



Journal of Business Cycle Measurement and Analysis

Edited by Bernd Schips



CIRET



K O F

**2007
Volume 3, No. 1**

Journal of Business Cycle Measurement and Analysis

Volume 3, No. 1

A Joint Publication of OECD and CIRET

ISSN 1729-3618
2007: Volume 3 (3 issues)



ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

CIRET



K O F

Centre for International Research on Economic Tendency Surveys

The CIRET Office is located at
KOF Swiss Economic Institute, ETH Zurich, Switzerland.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

© OECD 2007

No reproduction, copy, transmission or translation of this publication may be made without written permission. Applications should be sent to OECD Publishing rights@oecd.org or by fax 33 1 45 24 99 30. Permission to photocopy a portion of this work should be addressed to the Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, 75006 Paris, France, fax 33 1 46 34 67 19, contact@cfcopies.com or (for US only) to Copyright Clearance Center (CCC), 222 Rosewood Drive Danvers, MA 01923, USA, fax 1 978 646 8600, info@copyright.com.

Editorial

We welcome Prof. Bernd Schips as the new Editor-in-Chief of the JBCMA. Prof. Schips has been president of CIRET from 2000 to 2006 and Head of KOF Swiss Economic Institute at ETH Zurich from 1993 to 2005.

The current issue of the JBCMA is published with some delay. This is due to unforeseen difficulties in the submission, referee and revision processes, which have now been resolved. Unfortunately, a lack of resources at the CIRET office in the recent past also delayed publication. Since September 2006, the JBCMA is listed in the AEA electronic indexes, which are included in JEL on CD, e-JEL and EconLit. We expect this will help the journal to attract good papers.

In this issue, Richard Curtin gives several recommendations to enhance the predictive performance of consumer sentiment: by a new conceptualization, by new survey measurement techniques, by changes in the survey administration and by shifting the main focus of harmonization from the input questions to the output indices. Ard den Reijer identifies deviation cycles in the manufacturing industry of nine OECD-countries, measures the degree of synchronization and constructs a composite leading indicator. Daniel Smith and Allan Layton compare probability forecasts in Markov regime switching business cycle models and find that the Diebold and Mariano statistic tends to under-reject the null of "no difference in forecast accuracy".

The report by Richard McKenzie outlines the principles behind the development of the short-term economic statistics timeliness framework, explains its structure and reviews its current usage by statisticians. Finally, Péter Benczúr, Emin Muradov and Attila Rátfai report a number of stylized facts of cyclical fluctuations in CIS countries (former Soviet Union) and highlight among other findings, that private consumption is extremely volatile and government consumption is a very important, dominantly procyclical determinant of output.



Prof. Bernd Schips
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. Bernd Schips, ETH Zurich, Switzerland

Associate Editors

Daniel Bloesch, KOF Swiss Economic Institute,
ETH Zurich, Switzerland

Ronny Nilsson, OECD, Paris, France

Editorial Board

Chairman

Prof. Enrico Giovannini, Chief Statistician, OECD, Paris, France

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

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

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

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

Zbigniew Matkowski, Warsaw School
of Economics, Warszawa, Poland

Gian Luigi Mazzi, Eurostat, Luxembourg

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

Lars-Erik Öller, Stockholm University,
Stockholm, Sweden

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

Ataman Ozyildirim, The Conference Board,
New York, NY, USA

Giuseppe Parigi, Bank of Italy, Rome, Italy

Günter Poser, Former Editor-in-Chief of JBCMA,
Technische Universität Darmstadt, Germany

Gerd Ronning, University of Tübingen, Germany

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

Jan-Egbert Sturm, KOF Swiss Economic
Institute, ETH Zurich, Switzerland

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

Thomas M.F. Yeh, CEPD, Council for Economic
Planning and Development, Chinese Taipei

Contents

Consumer Sentiment Surveys: Worldwide Review and Assessment <i>Richard Curtin</i>	7
Deviation Cycles in Manufacturing: Business Cycle Measurement and Leading Indicators <i>Ard H.J. den Reijer</i>	43
Comparing Probability Forecasts in Markov Regime Switching Business Cycle Models <i>Daniel R. Smith and Allan Layton</i>	79
Report Improving Timeliness for Short-Term Economic Statistics <i>Richard McKenzie</i>	99
Report Cyclical Fluctuations in CIS Economies <i>Péter Benczúr, Emin Muradov and Attila Rátfai</i>	121
OECD and CIRET News – 2007/1	137

Consumer Sentiment Surveys: Worldwide Review and Assessment

Richard Curtin*

Abstract

Consumer sentiment surveys are regularly conducted in at least forty-five countries. The surveys are based on the premise that data on consumer expectations represent a leading indicator of future changes in the macro economy. A series of Granger causality tests indicated that measures of consumer sentiment both predict and are predicted by a wide range of economic variables, although the data exhibit substantial differences across countries and across variables. The predictive performance was not found to be directly related to differences in sampling, question wording, or index construction. Several recommendations are given to enhance the predictive performance of consumer sentiment. First, a new conceptualization of consumer sentiment is now warranted based on changes in the knowledge and sophistication of consumers, the increased availability of economic information, shifts in the composition of demand, and changes due to an aging population. New survey measurement techniques are needed as well. Preliminary data on numeric probability scales indicate that there are advantages and disadvantages to a shift away from the now commonly used verbal likelihood scales. Second, new analysis models are proposed that would disaggregate consumer spending into its components as well as disaggregate the consumer data into separate demographic and economic subgroups. Third, changes in the survey administration are proposed. The shift away from telephone surveys will be due to falling response rates, especially among younger adults; the shift toward internet surveys will be due to their greater capacity to handle much more complex measurement strategies. Finally, the very notion of worldwide harmonization must also evolve; shifting its main focus from the input questions to the output indices so as to encompass a greater range of diversity in measurement methodologies. The basic premise of the consumer sentiment surveys has been affirmed by past measurements, and with appropriate developments, the consumer sentiment surveys can provide even more robust and timely forecasts of trends in the macro economy in the future.

Key Words: Consumer Sentiment, Leading Indicators, Expectations, Survey Methodology

JEL Classification: C42, C82, D84, E27, E32

*

Director Surveys of Consumers, University of Michigan; curtin@umich.edu

The paper benefited from the detailed comments of two anonymous referees.

Résumé

Des enquêtes d'opinion auprès des consommateurs sont régulièrement menées dans quarante cinq pays au moins. Les enquêtes sont fondées sur la prémisse que les données concernant les attentes des consommateurs constituent un indicateur avancé sur les changements futurs de la macro économie. Une série de tests de causalité Granger a indiqué que les mesures de l'opinion du consommateur prévoient en même temps qu'elles sont prévues par un grand nombre de variables économiques, bien que ces données montrent des différences substantielles d'un pays à l'autre et d'une variable à l'autre. La performance de prédiction n'a pas semblé être directement liée aux différences d'échantillonnage, de formulation des questions ou de construction des indices. De nombreuses recommandations sont données afin d'améliorer la performance de prédiction de l'opinion du consommateur. Tout d'abord, une nouvelle conceptualisation de l'opinion du consommateur est dorénavant garantie. Elle est fondée sur les modifications de la connaissance et la complexité des consommateurs, la disponibilité accrue des données économiques, les modifications des éléments constitutifs de la demande ainsi que les changements dus à une population vieillissante. Il est aussi nécessaire de trouver de nouvelles techniques d'évaluation des enquêtes. Les données préliminaires sur les échelles de probabilité numériques indiquent l'existence d'avantages et d'inconvénients liés à un éventuel abandon des échelles de vraisemblance verbale que l'on utilise aujourd'hui. De plus, de nouveaux modèles d'analyse qui désagrégeraient les divers composantes des dépenses du consommateur et qui désagrégeraient aussi les données du consommateur en sous-ensembles démographiques et économiques distincts, ont été proposées. Par ailleurs, des modifications dans la gestion des enquêtes ont été proposées. L'abandon des enquêtes téléphoniques est dû à un taux de réponse décroissant, tout particulièrement auprès des jeunes adultes ; le choix des enquêtes sur Internet s'est fait grâce à leur plus grande capacité à gérer des stratégies d'évaluation plus complexes. Enfin, l'idée même d'une harmonisation mondiale doit aussi évoluer ; en déplaçant la priorité principale de la formulation des questions vers la création des indices de résultat afin de couvrir un plus large éventail de diversité concernant les méthodologies d'évaluation. La prémisse élémentaire des enquêtes d'opinion auprès des consommateurs a été affirmée par des évaluations passées, et avec des développements appropriés, ces enquêtes pourront fournir dans le futur des prévisions des tendances de la macro économie, dans les délais et encore plus solides.

1 Introduction

Consumer sentiment surveys are now regularly conducted in at least forty-five countries around the world, including nearly all developed economies as well as by many developing and emerging market economies. The basic premise of these surveys is that data on consumer expectations contain information about potential future changes in consumer spending and represent a leading indicator for the entire economy. Consumer spending accounts for one-half to two-thirds of all spending in market-based economies, so even small changes in the expenditures of households can have a major impact on the economy. Spending in the U.S. economy totalled \$12.5 trillion during 2005. Of that total, \$9.5 trillion was spent by consumers while investment spending by business firms was just \$1.3 trillion. Indeed, consumers invested more in homes, vehicles, and other durable goods than firms invested in new plants and equipment.

How successful are consumer surveys in providing advance information on potential changes in the macro economy? The overall evidence of its predictive track record is mixed; its performance is uneven across countries as well as across different economic measures within countries and over time.¹ Consumers' expectations are largely formed from information on the standard array of variables that economists typically use in their forecasting models. Thus, the repeated finding that these other economic variables often capture the information contained in consumer expectations should be no surprise. Consumer expectations data are intended to improve short-term forecasts; there is no credible use of the sentiment data for long term forecasts. Sentiment data improves near-term forecasts not just because the data is available sooner than the underlying economic data, but because it contributes independent information about the future course of the economy. Some have lamented that the contribution is small, but so too is the contribution from most other economic variables. The data from consumer surveys is meant to supplement not to replace other economic information.

There is evidence that the amplitude of business cycles have been reduced in OECD countries during the past few decades largely due to declining fluctuations in domestic demand (Dalsgaard, Elmeskov and Park, 2002). Consumers have been devoting an increased share of their incomes to purchases of services, which are much less volatile than expenditures on durable goods. Moreover, the ability of consumer sentiment to forecast cyclical changes in the economy is likely to be coincident with other economic variables

¹

The predictive ability of intentions data in cross-section analysis has been investigated by Klein and Lansing (1955), Mueller (1957), Tobin (1959), Juster (1964), Maynes (1967), and Dunkelberg (1972). Time-series tests were conducted by Mueller (1963), Friend and Adams (1964), Adams (1964; 1965), Suits and Sparks (1965), Ferber (1966), Maynes (1967), Dunkelberg (1969), Fair (1971), and Juster and Wachtel (1972), Garner (1981), Curtin (1984), Fuhrer (1988), Praet and Vuchelen (1989), Hall (1993), Blanchard (1993), Acemoglu and Scott (1994), Cochrane (1994), Huth, Eppright and Taube (1994), Carroll, Fuhrer and Wilcox (1994), Matsusaka and Sbordone (1995), Abderrezak (1997), and Eppright, Arguea and Huth (1998), Danthine, Donaldson and Johnsen (1998), Howrey (2001), Utaka (2003), Curtin (2003), Ludvigson (2004), and Golinelli and Parigi (2004).

during periods of stable economic growth, while the importance of sentiment increases near turning points or as the result of non-economic shocks (Keyfitz, 2004; Golinelli and Parigi, 2004). Insofar as the cycles in the economy have diminished, this may also mean that the forecasting ability of the sentiment index has also declined. Others have disputed this claim, finding that the ability of the sentiment index is not limited to special circumstances (Howrey, 2001).

To be sure, survey measures of economic expectations need to be improved to serve as more accurate indicators of near term changes in the economy. This paper focuses on the development of an overall strategy to guide those necessary changes. The review starts with a theoretical basis of the expectations measures, the measurement methodology, the questions that are now used, and the formulation of the sentiment indexes. Next, empirical tests are conducted for each country to determine how objective economic information influences the formation of consumer expectations, and how consumer expectations foreshadow subsequent economic developments. Based on an understanding of the strengths and weaknesses of the data, the paper recommends some fundamental changes for the future.

Although the influence of most demographic, economic, and institutional aspects of the economy have not changed, three must be highlighted: the changing economic concerns of an aging population, the more rapid transmission of economic information, and the impact of globalization. Aging populations will mean a rising share of consumers will become more concerned about the uncertainties in financial markets and less concerned about the uncertainties in the labour markets. The rapid transmission of information and the growing knowledge and sophistication of consumers will foster more active rather than passive responses on the part of consumers. Moreover, the continued globalization of national economies has meant that financial markets have become increasingly synchronized, and this could increase the transmission of economic shocks in the future.

Three recommendations are made to enhance the ability of consumer surveys to act as leading indicators of macroeconomic trends. First, a new conceptualization of consumer confidence is now warranted, and the revised definition requires new survey measures. New survey measures need to reflect advances in economic theory as well as changes in demographic characteristics and the economic sophistication of the population. The changes may entail modifications to some of the standard questions as well as the additions of some new questions. The goal of worldwide harmonization must yield to differences in implementations across countries so that differences in the characteristics and sophistication of the population can be reflected in the questionnaire. Such tailoring of the questionnaire would still allow full harmonization of the results for the analysis of business cycle developments in individual countries as well as across groups of countries.

Second, the analysis of consumer survey data must adopt a new analysis framework to optimally contribute to the prediction of near-term macroeconomic trends. The optimal design would be to disaggregate the major components of consumer purchases as well as disaggregate the survey data based on the household's demographic and economic characteristics. Predictions of economic behaviour would be done on the disaggregated components of consumer spending with the disaggregated population groups. Forecasts of macroeconomic trends would simply aggregate the resulting forecasts of the individual components.

Finally, rising pressure to reduce costs as well as increase the timeliness of the survey information will mean more diversity across countries in how the surveys are conducted. Internet surveys represent both substantial advantages as well as present significant challenges. The advantages of internet surveys include significant cost savings, the potential for higher quality data, and the near instantaneous availability of the survey results. The disadvantages of internet surveys include significant challenges to maintain the representativeness of the sample and the designation of the respondents. Strategies need to be developed that allows the combination of internet surveys with surveys based on more traditional data collection methods.

These improvements represent significant tasks that require extensive research and an extended period of time for implementation. Trends in current measures must be preserved and linked to trends in any new measures that are devised in the future. The current measures of consumer expectations were devised a half century ago in response to a different economic era. The measurement and analysis of consumer expectations must adapt to changed conditions or suffer the growing irrelevance of its unchanged methods. Given the potential usefulness of measures of consumer sentiment in the future, the only choice is to embrace change and devise robust and innovative solutions for the future.

2 Theoretical Background

The first measures of consumer confidence were developed by George Katona in the late 1940s as a means to directly incorporate empirical measures of expectations into models of spending and saving behaviour.² Katona's now famous dictum was that consumer spending depends on both their "ability and willingness to buy." By spending, he meant discretionary purchases; by ability, he meant the current income of consumers; and by willingness, he meant consumers' assessments of their future income prospects. Katona hypothesized that spending would increase when people became optimistic, and precautionary saving would rise when they became pessimistic.

Katona originally considered "consumer confidence" to be a broad measure of expected changes in income. It was not simply the expected size of a consumer's future income, but the certainty or uncertainty that was attached to those expectations. Thus, an important component of the definition of consumer confidence was that it encompassed both the expected level as well as the expected variance of income. To recognize this dual criterion, the dimension of consumer sentiment was defined as ranging from optimism and confidence to pessimism and uncertainty.

² Katona documented the development of his theories in a series of books: *Psychological Analysis of Economic Behavior* (1951), *The Powerful Consumers* (1960), *The Mass Consumption Society* (1964), and *Psychological Economics* (1975). For an overview of Katona's theoretical contributions, as well as a complete list of his publications and biographical information see Curtin (1983).

Discretionary purchases are the things that people buy that are not the necessities of life. All spending categories include discretionary items, with the key provision that discretionary purchases can be postponed. There is no reason to think that consumer confidence should play a role in determining the timing of necessary expenditures. The most commonly studied discretionary purchases are homes, vehicles, and household durables. These are relatively infrequent and large purchases, often involve the use of credit, and are typically planned purchases – just the type of purchases that would reflect both current economic conditions as well as expected future changes. Such expenditures are not properly viewed as current period consumption, but as investments that are intended to yield a future stream of consumption.

Katona foresaw the growing power of the consumer to shape the course of the macro economy resulting from two developments in the second half of the 20th century. First, growth in consumer income as well as asset holdings provided people with greater financial latitude. Income changes could no longer be expected to result in immediate and offsetting changes in consumption. This financial latitude meant that consumers became active decision makers, able to gauge the timing of their spending decisions to best serve their present and future needs. The second change involved the growing importance of consumer "investment" goods – purchases of homes, vehicles, and other large household durable goods. These large and infrequent purchases are more likely to be carefully deliberated decisions, the timing of the purchase can as easily be advanced as postponed, and these purchases frequently involve the risks associated with the use of credit or accumulated savings. As a result, the timing of spending decisions became increasingly dependent on consumers' expectations about future trends in income, employment, prices, and interest rates, and correspondingly less dependent on their past economic situation.

While the description of this paradigm reflects investment spending in advanced economies, Katona thought that the same reasoning applied to developing or emerging market economies. The key difference is that discretionary purchases in less developed economies are not big ticket purchases, but are typically inexpensive nondurable items (inexpensive at least by the standards of advanced economies). The underlying theory, however, remains the same. The same is true for the consumption of services by consumers in advanced economies, as some services can be considered discretionary and subject to expectations about future economic conditions. Katona included these qualifications to his basic theory.

