turning points. If the assumption of a time dependent stochastic term is used, when the process is in fact autocorrelated, the result is an increased false alarm probability. This effect of autocorrelation can be dealt with by adjusting the alarm limit. It is sometimes argued (see Ivanova et al. (2000)) that the effect of the autoregressive

parameters will largely be captured by the probabilities of remaining in the current state. Most suggested approaches for the detection of turns in business cycles assume independent observations conditionally on the state.

4.1.4 Trend

Many macroeconomic variables can be characterized as cyclical movements around a trend and sometimes it is useful to adjust for the trend. Adjusting for trend implies a data transformation, which may result in a distortion of the characteristics of the original series. Gordon (1997) warns against using other information from the data than that which is directly associated with the business cycle turning points. Trend removal is discussed by Canova (1998) and Canova (1999) and one conclusion is that statements concerning the turning points are not independent of the statistical assumptions needed to extract trends. Also, the turning points in classical and growth cycles have different interpretations.

In most HMM approaches and also the one considered here (PHM), differentiation is used as a way of adjusting for the trend. In the HMM approaches a long series is used in order to extract information about all earlier turns. However for the SRLin and MSR methods the knowledge on the last turn is utilized and thus the need for trend adjustment is less since only a short time series is analyzed.

4.1.5 Several leading indicators

By the common movement approach (Stock and Watson (1991) and Stock and Watson (1993)), a business cycle is characterized as the cyclical movement of many economical activities. Important information is contained in the relation between the turns of different indices. The common movements of coincident variables arise from an unobservable common factor (the overall state of the economy). Diebold and Rudebusch (1996) consider the common movements of coincident variables where the common factor depends on a hidden Markov chain with two states. Birchenhall et al. (1999) extract a business-cycle index from a vector of time series.

In Koskinen and Öller (2003) a joint vector of leading indicators with a common time of turn is monitored. When several variables have the same change point (or a known lag) Wessman (1998) demonstrates that the minimal sufficient alarm statistic is univariate. Thus a reduction to a univariate statistic is possible. For reviews on multivariate surveillance, see Wessman (1999) and Frisén (2003).

4.2 *Evaluation by data on the Swedish industrial production*

A common way to evaluate a method for turning point detection is to apply it on a set of real data. This is done in Andersson et al. (2002a) and here a brief summary of the results is given.

Seasonally adjusted quarterly data on the (logarithm of the) Swedish industrial production is used, from the period 1987Q2:1992Q2 (see Figure 1). For this period, according to official records (National Institute of Economic Research (1992)), the transition from expansion to recession occurs after 11 quarters (almost 3 years). All three methods (MSR, PHM and SRlin) give alarms earlier than the official times for this set of data, see Table 4.

Figure 1 **Quarterly data on the Swedish industrial production, adjusted for seasonality. The official time of change (according to National Institute of Economic Research) is marked with a vertical line.**

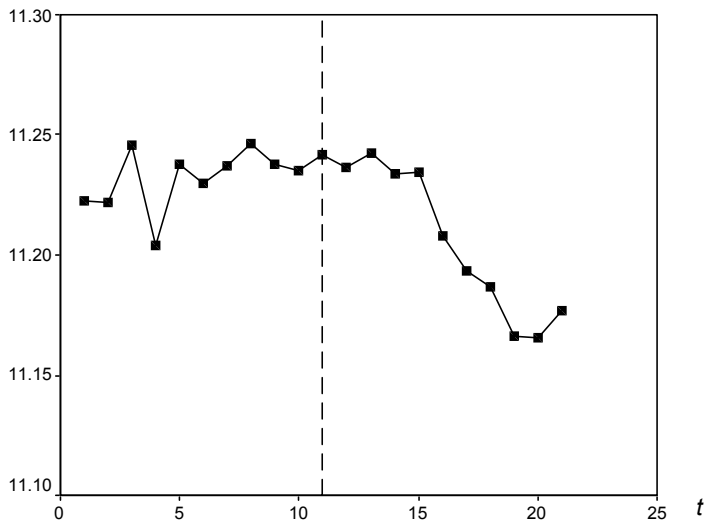


Table 4 **Time of alarm, in relation to official turning point time**

	SRlin	MSR	PHM
Time of alarm, t_A	10	4	7
Alarm time, in relation to turn, $t_A - \tau$	-1	-7	-4

SRlin and PHM use the assumption of a piecewise linear model, which fits less well since there is a plateau just before the turning point. McQueen and Thorley (1993) discuss this common phenomenon. If a plateau is an early indication of a coming recession, then the early alarms for the three methods can be considered to be an advantage, since they can be seen as warnings.

The model in a simulation study might not be representative of the process we want to study, whereas an actual data set certainly is representative. However, the actual data set might deviate stochastically from the process of interest. One difficulty with evaluation of the properties of a method by one sample of real data is the definition of the turning point. Another difficulty is to know whether the sample is an extreme result or typical. For a short period as the one used here this is very obvious. By using several real data sets (instead of just one) in the evaluation, the stochastic variation in the measures of evaluation would be reduced. However, these analyses must be independent (for example regarding the estimates of the parameters), in order to reduce the variance of the stochastic components.

5 Discussion

Three likelihood based methods (MSR, PHM and SRlin) have been presented. Differences and similarities between the methods have been pointed out.

The PHM method (contrary to the other two) does not assume knowledge of the type of the next turn, which has a major impact on the test statistic. Because this information is not used, there are frequent false alarms at early time points, but low alarm probability later. The results of this allocation of the alarm probability are short delay for alarm for early changes, but long delays for changes that occur late.

The intensity of a turning point is an important parameter. The most common assumption is that the intensity is constant and that the transition probabilities are constant over the cycles. A constant transition probability implies that the time of the turn has a geometric distribution, with the highest values at the early times, which is not in accordance with reality for business cycles. If the intensity is estimated from historical data, then the impact of the actual data would be reduced. It is very important that the monitoring system has the ability to detect a turning point also when this happens at an unexpected time. Thus, here we prefer to use a non-informative prior for the time of the turn (the SR approach).

Using a parametric model improves the performance of a method if the parameter information is reliable. But a wrongly specified parametric model has serious effects. The non-parametric approach, which only uses the change in monotonicity, has the advantage that it works also when reliable information on the parametric function is not available. Also important is that the non-parametric method does not assume that all expansion phases (or recession phases) have the same level and parametric shape. Thus, the very bad properties demonstrated for wrong specification of the slope give a warning. The safe way by the MSR method might be preferred.

A period of the Swedish industrial production is analyzed in Section 4, in order to compare the three methods. For this data set all three methods call an alarm before the turn, largely due to the plateau just before the transition. This is not necessarily a disadvantage. The drawback of evaluation by real data is that one is limited to one or a few replicates. Thereby it is hard to make inference about the certainty of the results.

References

- Andersson, E. (2002) Monitoring cyclical processes - a nonparametric approach. *Journal of Applied Statistics*, 29, 973-990.
- Andersson, E. and Bock, D. (2001) On seasonal filters and monotonicity. Research Report, Department of Statistics, Göteborg University, Sweden, 2001:4
- Andersson, E., Bock, D. and Frisén, M. (2002a) Some statistical aspects on methods for detection of turning points in business cycles. Research Report, Department of Statistics, Göteborg University, Sweden, 2002:7
- Andersson, E., Bock, D. and Frisén, M. (2002b) Statistical surveillance of cyclical processes: Detection of turning points in business cycles. Research Report, Research report 2002:8, Department of Statistics, Göteborg University, Sweden, 2002:8
- Birchenhall, C. R., Jessen, H., Osborn, D. R. and Simpson, P. (1999) Predicting US business-cycle regimes. *Journal of Business & Economic Statistics*, 17, 313-323.
- Canova, F. (1998) Detrending and business cycle facts. *Journal of Monetary Economics*, 41, 475-512.
- Canova, F. (1999) Does detrending matter for the determination of the reference cycle and the selection of turning points? *The Economic Journal*, 109, 126-150.
- Crowder, S. (1987) A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics*, 29, 401-407.
- Diebold, F. X. and Rudebusch, G. D. (1996) Measuring business cycles: A modern perspective. *Review of Economics and Statistics*, 78, 67-77.
- Domangue, R. and Patch, S. C. (1991) Some omnibus exponentially weighted moving average statistical process monitoring schemes. *Technometrics*, 33, 299-313.
- Franses, H. P. and Paap, R. (1999) Does Seasonality Influence the Dating of Business Cycle Turning Points? *Journal of Macroeconomics*, 21, 79-92.
- Frisén, M. (1986) Unimodal regression. *The Statistician*, 35, 479-485.
- Frisén, M. (1994) Statistical Surveillance of Business Cycles. Research Report, Department of Statistics, Göteborg University, Sweden, 1994:1 (Revised 2000)
- Frisén, M. (2003) Statistical surveillance. Optimality and methods. *International Statistical Review*, 71, 403-434.
- Frisén, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika*, 78, 271-80.
- Frisén, M. and Sonesson, C. (2001) Optimal surveillance by EWMA. Research Report, Department of Statistics, Göteborg University, Sweden, 2000:7
- Frisén, M. and Wessman, P. (1999) Evaluations of likelihood ratio methods for surveillance: Differences and robustness. *Communications in Statistics. Simulations and Computations*, 28, 597-622.
- Ghysel, E. and Perron, P. (1996) The Effect of Linear Filters on Dynamic Time Series With Structural Change. *Journal of Econometrics*, 70, 69-97.
- Gordon, S. (1997) Stochastic trends, deterministic trends, and business cycle turning points. *Journal of Applied Econometrics*, 12, 411-434.
- Hackl, P. and Westlund, A. H. (1989) Statistical analysis of "Structural Change". *Empirical Economics*, 14, 167-192.
- Hall, P., Marron, J. S. and Titterton, D. M. (1995) On partial local smoothing rules for curve estimation. *Biometrika*, 82, 575-587.
- Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357-384.
- Ivanova, D., Lahiri, K. and Seitz, F. (2000) Interest rate spreads as predictors of German inflation and business cycles. *International Journal of Forecasting*, 16, 39-58.
- Jun, D. B. and Joo, Y. J. (1993) Predicting turning points in business cycles by detection of slope changes in the leading composite index. *Journal of Forecasting*, 12, 197-213.
- Kim, C.-J. and Nelson, C. R. (1998) Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80, 188-201.

- Koskinen, L. and Öller, L.-E. (2003) A Classifying Procedure for Signalling Turning Points. To appear in *Journal of Forecasting*.
- Lahiri, K. and Wang, J. G. (1994) Predicting cyclical turning points with a leading index in a Markov Switching model. *Journal of Forecasting*, 13, 245-263.
- Lai, T. L. (1995) Sequential Change-point Detection in Quality-Control and Dynamical-Systems. *Journal of the Royal Statistical Society B*, 57, 613-658.
- Layton, A. P. (1996) Dating and predicting phase changes in the U.S. business cycle. *International Journal of Forecasting*, 12, 417-428.
- Li, D. T. and Dorfman, J. H. (1996) Predicting turning points through the integration of multiple models. *Journal of Business and Economic Statistics*, 14, 421-428.
- McQueen, G. and Thorley, S. (1993) Asymmetric business cycle turning points. *Journal of Monetary Economics*, 31, 341-362.
- National Institute of Economic Research (1992) Konjunkturläget Maj 1992. Research Report, National Institute of Economic Research, Stockholm, Sweden.
- Neftci, S. (1982) Optimal prediction of cyclical downturns. *Journal of Economic Dynamics and Control*, 4, 225-41.
- Öller, L.-E. (1986) A note on exponentially smoothed seasonal differences. *Journal of Business and Economic Statistics*, 4, 485-489.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) Order Restricted Statistical Inference, John Wiley & Sons Ltd.
- Shiryayev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 8, 22-46.
- Srivastava, M. S. and Wu, Y. (1993) Comparison of EWMA, CUSUM and Shirayayev-Roberts Procedures for Detecting a Shift in the Mean. *Annals of Statistics*, 21, 645-670.
- Stock, J. H. and Watson, M. W. (1991) A Probability Model of the Coincident Economic Indicators. In *Leading Economic Indicators: New Approaches and Forecasting Records* (Eds, Lahiri, K. and Moore, G. H.) Cambridge University Press, Cambridge, pp. 63-89.
- Stock, J. H. and Watson, M. W. (1993) A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience. In *Business Cycles, Indicators and Forecasting*, Vol. 28 (Eds, Stock, J. H. and Watson, M. W.) University of Chicago Press, Chicago, pp. 95-156.
- Wessman, P. (1998) Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods*, 27, 1143-1161.
- Wessman, P. (1999) Studies on the surveillance of univariate and multivariate processes. Department of Statistics, Göteborg University, Sweden, Göteborg, pp. 103.
- Westlund, A. H. and Zackrisson, U. (1986) On the prediction of structurally varying systems. *Technological Forecasting and Social Change*, 30, 63-72.
- Wetherhill, G. B. and Brown, D. W. (1991) *Statistical Process Control: Theory and Practice*, Chapman and Hill, London.
- Zarnowitz, V. and Moore, G. H. (1982) Sequential signals of recessions and recovery. *Journal of Business*, 55, 57-85.
- Zellner, A., Hong, C. and Min, C.-K. (1991) Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Pooling Techniques. *Journal of Econometrics*, 49, 275-304.

Coincident and Leading Indicators of Manufacturing Industry

Sales, Production, Orders and Inventories in Switzerland

Michael Graff^{*} and Richard Etter^{**}

Abstract

The Swiss Institute for Business Cycle Research regularly conducts business tendency surveys (BTS) amongst manufacturing firms. The information thus generated is available with a publication lead to the official Swiss sales, production, order and inventory statistics. It is shown that the survey data can be used to generate reasonably precise estimates of the reference series with leads of at least one quarter. Specifically, cross-correlations of quarterly series are computed to screen the data for pairs of highly correlated business tendency survey series and corresponding official statistics. All pairs are identified where the maximum correlation shows up simultaneously or with a lead of the survey based series and exceeds a given threshold. We then determine suitable subsets of indicator series with predetermined leads and compute their principal components. The resulting coincident as well as leading composite indicators for the reference series are closely correlated with the reference series from 1990–2000 and the correlations generally seem to hold for the out-of-sample period from 2001 to the present. Hence, apart from early availability, the survey data indeed comprises information on the Swiss economy's track in the near and medium term future. While the data refers to Switzerland, the methodology can readily be applied in other countries.

Key Words: Composite business cycle indicators; Principal components

JEL Classification: E37

^{*}

Reserve Bank of New Zealand, Wellington, New Zealand; graffM@rbnz.govt.nz

^{**}

Swiss Institute for Business Cycle Research (KOF), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland; etter@kof.gess.ethz.ch

Résumé

Le Centre de recherches conjoncturelles (KOF) en Suisse mène régulièrement des enquêtes de conjoncture auprès des entreprises manufacturières. Les informations qui en sont retirées sont disponibles avant la publication des statistiques officielles suisses des ventes, de la production, des commandes et des stocks. Les auteurs démontrent que les données issues des enquêtes autorisent des estimations raisonnablement précises des séries de référence permettant d'anticiper leurs fluctuations d'au moins un trimestre. A cet effet, ils calculent les corrélations croisées entre les séries trimestrielles afin de repérer les données appariées dans les séries issues des enquêtes de conjoncture et les statistiques officielles correspondantes. Toutes les paires où la corrélation atteint son niveau maximum et dépasse un seuil donné lorsque l'on considère les données relatives à la même date dans les deux séries ou celles concernant une période dans les séries officielles et une période antérieure dans la série dérivée des enquêtes sont ainsi mises en évidence. On détermine ensuite des sous-ensembles de séries d'indicateurs avec des horizons de prévision prédéterminés et on en calcule les principales composantes. Les indicateurs coïncidents et les indicateurs composites avancés ainsi obtenus pour les séries de référence présentent une forte corrélation avec ces séries de référence pour la période 1990-2000, corrélation qui semble valoir également pour la période allant de 2001 à aujourd'hui. En bref, outre le fait qu'elles sont disponibles avant les séries officielles, les données tirées des enquêtes donnent de fait des renseignements sur la trajectoire économique de la Suisse dans l'avenir proche et à moyen terme. Si les données se rapportent à la Suisse, la méthodologie peut facilement être reprise en vue de mener des analyses comparables pour d'autres pays.

1 Introduction¹

Industry sales, production, orders and inventories are key ingredients in assessing business conditions. Unfortunately, the usefulness of official data is usually hampered by a considerable compilation and publication lag. Hence, a fundamental issue for policy oriented research is to provide reasonably reliable coincident and – if possible – leading indicators. To this end, the Swiss Institute for Business Cycle Research (KOF)² is regularly conducting business tendency surveys (BTS) amongst Swiss manufacturing firms.

There is an obvious trade-off between early availability and precision of economic indicators. Specifically, business tendency surveys reveal first information as early as possible, whereas final official statistics are supposed to come as close as possible to the realisation of the economic process.³

Assuming that business tendency surveys and official statistics are the only two sources of information on a particular economic process, at any point t in time an observer can refer to the following set of information:

$$I_t = \{BTS(t)_t, BTS(t + \lambda)_t, OS(t - \mu)_t\},$$

where $BTS(t)$ refers to information from business tendency surveys on present conditions, $BTS(t + \lambda)$ to information from business tendency surveys on conditions with a lead of λ , and $OS(t - \mu)$ to final official statistics on past conditions with lag μ ($\lambda, \mu > 0$). The true, but ultimately unknown, realisation of economic conditions T can be represented by the elements of I as

$$\begin{aligned} T_t &= OS(t)_{t+\mu} + \varepsilon_{OS(t)_{t+\mu}} \\ &= f[BTS(t)]_t + \varepsilon_{BTS(t)_t} \\ &= g[BTS(t)]_{t-\lambda} + \varepsilon_{BTS(t)_{t-\lambda}}, \end{aligned}$$

where $\varepsilon_{BTS(t)_t}$ and $\varepsilon_{BTS(t)_{t-\lambda}}$ are the estimation and forecast errors for T_t from coincident and leading business tendency survey indicators, and $\varepsilon_{OS(t)_{t+\mu}}$ denotes the final error in the official data.

¹ We are indebted to Nils Bjorksten and to an anonymous referee for comments and suggestions. All remaining errors, of course, are ours, and the views expressed in this paper do not necessarily reflect those of the Reserve Bank of New Zealand or the Swiss Institute for Business Cycle Research.

² This acronym is derived from German "Konjunkturforschungsstelle".

³ Of course, since the compilation of statistics is costly, there is an optimum level of ignorance, which implies that even final official statistics do not necessarily come as close to the real world as technically possible. Nonetheless, official statistics certainly rely on a larger set of information than business tendency surveys.

Now, if both business tendency surveys and official statistics refer to the same empirical representations, they should be statistically correlated. Given this, since the survey data are available before their corresponding official statistics, it is important to have an idea about the extent to which the official data can be derived as functions $f[\text{BTS}(t)]$ from coincident survey indicators or $g[\text{BTS}(t+\lambda)]$ from leading survey indicators, respectively.