Katona did not anticipate, however, the vast impact of the information age on the economic decisions of consumers, the rapid growth in the knowledge and economic sophistication of consumers, the impact of the global economy on domestic economic conditions, and the extent to which consumers would base their spending decision on specific expectations rather than on general economic optimism.

A half-century ago most consumers viewed all aspects of the economy through the single dimension of how it affected their job and income prospects. Today, consumers in advanced economies are more likely to view their future economic prospects as dependent

on a much broader range of expectations, including expected returns on financial assets and real estate holdings, expected inflation rates, pension and health care entitlements, future taxes and government subsidies, as well as job and income prospects. Moreover, consumers are more likely to make distinctions about the direction and size of the expected change among these various economic factors, as well as distinctions between near and long term expected outcomes. Spending and saving decisions are responsive to these distinctions. As a result, there is no longer any presumption that a single index of consumer sentiment could be useful for the prediction of all types of spending and saving behaviour.

Presumably, given the differences in the stage of economic development across countries, along with greater economic development consumers would shift from a sole focus on income to the growing importance of the full array of economic expectations. Nonetheless, the pace of change along this dimension may be accelerated by the equal availability of economic information to anyone with an internet connection, whether they live in Indiana, Italy, India, or Indonesia. Greater information may also result in more rapid growth in economic sophistication. Moreover, the growth in information and sophistication may more rapidly diminish the power of a single index of consumer sentiment in favour of specific indexes for specific spending and saving decisions.

Along with the rise in the importance of a range of economic expectations, there is a growing demand from economists for more precise measures of expectations. As expectations have become more central components of economic models, the theoretical specifications of the desired measures have become more exacting. The ideal measure would specify the characteristics of the complete probability distribution for a precisely defined future event or outcome. Given an estimate of the entire probability distribution for each individual, measures of central tendency can then be supplemented by measures of dispersion to capture the uncertainty that individuals associate with that expectation.

Before I discuss the potential revisions to what is now measured, it is first necessary to document and assess the current survey methods.

3 Survey Methodology

Consumer expectations are typically measured by qualitative questions in population surveys conducted within each country. There are three general areas that are critical for the interpretation of the survey findings: the sample design, question wording, and transformations of the raw data into estimates of consumer sentiment.

The countries included in this analysis are listed in Table 1, along with the two-letter acronym used to identify the countries in later tables. There are representatives of all six inhabited continents, with most of the recent additions in Eastern Europe and Asia. There are other countries, such as India, that are now in the process of starting consumer surveys. In addition to the countries included in Table 1, there are a large number of other countries that have measured confidence sporadically. These countries include many Latin American

nations such as Chile, Colombia, Costa Rica, the Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Uruguay, and Venezuela. These surveys are typically conducted by marketing research firms and not under the direction of any governmental agency. In addition, consumer sentiment is measured in many cities and local areas around the world, none of which were considered for this paper.

Table 1 Countries conducting consumer tendency surveys

AR	Argentina	ES	Spain	NL	Netherlands
AT	Austria	FI	Finland	NO	Norway
AU	Australia	FR	France	NZ	New Zealand
BE	Belgium	HK	Hong Kong	PL	Poland
BG	Bulgaria	HU	Hungary	PT	Portugal
BR	Brazil	ID	Indonesia	RO	Romania
CA	Canada	IE	Ireland	RU	Russia
CH	Switzerland	IT	Italy	SE	Sweden
CN	China	JP	Japan	SI	Slovenia
CY	Cyprus	KO	Korea	SK	Slovak Republic
CZ	Czech Republic	LT	Lithuania	TH	Thailand
DE	Germany	LU	Luxembourg	TW	Taiwan
DK	Denmark	LV	Latvia	UK	United Kingdom
EE	Estonia	ML	Malaysia	US	United States
EL	Greece	MX	Mexico	ZF	South Africa

The ideal design for an international survey would be if all countries used a nationally representative sample, a common method for index construction and seasonal adjustment, a comparable number of interviews conducted at the same time each month and information on response rates and procedures used to construct non-response weights.

There are many variations of these standards in practice. The EU countries use the most consistent survey methodology across all members due to the guidance of the European Commission. Surveys conducted in other parts of the world are generally not harmonized, although the OECD has made progress in recent years in convincing countries to conform to the accepted standards. Nonetheless, there are considerable differences in the economic and population characteristics of the countries that now measure consumer sentiment, and these differences are likely to persist in the years ahead. The ideal design would still focus on the standardization of the measurement methodology, although it could allow variations in question wording that recognizes differences in the stage of development in the economy and in the population.

3.1 Sample design and interview methodology

Most countries have designed their sample so that every adult member of the population has an equal chance of being selected. Nationally representative samples are standard in all developed countries; in emerging economies the samples may be limited to urban centres, such as in China, Malaysia, and Mexico, for example. Such limitations of the national samples are not critical omissions if the areas excluded from the sample do not fully participate in the market economies of their countries.

Of greater importance is survey non-response, especially if there is a differential non-response rate across important population subgroups. Telephone samples have an added coverage problem, namely non-telephone households are excluded by definition. Information on response rates is provided for many of the surveys, but without a common definition, it is difficult to interpret whether a 100% response rate is better or worse than a 60% rate – the range recently reported for EU countries (Nilsson, 2000a). Sample weights are widely used to correct for differential non-response, or in some cases to align quota samples with actual population distributions. Most countries conduct the interviews by telephone, although some countries still conduct in-home interviews.

Some experiments have been done using the survey data collected in the United States which indicates that measures of consumer sentiment are not very sensitive to coverage or non-response bias. In fact, no significant time-series differences in the consumer sentiment index were uncovered when the overall response rate was reduced by two-thirds (Curtin, Presser, and Singer, 2001). This is of course no guarantee that the damaging impact of non-response bias will not be present in any other country or even in the United States in future surveys.

3.2 Questions used to measure consumer sentiment

There were several variations in the types of questions that are now being used to measure consumer confidence. In nearly all of the countries the questions focus on personal finances, general conditions in the economy including jobs and inflation, and intentions to spend or save. Of course, there are slightly different versions of these basic themes across the countries; nonetheless, all of the questions in all countries addressed similar topics. Indeed, the major differences involved variations in the frame of reference: some countries included questions about past as well as expected future changes, and some countries asked about time horizons of six months instead of the more typical twelve month reference period. It should be noted that the questions used to calculate the sentiment indexes are typically drawn from surveys that include a broader range of questions that could be used to supplement or reformulate the current sentiment index.

For the purposes of this discussion, I have separated the countries into three main groups largely based on differences in reference periods. The first group, accounting for more than half of all the countries, uses the harmonized EU questionnaire; the second group uses questions on both past and expected future changes as well as a longer time horizon; and the

final group uses a six month time horizon. These basic groups also differ in the number of question used in the overall confidence index, with the number ranging from three to six.

3.2.1 Question variant 1: EU harmonized

The European Commission governs the methodological standards for more than half of all countries that now conduct consumer tendency surveys. A recent review of the properties of the consumer sentiment index lead to the recommendation of a change from five questions (past and expected changes in general economic conditions, past and expected change in personal finances, and assessments of current buying conditions) to the following harmonized questions (Goldrian, Lindlbauer and Nerb, 2001):

- Expected change in financial situation over the next twelve months
- Expected change in general economic situation over next twelve months
- Expected change in unemployment over next twelve months
- Expected saving over next twelve months.

The sentiment index is defined as the arithmetic average of the four series. This new formulation was derived based on an analysis that used real private consumption as the criterion variable. The EU uses a weighting procedure to aggregate the indexes from the individual countries, with the weights depending on the relative size of total personal consumption expenditures.

3.2.2 Question variant 2: past and future changes

The second major grouping of countries generally follows the U.S. version of the questions. These countries are: Argentina, Australia, Brazil, Canada (Decima), Hong Kong, Indonesia, Mexico, New Zealand, South Africa, Russia, and the United States. Hong Kong and South Africa only use the three forward-looking expectation questions in their confidence indicator. These countries all generally use a form of the following questions:

- Personal financial situation over the past twelve months
- Expected change in personal finances in next twelve months
- Expected change in general economic conditions in next twelve months
- Expected change in general economic conditions in next five years
- Buying conditions for consumer durables

3.2.3 Question variant 3: six-month reference periods

The countries in this group include Canada (Conference Board), China, Japan, Korea, Malaysia, Taiwan, and the United States (Conference Board). The questions in these countries include the following:

- Expected change in family living standards in six months
- Expected change in income in six months
- Expected change in price level in six months
- Current and expected change in business conditions in six months
- Current and expected change in employment in six months
- Buying plans for durables in six months

No country's survey included questions on all of these topics; typically five questions were used. The single common element was the six month horizon.

3.3 Index formulations

Nearly all of the countries allow for five potential responses to their questions, ranging from very positive (PP), positive (P), neutral, negative (M) to very negative (MM). Typically, two steps are involved in the computation of the overall consumer confidence index for each country. First, for each question a balance score is computed based on the percentage distribution as:

$$B = \left(PP + \frac{1}{2}P \right) - \left(\frac{1}{2}M + MM \right)$$

The overall confidence index is then simply the average of the balance scores:

$$CCI = \sum_{i=1}^n B_i / n$$

For the few countries that coded three instead of five response categories – positive (PP), neutral, and negative (MM) – the same index formula applies except the half-weight terms are simply dropped. More generally, the index formula can be written as:

$$CCI_i = \sum_i w_i B_i^s$$

where w is the weight and B_i^s is the standardized balance score. The standardization is sometimes necessary to equalize the variance contributed by each component to the overall index. For example, an equal variance index would mean that each component question

would be normalized by subtracting the mean and dividing by the variance. The standardized balance score could also include the application of a smoothing factor, such as a moving average that represents the number of months to cyclical dominance (MCD). For the sentiment index, the MCD has been estimated at 2 – 3 months (Nilsson, 2000b). In the analysis included in this paper, the sentiment index that was used is the official published index; there was no attempt to reformulate any country's index data.

4 Interpreting Survey Responses

There are two approaches to determine what these questions measure. The first is an interpretation of the question wording; the second is an interpretation of the responses. In the first approach, the difference between what was asked and what was answered could be considered a form of measurement error if the question wording could be uniquely linked to an objectively observable answer. For example, expected changes in the unemployment rate could be compared with the actual change in the official unemployment rate. Even in this seemingly straightforward case the actual test is only approximate: for example, the EU question asked about the number unemployed, not conditioned on only counting those who are actively seeking work. Moreover, it is unclear whether the reference period of "over the next twelve months" means any instance of unemployment, the average level over the twelve months, or the number unemployed at the end of the period; it is also unclear whether people who work at part-time or temporary jobs while looking for a permanent job should be counted as unemployed. For almost all the other questions, the issues are even more difficult since unemployment has a much clearer objective counterpart than the questions on the "financial position of the household" or "general economic situation" and whether the household would "save any money."

The second approach simply uses an empirical standard to determine what the question measured. For example, it could be hypothesized that the consumer sentiment index measured expectations about the household's income and job prospects. This could be tested by the sentiment index's empirical relationship to changes in real personal disposable income and changes in the employment rate.

Indeed, testing survey measures by their relationship with objective economic measures has typically taken one of two forms: testing what sources of information consumers used to form their expectations, and testing what measures consumer expectations are able to predict. The first approach focuses on the inputs and the latter on the outputs of the decision process. While most have assumed that these two processes are independent, with expectations typically formed in advance of their use in decisions, consumers may be more likely to simultaneously form and use expectations. Indeed, consumers may only form explicit and specific expectations when they perceive a need for their use in making decisions. Prior to the actual decision, consumers may only form more generalized economic expectations – the type of expectations that the more general qualitative questions that are now asked are most likely to capture.

Another variant of these same issues has been framed in terms of whether consumer expectations are backward or forward looking, that is, do consumer expectations indicate what has happened in the past or indicate what will happen in the future. Clearly, expectations about prospective economic changes should incorporate information about the past as well as provide information about the future.

The advantage of directly observing the expectations of consumers is that they reflect private information in addition to the publicly available information on the economy. Moreover, it is this private information that represents the independent additional information that is the unique contribution of survey measures of consumer sentiment. Even for commonly publicized statistics, such as the unemployment rate, consumers rely to a greater extent on private information sources than on official information (Curtin, 2003).

4.1 Granger causality tests

A series of Granger causality tests were performed to determine whether the past history of an objective variable predicts the expectation or the expectation predicts the future changes in the same objective economic variable. Of course, the use of the term "causality" is not meant literally, but only to identify the widely used statistical test and should be interpreted as meaning that one variable has the ability to predict future changes in the other.

While there were differences in the question wording across countries, the differences were generally small. Moreover, for any difference in wording to affect the results of a time series analysis, the wording differences must prompt a changing pattern of responses, not simply a difference in levels across all surveys. Thus, for example, differences in the six month and twelve month reference periods would not be relevant for time series analysis if both economic expectations increased or decreased to approximately the same extent. Only if consumers consistently expected, for example, that their economic situation would improve during the next six months and then worsen would there be a chance for the differences to be observed in the time series. In addition, any such regular pattern of difference has been removed by seasonally adjusting the data. The same is true for differences in the wording of other questions.

For each country for which data could be obtained, a series of VAR regressions were run to test Granger causality between the sentiment index and a range of objective economic measure. The objective economic variables included variables generally considered to be determinants of sentiment, namely the unemployment rate, inflation rate, interest rate, and personal income, as well as measures of outcomes that are thought to be influenced by sentiment, such as consumption of durables, vehicle registrations, retail sales, total personal consumption, and gross domestic product (GDP). The selection of the variables was intended to cover a broad range of both informational inputs and behavioral outputs. All of these variables were not available for all of the countries, and the absence of data for any variable is indicated by a dash in the tables. Unfortunately, most of the missing variables related to consumption decisions, including purchases of durable goods and vehicle registrations.

Of the 45 countries identified as having consumer tendency surveys, eight of the countries had an insufficient number of data points on which to base the analysis (or I have simply been unable to obtain the information.) The analysis was thus limited to 37 countries. Quarterly data were used since monthly data was quite rare across most of the countries. All regressions included the latest data available when the analysis was done (usually the first quarter of 2004); the start dates for each equation differed depending on the availability of data for the sentiment index in each country and more often than not, the availability of the objective economic data.

All variables were defined as the year-to-year percentage change in the price adjusted series, with the exception of unemployment, inflation, and interest rates, which were defined as the year-to-year percentage point change. Five lags of all variables were entered into the regressions; the contemporaneous changes in the variables were not used in the regressions so that there was a true time separation between the survey measures and the objective economic variables. It is of course true that since the survey data are generally available well in advance of the objective economic measure they could be used to forecast the same periods economic outcome. Indeed, there are strong theoretical reasons to include the current period's survey measure in these regressions even if they were not available earlier than the objective data (Katona, 1951). Since the purpose of this analysis is not to test the predictive capacity of either the sentiment or the objective data, but to test the interconnections between an independently measured sentiment variable and objective data, a distinct time difference was favored.

All variables in all countries were transformed in the same way to facilitate comparisons across countries and across variables. There was no attempt to transform the data to match differences in the separate variables or by individual countries, or to maximize the observed relationship. These analyses were not designed to provide a definitive test of any specific hypothesis, but rather to generally ascertain the informational content of measures of consumer sentiment.

The results can be summarized as follows: measures of consumer sentiment show a significant correspondence with past changes in economic variables, and measures of consumer sentiment provide predictive information about future changes in those same variables. It should be no surprise that consumer sentiment provides both backward and forward looking information. This duality is as true for variables typically used as inputs to the formation of consumer expectations as it is for measures that reflect behavioral outcomes. A significant relationship between consumer sentiment and other economic variables, however, does not mean that the predictive power of either the sentiment variable or the economic variables was large. In fact, the amount of variance accounted for by the regressions that alternatively used the sentiment variables as an independent and as a dependent variable were quite small (generally well under 10%). There have been numerous analyses that focused on the issue of how much the sentiment data can reduce the mean square error in out of sample forecasts when all other economic variables are included in the regressions (see footnote 2 for a partial list). This analysis is focused on assessing the overall information content of the sentiment variables, its sensitivity to differences in measurement methodology, and its applicability across a wide range of countries and cultures.

4.2 Results by economic variables

Table 2 summarizes the results by variable.³ The first conclusion is that the information used by consumers to form their expectations is often the same information that the sentiment variable can predict. It should not be surprising that consumers view changes in employment, GDP, or even retail sales as indicators of strength or weakness in the economy and hence influence their level of economic sentiment, and in turn changes in sentiment forecast future changes in those same variables.

Whether you consider the fact that half of all tests on a given variable indicated a significant relationship represents a half-full or half-empty glass is of course a matter of perspective. I often find that psychologists are more impressed than economists, with the latter wondering why the relationship was not universally significant. Any shortfall can be attributed to one of two fundamental errors: the basic hypothesis is flawed or the relationship has been severely diminished due to measurement errors in either the sentiment or the economic variables.

The strongest and most widespread relationship was between unemployment and the sentiment index. In 50% of all countries there was a significant predictive relationship running from changes in the unemployment rate to changes in consumer sentiment, controlling for past changes in the sentiment variable. This indicates unemployment is a powerful determinant of changes in consumer sentiment. It was also true that in 62% of the countries a change in consumer sentiment proved to be a significant predictor in forecasting a future change in the unemployment rate, controlling for past changes in unemployment. Given the importance of wages and job security to the financial wellbeing of families, it is not surprising that whether consumers are optimistic or pessimistic is highly related to the future rate of unemployment; more surprising is that consumer expectations are often correct about year-ahead changes in the unemployment rate. This result supports the notion that consumer sentiment in large part reflects income expectations as well as uncertainty about future wage income.