Referring to this framework, our paper analyses how the information from the KOF business tendency surveys is related to the industrial statistics generated and published by the Swiss Federal Statistical Office (henceforth SFSO). Specifically, we outline a procedure to select a limited number of transformed series from the survey data to be combined into principal components that qualify as coincident or leading indicators of our five reference series – official data on sales, production, incoming orders, order backlog and inventories.⁴ Finally, we use the resulting indicator sets of the within-sample analyses of quarterly series from 1990:1 to 2000:4 and extend the estimations to simulated real-time estimations and forecasts for an out-of-sample period from 2001:1 to 2002:4. The out-of-sample performance is encouraging, and we conclude that the survey indicators are indeed a viable basis for the computation of coincident and leading indicators. Finally, note that while the data analysed in this paper refers to Switzerland, the suggested methodology can readily be implanted for comparable analyses in other countries.

2 Data and Univariate Transformations

Our reference series are the final⁵ official Swiss index series on industrial sales, production, orders and inventories, published quarterly by the SFSO ("Der Geschäftsgang im sekundären Sektor: Die Produktions-, Auftrags-, Umsatz- und Lagerindizes").

These series are given for various levels of aggregation according to the NOGA classification (corresponding down to the 4-digit-level to the NACE, Rev. 1, of the EU), and in many cases, we were able to find a matching sub aggregate in the KOF survey data. An extensive screening of the data by means of some 10,000 cross correlations revealed, however, that – for whatever transformation of either series – the narrower the aggregate, the lower is the correlation between the official statistics and the survey data (Etter/Graff 2001). From this strong regularity, we infer that a considerable amount of the noise in the 2- and 3-

⁴ In an earlier paper (Etter/Graff 2003), we conducted single and multiple indicator regression analyses to estimate the growth rates of the production and incoming orders indices. The predictive accuracy of these estimations, however, is impaired by multicollinearity between the regressors. In this paper, multicollinearity is not affecting the estimations, since principal component analysis is by construction identifying the common variance of a set of variables.

⁵ Before 2003, the official series were revised (sometimes substantially) up to four times. This year, the SFSO changed its publication policy and announced that from now on it will revise the last quarter only. Hence, compared to previous analyses (Etter/Graff 2001, 2003), we can now draw on comparatively long and up-to-date final official series. Given the fact that the provisional series now comprise only one additional quarter, we do no longer make an analytical distinction between provisional and final official series and analyse the final series only.

digit categories is cancelled out in the aggregate data. Therefore, we restrict this presentation to the aggregate level NOGA D (manufacturing industry), a subset of the secondary sector.⁶

Following their present definitions, SFSO data on the manufacturing industry have not been published before 1996. However, the last six years of the pre-1996 data have been recalculated within the statistical office and adapted to the current standards, so that there are now official index series going back as far as 1990:1. At the time of finishing this paper, the latest final official data (published at the end of June 2003) refer to 2002:4. Presently, we can thus refer to 13 years of industrial statistics, i.e. 52 quarterly observations.

For our within sample analyses, we draw on the final SFSO index series from 1990:1 to 2000:4, i.e. on 44 observations. Since in a growing economy, these series are non-stationary, and in addition to this, they are affected by seasonal factors, we refer to year-on-year growth rates (GR_Y), i.e. 43 transformed observations of the reference series (*Y*) only, where

$$GR_Y = \ln Y_t - \ln Y_{t-4}$$

The five reference series and their variable names are given in Table 1.

Table 1 **The reference series (n = 43)**

Series	Description
GR_SALES	Year-on-year growth rate of industrial sales index
GR_PRODU	Year-on-year growth rate of industrial production index
GR_I_ORD	Year-on-year growth rate of industrial incoming orders index
GR_ORD_B	Year-on-year growth rate of industrial order backlog index
GR_INVEN	Year-on-year growth rate of industrial inventories index

Data Source: Swiss Federal Statistical Office, Quarterly Industrial Statistics Indices, NOGA D, final values

The out-of-sample domain of the reference series ranges from 2001:1 to 2002:4, i.e. it covers 8 observations, which is rather short for formal analyses of the predictive properties of our models. This notwithstanding, these two years were characterised by a high degree of uncertainty with respect to business conditions in Switzerland (and elsewhere), so that they do serve well for a more informal evaluation of simulated real-time predictions versus realisations in a situation, where early yet reliable information on economic activity is highly welcome. Moreover, given the restriction that no data are available for the reference series before 1990, we decided rather not to sacrifice too many observations of the within-sample-

⁶ From the larger total "secondary sector", NOGA D excludes extracting industries, construction as well as energy and water supply, which are known to follow different patterns of business activity than manufacturing.

domain (which should cover at least two cycles, i.e. certainly no less than ten years, to pick up any potential regularity in the data) in favour of the out-of-sample domain.

Table 2 Questions of the KOF business tendency survey in the manufacturing industry

Notation	Question
Monthly survey	
1a	Incoming orders compared to previous month
1b	Incoming orders compared to the month of the previous year
2a	Order books compared to previous month
2b	Order books, judgement
3a	Production compared to previous month
3b	Production compared to the month of the previous year
4a	Stocks of intermediate products compared to previous month
4b	Stocks of intermediate products, judgement
5a	Stocks of finished products compared to previous month
5b	Stocks of finished products, judgement
6a	Expected incoming orders (next three months)
6b	Planned production (next three months)
6c	Planned purchase of intermediate products (next three months)
7	Expected midterm development (beyond the next three months)
BA*	Business activities
PL*	Plans of firms
Quarterly survey	
Q1a	Employment compared to previous quarter
Q1b	Employment, judgement
Q2a	Technical capacities compared to previous quarter
Q2b	Technical capacities, judgement
Q2c	Degree of capacity utilisation, percentage
Q4	Profit situation compared to previous quarter
Q8b	Expected prices for intermediate products
* Aggregated index	

For potential indicators, we draw exclusively on the extensive KOF industrial business tendency survey data base, which is derived from monthly and quarterly surveys of the Swiss manufacturing industry. The items of the KOF questionnaires, which constitute the basis for

our indicator series, are given in Table 2.⁷ The latest data referred to in this study are from the surveys for April 2003 on a monthly, and for 2003:1 on a quarterly basis.

There are three possible answers to the qualitative questions. The appraisals may be stated as "too high", "normal", "too low", and the changes as "up", "unchanged" and "down". The responses from the questionnaires are weighted⁸ and aggregated to form percentages of each response category of the total. Moreover, the difference between the positive and the negative shares (commonly called "balance"), is calculated to reduce the information into a single index number.⁹ Below, we do consider not only the balance indicator (which by construction ignores the information revealed by the shares of respondents ticking "normal" or "unchanged"), but also each of the original three response categories. Since this is hardly ever done in comparable analyses,¹⁰ we dispose of a larger set of potential coincident and leading indicators than most other studies. (The series derived from KOF data carry the following suffixes for identification of the quantification method: "+" : too high/up, "=" : normal/unchanged, "-" : too low/down and "±" : balance.)

Before linking the monthly KOF series (X) to the reference series, they had to be aggregated into quarterly series, which was done by computing quarterly mean values $A(X)$ and, alternatively, by referring to their end values $E(X)$, where $A(X)_Q = (X_{m1} + X_{m2} + X_{m3})/3$ and $E(X)_Q = X_{m3}$. Moreover, given the pronounced seasonality in some of the survey data, the KOF series are referred to alternatively as unfiltered (u) and seasonally adjusted series (s: Census-X12, option X11). Finally, since what might matter for predictive purposes could be the levels as well as the changes of the survey based indicators (with changes potentially being better than levels to capture accelerations or decelerations of economic activity), we compute the first differences with respect to the preceding quarter $\Delta 1X = X_t + X_{t-1}$ as well as the seasonal differences with respect to last year's corresponding quarter $\Delta 4X = X_t + X_{t-4}$.

To summarise the transformations and their notation, for each item X in the monthly survey questionnaires, there are four alternative quantifications (suffixes: +, =, - or ±), two alternative aggregations into quarterly series (prefixes: A or E) as well as original and seasonally adjusted alternatives (suffixes: u or s). The number of potential indicator series is multiplied further by three, since all of these are evaluated alternatively in levels, differences (prefix: $\Delta 1$) and seasonal differences (prefix: $\Delta 4$). Accordingly, one single item in the questionnaire can result in a maximum of 48 ($4 \times 2 \times 2 \times 3$) quantitative series. The notational syntax is (where applicable): difference operator – aggregation – question as in Table 2 (in brackets) – quantification – filter.

⁷ For the questionnaires see <http://www.kof.ethz.ch>.

⁸ The weights on the micro-level are the firms' employment; weights on aggregate level are demographic data.

⁹ This method of extracting relevant information is widely used. For a discussion of its properties, see Dasgupta/Lahiri (1992).

¹⁰ For one of the rare exceptions, see Entorf (1993).

Last but not least, we have to be aware of potential boundary instability of our time series. While this affects neither the reference series – which are growth rates of final official numbers – nor the raw numbers of the indicator series – which are final from the beginning and hence not subject to revisions – the Census-X12, option X11, seasonal filter is asymmetric, so that at the boundary it may take some time until the latest entries converge to their final values. We deal with this issue in the following way: For the within-sample analyses we use the final data from 1990:1 to 2000:4. For the out-of-sample analyses, however, we follow a stepwise procedure and draw on the indicator series as they would have presented themselves at any given point in time, had they been referred to in order to derive the most up-to-date estimation or forecast of the reference series. In other words, we are strictly simulating real-time estimations and forecasts.

3 Previous Empirical Work

The search for coincident and leading indicators of economic activity has long attracted the attention of applied economists. Traditionally this line of research concentrates on quantitative statistical indicators, but qualitative indicators have more recently been receiving a fair share of attention. In this section, we discuss some studies that relate closely to our work.

Moore et al. (1994) analyse the behaviour of the cumulated balance indicators of different surveys in the CIBCR Leading Index, which is a highly reliable indicator for detrended economic development in the United States. They find that qualitative survey indicators are less noisy than quantitative economic indicator series, but the lead of the former is shorter than that of the latter. Takala/Tsupari (2000) take a closer look at balance indicators from Finland in total manufacturing industry and in some important sectors and compare them with year-on-year growth rates of industrial production. They refer to cross-correlations and Granger causality tests to identify leading indicators. Finally, cross-spectral analyses are conducted to find the optimal time lead. As those mentioned above, most studies consider only balance indicators of survey data in the search for coincident or leading indicators of economic reference series. An exception is Entorf (1993), who extends the search and includes – apart from the usual balance indicators – the positive, the negative as well as the "no change" shares of responses. In this study, the negative share of responses yield the best results, hence it makes sense not to confine the search for indicators from survey data to the balance indicators. Fritsche (1999) analyses German business tendency survey data. He starts with a graphical interpretation of the leading properties of expectation indicators at economic turning points. Cross-correlation and Granger causality tests are run to confirm the lead-lag structures. However, since a detrended reference series was used, it is difficult to assess the exact lead of the indicator series.

The general picture that emerges from this literature is that a considerable number of potential leading indicators perform well on average over an extended time span, but that their prognostic power is severely reduced in the vicinity of turning points, where from a policy

perspective it is most important to have reliable signals of the state and the development of the economy. Therefore, Koskinen/Öller (2002) direct their attention to the performance of leading indicators at turning points. They refer to hidden Markov model probability forecasts of changes between two regimes for data from Sweden and the United States. However, this method cannot be applied for point estimates.

De Mulder/Dresse's (2002) search for coincident and leading indicators of GDP on the basis of survey data for Belgium is in many respects comparable to our work. Their aim is not only to find indicators for the qualitative development of GDP and the corresponding turning points, but in addition to derive quantitative estimations and forecasts of the GDP growth rate. To this end, they isolate the common component of selected survey data. For the idiosyncratic component of GDP, an estimation of an ARIMA representation was fitted, however, this last step of their analyses did not prove successful.

For Switzerland, an earlier attempt to derive numerical estimates of the Swiss official statistical production index by means of KOF survey data failed to find sufficiently close correlations. This was attributed – at least in part – to the dubious quality of the reference series, the BIGA production index¹¹ from 1980:1–1989:3 (Stalder 1998). Since official data, which are adapted to the new SFSO production index, have not been calculated for the years before 1990, there is no direct way to address Stalder's assumption. If, however, the post-1990 official statistics series are less plagued by data problems than the pre-1990 series, the correlation between OS data and quantifications of survey data may have improved, thereby providing a better basis for BTS → OS estimations. Marty's (1997) reference series was Swiss GDP, for which he analysed – among other series – the KOF survey indicators. After an initial selection process by cross-correlation analysis and graphical interpretation of turning points, the remaining indicators were combined into a principal component to aggregate them into a leading indicator. The resulting "KOF economic barometer" is now widely referred to in Switzerland as a reliable short and medium term indicator of turning points in GDP growth. However, this indicator focuses exclusively on GDP dynamics and turning points and does not make any reference to levels. Besides, Marty's study does not deal with industrial statistics, so that a new look at the data is in order.

4 Empirical Analysis

The empirical analysis establishes a methodology for selecting coincident as well as leading indicators for the official Swiss industrial statistics reference series and combining them into composite indicators that are appropriate specifications of $f[\text{BTS}(t)]$ and $g[\text{BTS}(t + \lambda)]$. This helps to equip analysts as well as policy makers with timely information on crucial trends of economic activity. Since this study is exclusively concerned with

¹¹ Recall that official industrial statistics in Switzerland underwent major changes in the 1990s. In this context, "BIGA production index" denotes the former series.

estimates $BTS \rightarrow OS$, we refrain from the use of autoregressive forecasts within a given OS series as well as cross-references within different official series.¹²

4.1 Cross-Correlations

The first step of the analysis consisted in extensive data mining. To this end, we computed all possible pairwise permutations of cross-correlations to screen the data for highly correlated KOF and SFSO series. From these, we picked all pairs where the maximum correlation showed up simultaneously or with a lead of the KOF data. The selection threshold was set at $|r| \geq 0.7 \Leftrightarrow r^2 \geq 0.5$, which resulted in numerous hits for the reference series except for the inventory index. Now, aggregate data on inventories are notoriously prone to capturing statistical residual components rather than what they are meant to measure, which might explain the comparatively poor performance of the survey indicators with respect to this series. On the other hand, inventories play a prominent role in theoretical analyses of business cycles, so that we were reluctant to drop this reference series altogether. Therefore, we lowered the threshold for the inventory series to $|r| \geq 0.5$.

Then, for the five reference series and all leads (λ) with $3 \geq \lambda \geq 0$, the cross-correlogrammes were listed in descending order by the absolute value of the closest correlation. Hence, after this initial step of analysis, we are equipped with a complete map of those coincident and leading indicator series, which would reproduce the largest shares of the reference series' variance in the within sample domain.¹³

For the next step of our study, the principal component analysis, we needed to have at least two qualifying indicator series for any pre-determined lag, given a specific reference series. This was the case for $\lambda=0$ and $\lambda=1$ with respect to all our reference series. However, for $\lambda=2$ referring to GR_PRODU and GR_I_ORD we failed to find at least two qualifying indicator series, and only GR_ORD_B allows further multivariate analyses for $\lambda=3$. For the purpose of this paper, we therefore restricted our further presentation to coincident indicators as well as leading indicators with $\lambda=1$.

Table 3 shows the correlations for the indicator series with the highest absolute within-sample correlation. Since this is by definition equal to the R that would result from a regression of the corresponding reference series on the indicator series (coincident or lagged

¹² As the data generation process within the statistical office is (at least in part) a black box for outsiders, the last point is essential: As far as autoregressive techniques and cross reference are practised, outsiders can improve the precision of their estimates with respect to a reference series by same means, but the results would be partly tautological. Accordingly, this study does not aim at finding specifications of forecast models which reveal the best fit with respect to a given OS series – these would probably include autoregressive elements as well as cross references with the official data corpus –, but to aggregate a set of business tendency survey indicators into single indexes which can be used as coincident or leading indicators of the reference series.

¹³ The detailed results are tabled in Etter/Graff (2001) and available from the authors upon request.

accordingly), the tabled correlations can serve as a first benchmark for the following multi-indicator estimations.

Table 3 **Highest absolute correlations, $1 \geq \lambda \geq 0$**

OS series	KOF BTS series	$ r $ max.	Lead BTS→OS (in quarters)
GR_SALES	E(7)–s	–.869	1
	A(1a)–s	–.900	0
GR_PRODU	E(6b)±s	.863	1
	A(1b)–u	–.878	0
GR_I_ORD	E(6a)±s	.851	1
	E(1b)±u	.891	0
GR_ORD_B	E(1b)–u	–.878	1
	A(3b)+u	.822	0
GR_INVEN	E(4a)+s	.636	1
	A(4a)=u	–.696	0

For variable notation, see section 2; e.g. E(7)–s denotes quarterly end-value of item 7, minus-share, seasonally adjusted

The highest correlations with a lead of one quarter to the reference series of sales, production and incoming orders were realised by the survey questions on expectations (for production: monthly question, third month of quarter taken as quarterly value, planned production, balance indicator, seasonally adjusted). The order books are led by the incoming orders. Changes of inventories of intermediate products lead the yearly changes of inventories of final products. For the coincident indicators, questions on past changes yielded the highest correlation with values between 0.8 and 0.9, except for inventories. Apart from delivering highly correlated series, these results are plausible and obviously identify meaningful rather than spurious correlations.

4.2 Principal Component Extractions

The selection of a single leading or coincident indicator is straightforward for any of our reference series: The indicators given in Table 3 show the best within-sample performance and the coefficient of determination that we would get from a bivariate regression of the reference series on any of the tabled indicators¹⁴ is already known from the correlation

¹⁴ In a previous study (Etter/Graff 2001), we referred to multiple regressions, but – with the above mentioned problems in mind – we restricted the number of regressors to two, which certainly does not exploit all the relevant information from the survey data.

analyses. Hence, referring to these pairs, rather precise ex post-estimates of OS data are feasible through OLS-regressions of Y on just one coincident or – though to a lesser extent – leading series from the KOF business tendency surveys X :

$$Y_t = \beta_0 + \beta_1 X_{t-\lambda} + \varepsilon_t$$

To improve the within-sample fit, we can include additional "second-best" survey based series as regressors, but multicollinearity will soon render the regression parameters too imprecise for sensible out-of-sample estimates, when the pattern of multicollinearity is random and specific to the within-sample domain. In this case, the resulting overfitting to an ultimately meaningless random pattern will impair the accuracy of the fitted values in the forecast period.

On the other hand, for substantial as well as statistical reasons, it is unlikely that a single indicator series will bear sufficient information to secure informational efficiency in the process of estimating and forecasting a reference series. Moreover, our survey based indicators certainly comprise a considerable share of noise, so that at this stage of our analysis, we refer to a statistical method that is designed to identify and combine the common variance of a chosen set of indicator series into a new synthetic variable. The procedure chosen here is principal component (PC) analysis. If the variance of a given set of pre-selected indicators, which are closely correlated to a given reference series with a pre-determined lag, can reasonably well be represented by one principal component only,¹⁵ this first component will serve as the basis to derive our final indicator.¹⁶ Specifically, for any reference series and both $\lambda = 0$ and $\lambda = 1$, we proceed as follows:

1. Find all series that pass our initial threshold criterion; if several transformations of one and the same item from the questionnaires are amongst the candidates, select only the ones that yield the closest correlation with the reference series;
2. table the selected BTS based series in descending order to get a list of at least two and at most 23 (i.e. the number of items analysed) properly transformed indicator series;

¹⁵ That is, if the correlations between the desired representations are high, but measurement errors and stochastic shocks in the data for the individual indicators have little common variance. Technically, we accept this condition as fulfilled, if no eigenvalue, except for the first component, exceeds unity.