It is of some importance to emphasize that when predicting future changes in the unemployment rate, for example, the sentiment variable must have accounted for the future changes in the unemployment rate over and above what could be anticipated based on five lags of the unemployment rate. The significance of this relationship is due to private information on potential changes in the unemployment rate (Curtin, 2003).

³ The full set of tables giving the significance levels for Wald tests for Granger causality for each variable within each country are available from the author.

Table 2 Proportion of Granger causality tests that indicated a significant relationship by variable

Variables	Objective Predicts Sentiment	Sentiment Predicts Objective
Unemployment Rate	50%	62%
Consumer Price Index	29%	26%
Short-term Interest Rates	40%	31%
Personal Income	56%	48%
Gross Domestic Product	59%	54%
Total Personal Consumption	36%	51%
Retail Sales	39%	55%
Expenditures for Durables	54%	45%
Vehicle Registrations	23%	46%

The next highest correspondence was between Gross Domestic Product (GDP) and sentiment. Past changes in GDP had a significant impact on sentiment in 59% of the countries, while past changes in sentiment were significantly related to future changes in GDP in 54% of the countries. Given that consumer spending accounts for over half of GDP, these relationships are not surprising. The significant links to future changes in GDP reflect the component questions about the overall economy and have also benefited from private as well as public information.

Changes in personal income indicated a comparable performance on these tests as well: past changes in income predicted future sentiment in 55% of the countries, and past changes in sentiment predicted future changes in income in 48% of the countries. These results again support the notion that – at the present time – sentiment is a broadly based measure of expected income prospects.

Sentiment was more successful in predicting future retail sales and new vehicle registrations. Sentiment anticipated changes in retail sales in 55% of the countries, and anticipated changes in vehicle registrations in 46% of the countries. Perhaps the most difficult to understand is the performance of durable goods expenditures. To be sure, quarterly data on expenditures for durable goods could not be located for most countries. Indeed, I was only able to locate quarterly data for durable expenditures in just eleven of the 37 countries. This is a serious drawback of any analysis of consumer sentiment index, a variable that the index was originally designed to predict. Durable expenditures predicted sentiment in 55% (6) countries, and sentiment predicted subsequent durable expenditures in 45% (5) of the countries – a difference of just one country. Nonetheless, it would seem to be a natural hypothesis that this relationship should be stronger than others, say for example, with GDP.

4.3 Results by country

Another view of the data also provides interesting results, with the findings summarized by country rather than by variable. It could be hypothesized that most of the significant relationships were associated with a subset of the countries, for example, the countries that developed market economies earlier or countries that have measured consumer expectations for a longer time period. Other hypotheses are related to differences in question wording, whether a six or twelve month horizon was used, and so forth. After an examination of the data, these hypotheses can be rejected. To be sure, some cannot be rejected with a good deal of confidence given that there were very few observations on which to base the comparisons.

Overall, I was unable to discern any pattern in the results. For example, the few countries that recorded no significant relationships could be due to the fact that some of the nine variables were missing; indeed, all of the countries that had no significant relationships included fewer than six of the variables.

Table 3 Proportion of Granger causality tests that indicated a significant relationship by country

Country	Objective predicts sentiment	Sentiment predicts objective	Country	Objective predicts sentiment	Sentiment predicts objective
AR	67%	50%	IT	25%	25%
AT	44%	33%	JP	25%	38%
AU	63%	88%	KO	25%	63%
BE	63%	25%	LT	50%	0%
BR	33%	17%	ML	0%	60%
CH	50%	38%	NL	38%	50%
CN	80%	60%	NO	56%	22%
CY	50%	0%	NZ	25%	50%
CZ	43%	29%	PL	20%	20%
DE	33%	67%	PT	43%	29%
DK	44%	33%	SE	56%	11%
EL	0%	17%	SI	17%	33%
ES	25%	88%	SK	29%	29%
FI	22%	33%	TH	17%	83%
FR	33%	33%	TW	20%	0%
HK	57%	57%	UK	33%	44%
HU	43%	57%	US	89%	56%
ID	75%	25%	ZF	43%	71%
IE	75%	75%			

A summary of these results are included in Table 3. Australia, Ireland, and the United States had the highest percentages of significant relationships, closely followed by Argentina, South Africa, China, Hong Kong, Spain, Germany, Hungary, and Thailand. The ability to predict the sentiment variables was much higher than the ability to predict the object variables in Sweden, Belgium, Indonesia, Norway, and Portugal. The sentiment index was much more likely to predict the objective variables in Spain, Germany, Australia, Korea, South Africa, and New Zealand. The smallest relationships were found for Greece, Taiwan, Brazil, France, Finland, Italy, and Slovenia.

This assortment of results cuts across the different questions used to construct the indexes as well as the different time horizons and whether the index included both backward and forward looking questions. The results were also independent of whether it was a developed or emerging economy, if the country had recently started the surveys or the surveys were longstanding, and so forth.

The data appear to suggest that a country's past economic history, the exact questions used to measure consumer sentiment, and the presumed differences in knowledge and sophistication of the populations did not produce corresponding variations in the analytic results across countries. Nonetheless, with so few cases for some countries, this analysis must be considered tentative.

5 New Agenda for Consumer Sentiment Research

It should be clear that the analysis was not intended to assess the ability of current measures of sentiment to act as a leading indicator of near term changes in the macro economy. Nor did the analysis attempt to assess how much variance the sentiment variable could explain in the subsequent changes in the objective variable – even when sentiment was a significant predictor, the explained variation was quite small in most cases.

Nonetheless, the data clearly indicate that sentiment measures contain a good deal of information about future trends in the macro economy. The motivation underlying this analysis was to identify a strategy for developing measures of consumer expectations that are better suited for short term forecasting purposes. While I do not doubt that a refinement of current methods would improve the forecasting performance of the data, it is now time to consider a more substantial break with the past. While the analysis did show some differences across countries and across variables, these differences were not primarily related to variations in sampling, question wording, or index construction. Indeed, the data makes a compelling case that refinements in the current methodology are unlikely to affect its performance as a leading economic indicator.

To be sure, I am not advocating that the current methodology be discarded and new measures adopted in the immediate future. The modifications that will be necessary during the decades ahead, however, will be far reaching. The required program of research is to determine what should be measured and how it should be measured. The goal is to transition

the surveys from the old to the new methodology in a seamless manner, using a series of incremental changes that maintains the comparability of the data over time. The challenge faced in the planning process is to design a series of robust experiments that can independently test the impact of each potential change prior to its implementation.

The three main phases of the transition research program involve what topics should be measured, how they should be measured, and how the data should be modeled in analysis. The theory underlying the sentiment index was developed in a different and simpler era, and its aim was to anticipate rather general trends in consumer spending in the context of a closed economy. It is sometimes assumed that since the goal of forecasting cyclical changes in spending has not changed, the questions used to determine consumer sentiment should remain constant. While this view is generally true, it is not entirely correct.

While income and employment prospects are still the main driving force, there is now a greater range of economic variables on which consumers base their decisions, and these additional economic variables will continue to grow in relative importance in the decades ahead. The borders of nation states no longer restrain information flows nor limit the direct purchase of goods or services by consumers. The economic sophistication and education levels of the population have increased substantially in recent decades. Local economies are becoming virtual global economies as consumers use the internet to seek out information about the price and quality of products and services from around the world. Consumers now more often incorporate into their expectations information that originates outside of their nation's borders, whether due to links from trade or financial markets, or due to more formal arrangements like the EU.

Moreover, aging populations have promoted a shift from the acquisition of consumer durables and the incurrence of debt toward the accumulation of assets and the preparation for retirement. This trend has also accelerated the underlying shift in demand from goods to services. As a result, the composition of cyclical changes in spending and saving behaviour are also undergoing fundamental shifts.

These are significant changes that require a new paradigm for measures of consumer sentiment. Economists have recently embraced a broader range of theoretical models that allow deviations from the classical assumptions of rational decision making. Models are now often based on bounded rationality, limited computational capacities, use of decision heuristics, the impact of framing on decisions, and the persistence of biases. While these models are likely to lead to more accurate predictions, they are much more complex to implement than models based on the traditional assumptions. Moreover, questions have again arisen about causal structures and self-fulfilling prophecies given the renewed recognition that consumers' expectations may themselves influence future economic conditions.

Research on how expectations should be measured involves both the question wording as well as the format in which the interviews are likely to take place. The University of Michigan has begun experiments with probability based measures of expectations, measures that have the capacity to yield estimates of the mean expectation as well as the uncertainty

with which the expectation is held. Although such questions may be difficult to ask and answer in telephone surveys, interviews conducted via the internet offer expanded possibilities in how the questions can be asked and the form in which the answers are recorded, including non-verbal graphical responses.

5.1 New sources of uncertainty

While the main focus of most consumers is now on job and income prospects, a wide range of other factors have grown in importance in recent years. This shift in emphasis will accelerate as a large proportion of the population will begin to retire in the coming decades. As people near retirement age, the importance of labour income declines while the importance of assets and pension entitlements rise. As a consequence, people will increasingly shift their attention from income uncertainties originating in the labour markets to uncertainties about future rates of return on financial assets. Moreover, compared with variations in labour market conditions, variations in the value of financial assets are likely to exhibit more abrupt and larger changes in response to both local and global financial markets. Little is known, however, about how consumers form expectations about future real rates of return on various assets.

Uncertainty about future rates of return on assets will prompt consumers to engage in precautionary declines in spending during retirement in much the same way as it prompted precautionary increases in savings prior to retirement. Understanding how people assess uncertainty must incorporate information on how people form expectations about future needs and how those expectations change as they age. Expectations about longevity and disability, in vivo transfers and bequests, as well as changes in expectations about the provisions of private and public pensions and health care programs will all play a more important role.

Given that pension entitlement programs will represent a major share of the incomes of the elderly, some have contended that there is little reason to expect income uncertainty during retirement to have a significant impact on the macro economy. By this same reasoning, it could also be argued that uncertainty about future labour income would not represent a significant problem since even at the worst of times the overwhelming majority remain employed. In addition, this line of reasoning ignores the potential impact of inflation, particularly changes in relative prices that are to the disadvantage of the elderly since not all pension entitlements are fully indexed to inflation nor indexed to the prices of goods and services generally purchased by the elderly.

Moreover, compared with variations in labour market conditions, valuations of financial assets are likely to exhibit more abrupt and relatively larger changes. Shocks in global financial markets are more rapidly reflected in domestic financial markets than in domestic labour markets. To be sure, consumers do not react to every change in asset prices, especially when their investment horizons are long, but among the elderly the horizon is bound to be shorter. Additionally, little is known about the thresholds or conditions which

prompt changes in expectations about future rates of return, whether those factors operate in an asymmetric fashion, how time preferences and risk aversion change as people age, and what sources of information have the most influence on their assessments.

5.2 Low probability events

Loss-aversion theory may help to explain why recoveries in confidence are slow to develop but declines occur abruptly. There have been occasions during the past fifty years that rapid and significant declines in confidence have been caused by unexpected events, and at other times those same unexpected events have only caused a temporary ripple in consumer confidence. While it has been repeatedly observed that apparently similar economic developments have had quite dissimilar effects on consumers' expectations, no theories have yet to be advanced to explain these differing reactions. Research needs to be directed toward understanding the conditions under which similar economic shocks produce quite dissimilar results.

Such theoretical developments are likely to involve the affective as well as cognitive aspects of consumer sentiment. It is the affective components of economic expectations – whether the content of the expectations are judged favorably or unfavorably – that serve to integrate diverse pieces of economic information. Moreover, it is the affective components that enable waves of optimism or pessimism to sweep across the population with great speed.

Whereas slow change allows gradual adaptation, rapid change prompts a disabling sense of uncertainty and disengagement. The inaction caused by such abrupt changes may generate a self-reinforcing process which acts to accentuate rather than mitigate the underlying source of uncertainty and, hence, act as a self-fulfilling prophesy. More generally, attention is being given to how consumers evaluate and incorporate information about low probability events that have potentially large consequences. This task is as important as it is challenging since it directly focuses on the extreme tails of probability distributions where prior research on economic expectations is almost entirely absent.

5.3 Improved measures of expectations

As expectations have become more central components of economic models, the theoretical specifications of the desired measures have become more exacting. The ideal measure would specify the characteristics of the complete probability distribution for a precisely defined future event or outcome (Manski, 2004). Given an estimate of the entire probability distribution for each individual, measures of central tendency can then be supplemented by measures of dispersion to capture the uncertainty that individuals associate with that expectation.

To measure the complete probability distribution of a given expectation requires more detailed and more complex questions. These more demanding questions require a high

degree of motivation on the part of respondents to provide accurate answers, often require respondents to access a greater amount of information from memory, and demand that respondents have greater conceptual and computational skills. These higher demands may cause respondents to refuse to answer the questions resulting in more missing data or cause respondents to simply guess or otherwise provide unconsidered answers resulting in greater measurement errors. In addition to the added cognitive burden, research has indicated that probability measures may reflect systematic biases, including the overestimation of low probability events and the underestimation of high probability events.

5.3.1 Experimental measures

To assess the potential of probability measures, the University of Michigan has begun experimenting with a limited number of probability questions to determine an optimal format of the questions, to assess the cognitive burden of various questions, and to determine how to best minimize any systematic biases in the responses. Thus far the tests have shown mixed results.

On the positive side, probability questions were easily adapted to phone interviews and respondents did not find them too onerous to provide an answer. Eleven probability questions have been asked during the past several years covering three core areas: employment, income, and financial investments. Each questions asked respondents to assess the probability that the:

- Y1: Household income will increase by more than overall prices during next five years
- Y2: Personal income will increase during next twelve months
- Y3: Personal income will increase by more than 10% during year ahead
- Y4: Personal income will increase by more than overall prices during the year ahead
- Y5: Income from social security and job pensions will be adequate for retirement
- J1: Respondent or spouse will lose a job within the next five years that they wanted to keep
- J2: Respondent will lose a job within the next twelve months that they wanted to keep
- J3: Respondent will get a new job at least as good in terms of wages and benefits
- M1: Mutual fund of \$1,000 will increase in value during next twelve months
- M2: Mutual fund of \$1,000 will increase by at least 10% during next 12 months
- M3: Mutual fund of \$1,000 will increase by more than overall prices during next 12 months

Note that all the questions asked about "one-sided" probabilities: the probability of a job loss, the probability of an increase in income, or the probability of a gain in the value of mutual funds. The questions did not attempt to determine the probability distribution across all potential outcomes. Rather the questions focused on the probability of a single outcome

(for example, a 10% increase in income). The first set of experiments thus fell short of specifying the entire probability distribution. This decision was based on the trade-off between the much larger number of questions that would be needed to measure the entire probability distribution of a single topic and the number of topic areas that could be covered in the experiment. The time period each question was asked is shown in Table 4, along with the mean and median of the replies, the extent of missing data and the number of cases.

Table 4 **Summary statistics for probability measures of expectations**

	Probability that . . .	Time Period	Mean	Median	% Miss	Cases
Y1:	Household income will increase by more than overall prices during next five years	1998:01 – 2006:09	40.9	40.4	3.0%	52642
Y2:	Personal income will increase during next twelve months	2002:06 – 2006:09	51.1	50.1	2.2%	26101
Y3:	Personal income will increase by more than 10% during year ahead	2002:06 – 2003:05	30.2	19.0	3.9%	6015
Y4:	Personal income will increase by more than overall prices during the year ahead	2003:06 – 2004:09	36.2	30.1	2.8%	8043
Y5:	Income from social security and job pensions will be adequate for retirement	1998:01 – 2006:09	35.3	29.2	3.4%	52642
J1:	Respondent or spouse will lose job within next five years that they wanted to keep	1998:01 – 2006:09	18.9	9.7	1.3%	52642
J2:	Respondent will lose a job within next twelve months that they wanted to keep	2002:06 – 2004:09	15.0	4.9	1.1%	9052
J3:	Get a new job at least as good as job lost in terms of wages and benefits	2002:06 – 2004:09	47.0	49.8	2.3%	9052
M1:	Mutual fund of \$1,000 will increase in value during next twelve months	2002:06 – 2006:09	48.9	50.0	6.2%	26101
M2:	Mutual fund of \$1,000 will increase by at least 10% during next 12 months	2002:06 – 2003:05	31.1	25.0	7.6%	6015
M3:	Mutual fund of \$1,000 will increase by more than overall prices during next 12 months	2003:06 – 2004:09	40.3	49.5	6.0%	8043

It should also be noted that it is unlikely that the same probability distributions would be measured if the frame of reference was shifted from gains to losses (Kahneman and Tversky,

1979). For example, it is likely that asking about the probability of losing a job is not the inverse of asking about the probability of keeping a job, or asking the probability of income or investment gain is not equivalent to asking about losses. It is likely that responses to questions about gains and losses are not symmetrical. For these initial experiments, the questions referred to the most common frame of reference for the respondents, namely losses about jobs and gains about income and investments.

5.3.2 Incomplete use of probability scales

The extent of "don't know" responses is typically taken as an indication of the difficulty of the question, and in this case the relatively low missing data rates suggest that most respondents did not find these questions too difficult (see Table 4). For the probability questions about job prospects, missing data rates ranged from 1.1% to 2.3%, missing data rates on income prospects were higher, ranging from 2.2% to 3.9%, and missing data rates for the questions on returns to investments in mutual funds were the highest, ranging from 6.0% to 7.6%. This pattern was not surprising and indicates the greater difficulty in estimating stock market returns than income or job prospects. Nonetheless, the relatively low levels of missing data indicate that these questions were not perceived to being too difficult to answer.