¹⁶ Note that this method, which in our set-up amounts to the identification of the co-variance of selected time series, captures some of the spirit of Burns/Mitchell's (1946) notion of the business cycle, which is the co-movement of a number of economic series; see also Stock/Watson (1989) and Forni et al. (2000).

3. run a series of principal component analyses starting with the first two indicator series, then add the third series, and so forth until all qualifying series from step 2 are included, where the estimation runs from 1985:1 to 2000:4;¹⁷
4. check that all principal component analyses result in one common factor only and save the factor values;
5. regress the reference series on the properly lagged alternative factor values and order these regressions by their coefficients of determination R^2 ;
6. from step 5, select as the resulting combined indicator series the principal component with the number of indicators that yields the best within sample-fit.

The resulting combined indicator series are presented in Table 4. The combined PC based regressions perform better than any of the corresponding bivariate regressions (column 3 displays the PC regression R , column 4 repeats the maximum $|r|$ from the correlation benchmark estimator from Table 3). Accordingly, a restriction on one indicator series alone does not make efficient use of the information from the survey data.

As reported in the table, the combined coincident indicators for sales and incoming orders are the ones that are most closely related to their reference series ($R > 0.91$). Nevertheless, the PC indicators for production and of order books are likewise performing reasonably well ($R > 0.83$) for either of the two pre-determined lead-lag structures. The poorest, yet not completely disappointing, performance ($R = 0.72$) is found for the inventory series. Moreover, as expected, the coincident indicators generally perform slightly better than the corresponding leading indicator series.

Note that the survey based time series that our selection procedure identified as suitable indicators to be combined into the various principal components comprise 1. all four quantifications, 2. the two alternative aggregations to quarterly series, 3. original and seasonally adjusted series as well as 4. series that are transformed with either of the two difference operators. Accordingly, given this data, it makes sense not to rely on a single quantification (e.g. balance) and a restricted set of transformations, but rather to let a wider set of transformed series enter at the first stage.

The time series plots (see Graphs 1 to 10 in the Appendix) visualise our combined indicators' ex-post within-sample fit as well as the resulting estimates for the prediction period from 2001:1 to the present. Note that there are two charts for each reference series, referring to the estimations based on their best combined coincident indicator ($\lambda = 0$), and on their best combined leading indicator ($\lambda = 1$), respectively. The vertical lines delimit the within-sample-domains from the estimation periods.

¹⁷ We go back to 1985 for the computation of the principal components, since these refer to intercorrelations of the survey data only and there is no reason to restrict the period of analysis to the availability of the reference series. Nota bene: as a by-product, this would allow to reconstruct the reference series back to 1985.

Table 4 **Combined principal component indicator series, $1 \geq \lambda \geq 0$**

OS series	Lead BTS→OS (in quarters)	number of BTS series selected for PC	R	r max. from Table 3
GR_SALES	1	7	.897	-.869
	0	2	.916	-.900
GR_PRODU	1	3	.873	.863
	0	4	.888	-.878
GR_I_ORD	1	3	.887	.851
	0	5	.913	.891
GR_ORD_B	1	6	.898	-.878
	0	7	.834	.822
GR_INVEN	1	none		.636
	0	2	.722	-.696

Reference series and BTS indicator series selected for PC extraction
(BTS series ordered from highest to lowest correlation with reference series)

GR_SALES,	$\lambda = 1$: E(7)-s, E(6a)-s, E(6b) $\pm$ s, E(2a)-s, E(6c) $\pm$ s, $\Delta 4A(3a)+s$, $\Delta 4A(1a)-s$
GR_SALES,	$\lambda = 0$: A(1a)-s, E(BA) $\pm$ u
GR_PRODU,	$\lambda = 1$: E(6b) $\pm$ s, E(6c) $\pm$ s, E(6a)-s
GR_PRODU,	$\lambda = 0$: A(1b)-u, E(BA) $\pm$ u, (Q4) $\pm$ s, E(3b)-u
GR_I_ORD,	$\lambda = 1$: E(6a) $\pm$ s, $\Delta 4A(6b)+u$, $\Delta 1A(BA)\pm u$
GR_I_ORD,	$\lambda = 0$: E(1b) $\pm$ u, A(1a)-s, (Q4) $\pm$ s, A(2a) $\pm$ s, $\Delta 4E(BA)\pm u$
GR_ORD_B	$\lambda = 1$: E(1b)-u, $\Delta 4E(2b)\pm u$, $\Delta 4A(BA)\pm u$, $\Delta 4A(3b)+u$, (Q4) $\pm$ s, A(5b) $\pm$ u
GR_ORD_B	$\lambda = 0$: A(3b)+u, A(BA) $\pm$ u, (Q1a) $\pm$ s, E(2b)-u, A(4b)=u, (Q1b) $\pm$ u, $\Delta 4(Q2b)+u$
GR_INVEN,	$\lambda = 1$: only one qualifying series E(4a)+s, see Table 3
GR_INVEN,	$\lambda = 0$: $\Delta 4A(4a)=u$, $\Delta 4E(4b)=u$

Looking at the within-sample-domains, it is obvious that our ex-post estimates are closely tracing their reference series, the only exception being the inventory series. As discussed above, the poor performance of the inventory estimations might be due to the dubious validity of the reference series, though in this study we cannot assess this conjecture directly. However, given this empirical finding, we can easily explain another finding, namely why the sales estimates are superior to those for the production index: Production is basically derived from sales by means of deflators of its sub-components, which introduces a potential source of noise as well as bias and can thereby easily fail to trace real production. Moreover,

recall that for every period, sales + change of inventories = production, so that through this identity any noise from the inventory series will affect the production series.

Finally, let us look at the out-of-sample domains. At the time of writing, based on our coincident indicator series ($\lambda = 0$) we can compute estimations up to the first quarter of 2003, and based on our leading indicator ($\lambda = 1$) we arrive at predictions until 2003:2. Recall however that the reference series end with their last final values for 2002:4, so that we have only eight observations to assess the out-of-sample properties of our estimating procedure. To the right of the vertical lines, we plot the known out-of-sample realisations of the reference series together with their corresponding forecasts from 2001:1 to the last available data point. Note that we reproduce two slightly different forecasts in one and the same graph, based on the real-time values and, alternatively, on the ex-post final values of the underlying BTS indicator series. Though our BTS indicators are not subject to revisions so that the original series are final, a difference results from seasonal adjustment, which is not stable at the boundary. The only exception is the composite indicator for GR_INVEN ($\lambda = 0$), which includes unadjusted data only, so that there is no distinction between final and real-time.

From a visual inspection, it is probably fair to conclude that the correlations between the composite indicator series and their reference series do indeed persist beyond 2001; they appear weak after 2001 for the inventory series only, but for this series, it was hard to find any suitable indicator series for the within-sample domain in the first place. Moreover, a close inspection reveals that generally the ex-post indicator performs slightly better in tracing the out-of-sample movements of the reference series than the corresponding real-time series, though the differences are obviously not fundamental but rather minor. We are inclined to conclude that this finding adds further confidence to our analyses: Since the ex-post estimators refer to a larger set of information (precisely: to longer time series, which affects the seasonal adjustment of the series in question), a sensible methodology should produce better results with these as compared to results obtained with the real-time series.

In Table 5, we present a first numerical assessment of the out-of-sample properties. The first column shows the correlations of the simulated real-time forecasts with the eight presently available values of their reference series. The second column enables to compare these with the ex-post estimates that we get if we draw on the whole set of information available now, i.e. with the seasonal filters calibrated with the series ending in 2003:1, and not at the time of simulation. The numerical analyses confirm that the real-time simulations are performing somewhat poorer than the ex-post estimation (the only, albeit hardly perceivable exception being GR_I_ORD, $\lambda = 0$). Though the differences are rather small, and moreover, the number of out-of-sample data points is too restricted to consider statistical tests of significance of the difference between the corresponding correlation coefficients, this stresses the fact that when performing out-of-sample simulations, one should be careful to refer to real-time data only. The third column repeats the within-sample correlations from Table 4. Unsurprisingly, the out-of-sample correlations are different from the corresponding within-sample correlations, but they do not generally appear less close than the within-sample relationships, and there are no really remarkable differences, with the exception of

the notoriously problematic inventory index series for $\lambda=0$, where the out-of-sample correlation is insignificant.

Table 5 **Evaluation of out-of-sample estimates, $n = 8$, $1 \geq \lambda \geq 0$**

OS series	R real-time	R ex-post	R within-sample	Lead BTS→OS (in quarters)
GR_SALES	.729	.733	.897	1
	.774	.811	.916	0
GR_PRODU	.724	.740	.873	1
	.884	.889	.888	0
GR_I_ORD	.911	.908	.887	1
	.869	.880	.913	0
GR_ORD_B	.904	.918	.898	1
	.966	.975	.834	0
GR_INVEN	.806	.869	.636	1 ^a
	.108	.108	.722	0 ^b

a Single indicator regression (regressor: $E(4a)+s$).

b Principal component regression, no seasonally adjusted regressors.

5 Summary and Conclusion

Extensive analyses of Swiss quarterly data on industrial activity were conducted. Using the official statistics as the reference series, a large set of business tendency survey based series are referred to as potential indicators, comprising not only the common balance indicator, but the percentage part of all three response categories. Several transformations of these survey data series were also considered.

The official statistics were revised in 1996 and hence permit a new look at their interrelatedness with the survey data. Specifically, based on the within-sample-period from 1990:1 to 2000:4, we found high correlations between survey data and corresponding official statistics with leads of the former of up to three quarters. Yet, the highest correlations were generally found for coincident indicators and leads of one quarter, so we restricted further analyses to these. The closest correlations with a lead to reference series of sales, production and orders were found to refer to series from survey questions on expectations rather than present conditions. For the coincident indicators, questions on past changes yielded the highest correlation, which is obviously in line with economic plausibility.

Then, a selection algorithm was developed to determine suitable subsets of indicator series with predetermined leads and compute their principal components. These are exclusively based on business tendency surveys. The coincident, and to a slightly lesser extent, the leading indicators, provide reasonably reliable estimators for Swiss manufacturing sales, production and orders in the industry. The results for inventories are somewhat poorer.

Finally, we traced the performance of our coincident and leading indicator in the out-of-sample domain (2001:1 to 2002:4). To get a correct picture, we used a step-by-step real time method and referred only to data that was available at the time of the simulated forecast. Given that we did not find clear evidence of within-sample-correlations breaking down, our results are reassuring and show that the proposed composite indicators are indeed useful for policy oriented analyses of present and future economic conditions.

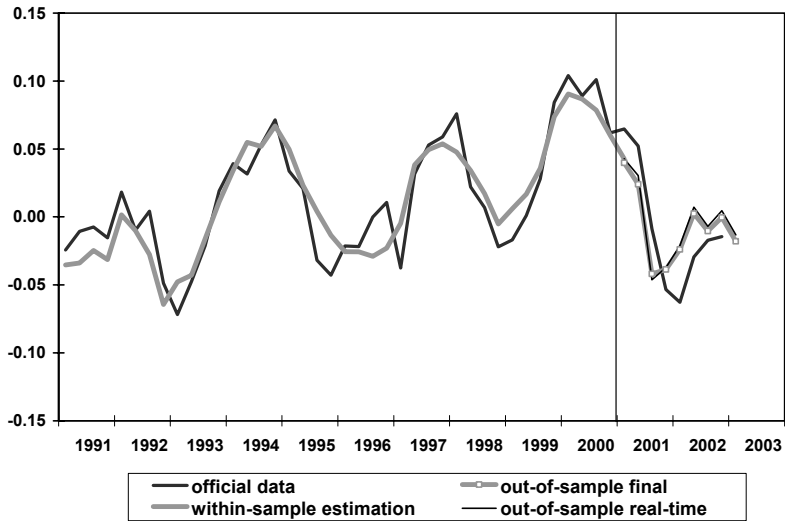
All analysed data refer to Switzerland. However, we emphasise that the methodology outlined in this paper can readily be used for comparable analyses in other countries.

References

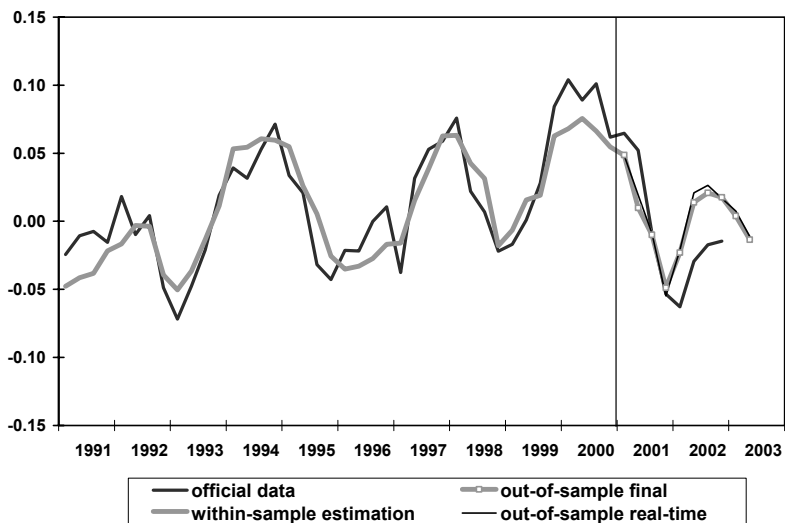
- Burns, A. and W. C. Mitchell (1946): *Measuring Business Cycles*, National Bureau of Economic Research, New York.
- Dasgupta, S. and K. Lahiri (1992): A Comparative Study of Alternative Methods of Quantifying Qualitative Survey Responses Using NAPM Data, *Journal of Business & Economic Statistics*, Vol. 10, pp. 391–400.
- De Mulder, J. and L. Dresse (2002): The Construction of Leading and/or Coincident Indicators for Belgian Economic Activity, Paper Presented at the 26th CIRET Conference, Taipei, October.
- Entorf, H. (1993): Constructing leading indicators from non-balanced sectoral business survey series, *International Journal of Forecasting*, Vol. 9, pp. 211–225.
- Etter, R. and M. Graff (2001): Untersuchung des Zusammenhangs zwischen der PAUL-Statistik des BFS und den Konjunkturtestdaten der KOF in den Bereichen Industrie und Bauhauptgewerbe (1990–2000), Konjunkturforschungsstelle der ETH Zürich.
- Etter, R. and M. Graff (2003): Estimating and Forecasting Production and Orders in Manufacturing Industry from Business Survey Data. Evidence from Switzerland, 1990–2003, *Swiss Journal of Economics and Statistics*, Vol. 139, No. 4 (to appear shortly).
- Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2000): The Generalized Dynamic-Factor Model: Identification and Estimation, *Review of Economics and Statistics*, Vol. 82, pp. 540–554.
- Fritzsche, U. (1999): Vorlaufeigenschaften von Ifo-Indikatoren für Westdeutschland, DIW Discussion Papers Nr. 179, Berlin.
- Koskinen, L. and L. E. Öller (2002): A Classifying Procedure for Signalling Turning Points, Paper presented at the 26th CIRET Conference, Taipei, October.
- Marty, R. (1997): Die Vorlaufeigenschaften der Indikatoren des KOF/ETH-Industrietests, in: *Konjunktur*, Nr. 9, 60. Jahrgang, Konjunkturforschungsstelle der ETH Zürich, pp. 11–23.
- Moore, G. H., Klein, P.A. and L. Hiris (1994): The Role of Surveys in Developing Promptly Available Indexes for Forecasting, in: *The Explanatory Power of Business Cycle Surveys*, Papers presented at the 21st CIRET Conference, edited by K. H. Oppenländer and G. Poser, Avebury, Aldershot, pp. 217–239.
- Stalder, P. (1989): Verfahren zur Quantifizierung qualitativer Konjunkturtestdaten, KOF-Arbeitspapier Nr. 26, Dezember 1989, Zürich.
- Stock, J. H. and M. W. Watson (1989): New Indexes of Coincident and Leading Economic Indicators, *NBER Macroeconomics Annual*, Vol. 1989, pp. 351–394.
- Swiss Federal Statistical Office (quarterly): *Der Geschäftsgang im sekundären Sektor: Die Produktions-, Auftrags-, Umsatz- und Lagerindizes*, Bern.
- Takala, K. and P. Tsupari (2000): The predictive power of Finnish business survey among industrial sectors, in: *Use of Survey Data for Industry, Research and Economic Policy*, Selected Papers presented at the 24th CIRET conference, Wellington, New Zealand, 1999, edited by K. H. Oppenländer, G. Poser and B. Schips, p. 583-612, Ashgate, Aldershot.