A closer inspection of the data does raise more concerns. More troublesome is that the full range of probability scale was not utilized by respondents. Across the eleven questions just 2% to 6% of all respondents gave answers that did not indicate near certainty (probabilities of 0% to 5% or 95% to 100%) or were not rounded responses (multiples of 10 or 25 and 75). Indeed, about half of all respondents gave certain or near certain responses or responded 25%, 50% or 75%, a virtual five point scale. Most of the rest of the responses were multiples of 10.

Why did the vast majority of respondents choose so few points out of the range from 0 to 100? The most common answer is that the responses represent "rounded" estimates. There have been several hypotheses suggested to account for this pattern. Given the potentially high cognitive burden of these questions, some have suggested that respondents simply report the first satisfactory answer that comes to mind and these estimates are often rounded (Krosnick, 1999). Others have suggest that digit preference is a widespread phenomenon, exhibited by nearly all responses to open-ended numeric questions of all types (Baker, 1992; Edouard and Senthilselvan, 1997). The digits that are favored are precisely those that constituted the vast majority of the responses to the probability questions, namely multiples of 10 plus 5, 25, and 75. Others have suggested that people retrieve the most accessible data from memory, and that is the mode rather than the mean or median of the underlying probability distribution (Lillard and Willis, 2001). This "modal choice" hypothesis suggests that the data that is easiest to retrieve is where the probability distribution reaches its peak.

Table 5 Summary statistics for probability measures of expectations

Probability that . . .	Certain (0% or 100%)	Near Certain (1-5% or 95-99%)	Quarter Rounding (25%,50%, 75%)	Ten Rounding (10,20,30, 40,60,70, 80,90)	All Other Responses	Total
Y1: Household income will increase by more than overall prices during next five years	18	4	27	48	3	100%
Y2: Personal income will increase during next twelve months	32	7	20	38	3	100%
Y3: Personal income will increase by more than 10% during year ahead	36	7	17	37	3	100%
Y4: Personal income will increase by more than overall prices during the year ahead	26	7	23	40	4	100%
Y5: Income from social security and job pensions will be adequate for retirement	23	6	22	45	4	100%
J1: Respondent or spouse will lose job within next five years that they wanted to keep	45	5	14	34	2	100%
J2: Respondent will lose a job within next twelve months that they wanted to keep	45	12	10	31	2	100%
J3: Get a new job at least as good as job lost in terms of wages and benefits	21	5	25	45	4	100%
M1: Mutual fund of \$1,000 will increase in value during next twelve months	10	6	31	47	6	100%
M2: Mutual fund of \$1,000 will increase by at least 10% during next 12 months	15	8	25	47	5	100%
M3: Mutual fund of \$1,000 will increase by more than overall prices during next 12 months	11	7	31	46	5	100%

Economists have interpreted such results as exhibiting a "near rationality," whereby the rounding represents the level of precision that is associated with differences that matter to the respondent. A more traditional economic interpretation involves balancing costs and benefits: people want to minimize computational costs so they compute detailed probabilities only to

the extent that the finer detail would affect their decisions. These explanations also favour rounded numbers.

There is another explanation emerging from neuroscience that asserts that people do not possess such fine detail but hold coarser probabilities in memory. In this explanation rounding does not reflect the additional cost of cognitive processing but reflects the true state of their knowledge as most people only maintain "rounded" numbers in memory (Dehaene, 1997). This hypothesis has other implications, including that probabilities would not generally adjust to small changes but only after some threshold was reached, and that sudden and unusually large change in probabilities could result when the new information exceeded the threshold.

5.3.3 Errors in probability assessments

The experimental questions allowed another test of how well the respondents understood the underlying task. Respondents were asked about the probability that their income would increase during the year ahead and then asked about the probability that their income would increase by more than 10% during the year ahead (questions Y2 and Y3). A conjunctive error appears if a respondent judged the probability of getting any increase in income was lower than the probability of getting a 10% increase in income. There are two possible exceptions to this test of the conjunction fallacy. The first is if a respondent expected an income decline with certainty, then the probability of any increases as well as the probability of an increase of 10% could be zero; the second is if the respondent expected an increase of more than 10% with certainty, then both probabilities could be 100%.

When the two responses were compared on more than five thousand cases, 53% of all replies were consistent in that respondents assigned a lower probability to getting a 10% increase than to getting any income increase at all. Another 14% reported zero probabilities to both questions, and 6% reported that both probabilities were 100%. If these are also assumed to be consistent, a total of 73% provided consistent probabilities (see Table 6).

Inconsistent replies to the two probability questions were given by a total of 27%, but the majority of these errors were recorded when respondent assigned the same probability to both expecting any increase as well as expecting a 10% increase. Just 8% of the sample gave different probabilities that were inconsistent. Some of these equal probabilities may have resulted from the tendency of respondents to round their answers and may not reflect a true inconsistency. Nonetheless, providing equal probabilities to two questions that span a relatively large difference in income expectations seems to indicate that these respondents did not understand the underlying notion of probabilities or even of percentages.

The same overall pattern was found to the questions on the probability of any positive increase in a mutual fund and the probability of a 10% gain (questions M1 and M2). Although the majority of respondents (54%) provided consistent probabilities, a somewhat larger percentage reported inconsistent probabilities to the mutual fund question compared with income expectations. Again, most of the inconsistent replies involved identical probabilities.

Inconsistent replies that were not equal were recorded twice as frequently for mutual fund investments compared with income expectations (16% versus 8%). While rounding and digit preference may be expected to be even more likely responses to questions about future returns on investments, it is still rather disappointing that even a 10% gain was judged equal to any gain at all by one-quarter of all respondents.

Table 6 Consistent and inconsistent probability responses to questions on expected income gains and expected gains in the value of mutual funds

	Personal Income		Mutual Fund	
	10% Gain	Real Gain	10% Gain	Real Gain
Consistent with probability of any increase in income/mutual fund – Total	73	69	60	56
Both 0%	14	12	5	2
Both 100%	6	5	1	2
Other consistent replies	53	52	54	52
Inconsistent with probability of any increase in income/mutual fund – Total	27	31	40	44
Equal probabilities	19	20	24	27
Other inconsistent replies	8	11	16	17
Total	100%	100%	100%	100%
Cases	5724	7731	5438	7359

The questions on the probability of income and mutual fund increases were also followed by an alternative set of questions about real gains, that is, the probabilities of gains after an adjustment was made for changes in the overall level of inflation (questions Y2 and Y4 and questions M1 and M3). The consistency test is much less rigorous in this case since the level of inflation was relatively low; at the time these data were collected the inflation rate was about 2.5% (CPI-u). Surprisingly, the number of consistent replies was quite close to the 10% experiment for both the income question (52% versus 53%) and the mutual fund question (54% versus 52%). Moreover, similar proportions gave equal non-zero probabilities to both the 10% experiment as well as the comparison to inflation adjusted gains. For the question on personal income, 27% gave equal probabilities for any gains and an inflation adjusted gain, which was identical to the percentage in the 10% experiment. Even for the mutual fund experiment, the two percentages giving equal probabilities were quite close (24% versus 27%). Again the data suggest that either the probabilities given by respondents were rounded and cover a broad range or that a significant number of respondents didn't understand the concept of probabilities.

The data on whether the respondent gave consistent or inconsistent replies can also be examined within demographic subgroups. The standard hypotheses are related to age, education and income. The higher the education, the higher the cognitive capacity; advancing age tends to diminish cognitive capacity, and higher income households presumably have greater experiences with evaluating economic information. Table 7 indicates that the replies for the questions on the probability of any sized income increase compared with the probability of a 10% income increase, and Table 8 show similar data for the replies concerning mutual funds. The other two comparisons between nominal and real gains are not shown but exhibited similar patterns.

Table 7 Consistent and inconsistent replies to probabilities of income increases by demographics

Probability of Any Sized Gain versus 10%+ Gain in Income	Consistent Replies				Inconsistent Replies			Total	Cases
	Both 0%	Both 100%	Other Consis.	Sub- total	Equal Prob.	Other Incon.	Sub- total		
Age of Respondent									
18-34	5	11	53	69	21	10	31	100%	1354
35-64	12	5	56	73	20	7	27	100%	3400
65 or older	33	1	42	76	15	9	24	100%	937
Education of Respondent									
High School or less	18	4	46	68	21	11	32	100%	1928
Some college	13	6	52	71	20	9	29	100%	1301
College graduate	12	7	58	77	17	6	23	100%	2472
Income of Household									
Bottom third	19	5	42	66	23	11	34	100%	1397
Middle third	13	5	55	73	19	8	27	100%	1851
Top third	10	6	61	77	17	6	23	100%	2022

The demographic data on income expectations do indeed indicate that giving a zero probability to both questions is most common among those over age 65, reported by 33% compared with just 5% among those under 35. Similarly, probabilities of 100% were recorded more often by those under 35 than those over 65 – 11% versus 1%. Overall, consistent replies were more frequently given by younger, more educated, and by those with higher incomes.

The demographic distributions for the question on mutual fund gains show generally similar results, although the differences among demographic groups were generally wider. Among the least educated, 51% provided inconsistent replies for mutual fund gains compared with just 32% among those with college degrees. This same difference is mirrored by income groups as well. Somewhat surprisingly, there was little difference across age groups for the mutual fund probabilities.

Table 8 Consistent and inconsistent replies to probabilities of gain in mutual funds by demographics

Probability of Any Sized Gain versus 10%+ Gain in Income	Consistent Replies				Inconsistent Replies			Total	Cases
	Both 0%	Both 100%	Other Consis.	Sub- total	Equal Prob.	Other Incon.	Sub- total		
Age of Respondent									
18-34	3	1	56	60	22	18	40	100%	1294
35-64	5	1	54	60	25	15	40	100%	3284
65 or older	11	1	47	59	24	17	41	100%	828
Education of Respondent									
High School or less	6	2	41	49	30	21	51	100%	1759
Some college	5	1	55	61	23	16	39	100%	1239
College graduate	5	1	62	68	19	13	32	100%	2415
Income of Household									
Bottom third	6	1	47	54	26	20	46	100%	1291
Middle third	6	1	51	58	24	18	42	100%	1776
Top third	4	1	61	66	22	12	34	100%	1983

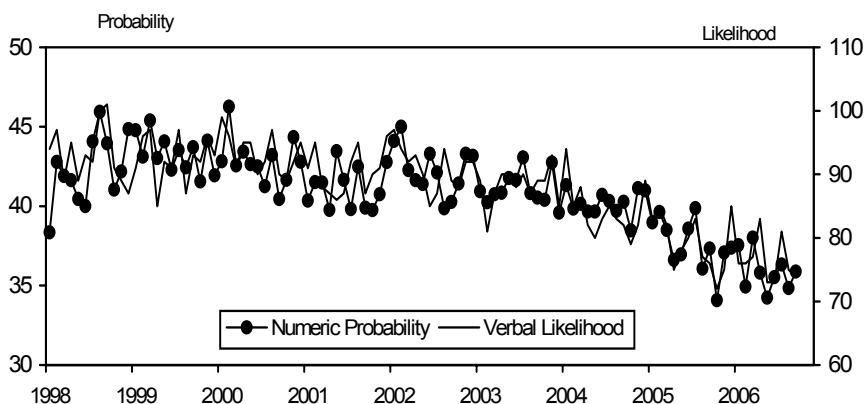
5.3.4 Aggregate comparison of probability and verbal likelihood scales

While no rigorous comparison between the results based on probability scales and verbal likelihood scales is possible at the level of the individual respondent, some time series comparisons are possible. Given that the data are intended to act as a leading indicator of changes in the macro economy, the time series tests may be the most appropriate if not the most rigorous test of the underlying theory. Unfortunately, only a few questions have been asked over a long enough periods for time series testing (questions Y1 and J1).

Figure 1 shows the correspondence between the probability of real income gains (Y1) and a verbal likelihood question about real income expectations. Instead of asking about the probability of real income gains, the verbal likelihood question is a typical three point scale that asks whether their incomes are expected to increase faster than prices, prices are expected to increase faster than their incomes, or if both are expected to increase at the same rate. The correspondence is quite close between these two series (time series correlation of 0.8). The close correspondence is all the more surprising given the significant difference between the two questions: the probability question asks about the next five years and the verbal likelihood questions about the year ahead; the probability scale only asks about expected gains not expected gains and declines in real income; and the probability scale allows 100 different gradations of responses compared with just one-point "increase" code on the verbal likelihood scale. This comparison implies that either respondents failed to

understand these crucial differences in the wording of the questions or that respondents were either unable or unwilling to expend the resources to make the required distinctions.

Figure 1 **Real Income expectations: Numeric probability versus verbal likelihood measures**

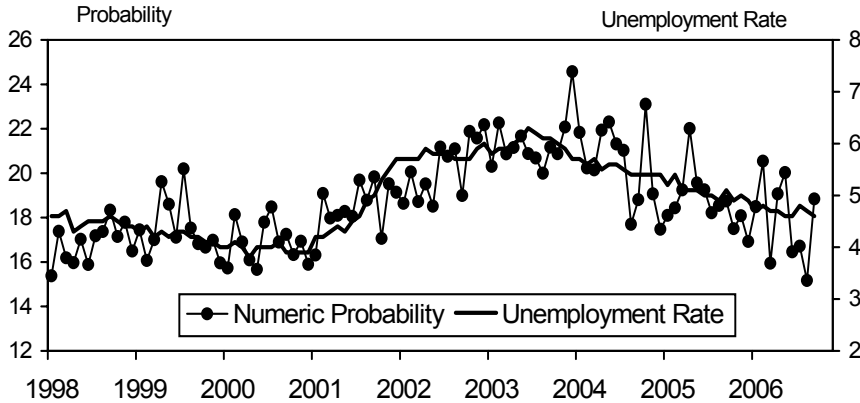


The job probability measure (J1) was compared with the actual unemployment rate to determine the correspondence of the survey data with the actual developments in the U.S. unemployment rate. The contemporaneous time-series correlation with the probability of job loss and the unemployment rate was 0.78 (see Figure 2). This rather high correlation was recorded despite the fact that the probability measure asked about job losses over the next five years. Although the question on the probability of job loss during the year ahead (J2) would seem to have a greater correspondence with the current national unemployment rate, the time series correlation was much lower at just 0.15 (and insignificant with 28 observations).

Aside from the issue of whether the survey data led or lagged the unemployment rate, the survey measure seems to be quite volatile when compared with the unemployment rate. Indeed, in the last year or so the proportion of false to true signals seems particularly high (see Figure 2). The same type of instability is found if the data are averaged into quarterly observations.

The basic problem appears to be that the sampling errors have overwhelmed any real change in how respondents judge the probability of a job loss. Even if the measure was an accurate leading indicator of changes in the unemployment rate, its instability would mean that the series would give nearly as many false signals as true indications of change in joblessness. Indeed, during the past several years, the probability of job loss repeatedly recorded reversals in trend, with the size of the reversals equal to about half of the range from its peak to the trough values.

Figure 2 Probability of job loss versus national unemployment rate



The interquartile range of the time-series realizations of probabilities about future job losses was just 3.1. The average standard error of the monthly time-series observations was 1.1. Thus, the interquartile range of the job probability measure is less than 3 times the standard error of the monthly estimates. In comparison, the monthly data for the Index of Consumer Sentiment has an interquartile range that is more than 10 times the size of the standard error of the estimate. Even when the data are averaged to quarterly observations (1500 cases), the interquartile range remained at 3.1 and the standard error of the quarterly observations was reduced to 0.65, or just under 5 times the estimate's standard error. This suggests that very large sample sizes would be needed for more precise estimates of change; for example, a sample of 10000 observations would reduce the standard error of the estimate to 0.25 or about 12 times its interquartile range.

Although these experiments are not complete, the preliminary results are as challenging as they are promising. The comparison of the probabilities of real income change and the verbal likelihood scales indicated that these two approaches yield nearly identical results. While some may take this as a reason to adopt the probability measures, the data also indicated errors in the probability judgements. Moreover, the results for the question on the probability of job loss indicated some substantial methodological problems, including a limited response range as well as a relatively high instability in the responses. Unfortunately, this amounts to a one-sided comparison since the more traditional likelihood measures were not subjected to such scrutiny.

These difficulties certainly do not recommend an immediate and widespread shift from verbal likelihood to numeric probability measures. Nonetheless, some measures may be usefully asked in the probability format. The identification of which areas hold the most promise for probability questions remains a high priority task. Given the long lead times required to develop and test new survey measures, such experiments should be encouraged and funded in the coming years.

5.4 Development of a new analysis strategy

Expanding the research potential of the survey data is also part of our development plans. I will mention just two shortcomings of the most common analytic models, both involving aggregation. First, most models do not disaggregate the major components of consumer demand; second, most models do not disaggregate household data into demographic and economic subgroups to make those forecasts.

Macro models that use "consumer sentiment" for all purposes miss one simple fact: no single index of consumer expectations can be devised to accurately predict all types of consumer expenditures at all times. Consumer expectations for interest rates, inflation, income, jobs, stock prices, home values, economic policies, to name just a few, have moved independently and often in opposite directions during the past several years. Although these are all factors that determine people's expectations about their future real income, each of these factors has a very different impact on spending on homes, vehicles, household durables, travel, and so forth. More accurate forecasts require disaggregating consumer expenditures into its major components and then utilizing the most appropriate information on consumer expectations to forecast each component. The individual forecasts of the component series can then be combined to produce estimates at higher levels of aggregation.