Appendix

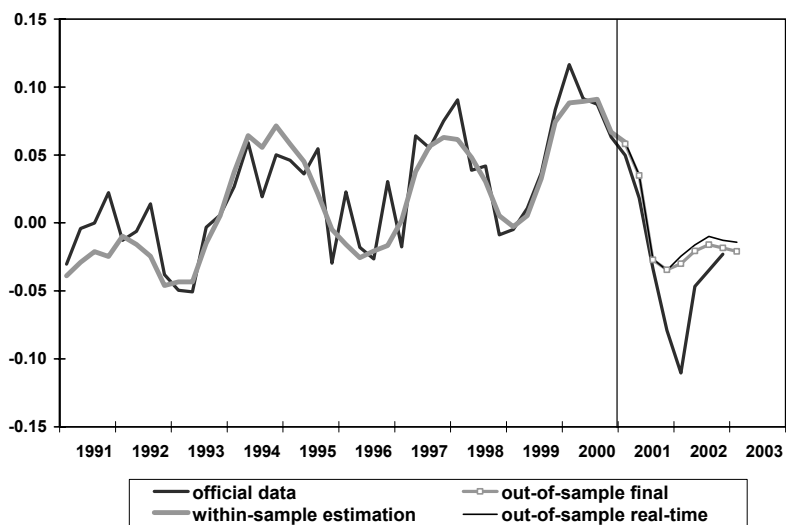
Graph 1 GR_SALES, $\lambda = 0$



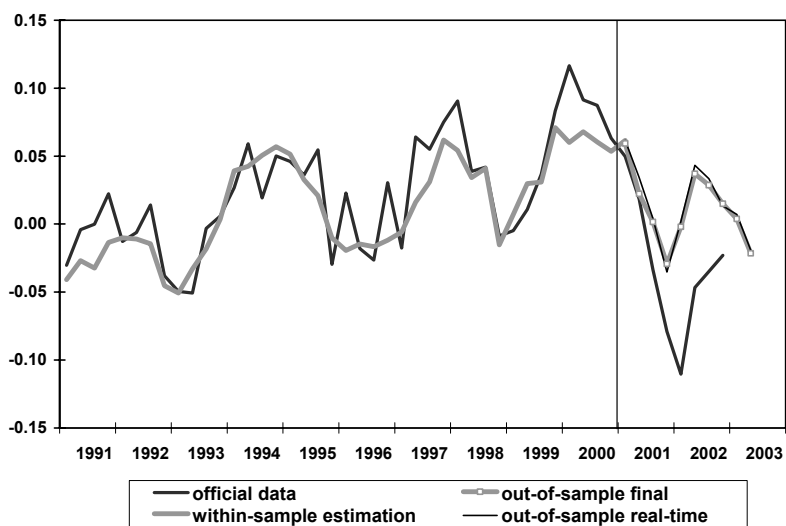
Graph 2 GR_SALES, $\lambda = 1$

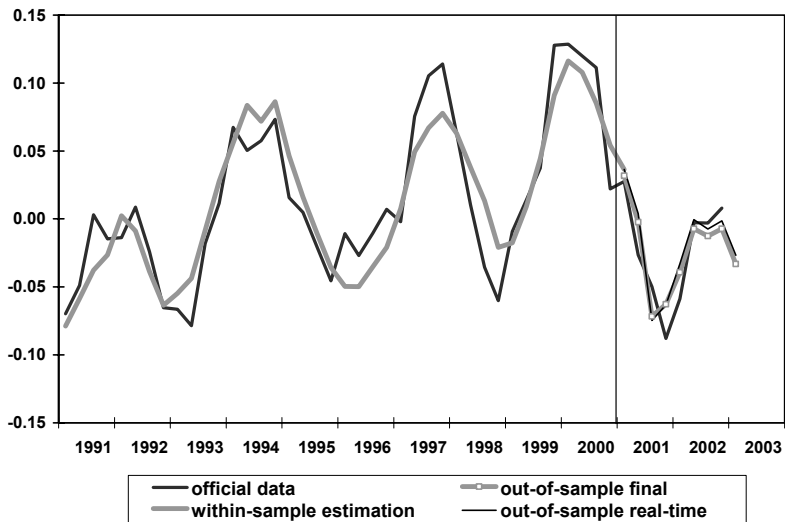
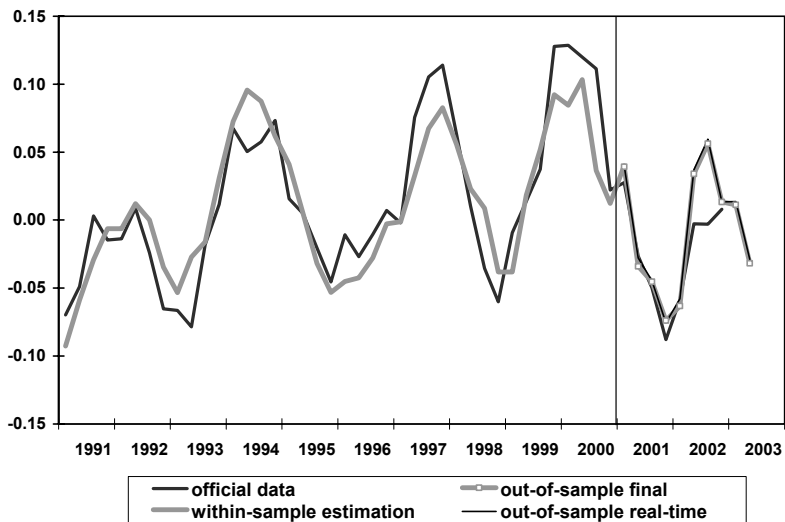


Graph 3 **GR_PRODU**, $\lambda = 0$

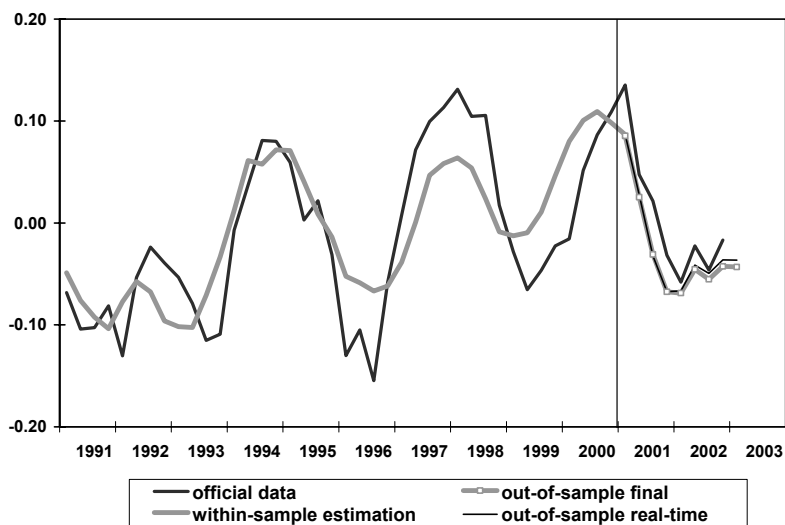


Graph 4 **GR_PRODU**, $\lambda = 1$

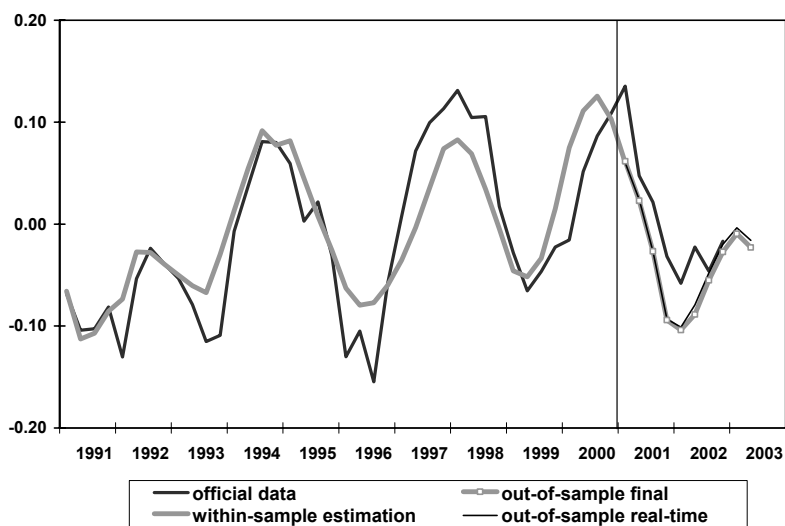


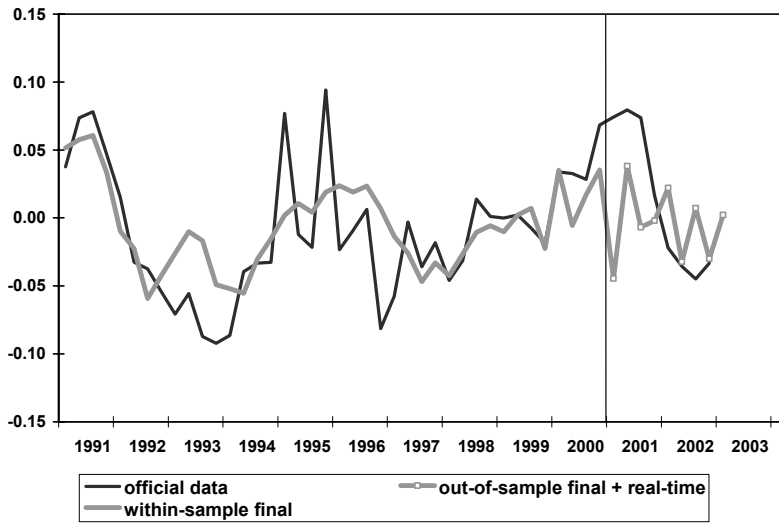
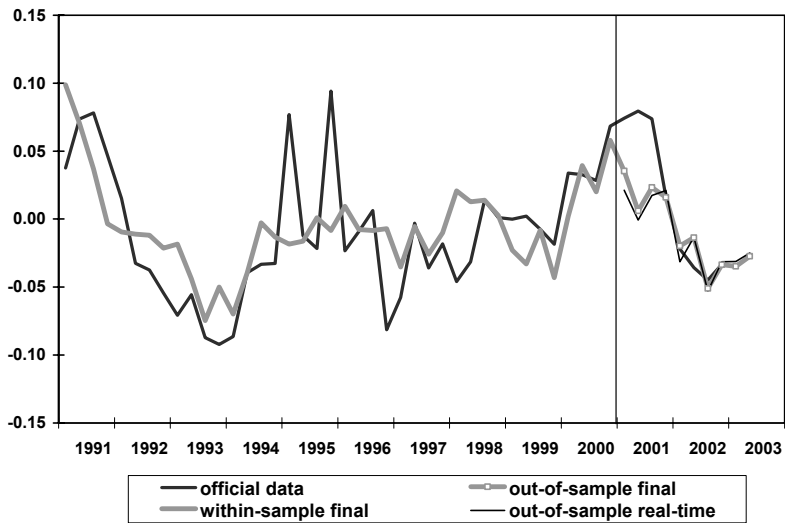
Graph 5 $GR_I_ORD, \lambda = 0$ Graph 6 $GR_I_ORD, \lambda = 1$ 

Graph 7 **GR_ORD_B**, $\lambda = 0$



Graph 8 **GR_ORD_B**, $\lambda = 1$



Graph 9 GR_INVEN, $\lambda = 0$ Graph 10 GR_INVEN, $\lambda = 1$ 

OECD and CIRET News – 2004/1

OECD Short-term Economic Statistics

Key Developments

134

Call for Papers

27th CIRET Conference, September 15–18, 2004, Warsaw, Poland

137

OECD Short-term Economic Statistics

Key Developments

The OECD has been working on a number of key developments in short-term economic statistics over the last twelve months for both quantitative and qualitative indicators.