Cyclical changes in total personal consumption expenditures are dominated by spending on durable goods, and it is for these discretionary purchases that consumer expectations were originally designed to predict. To be sure, some spending on non-durables and services represent discretionary purchases, but most of the expenditures in these categories are relatively insensitive to changes in expectations. For these expenditures, the best use of the survey data would be to utilize the data to generate forecasts of the more usual explanatory variables, such as income, employment, and prices in advance of their official release by government agencies. As shown by the analysis in this paper, the sentiment data contains reliable information about future changes in these variables.

Consumer expectations can only provide reliable information on the near-term future; they are useless for long term forecasts. Without doubt, in the long run income and consumption form a cointegrating relationship, with consumption a relatively constant proportion of income. Long term growth in consumption is tantamount to long term growth in income. It is only in the short run that the growth rates in consumption and income diverge. Indeed, the permanent income and life-cycle hypotheses provide for compelling reasons why they should diverge, namely so that consumers can maximize lifetime utility by smoothing consumption. Moreover, when a particular divergence has been observed, sooner or later the divergence will reverse itself, as is true of any cointegrating relationship. Indeed, to admit to the ephemeral nature of consumer sentiment is not to acknowledge a theoretical deficit; rather, it is the very strength of consumer expectations for the analysis of business cycles since it indicates the direction, the timing, and the extent of the divergence.

To capitalize on this aspect of consumer sentiment data, the data should be disaggregated by demographic and economic subgroups. Analysis based on particular income, age, or regional groups has already shown promise. For example, some purchases are dominated by younger consumers, by higher income consumers, by urban consumers, and so forth. Such disaggregated

forecasts can be recombined before publication. The disaggregated data are immediately released for income, age, and regional subgroups to promote the use of these new forecasting models.

This same goal could be approached from an entirely different angle. Consumer survey data are typically published so that the results are representative of all households in a particular country. No recognition is given to the differential economic impact of these households as the highest income household is given the same weight in the index as the lowest income household. Case weights could be devised to account for the differential economic status of individual households. The EU now does something similar when combining the data from different countries to form a sentiment index for the entire community; only in that case each country's data is weighted by the proportion of total spending. The same type of procedure could be applied within countries as well as across countries. Thus, for example, for the income and age groups that spent the highest amounts on discretionary purchases would be given correspondingly higher weight in computing the index within each country.

6 Changes in Sample Design and Survey Administration

Along with the challenges of developing new questionnaire content, the surveys face increasing pressure to change the sample design and the means by which the surveys are administered. Just as few argued that telephone surveys were superior to face-to-face in-home surveys thirty years ago, few would now argue that internet surveys are superior to telephone surveys. The outcome of the same comparison made ten or twenty years from now, however, is not obvious.

Change is sometimes motivated by the push from deteriorating conditions, and sometimes by the pull of potential improvements. The move from in-home surveys to telephone surveys was almost entirely due to push from rising costs; there was little if any pull from the greater research potential of telephone surveys. The shift away from telephone surveys will not be primarily determined by higher costs but by the push of falling response rates; the shift toward the internet will also be due to the ability of internet surveys to handle much more complex research measurements. Just like the first steps toward the use of the telephone, the initial uses of the internet will be to collect experimental and supplemental data and then to gradually over time become a dual-mode survey. Much of the learning about how to best use internet surveys will be accomplished by the more rapid development of this tool for business surveys.

The pull of internet surveys is substantial for both respondents as well as survey organizations. Respondents could view online information about the research project, check the project's credentials, consult privacy and confidentiality pledges and other information about the protection of human subjects, and complete the interview at a time that best meets their own schedule. Survey organizations could utilize more complex questionnaires, multi-media content, design questions and response codes using graphical as well as verbal methods, provide background information, provide definitions and question instructions, and automatically generate additional probes to clarify and resolve any inconsistencies.

The promise of internet administered surveys is as great as the methodological innovations that are required to produce scientific estimates. There are several significant problems that must be resolved before the internet can be used to conduct interviews. These issues include the lack of coverage of the entire population, the inability to select a representative sample of households, and the inability to verify which person within the household actually answered the questions. The most likely first step would be to utilize a dual-mode design whereby all traditional survey design features are maintained and respondents are given the choice to either complete the interview on the phone or on the internet. This narrows the potential differences to how responses to the same question differ depending on how the interview was administered.

7 Concluding Remarks

Change is the only constant in economics. There is a growing recognition that there has been a substantial change in how people evaluate and utilize economic information in managing their financial affairs. The changes have been far reaching, expanding the conceptualization of consumer sentiment and challenging the traditional measurement methodology. As populations age, concerns shift from the uncertainties of the labour market to uncertainties about future real rates of return on financial assets including their pensions. Moreover, as more countries measure consumer sentiment at earlier stages in their economic development, there is a growing gap across countries in the economic sophistication of consumers. The old research paradigm must adapt in order to advance our understanding of the role of the consumer in shaping macro economic trends during the 21st century.

No specific changes in question wording or survey methodology have been proposed, although such changes will certainly be required. What I have proposed is a research agenda to help guide the future development. Explicit recommendations have been avoided because the very notion of harmonization must also evolve so as to encompass more diverse measurement methodology. We must abandon the notion that one index will be suitable for all predictive purposes, that the same questions will be suitable in all countries, and that all interviews need to be conducted in a single interviewing mode rather than allow the mode of interviewing to vary in order to achieve the highest response rate. While there still needs to be a clear standard for the integration of the findings from sentiment surveys, the methods of harmonization must be shifted from the input questions to the output indices.

A broad range of independent research projects will make the necessary advances in theory and survey methodology. A broad diversity in approaches should be encouraged as the best means to promote scientific advancement. While there is no urgency to produce new results in the immediate future, such fundamental changes mean that new data must be collected and tested for some years before being widely adopted. There is no better time than the present to begin this challenging research agenda.

References

- Abderrezak, Ali. 1997. "Consumer expectations and cyclical durations," *Quarterly Review of Economics and Finance* 37:4, pp. 843-54.
- Acemoglu, Daron and Andrew Scott. 1994. "Consumer confidence and rational expectations: Are agents' beliefs consistent with the theory?" *Economic Journal* 104, pp. 1-19.
- Adams, F. Gerard. 1964. "Consumer attitudes, buying plans, and purchases of durable goods: A principal components, time series approach," *Review of Economics and Statistics* 46:4, pp. 347-55.
- Adams, F. Gerard. 1965. "Prediction with consumer attitudes: The time series - cross section paradox," *Review of Economics and Statistics* 47:4, pp. 367-78.
- Baker, Michael. 1992. "Digit Preferences in CPS Unemployment Data," *Economic Letters* 39:1, pp. 117-21.
- Blanchard, Olivier (1993), "Consumption and the recession of 1990-1991," *American Economic Review* 83, no. 2: 270-4.
- Carroll, Christopher D.; Jeffrey C. Fuhrer and David W. Wilcox. 1994. "Does consumer sentiment forecast household spending? If so, why?" *American Economic Review*, pp. 1397-408.
- Cochrane, John H. 1994. "Shocks," *Carnegie-Rochester Conference Series on Public Policy* 41, pp. 295-364.
- Curtin, Richard. 1984. "Consumer attitudes for forecasting," in *Advances in Consumer Research*, Thomas C. Kinnear, ed. Association for Consumer Research, pp. 714-17.
- Curtin, Richard T. 1983. "Curtin on Katona. Contemporary Economists in Perspective," Henry W. Spiegel and Warren J. Samuels, eds. New York: Jai Press, pp. 495-522.
- Curtin, Richard, Stanley Presser and Eleanor Singer. 2001. "The Effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly* 64:4, pp 413-28.
- Curtin, Richard. 2003. "Unemployment Expectations: The Impact of Private Information on Income Uncertainty," *Review of Income and Wealth* 49:4, pp 539-54.
- Dalsgaard, Thomas; Jorgen Elmeskov and Cyn-Young Park. 2002. "Ongoing Changes in the Business Cycle: Evidence and Causes," OECD.
- Danthine, Jean Pierre; John B. Donaldson, and Thore Johnsen. 1998. "Productivity growth, consumer confidence and the business cycle," *European Economic Review* 42, pp. 1113-40.
- Dehaene, S. 1997. *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.
- Dunkelberg, W. C. 1969. *Forecasting consumer expenditures with measures of attitudes and expectations*. Michigan: University of Michigan.
- Dunkelberg, W. C. 1972. "Impact of consumer attitudes on behavior: A cross-section study," in *Human Behavior in Economic Affairs*, Burkhard Strumpel, James N. Morgan and Ernest Zahn, eds. New York: Elsevier.
- Edouard, L. and A. Senthilselvan. 1997. "Observer Error and Birthweight: Digit Preference in Recording," *Public Health* 111:2, pp. 77-9.
- Eppright, David R.; Nestor M. Arguea and William L. Huth. 1998. "Aggregate consumer expectation indexes as indicators of future consumer expenditures," *Journal of Economic Psychology* 19, pp. 215-35.
- European Commission. 2004. "The Joint Harmonised EU Programme of Business and Consumer Surveys," OECD, Directorate General Economic and Financial Affairs.
- Fair, Ray C. 1971. "A Short-Run Forecasting Model of the United States Economy," in *A Short-Run Forecasting Model of the United States Economy*, Ray C. Fair. Lexington, ed. Massachusetts: Heath.
- Ferber, Robert. 1966. "Anticipations, statistics and consumer behavior," *American Statistician*, pp. 20-4.
- Friend, Irwin and F. Gerard Adams. 1964. "The predictive ability of consumer attitudes, stock prices, and non-attitudinal variables," *Journal of the American Statistical Association* 59, pp. 987-1005.
- Fuhrer, Jeffrey C. 1988. "On the information content of consumer survey expectations," *Review of Economics and Statistics* 70:1, pp. 140-4.
- Garner, A. 1981. "Economic determinants of consumer sentiment," *Journal of Business Research* 9, pp. 205-20.
- Garner, A. 1991. "Forecasting consumer spending: Should economists pay attention to consumer confidence surveys?" *Economic Review*, pp. 57-71.

- Goldrian, G.; J.D. Lindlbauer and G. Nerb. 2001. "Evaluation and Development of Confidence Indicators Based on Harmonised Business and Consumer Surveys," OECD, Directorate General Economic and Financial Affairs.
- Golinelli, Roberto and Giuseppe Parigi. 2004. "Consumer Sentiment and Economic Activity: A Cross-Country Comparison," *Journal of Business Cycle Measurement and Analysis* 2:1, pp. 147-70.
- Hall, Robert E. 1993. "Macro theory and the recession of 1990-1991," *American Economic Review* 83:2, pp. 275-9.
- Howrey, E. P. 2001. "The Predictive Power of the Index of Consumer Sentiment," *Brookings Papers on Economic Activity* 1, pp. 175-216.
- Huth, William L.; David R. Eppright and Paul M. Taube. 1994. "The indexes of consumer sentiment and confidence: Leading or misleading guides to future buyer behavior," *Journal of Business Research* 29, pp. 199-206.
- Juster, F. Thomas 1964. *Anticipations and Purchases*. Princeton: Princeton University Press.
- Juster, F. Thomas. 1969. "Consumer anticipations and models of durable goods demand," in *Economic Forecasts and Expectations*, Jacob Mincer, ed. New York: Columbia University Press for NBER.
- Juster, F. Thomas and Paul Wachtel. 1972. "Anticipatory and objective models of durable goods demand," *American Economic Review* 62:4, pp. 564-79.
- Kahneman, Daniel and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk," *Econometrica* 47:2, pp. 263-92.
- Katona, George. 1964. *The mass consumption society*. New York: McGraw-Hill.
- Katona, George. 1960). *The powerful consumer: Psychological studies of the American economy*. New York: McGraw-Hill.
- Katona, George. 1951. *Psychological analysis of economic behavior*. New York: McGraw-Hill.
- Katona, George. 1975. *Psychological economics*. New York: Elsevier.
- Keyfitz, Robert. 2004. "U.S. Consumers and the War in Iraq," *Business Economics* 39:2.
- Klein, L. R. and J. B. Lansing. 1955. "Decisions to purchase consumer durable goods," *Journal of Marketing* 20, pp. 109-32.
- Krosnick, Jon M. 1999., "Survey Research," *Annual Review of Psychology* 50, pp. 537-67.
- Lillard, Lee and Robert Willis. 2001. "Cognition and Wealth: The Importance of Probabilistic Thinking," Michigan Retirement and Research Center, Working Paper no. 2001-007.
- Ludvigson, Sydney. 2004. "Consumer Confidence and Consumer Spending," *Journal of Economic Perspectives* 18:2, pp 29-50.
- Manski, Charles F. 2004. "Measuring Expectations," *Econometrica* 72:5, pp. 1329-76.
- Matsusaka, John G. and Argia M. Sbordone. 1995. "Consumer confidence and economic fluctuations," *Economic Inquiry* 33, pp. 296-318.
- Maynes, E. Scott. 1967. "An appraisal of the consumer anticipations approaches of forecasting," American Statistical Association, Proceedings of the Business Economics Statistical Section of the American Statistical Association, pp. 114-23.
- Mueller, Eva. 1957. "Effects of consumer attitudes on purchases," *American Economic Review* 47:6, pp. 946-65.
- Mueller, Eva. 1963. "Ten years of consumer attitude surveys: Their forecasting record," *Journal of the American Statistical Association* 53, pp. 899-917.
- Nilsson, Ronny. 2000a. "Consumer Surveys: Methodology, Analytic Use, and Presentation of Results," OECD.
- Nilsson, Ronny. 2000b. "Confidence Indicators and Composite Indicators," paper presented at CIRET Conference, Paris.
- Praet, P. and J. Vuchelen. 1989. "The contribution of consumer confidence indexes in forecasting the effects of oil prices on private consumption," *International Journal of Forecasting* 5, pp. 393-7.
- Suits, Daniel B. and Gordon R. Sparks. 1965. "Consumption regressions with quarterly data," in *Brookings Quarterly Econometric Model of the United States*, J. S. Duesenberry and others, eds. Chicago: Rand McNally.
- Tobin, James. 1959. "On the predictive value of consumer intentions and attitudes," *Review of Economics and Statistics* 41:1, pp. 1-11.

Deviation Cycles in Manufacturing: Business Cycle Measurement and Leading Indicators

Ard H.J. den Reijer^{*}

Abstract

The deviation cycles in the manufacturing industry of nine OECD-countries are identified by applying the Christiano-Fitzgerald band-pass filter. Turning points, low- and high-growth phases and other descriptive statistics are derived from these deviation cycles. A regression based test statistic is applied to test for duration dependence. Moreover, the international linkage between the cyclical motions in the manufacturing industry of two countries is investigated by measuring the degree of synchronisation. In addition to measuring the cyclical fluctuation, a composite leading indicator is constructed which replicates and predicts the deviation cycle in the manufacturing industry. This composite leading indicator is a single index composed of economic, financial and expectation variables possessing leading properties.

Key Words: Business cycles, Turning points, Leading indicators, Band-pass filter

JEL Classification: C82, E32, E37

^{*}

De Nederlandsche Bank (DNB), Economics and Research Division (Amsterdam, The Netherlands); a.h.j.den.reijer@dnb.nl

I would like to thank Claudia Kerkhoff and Peter Keus for statistical assistance and Bouke Buitenkamp, Leo de Haan, Lex Hoogduin, Marga Peeters, Maarten van Rooij, Ad Stokman and participants at the internal lunch seminar for useful comments and fruitful discussions. Moreover, I would like to thank Bertrand Melenberg, Franz Palm, three anonymous referees, participants at the ESEM 2003 conference the January 2004 SOM workshop on Business Cycle Analysis at the university of Groningen, especially Robert Inklaar, and seminar participants at Sveriges Riksbanken.

The views expressed are those of the individual author and do not necessarily reflect official positions of De Nederlandsche Bank.

Résumé

Les cycles de déviation dans le secteur de l'industrie manufacturière de neuf pays membres de l' OCDE sont identifiés grâce à l'application du filtre passe-bande Christiano-Fitzgerald. Les points de retournement, les périodes de forte et de faible croissance économique ainsi que d'autres statistiques descriptives sont un dérivé de ces cycles de déviation. Un test statistique fondé sur une régression est appliqué pour tester la dépendance à la durée. De plus, le lien international entre les mouvements cycliques de deux pays au sein du secteur de l'industrie manufacturière est examiné en mesurant le degré de synchronisation. Outre le fait de mesurer la fluctuation cyclique, un indicateur composite avancé, a été construit afin de suivre et de prévoir le cycle de déviation au sein de l'industrie manufacturière. Cet indicateur composite avancé est un indice unique composé de variables économiques, financières et de prévision possédant des caractéristiques d'anticipation.

1 Introduction

Under the second pillar of the European Central Bank's (ECB) monetary policy strategy, information on the cyclical position of the economy is an important element in the assessment of the outlook for future price developments. The cyclical instability of aggregate economic behaviour will sooner or later be reflected in similar patterns across various macroeconomic time series. The cyclical motions of the variables are mostly simultaneous or in a rapid succession of one another. This phenomenon of leading and lagging cyclical behaviour over time across macroeconomic variables can be exploited for the measurement and forecasting of the conjunctural position of the economy. The study of business cycles, which is of cyclical fluctuations in economic activity, goes back to the seminal contribution of Burns and Mitchell (1946), one of the earliest writings on the business cycle. As a starting point, it seems worthwhile to recall their definition:

"Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own."

This paper builds on the tradition of business cycle indicator research at De Nederlandsche Bank, see for instance Fase and Bikker (1985), Bikker and De Haan (1988), Berk and Bikker (1995) and De Haan and Vijselaar (1998). The first aim of this study is to identify cycles in the manufacturing industry for 9 OECD-countries: the Netherlands, Belgium, Germany, France, Italy, Spain, the United Kingdom, the United States and Japan. We will construct indicators for the nine OECD-countries with special attention to the Netherlands. Moreover, we will provide for each country descriptive statistics like the dating of cyclical turning points, moment properties like the amplitude describing the severity of the recessions and the strength of the booms. Moreover, we will test for duration dependence of low- and high-growth periods. These statistics are derived for each country separately. In addition, the international linkages of the cyclical motions in the manufacturing industries among countries can be measured by the fraction of time the cycles are simultaneously in an upturn and/or in a downturn. The second aim of this study is to forecast the cyclical motions in the manufacturing industries by exploiting the early signals provided by leading indicator variables. The cyclical movement of a broad set of macroeconomic variables will be evaluated on statistical grounds and accordingly selected to form the basis for the composite leading indicator. The selection criteria are a sufficient degree of similarity both in terms of correlation and in terms of turning points and, secondly, leading properties with respect to the cycle in the manufacturing industry.

2 Measuring Business Cycles

Business cycles can be described in general terms as the oscillating behaviour of economic activity. The business cycle consists of a peak in economic activity, a period of recession, followed by a trough and a period of expansion. Subsequent business cycles are therefore separated by the business cycle turning points. A recession of the classical cycle is by definition a period with a decline in the level of economic activity, see Harding and Pagan (2001, 2002, 2005). In effect, negative growth is a sufficient condition for a classical recession. A recession of the deviation cycle, or a low-growth phase, means that the level of economic activity is below its trend, or potential, level. Given that the potential trend growth is positive, classical cycle recessions are always a subset of deviation cycle recessions¹ and there can be multiple classical cycle recessionary episodes within one single period of deviation cycle recession. The durations of low- and high-growth periods of deviation cycles are symmetric: on average is the economy half of the time in recession. Contrarily, the duration of classical cycle recessions are typically much shorter than expansions. Since periods of declining economic activity were relatively scarce in Western Europe since the 2nd World War, deviation cycles are more in line with the fluctuations in the levels of employment and unemployment. Moreover, deviation cycles, or output gaps, gained much policy relevance for Taylor-rule driven monetary policy and for cyclically adjusted government budget deficits.

As the trend cannot be observed directly, the strategic choice concerning the deviation cycle is the methodology to filter the trend, and implicitly the cycle, out of a time series of observations. Canova (1998) provides an overview of the comparative properties of cyclical components obtained by using different trend estimators on a couple of commonly used macroeconomic time series. He concludes that the existence of a set of business cycle stylized facts that is robust to varying trend estimators is misleading, because different concepts of a business cycle, and correspondingly different trend estimators, generate different economic objects which not necessarily have similar properties. Basically two concepts of a business cycle can be distinguished, which originate from Koopmans' (1947) 'measurement without theory' debate on whether the economic considerations underlying a statistical based decomposition should be modelled explicitly? For the econometric model based approach, the optimal filter is the signal extracting device, which minimizes the mean squared error to the actual data generating process. Examples of filters based on econometric models are the exponential smoothing filter, the HP-filter and more recently, the Butterworth-filter known from electrical engineering, see Pollock (2000) and Gomez (2001).

¹ If a variable x_t , say GDP, admits the log-additive decomposition: $x_t = u_t + \psi_t$, where u_t denotes the trend and ψ_t denotes the deviation cycle, then growth, Δx_t , can be decomposed into: $\Delta x_t = \Delta u_t + \Delta \psi_t$, that is potential growth Δu_t and cyclical change $\Delta \psi_t$. Deviation cycle recessions, $\Delta \psi_t < 0$, correspond to periods of growth below potential growth: $\Delta x_t < \Delta u_t$, which then encompasses the classical recessions, that is $\Delta x_t < 0$ (given that potential growth is at least positive, that is $\Delta u_t > 0$).

The statistical based approach assumes the existence of an ideal band-pass filter, often characterized by isolating a pre-specified frequency interval, which is then used as a benchmark for finite sample approximations. The choice of the ideal band-pass filter as a reference model is justified by the definition of Burns and Mitchell (1946) of business cycles as recurrent waves lasting longer than a minimum and shorter than a maximum duration. This wave length interval corresponds one-to-one with the business cycle frequencies. Examples of filters based on the ideal band-pass filter are Baxter and King (1999) (BK-filter), Christiano and Fitzgerald (2003) (CF-filter) and Pedersen (2001).

These papers focus on the optimal approximation to the ideal filter of a finite sample of data that is possibly non-stationary and auto-correlated. Harvey and Trimbur (2003) combine both approaches within a unified framework. They show that the least squares linear estimator of an Unobserved Components model (UC-model, cf Harvey and Jaeger, (1993)) corresponds to a generalized Butterworth filter, which is a specific type of band-pass filter. Conventional econometric specification of the UC-model implies a sub-optimal parameterization of the Butterworth-filter, as the ideal band-pass filter emerges only as the limiting case. Equivalently, the implied cyclical model corresponding to the ideal band-pass filter has no direct economic interpretation. A more general disadvantage of UC-model based filtering is that adding additional new data may result in a higher likelihood for an alternative model specification. Since the decomposition in trend and cycle is model specific, the alternative model can then induce a radical change in the historical shape of the business cycle. This unreliability of the output gap estimates is undesirable from a practical policy making perspective. Orphanides and Van Norden (2002) conclude that the reliability, that is the extent to which output gap estimates are revised over time as more information arrives, is quite low for a broad class of detrending methods. As more data are released however, the estimate of the non-model based band-pass filtered cycle will eventually converge to its final value as the approximation converges to the ideal band pass, regardless of the specific choice of the approximation filter. Therefore, this study defines the cycle to be the optimal band-pass filtered component of the observed variable and remains agnostic about the structural model.

The BK-filter is an approximation to the ideal band-pass filter under side constraints and renders a series stationary without inducing a phase shift. So, it does not shift the filtered cycles forward or backward in time. The CF-filter is a weighted approximation, which takes into account the empirical property of macroeconomic data known as Granger's (1966) typical spectral shape. The approximation of the ideal band-pass filter at the lower frequencies is consequently more important than at the higher frequencies. Schleicher (2004) confirms theoretically and empirically the optimality of the CF-filter, especially concerning the end-of-sample performance for current analysis.² For these reasons, this study will use the CF-filter to extract cycles from time series.

² Even though the CF-filter might show the best end-of-sample performance, this filter exhibits considerable more instability if it is applied to not-seasonally adjusted data. The (seasonal) instability is corrected for by averaging over (maximum) half a cycle (which is 60 months). More precisely: the definitive deviation cycle indicator $DCI_{t-i|t}^{def}$ for $i=1, \dots, 60$ for the period $t-i$ given data up until period t , equals to: $DCI_{t-i|t}^{def} = \sum_{j=1}^i DCI_{t-i|t-i+j}$, where DCI_{t-i+j} is the CF-filtered cyclical component at period $t-i+j$.

A starting point in defining the range of business cycle periodicities is given in the introduction's quote of Burns and Mitchell (1946) saying between one year and ten or twelve years. Stock and Watson (1999) and Baxter and King (1999) apply the BK-filter using the convention of business cycles as variations in GDP between 6 and 32 quarters. This is based on the NBER business cycle reference dates, for which the shortest cycle since 1858 lasted 6 quarters and the longest ones are mostly shorter than 8 years. However, the latest two US-cycles lasted more than eight years. Agresti and Mojon (2001) use for the euro area business cycle the BK-filter with a periodicity band of 1.5 years and 10 years. They arrive at this higher upper bound by stating that the euro area experienced only three recessions over the last thirty years. This study adopts the 1.5-10 years convention of business cycle frequencies³, see also Kranendonk *et al.* (2004) concerning the Netherlands.

3 Detecting and Dating Dutch Deviation Cycles

GDP is the aggregate statistic of all economic activity and provides the broadest coverage of the economy. As the accounting identity holds, GDP is the aggregate of both production and expenditures and of income. Deviation cycles as recurrent periods of above- and below-trend growth are visible in many economic activities and should therefore be visible in GDP as the aggregate of all activity as well. However, not all economic activity is affected simultaneously by the deviation cycle. Some macroeconomic variables lead the cycle, a feature that will be exploited in the next section, and other variables are coincident or lag the cycle. The non-synchronized cyclical behaviour of some sub-components partly cancels in the aggregate GDP. For instance, the expenditure component government consumption is well-known to act as an automatic stabilizer of the general cycle due to counter-cyclical and lagging patterns of tax revenues and social security payments. The expenditure component exports tends to pro-cyclically lead the general cycle, particularly so for a small open economy as the Netherlands.

Since the cyclical motion of GDP does not parallel the ones of its underlying components, the business cycle dating committees of the European CEPR⁴ and the American NBER⁵ look at more macroeconomic variables for dating the turning points of the classical cycle. Apart from GDP, both committees analyse industrial production and total employment. The CEPR committee considers moreover investment and consumption while the NBER committee considers instead real disposable income and the wholesale and retail sales, see Hall *et al.* (2003). Business cycle committees officially establish a turning point of the classical cycle only after a considerable time period has elapsed since its occurrence. Since the cyclical indicator is used for current analysis policy purposes, it is of practical

³ Noteworthy, Ravn and Uhlig (2002) show that the HP-filter is equivalent to a one sided band-pass filter. The standard parameter value of $\lambda=1600$ for quarterly data filters out cycles of duration longer than 8 years.

⁴ <http://www.cepr.org/data/Dating>

⁵ <http://www.nber.org/cycles.html>

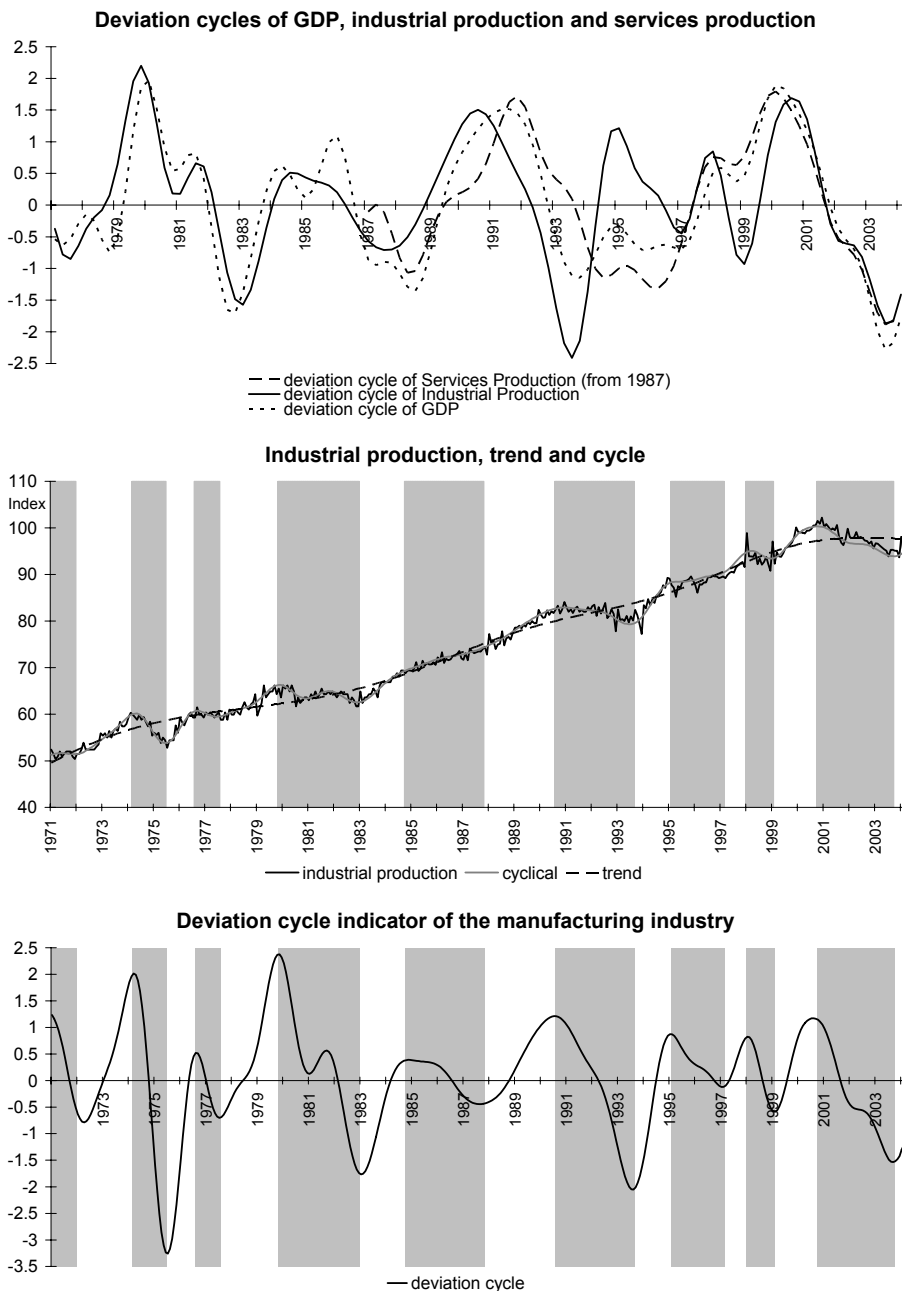
importance to base the indicator on monthly data that are released with a short publication delay and are only to a minor extent subject to revisions.

Stock and Watson (1989) provide a formal statistical framework to construct a coincident indicator for the United States based on the four macroeconomic variables analysed by the NBER dating committee. The indicator closely resembles The Conference Board's (2000) (TCB) one, which is a contemporaneous weighted average of the same four variables. TCB does not straightforwardly construct similar indicators for other countries due to data availability and issues of data definitions. The international indicators of TCB rely on proxy data such as surveys for sales and unemployment rates for payroll employment. Moreover, personal income data is notoriously tedious to obtain at a high sampling frequency without a substantial publication delay. Only the industrial production series is an internationally widely available and comparable statistic.

In accordance with practices at the OECD (2002) and at De Nederlandsche Bank, see Berk and Bikker (1995), this study will use the manufacturing output volume as the reference series from which the cyclical indicator is filtered. Like GDP, industrial production is a main focus variable of both the CEPR and NBER business cycle dating committees. However, industrial production is available on a higher, that is monthly, frequency, the publication lag is considerably shorter and the revision of the numbers after the first publication is smaller than for GDP. The standardized CF-filtered deviation cycles of GDP, services production and industrial production on a quarterly basis for the Netherlands are displayed in Figure 1. The industrial production is, with a cross-correlation with cyclical GDP of 0.77, only approximately cyclically representative. Especially the mid-nineties show discrepancy between the cycles of services production and industrial production. These two types of production constitute 70% respectively 25% of GDP and agriculture producing the remaining 5% for the highly developed countries considered in this paper.

The periods between a peak and a trough of the deviation cycle of manufacturing industry are the low-growth phases and are indicated in the graphs of Figure 1 by the shaded areas. The collection of peaks and troughs, the turning points, are derived from the deviation cycle according to the procedure outlined below.

Equivalent to the standard calculus rule for finding extreme values, i.e. setting the first derivative of a function with respect to a variable equal to zero, a local peak in (the log of) a macroeconomic time series variable, x_t , occurs at time t if x_t exceeds values x_s for $t-k < s < t$ and $t+k > s > t$, where k delineates some symmetric window in time around t . This simple idea is the basis of the NBER procedures summarized in the Bry and Boschan (1971) dating algorithm. In practice, the Bry-Boschan algorithm with $k=5$ for monthly data also applies some censoring rules ensuring a minimum cycle duration of 15 months and alternation of peaks and troughs. This algorithm has been applied on time series data in levels and so produces the dating of classical cycles. Now consider the deviation cycle variable $DCI_t = y_t - g_t$. The variable DCI_t is now the deviation of the log of a macroeconomic time series variable, x_t , from its long run CF-trend, g_t , and a similar turning point algorithm can be applied on this deviation series.

Figure 1 **Deviation cycles of the Netherlands**

The Bry-Boschan algorithm has been adapted by Harding and Pagan (2002) to date growth cycles. We perform our dating in this spirit and apply an algorithm that obeys the following rules:

- peaks and troughs must alternate,
- a peak (trough) must represent above (below) trend growth and so the corresponding $DCI_t > (<)0$,
- the turning point must be a local optimum:

$$\begin{aligned} &\text{a peak at } t \text{ if } (DCI_{t-p}, \dots, DCI_{t-1} < DCI_t > DCI_{t+1}, \dots, DCI_{t+q}), DCI_t > 0 \text{ for } t' \in [-p, \dots, q], \\ &\text{a trough at } t \text{ if } (DCI_{t-p}, \dots, DCI_{t-1} > DCI_t < DCI_{t+1}, \dots, DCI_{t+q}), DCI_t < 0 \text{ for } t' \in [-p, \dots, q]. \end{aligned} \quad (1)$$

Since the CF-filter parameters are set such that the shortest possible cycle lasts at least 18 months, no two peaks or troughs can coexist within this time frame. Moreover, a deviation cycle is approximately symmetric; the cycle is roughly half of the time below trend level. The study therefore sets $p = q = 9$, which defines a symmetric window consistent with the a priori definition of the minimum duration of the deviation cycle of 18 months. The dating procedure provides the collection of dates at which the deviation cycle reaches a peak or a trough. A low-growth (high-growth) period is straightforwardly defined as the period between a peak (trough) and a trough (peak).