Quantitative indicators

Work for these indicators has arisen out of discussions at the annual meeting of the Organisation's Short-term Economic Statistics Expert Group (STESEG), which comprises delegates from the national statistical institutes of most OECD Member countries and a small number of international organisations. The June 2003 STESEG meeting discussed the extensive preliminary outputs of the three task forces created following the 2002 meeting. The task forces dealt with three key issues for the future evolution of short-term economic statistics, namely:

- continued development of indicators for the services sector;
- the development of international standards for the presentation of data and metadata;
- development of new methods to improve the timeliness of short-term economic statistics.

Services: The need to provide national and international users with a broader array of short-term indicators for the service sector has been recognised for some time and whilst there has been progress in some countries, overall, the actual availability of such indicators across the entire OECD membership is still patchy. Work on the development of international guidelines and recommendations and the identification of best practice is currently underway on several fronts including trade in services, price indices and services production indicators. Discussion at the United Nations Statistical Commission recognised the need to ensure that such development work was co-ordinated and to this end, the Commission requested the OECD to play an active role in providing leadership and direction in the development of services statistics. STESEG work, which will initially focus on the development of standards for infra-annual (particularly monthly) indices of service production, is a key element of this co-ordinated activity.

Data and metadata presentation: The major sources of difference in the comparability of data, both between countries and over time, stem from conceptual, classification and definitional differences, and operational differences flowing out of differences in data collection and processing practices. However, even where such differences are minimal, data

disseminated by different countries, etc. may actually *look* different because of the use of different data presentation practices. Similarly, data for identical subjects disseminated by various international organisations, although coming from the same national source, may also appear different due to differing presentation practices and data transformation processes used at the international level. STESEG work on data presentation focuses on the articulation of recommendation and guidelines for different types of data (absolute figures, indices, growth rates) and forms of data (original, seasonally adjusted and trend-cycle) as well as on a small number of key data presentation practices that impact on data consistency both over time and between countries. Such practices include data revision, presentation of series breaks, provision of information on sampling and non-sampling errors, presentation of metadata, etc.

Work in this area will take the form of the preparation of a draft manual over 2004 which will incorporate guidelines and recommendations from existing international standards where they exist and the formulation of new standards where necessary. Following the preparation of the draft manual there will be a round of consultation with both other international organisations and national agencies in order to obtain more extensive input for the preparation of a final version.

Improvements in timeliness: In recent years there has been increased user demand for improvements in the timeliness of short-term economic indicators. In response to this demand, national statistical agencies have been at the forefront in implementing new practices and processes that have resulted in considerable reductions in the delay between the end of the data reference period and the availability of data to users. Future STESEG work on timeliness will focus on the development a framework of current best practice that will allow national agencies to more readily identify the relevance and suitability of existing approaches and possible tradeoffs between further improvements and data accuracy. Task force work over the remainder of 2003 and early 2004 will finalise the desired content of the framework and identification of material that should be referenced in the framework.

Further background information on these three activities are available on the STESEG website at www.oecd.org/std/meeting-papers.

Qualitative indicators

Joint European Commission – OECD workshop: The OECD and the European Commission (EC) are jointly organizing a Workshop on International Development of Business and Consumer Surveys for statistical and research institutions in OECD Member countries, European Union (EU) Members, Candidate countries and selected non-Member countries (Brazil, China, India, Indonesia, Russia and South Africa). The workshop will be held on 20-21 November 2003 in Brussels, Belgium. This meeting will be held back-to-back with a one day meeting organised by the EC for EU and Candidate countries.

The main objectives of the workshop are to:

- Strengthen co-operation between the European Commission, OECD and participating national agencies in the area of business tendency and consumer opinion surveys, particularly between agencies located in European and Candidate countries and those in non-EU OECD Member countries and in large non-member economies.
- Develop appropriate guidelines and recommendations in the collection and dissemination of survey data and encourage the development of internationally comparable business and consumer survey data.
- Provide a forum for sharing experiences on progress with harmonization of their surveys, the analytical use and interpretation of survey results for short-term monitoring and the construction of confidence or sentiment indicators, etc.

Information on the outcomes of this meeting will be provided in future editions of this Journal.

Expansion of OECD data base on qualitative indicators: The OECD has been publishing a range of business tendency and consumer opinion surveys for Member countries and selected non members in its monthly Main Economic Indicators for some time. By and large, the tendency data published has largely been restricted to the industry sector and to meet the demand for indicators for the services sector the OECD is currently engaged in a program to significantly expand the databases of business and consumer tendency surveys to include construction, retail trade and other services sectors, and more data for selected non-Member countries. This expansion has raised the need for a meeting with contact persons from participating institutions to discuss the content of the OECD global database on business and consumer tendency surveys, the need for harmonization of variables, presentation and dissemination issues, including methodological transparency, and use of business and consumer tendency surveys results in cyclical analysis, etc. The need for improved dialogue and contact with national data producers of these statistics will be met through the joint EC-OECD workshop described above. In order to minimize the reporting burden on national agencies a key feature of the OECD database enlargement process will be co-ordinated collection with the European Commission.

Call for Papers

27th CIRET Conference, September 15–18, 2004, Warsaw, Poland on Economic Tendency Surveys and Cyclical Indicators

Hosted by the Research Institute of Economic Development (RIED),
at Warsaw School of Economics (WSE)

Sponsored by the National Bank of Poland and the Central Statistical Office

Special Topic:

Economic Situation in Countries of Central and Eastern Europe

The overall aim of CIRET conferences is to encourage and improve communication, exchange and co-operation between academics and practitioners who conduct economic or social surveys, analyse survey data and develop or make use of cyclical indicators. CIRET, the Centre for International Research on Economic Tendency Surveys, is also a forum for discussion and application of new methodological developments and their results.

This call invites papers related to a broad range of topics to be investigated based on several types of surveys (business tendency surveys, consumer surveys, investment and innovation surveys, etc.) or based on the development and use of cyclical indicators, including methodological aspects. Within this framework, all types of contributions are welcome.

Topics

I. Short-term Economic Surveys and Indicators

- Business Tendency Surveys (including Investment Surveys)
- Consumer Surveys
- Composite and Leading Indicators
- Ad hoc Surveys

Sub Topics (on each topic of I.)

- New Surveys or New Cyclical Indicators
- Use of Survey Data or Cyclical Indicators for Micro- or Macroeconomic Analysis
- Use of Survey Data or Cyclical Indicators for Preparing Decisions in Firms and Policy
- New Methods

II. Surveys Related to Structural Aspects of the Economy

- Innovative Activity
- Information and Communication Technology (ICT)
- Organisational Change
- Labour Market Analysis

III. Special Topic

- Economic Situation in Countries of Central and Eastern Europe

The special topic invites papers dealing with the economic situation of countries in Central and Eastern Europe. Subjects relate to cyclical and structural aspects of the economies and may cover implications of EU enlargement. Cyclical analysis should be preferably based on short-term surveys and/or indicators.

Submission Procedure

Abstracts

Deadline: Mid of February 2004

The abstract (300-500 words) should include: Name(s) of author(s), affiliation, address, phone number, fax, e-mail address, title of abstract and keywords.

Please send your abstract by filling out the **form for abstracts** on:

<http://www.ciret.org/callforpapers>

Letters of acceptance will be sent to the corresponding author by end of March 2004.

Papers

Deadline: End of June 2004

Papers should use the **template** for MS-Word or the corresponding style definitions on:

<http://www.ciret.org/callforpapers>

Please send your submission in MS-Word- or Adobe-PDF-Format to:

papers@ciret.org

For other transfer see:

<http://www.ciret.org/callforpapers>

Additional information on the conference is available on the CIRET Web site:

<http://www.ciret.org>

see "Conferences, Warsaw 2004"

Contact

CIRET Office
c/o KOF
Swiss Institute for Business Cycle Research
ETH Zentrum WEH
8092 Zürich
Switzerland

Phone: +41 1 632 42 38
Fax: +41 1 632 11 50

<http://www.ciret.org>
info@ciret.org

Aims and Scope

The Journal of Business Cycle Measurement and Analysis is jointly published by the OECD and CIRET to promote the exchange of knowledge and information on theoretical and operational aspects of economic cycle research, involving both measurement and analysis.

The OECD has developed various cyclical indicators using quantitative and qualitative information from its member countries. Information about the work of OECD in the field of short-term economic statistics may be found at www.oecd.org/std.

CIRET, the Centre for International Research on Economic Tendency Surveys, is a global forum for economists and institutions which conduct and analyse business and consumer surveys. CIRET holds international conferences every other year. Information about CIRET may be found at www.ciret.org.

The scope of the Journal of Business Cycle Measurement and Analysis encompasses:

- Analysis of cyclical fluctuations
- Business cycle specification, definition and classification
- Statistical approaches to the development of short-term economic statistics and indicators
- Business tendency, investment and consumer surveys
- Use of survey data or cyclical indicators for business cycle analysis

Researchers in the field of business cycle measurement and analysis are encouraged to submit their papers to the Editor-in-Chief, Professor Günter Poser (editor-in-chief@ciret.org). All papers undergo a referee process.

Notes for authors and a template for manuscripts are available on www.ciret.org/jbcma or www.oecd.org/std/jbcma.

Correspondence about editorial and other matters related to the journal may also be addressed to Daniel Bloesch at the CIRET Office (jbcma@ciret.org) or Ronny Nilsson at the OECD (ronny.nilsson@oecd.org).