Applying the dating rules (1) to a couple of different deviation cycles for the Netherlands results in the collection of deviation cycle turning points which are represented in Table 1. It shows both the results of filtered GDP on quarterly basis and of filtered industrial production on monthly basis. Moreover, the filters applied are the CF-filter used in this study, the HP-filter used by De Haan and Vjselaar (1998) and the Phase Average Trend (PAT) filter used by the OECD (2002). The last column of the table shows the chronology of turning points by the OECD of the Dutch deviation cycle. This chronology results from the comparison of turning points of PAT-filtered GDP and industrial production combined with expert judgement by country specialists. As shown in the last row, the turning points of both the CF- and the HP-filtered industrial production series produces deviation cycle turning points, as defined by (1), which closely match the OECD's final chronology. The mean absolute difference of the timing of the turning point dates as compared to this chronology is close to 3 months for all filtered series considered.

Table 1 **Dutch deviation cycle turning points of industrial production and GDP**

	Industrial Production index (manufacturing)			GDP		Dating ¹⁾²⁾
	Hodrick- Prescott ³⁾	Christiano- Fitzgerald ⁴⁾	Phase Average Trend ¹⁾	Christiano- Fitzgerald ⁴⁾	Phase Average Trend ¹⁾	
Through	67m07	67m06	67m05		67q2	67m08
Peak	70m09	69m12	70m03		70q4	70m06
Through	72m06	72m01	71m12		72q3	72m06
Peak	74m04	74m03	74m08		74q3	74m01
Through	75m07	75m07	75m08		75q1	75m07
Peak	76m11	76m08	76m09			[76m09]
Through	78m01	77m08	78m05	79q1		[78m05]
Peak	79m11	79m11	79m11	80q2	79q4	79m10
Through	83m05	83m01	82m11	83q1	82q4	82m12
Peak	85m12	84 m10	85m01	86q2	86q2	84m10
Through			86m05			[86m05]
Peak			87m01			[87m01]
Through	87m07	87m11	88m04	88q4	88q1	87m08
Peak	90m05	90m08	91m02	91q4	90q4	90m12
Through					91q3	
Peak					92q1	
Through	93m08	93m09	93m06		93q4	93m12
Peak	95m06	95m02	95m06		95q4	95m02
Through	96m08	97m03	98m08	97q2	96q4	96m12
Peak	98m03	98m01			98q1	[98m01]
Through	99m01	99m02			99q1	[98m12]
Peak	00m10	00m10	00m06	00q2	99q4	00m06
Through	04m01	03m10		03q4		
MAD ⁵⁾	3.3	2.6	2.7			

1) OECD (2002)

2) [] = minor cycle

3) De Haan and Vijsselaar (1998)

4) with lower bound and upper bound of 1.5 and 10 years respectively

5) *MAD* = mean absolute difference in months of the timing of the turning point dates as compared to the OECD's chronology that is shown in the last column.

4 The Empirics of International Deviation Cycles in Manufacturing

The indicators of the deviation cycle in the manufacturing industry of the nine countries are calculated by applying the CF-filter and standardising the resulting cyclical series. Subsequently, the dating is performed by applying the dating rules (1) to the indicators. The successive above- and below-trend growth periods are shown together with the deviation cycle indicators (*DCI*) in the manufacturing industry in Figure A1 in the appendix. The deviation cycles are represented by the bold lines and the low-growth phases by the shaded areas. The construction of the composite leading indicators, which are shown in the graphs by the dashed lines, will be explained in the next section. In order to derive descriptive statistics and measure synchronization among the deviation cycles in the manufacturing industry, we construct a state variable S_t , which equals one in high-growth periods and zero otherwise. The information $\{DCI_t, S_t\}$ provides for each country a picture of the characteristics of the cyclical motion and can be used to produce descriptive statistical measures as duration, amplitude, steepness and sharpness of high- and low-growth phases, see for instance Harding and Pagan (2002). Moreover, the information $\{DCI_t, S_t\}$ can be used to assess synchronization among deviation cycles in the manufacturing industry of different countries, see Harding and Pagan (2005).

The most commonly used descriptive statistics are the standard deviation, skewness and kurtosis. The standard deviation measures the variability of the indicator. Skewness measures the one-sided fattedness of the probability distribution of the indicator. So, it measures whether high-growth phases are relatively more exuberant than low-growth phases are severe. Kurtosis measures the two-sided fattedness and indicates whether large cyclical movements are relatively more common.

The state variable S_t separates above- and below-trend growth phases. A peak (trough) occurs at time t when $S_t = 1$ ($S_t = 0$) and $S_{t+1} = 0$ ($S_{t+1} = 1$). A high- (low-)growth phase consists of the period beginning the month after the date of a trough (peak) and ending at the date of a peak (trough). High growth phases and TP (trough-to-peak) and low-growth phases and PT (peak-to-trough) will be used interchangeably. The average duration⁶ of high- and low-growth phases⁷ are:

$$DUR_{TP} = \frac{\sum_{t=1}^T S_t}{\sum_{t=1}^{T-1} (1 - S_{t+1}) S_t}. \quad (2)$$

⁶ In calculating average of high- and low-growth phases one has to deal with uncompleted phases at both ends of the sample. The statistics will only be based on completed phases and therefore the summation runs from the beginning of the first completed phase until the end of the last one rather than over $1, \dots, T$.

⁷ The corresponding formulas for low-growth phases, PT, are obtained by replacing S_t by $(1 - S_t)$.

The numerator in (2) measures the total time spent in low-growth phases and the denominator counts the number of peaks. The average amplitude straightforwardly equals:

$$AMP_{TP} = \frac{\sum_{t=1}^T S_t \Delta DCI_t}{\sum_{t=1}^{T-1} (1 - S_{t+1}) S_t} . \quad (3)$$

Thinking of a phase as a triangle with as base its duration and as height its amplitude, then the steepness of a phase is naturally measured by the slope of the triangle:

$$STEEP_{TP} = AMP_{TP} / DUR_{TP} . \quad (4)$$

In addition to looking at the steepness over the phases, one might want to compare the shape of the cycle in the early part of the cycle to that in the later part. A simple way of capturing this shape is to divide a phase into two parts and to compare the average growth rates in the first part to those in the second part. This statistic is called sharpness and essentially measures a degree of convexity of the indicator, which is the relative steepness of the early stage in the cycle. Let i be the i^{th} phase and d_i be the duration of this phase, starting at time t_i . The measure of the shape of a single phase can then be defined as:

$$v_i = \frac{1}{\lfloor d_i/2 \rfloor} \left[\sum_{k=1}^{\lfloor d_i/2 \rfloor} (DCI_{k+t_i} - DCI_{k+t_i-1}) - \sum_{k=\lfloor d_i/2 \rfloor+1}^{d_i} (DCI_{k+t_i} - DCI_{k+t_i-1}) \right] . \quad (5)$$

Thus, if in a TP-phase the growth is faster in the first half of its duration than in the second half, then this measure will be positive. If in a PT-phase the decline is slower in the first half of its duration than in the second half, then this measure will be positive. If a cyclical indicator shows these two positive measurements it implies that the peaks are relatively long lasting and rounded and the troughs are relatively short lasting and pointed.

Averaging (5) over all similar phases can be defined as:

$$SHARP_{TP} = \sum_i S_{t_i} \left(\frac{d_i}{\sum_t S_t} \right) v_i . \quad (6)$$

In equations (6), S_{t_i} serves as an indicator function which equals one if the i^{th} cyclical phase starting at time t_i is a high-growth phase. Moreover, the equation is a weighted moving average of (5) over respectively high- and low-growth phases. The weight for a particular high-growth phase is the duration of that phase divided by the total time spent in

TP-phases over the sample. So, this weighting scheme gives more weight to longer phases and the sum of the weights is equal to one.

The descriptive statistics for the nine deviation cycle indicators of the manufacturing industry are presented in Table 2. Almost all countries tend to display kurtosis indicating that large cyclical movements in the manufacturing industry are relatively common. The Netherlands and to a lesser extent Belgium stand out on stability, but not on moderateness. Stability is revealed by the volatility of the cyclical behaviour around the trend and by the amplitude. This is the absolute deepness of a cycle decomposed into the ones belonging to low- and high-growth phases. Both the Netherlands and Belgium show the lowest numbers for all nine countries on the standard deviation and the amplitude. Moderateness is revealed by the steepness statistic. Steepness can be interpreted as the motion deepness of a low- or high-growth phase like amplitude can be seen as the absolute deepness of a low- or high-growth phase. For both phase types, the Netherlands and to a lesser extent Belgium show among the highest numbers for the steepness statistic indicating that the deviation cycle of the manufacturing industries for both countries move quickly from one turning point to another. Japan's manufacturing sector acts as the mirror image of the ones of these two countries by showing the highest cyclical volatility and the highest cyclical severity both in terms of amplitude and steepness for both PT- and to a lesser extent TP-phases. The deviation cycle of the manufacturing industry of the U.S. also display high cyclical volatility and severity. The negative skewness indicates contraction phase outliers indicating at relatively severe recessions. The large slowdown sharpness and the large acceleration sharpness for the U.S. reveals a pattern of smooth rounded peaks and pointed deep troughs. Italy's manufacturing industry demonstrates the largest positive skewness and among the highest peaks, that is the average amplitude of the high-growth phase. Both statistics indicate large upward cyclical outliers. The quite moderate sharpness statistic reveals that the Italian cyclical pattern surrounding turning points is particularly smooth. Germany's manufacturing industry shows a cycle with a relatively long TP-phase compared as to the PT-phase. On the other hand the motion of the low-growth phases is relatively fast according to the steepness statistic and relatively deep according to the negative skewness. The sharpness statistic reveals that the German turning points are the most sharply pointed ones of the nine countries. Therefore, the German indicator looks somewhat like a saw tooth.

The descriptive statistics show the idiosyncrasy of the deviation cycle of each individual country's manufacturing industry as compared to a sinusoid cyclical motion. This stylized cyclical pattern implies a deterministic periodicity. The periodicity of the cycle is the amount of time between two peaks, or two troughs. A complete cycle covers therefore two consecutive phases, which is either a PT-TP or a TP-PT period. The theoretical literature considers the cycle as the total sum of sub-cycles identified by theoretically founded durations. Impulses from inventory holdings, (infrastructural) investments and technological breakthrough innovations are the driving forces behind respectively the Kitchin, Juglar, Kuznets and the Kondratieff cycles.

Table 2 **Summary statistics for the deviation cycle indicators of the manufacturing industry**

		NLD	BEL	FRA	DEU	ITA	ESP	JPN	GBR	USA
Standard deviation		1.67	2.16	2.50	2.64	2.67	2.33	3.38	2.47	2.82
Skewness		-0.38	-0.0091	-0.56	-0.25	0.21	-0.68	0.16	-0.13	-0.35
Kurtosis		3.51	4.08	4.78	2.68	3.25	3.69	2.67	3.14	3.55
Duration	TP	26.00	22.00	24.83	24.90	33.40	38.83	26.00	36.75	27.00
	PT	22.00	28.13	22.33	28.00	33.50	30.67	39.00	33.40	36.50
Amplitude	TP ¹⁾	2.87	2.67	2.44	2.93	3.40	3.02	3.39	3.00	2.74
	PT	2.42	2.38	2.59	2.31	2.83	2.97	3.12	2.39	2.51
Steepness	TP ¹⁾	0.11	0.12	0.098	0.12	0.10	0.078	0.13	0.098	0.10
	PT	0.11	0.085	0.12	0.083	0.085	0.097	0.080	0.072	0.069
Sharpness	TP	0.015	-0.017	-0.038 ²⁾	-0.023	-0.66 ²⁾	0.053	-0.14 ²⁾	0.14 ²⁾	0.083
	PT	0.015	-0.012	-0.042 ²⁾	-0.017	-0.55 ²⁾	0.067	-0.097 ²⁾	0.10 ²⁾	0.061

Notes: The statistics are calculated over the sample 1965:1-2001:9 and represent the deviations from trend series which are not standardised. The mean of the deviation cycle is by construction approximately zero. So, standardising approximately only implies dividing the deviation time series by its standard deviation. PT means the phase Peak-to-Trough and is therefore synonym to low-growth phase. Equivalently, TP is synonym to high-growth phase. The skewness measure is $u_3/(u_2)^{1.5}$ and the kurtosis measure is $\mu_4/(\mu_2)^2$, where μ_r is the r^{th} central moment. The skewness of a symmetrical distribution, such as the normal distribution, is zero. Similarly, the kurtosis of the normal distribution is 3.

1) The numbers are presented with opposite signs for the sake of comparison with periods of high growth.

2) These numbers are multiplied by 100.

The elapsed time between two consecutive peaks would then be a fixed period T , for which the duration depends on the type of the subcycle. So, if DCI_t is identified as a peak, then $DCI_{t+\tau}$ must also be a peak with fixed duration time of $\tau = T$ periods, whereas in case of weak periodicity τ is considered as a stochastic duration variable with a mean of T . For $\tau > T$, the probability that the cycle switches to the other phase increases then with the ongoing duration of the current phase. Diebold and Rudebusch (1990) refer to the feature that the probability of a turning point is some function of the age of the cycle as duration dependence. Their evidence for the existence of some duration dependence is corroborated by Ohn *et al.* (2004) using novel discrete-time test statistics. A minimum duration on the complete deviation cycle in the manufacturing industry of 18 months is imposed in this study by the parameter settings of the CF-filter. The appendix shows that Ohn's *et al.* (2004) discrete test statistic is still valid to test duration dependence if the cycle is subject to minimum duration requirements. The hypothesis of no duration dependence is for both low- and high-growth phases for almost all countries rejected, according to the reported statistics in Table 3.

Table 3 Statistics of duration dependence of the deviation cycle in the manufacturing industry

	NLD	BEL	FRA	DEU	ITA	ESP	JPN	GBR	USA
TP	-4.35	-7.98	-1.93	-5.13	-1.25*	-3.74	-4.53	-4.48	-7.58
PT	-6.04	-2.75	-1.86	-3.82	-2.71	-2.71	-5.91	-2.8	-4.62

Note: The sample period is 1965:1-2004:2. The numbers are the Newey-West t-statistics for the null hypothesis of no duration dependence. PT means the phase Peak-to-Trough and is therefore synonym to low-growth phase. Equivalently, TP is synonym to high-growth phase.

The next step after examining the characteristics of the deviation cycle in the manufacturing industry is to analyze the relationships between them. More precisely, we want to examine how closely the cycles of individual countries are synchronized to one another. Harding and Pagan (2001) propose the index of concordance, which measures the fraction of time that two cycles spend in the same phase. Let S^i and S^j denote the state variables of the deviation cycle in the manufacturing industry of countries i and j . The index of concordance is defined as:

$$IC_{ij} = \frac{1}{T} \sum_{t=1}^T \{S_t^i S_t^j + (1 - S_t^i)(1 - S_t^j)\}. \quad (7)$$

For all nine countries the results for the index of concordance are presented in the upper triangle of Table 4. The italic numbers in the lower triangle are the cross correlation coefficients of the two concerning deviation cycles in the manufacturing industry.

Six of the nine countries examined in this study are part of the euro area and account for more than 90% of euro area GDP. These six countries can form 15 unique couples⁸ which provides 15 indices of concordance (ICs) measuring euro area synchronization of deviation cycles in the manufacturing industry. The average of these 15 indices is 0.74 and the average correlation is 0.66. As can be calculated from Table 3, the manufacturing industries of two euro area countries are on average roughly three quarters of the time in the same conjunctural state. Both the correlation and the IC of the big euro area countries Germany and France are both 0.68 and those between Germany and Italy only 0.32 respectively 0.65. The average IC and correlation of the potential euro area member the United Kingdom and the five euro area countries equal 0.64 respectively 0.56, which is even worse than the average IC of 0.67 between the United States and the euro area countries and slightly better than the average correlation of 0.54 between these two economic blocks. Moreover, the IC and the correlation between the United Kingdom and the United States equal 0.75 and 0.71 respectively. This indicates that the United Kingdom's deviation cycle in the manufacturing industry is more synchronized with the United States' one than with the cycle of the euro area.

⁸ One could also construct an index of concordance which only measures the simultaneity of more than two countries together. It then measures the fraction of time that all countries are in the same state. For example, the concordance of Germany, France and Italy together is 0.57

Table 4 **Index of concordance and correlation between indicators of deviation cycles in the manufacturing industry**

	NLD	BEL	FRA	DEU	ITA	ESP	JPN	GBR	USA
NLD		0.81	0.68	0.81	0.70	0.67	0.82	0.63	0.63
BEL	<i>0.82</i>		0.77	0.68	0.71	0.8	0.71	0.68	0.71
FRA	<i>0.76</i>	<i>0.86</i>		0.68	0.81	0.86	0.61	0.59	0.65
DEU	<i>0.80</i>	<i>0.66</i>	<i>0.67</i>		0.65	0.69	0.76	0.66	0.67
ITA	<i>0.64</i>	<i>0.64</i>	<i>0.66</i>	<i>0.31</i>		0.76	0.68	0.62	0.63
ESP	<i>0.58</i>	<i>0.67</i>	<i>0.81</i>	<i>0.47</i>	<i>0.61</i>		0.65	0.65	0.74
JPN	<i>0.74</i>	<i>0.76</i>	<i>0.65</i>	<i>0.52</i>	<i>0.65</i>	<i>0.52</i>		0.63	0.65
GBR	<i>0.66</i>	<i>0.72</i>	<i>0.58</i>	<i>0.37</i>	<i>0.47</i>	<i>0.53</i>	<i>0.60</i>		0.75
USA	<i>0.55</i>	<i>0.6</i>	<i>0.59</i>	<i>0.40</i>	<i>0.51</i>	<i>0.57</i>	<i>0.58</i>	<i>0.71</i>	

Notes: The sample period is 1965:1-2001:9. The indices of concordance are represented in the upper triangle of the table. The correlation coefficients are represented in *italic* in the lower triangle of the table.

5 Leading Indicators

The deviation cycles provide a signal for the current conjunctural state in the manufacturing industry. Cyclical motion in the economy is reflected in various macroeconomic time series. Depending on the timing behaviour of their cyclicity vis-à-vis the deviation cycle in the manufacturing industry, variables can be classified as leading, coincident or lagging. Leading variables give by nature an early signal on the conjunctural position. We will exploit this feature by selecting a set of leading variables and transform them into a single composite leading indicator that replicates and predicts the deviation cycle in the manufacturing industry. The combination of leading indicators is useful to pick up signals from different sectors of the economy. Stock and Watson (2003) conclude in their empirical literature review on the usefulness of financial indicators to forecast GDP growth and inflation that some asset prices predict inflation or output growth in some countries during some periods, but which series predicts what, when, and where is difficult to predict. As the forecasting literature moreover suggests that the simplicity of a model often implies forecasting power, this study aims to predict the deviation cycle in the manufacturing industry in a non-model based framework. Marcellino (2006) notes the overall good forecasting performance of the simple non-model based leading indicators concerning the latest two recessions in the U.S. in an overview of the different approaches and methods for the construction, use and evaluation of leading indicators, both in the academic literature and in the forecasting practice.

Potential leading variables that can be expected to lead and predict the general business climate are gathered and will be called basic indicators. The set of potential basic indicators is then screened in the spirit of the formalized scoring system of Moore and Shiskin

(1967). The collection of basic leading indicators should be a balanced representation of the total economy. Each basic indicator should be economically plausible, so it possibly causes the deviation cycle in the manufacturing industry or it reacts quicker to shocks. Moreover, the series should be statistically reliable, that is a long history of observations with as few interruptions and definition alterations as possible. The data should be available with a minimum publication delay and not be subject to substantial revisions after the first publication. The basic indicator should conform to the deviation cycle; it has good forecasting properties, not only at peaks and troughs. Finally, the basic indicator should possess a consistent timing as a leading indicator; systematically anticipate peaks and troughs with a rather constant lead time. The cyclical parts of all basic indicators are compared to the deviation cycle indicator of the manufacturing industry in order to determine the numbers of months by which the basic indicators lead the cycle, i.e. the timing. The measure of this timing is the lead in number of months corresponding to the maximum cross correlation coefficient between the deviation cycle indicator and the lagged basic leading indicator (BI):

$$\rho(l) = \frac{1}{T} \sum_t (DCI_t - \overline{DCI})(BI_{t+l} - \overline{BI_t}) / \left(\sum_t (DCI_t - \overline{DCI})^2 \sum_t (BI_{t+l} - \overline{BI_t})^2 \right)^{0.5}, \quad (8)$$

$$l^* = \arg \max_l \rho(l). \quad (9)$$

In equation (8) $\overline{DCI_t}$ and $\overline{BI_t}$ are the averages of DCI_t and BI_{t+l} respectively, over t . The magnitude of the correlation coefficient (8) measures the degree of cyclical similarity of the deviation cycle indicator of the manufacturing industry and the basic indicator. Moreover, the corresponding lead of l^* is determined in (9) and represents the optimal lead of the basic indicator. Analogous to Berk and Bikker (1995), we screen a set of potential leading variables consisting of about 40 monthly time series for each country on the criteria of a cross correlation $\rho(l^*)$ of at least 0.5 and a corresponding maximum lead l^* of at least 5 months. These boundaries are chosen as high as possible under the restriction that a sufficient number of basic leading indicators are selected. The selection statistics for the selected basic leading indicators are presented in Table 5. The first number in each cell is the optimal fine tuned lead l^* and the second number is the maximum cross correlation $\rho(l^*)$ for the concerning leading indicator and country. The selected leading indicators are of an economic, financial and monetary nature and confidence and expectations surveys.

Of the economic flow variables consumption, investment, exports and imports only the consumption in Japan shows a leading cyclical pattern. Consumption can straightforwardly be seen as a demand pull trigger for industrial activity. Inventories are a result of non-matching economic flows and for most countries the storage of final products provides an early signal for production activity. According to De Haan and Visselaar (1998) there exists a one-way Granger-causality from the storage of final products towards industrial production for the Netherlands. Other types of inventories are those at the start of the production process,

like the level order positions and issued building permits. Both variables pro-cyclically lead the deviation cycle in the manufacturing industry.

Prices and wages are a fundamental part in every economic model describing cyclical fluctuation. Following Stock and Watson (1999), a general pattern emerges of leading, countercyclical price levels and lagging, pro-cyclical rates of inflation. The nominal wage index exhibits a pattern quite similar to consumer prices, probably due to the contractual indexing of wages. Consumer prices, sales prices, total wages and hourly wages emerge in this study as cyclical leaders. Moreover, labour costs per unit product are a combination of wages and productivity and on average leads the cycle countercyclically. The same pattern shows up for the world market prices of commodities, which basically is a cost variable for production.

Dominant classes of leading variables concerning business cycles are those specifically related to the future. Survey variables that pro-cyclically lead the general cycle are consumer and producer confidence, expected future industrial production, prospects total economy, judgment of order arrivals and the IFO-indicator⁹ for Germany. Financial variables deal with consumption and investment possibilities and therefore also with the choices made by private and public agents. As the sum of discounted expected future cash flows, share prices are, like surveys, future linked variables. Interest rates are a cost of capital and therefore partly determine the budget constraint. Both short term and long term interest rates lead the cycle with positive interest rates associated with cyclical declines in output. The lead of the short term interest rate is approximately 1.5 years for most countries to over two years for Japan, see Table 5. The connection between short term interest rate movements and the output gap is well established in the Taylor rule. The spread between the short and long term interest rates has long been recognized as a leading indicator. However, we prohibit taking this variable into the model together with both the short and the long term interest rates.

Monetary aggregates play an important role in the determination of price levels since nominal frictions in the economy can cause movements of real quantities. Following the time inconsistency literature, in the short run a monetary expansion creates a boost in employment and output that causes in the long run only a rise in the price level. We consider three monetary aggregates if available, namely M1, M2 and M3 and test them both in nominal and real terms and in levels and growth rates. However, we find only for Spain a monetary aggregate functioning as a leading indicator. Stock and Watson (2003) note that the contemporaneous cross correlation between the cyclical components of nominal M2 and output has drastically declined since the early eighties for the United States.

The aim is to transform the selected basic leading indicators into a single composite index which replicates and predicts the deviation cycle indicator of the manufacturing industry. The replication property is ensured by the minimum correlation requirement. The prediction property is ensured by the criterion of minimum lead. A single composite index is then constructed as an optimal linear combination of normalized and synchronized selected basic leading indicators. The normalization procedure makes indicators exhibiting weak

⁹ We choose a sub index of the IFO-indicator which refers to future developments.

cyclical variation, for instance wages, comparable to indicators exhibiting strong cyclical variation, for instance confidence indicators and prevents the latter ones from dominating the former ones in a single composite index. The selected basic leading indicators are synchronized with the deviation cycle indicator of the manufacturing industry by shifting them forward with their respective leads as determined by (9) and reported in Table 5. Shifting a basic leading indicator variable forward by a number of months equal to its lead means that the shifted basic leading indicator variable no longer leads, but instead coincides with the cyclical phase of the manufacturing industry. Because of the shifting, the series extends beyond the observation period and provides a predictive signal for the near future cyclical development. Each selected basic leading indicator variable provides its own early signal. A one-dimensional single signal is obtained by constructing an optimal linear combination of the individual normalized and synchronized basic leading indicator variables by means of principal component analysis. The first principal component reflects the composite leading indicator since it captures the maximum common variation across the set of leading indicators.

According to Moore and Shiskin's (1967) formalized criteria, the composite leading indicator is not only supposed to reflect the motion of the deviation cycle indicator of the manufacturing industry, but also to provide an early signal of upcoming turning points. As a linear combination of synchronized basic variables, the composite leading indicator's turning points not necessarily coincide optimally with the corresponding ones of the deviation cycle indicator of the manufacturing industry. The lead of the composite leading indicator will therefore be fine tuned by matching the corresponding turning points of both indicators and minimizing a distance metric. The turning points of both indicators are dated using the procedures as stated in (1). The metric is the sum of the differences measured in months between all corresponding peaks and troughs of both indicators. The optimal leads of the synchronized composite leading indicators and their corresponding cross correlations with the deviation cycle indicators of the manufacturing industries are displayed in the last two rows of Table 5. Moreover, both indicators of each country are graphed in Figure A1 in the appendix. The shaded areas in the graphs correspond to the below-trend growth periods S_t . Table 6 presents the dating of the cyclical turning points of both indicators for all countries. The composite leading indicator sometimes misses turning points or sometimes gives false alarms, which means that the leading indicator signals a turning point that is not present in the deviation cycle indicator of the manufacturing industry. Missed signals and false alarms are mostly connected to minor cycles.

Table 5 Results of selection leading indicators

	NLD	BEL	FRA	DEU	ITA	ESP	JPN	GBR	USA
Consumer confidence			5 0.60	6 0.63			9 0.51	12 0.75	
Consumer price index*	8 0.54	17 0.72			9 0.79		16 0.53	14 0.83	
World market prices commodities*					18 0.61	15 0.55			14 0.84
M1*						22 0.78			
Equity price			14 0.62						9 0.52
Short term interest rate*	19 0.51	18 0.93		12 0.61	15 0.80		27 0.64	18 0.88	18 0.80
Long term interest rate*	18 0.53	16 0.57			12 0.68		25 0.53		15 0.71
IFO	5 0.74	7 0.63		5 0.91	9 0.60	6 0.60			
Expected business activity	6 0.84								
Expected future industrial production								8 0.94	
Storage final products*	6 0.59		5 0.61	5 0.90		6 0.70	13 0.70	10 0.92	9 0.77
Prospects total economy					10 0.72			11 0.95	
Judgement of order arrivals			5 0.70	6 0.85	15 0.73	6 0.70			
Level order position			5 0.82	5 0.91					
Hourly wage industry*	6 0.57	12 0.78		5 0.88	11 0.89		11 0.56		8 0.73
Domestic sales prices*		9 0.56			15 0.73		23 0.68	13 0.82	7 0.90
Producer confidence			5 0.74						
Consumption							10 0.67		
Total wages*								16 0.77	
Labour costs per unit product*		7 0.92	12 0.63						
Composite leading indicator	6 0.81	9 0.66	3 0.81	7 0.78	6 0.84	7 0.77	7 0.68	5 0.87	9 0.83

Table 5 Results of selection leading indicators (continued)

Notes:	The first number in each cell is the optimal fine-tuned lead (9) and the second number is the maximum correlation (8) of the corresponding basic leading indicator variable with the deviation cycle indicator of the manufacturing industry, <i>DCI</i>. The variables marked with a "*" counter-cyclically lead the cycle instead of pro-cyclically. The statistics presented in this table are, except for the Netherlands, based on the cyclical motions of the variables filtered with a lower bound of 36 instead of 18 months and an upper bound of 120 months. The composite leading indicator, <i>CLI</i>, is the weighted average of the synchronized basic leading indicator variables. The lead of the <i>CLI</i> is adjusted such that its turning points match most closely with the corresponding ones of the <i>DCI</i>. The resulting optimal lead and the corresponding cross-correlation of the fine-tuned <i>CLI</i> with the <i>DCI</i> are presented in the last two rows.
---------------	---

The mean and the standard deviation of the differences in months between the timing of the corresponding turning points of the *DCI* and *CLI* are presented in the last rows of Table 6. The absolute value of the mean difference is by construction smaller than a half due to the fine tuning, which optimally matches the corresponding turning points of both indicators. While on average the match is optimal for each country, the standard deviation of the differences reveals the variation in the lead/lag time for the turning point dates. The standard deviation is for most countries of the same order of magnitude as the lead of the *CLI*. The lead time of the *CLI* lays therefore grosso modo just outside the one standard deviation confidence interval of the difference in months between the corresponding turning points of both indicators. Therefore, more or less 85% of all turning points are signalled by the *CLI* during its lead time, so before the *DCI* does. Only Italy is a special case, the lead of the *CLI* is larger than two times the standard deviation. The *CLI* is signalling almost all turning points before the *DCI* does. Moreover, the Italian *CLI* never missed a turning point or gave a false alarm.

6 Conclusion

The cycles in the manufacturing industry of nine OECD-countries are identified as deviations from trend. We use the Christiano-Fitzgerald band-pass filter to estimate the unobservable trend and the cyclical component of a time series. We adopt the convention that the business cycle frequencies consist of all cycles with duration longer than 18 months and shorter than 10 years. From the deviation cycle in the manufacturing industry, we derive the cyclical turning points, low- and high-growth periods and summary statistics describing features like amplitude, steepness and duration of the cycle for each country.

The manufacturing industries in the Netherlands and, to a lesser extent, Belgium stand out compared to the nine countries on stability by showing a relatively modest cyclical spread around the trend. Their cyclical motions are however the least moderate so that the manufacturing industries of both countries move quickly from low- to high-growth phases and vice versa. Japan acts as the mirror image by showing the largest cyclical swings. The U.S. reveals a pattern of smooth rounded peaks and pointed deep troughs. The hypothesis of no

duration dependence is rejected for nearly all deviation cycles in the manufacturing industries for all countries in both high- and low-growth phases. The international linkage between the manufacturing industries is explored by calculating the fraction of time the two countries are both in the same phase. This statistic shows that manufacturing industry in the United Kingdom is more synchronised with the one in the United States than with the one in the euro area. Moreover, the average synchronisation between the United Kingdom and the euro area countries is lower than the average synchronisation of the euro area countries with one another.

In addition to measuring cycles we constructed for each country a single composite leading indicator, which replicates and predicts the deviation cycle in the manufacturing industry. The composite leading indicator is based on economic, financial and survey variables possessing leading properties. These variables are selected from a set of candidate variables for each country. The variables that have been selected for five or more countries are the short term and long term interest rates, the storage of final products, the hourly wages, the domestic sales prices, the IFO-indicator for Germany and the consumption price index. The lead of the composite leading indicator is determined such that the dates of the turning points match most closely with the corresponding turning point dates of the deviation cycle indicator of the manufacturing industry. Moreover, the majority of the turning points are signalled by the composite leading indicator within its lead time.

Table 6 Dating of turning points

	NLD		BEL		FRA		DEU		ITA	
	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>
Trough	67m6		67m5		68m4		67m6		69m11	
Peak	70m1		69m8		69m6		69m12		70m8	
Trough	72m4		71m11		71m9		72m1		71m7	71m10
Peak	74m3	74m4	74m4	73m12	74m2		73m11		74m1	73m12
Trough	75m7	75m12	75m7	76m4	75m7		75m6	75m8	75m8	75m9
Peak	76m8		76m9		76m10	76m10		77m1	76m11	76m10
Trough	77m7		77m8		77m9	77m10		78m2	77m12	77m11
Peak	79m11	79m11	79m11	80m7	79m11	79m12	79m12	80m1	80m3	79m12
Trough			81m2							
Peak			81m12							
Trough	83m1	83m4	82m11			80m4	82m12	83m1	83m3	83m5
Peak	84m11	84m9	84m1			82m6	85m9	84m7		
Trough	87m8	86m12	86m12	86m3	85m4	85m6	87m11	88m3		
Peak	90m7	90m1	89m5	89m6	89m6	89m8	91m9	91m1	89m7	89m6
Trough	93m8	93m11	93m1	92m4	93m7	93m8	93m8	93m10	93m6	93m7
Peak	95m1	95m7	95m7	95m7	95m2	95m3		95m7	95m8	95m3
Trough	97m1	96m9	96m6	96m6	96m9	96m10		96m11	96m11	96m12
Peak	98m1	98m5	98m1	98m4	97m12	98m6	98m4	98m8	97m12	98m6
Trough	99m2	99m9	99m4			99m7	99m5	99m11	99m3	99m7
Peak	00m8	00m11	00m8			00m9	00m10	01m1	00m12	00m8
Trough	03m9	02m5		02m4	03m5	02m3		02m6		02m5
False signals	0		0			4		4		0
Missed turnings	2		8			0		0		0
MEAN	-4/14		-1/9*			1/10		-2/10*		3/14*
STDV	6.4		6.5			5.2		7		3

Table 6 **Dating of turning points (continued)**

	ESP		JPN		GBR		USA	
	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>	<i>DCI</i>	<i>CLI</i>
Trough	68m03				68m07		67m07	
Peak	69m07	70m03	70m03		69m08		69m07	
Trough	71m07	71m11	71m12		72m01		70m11	
Peak	74m02	73m07	73m10	74m5	73m05		74m01	73m04
Trough	75m10	75m10	75m06	76m4	75m10		75m07	75m11
Peak		77m06						
Trough		78m07						
Peak	79m12	79m10	80m01	80m6	79m09	79m09	79m05	79m05
Trough					81m02	81m04	80m08	
Peak							81m04	
Trough	82m05	81m04	83m01	82m5			82m11	83m03
Peak	83m05	83m01	84m12		85m02	84m06	84m06	84m08
Trough	85m05	85m07	87m02		86m06	86m10	86m08	85m12
Peak	89m09	89m09	91m02	89m6	89m01	88m12	88m07	88m02
Trough	93m04	93m10	93m11	93m4	91m10	91m09	91m07	91m12
Peak	95m03	95m06			95m01	95m06	95m01	95m01
Trough	96m06	96m11				96m09	96m04	
Peak	98m04	98m06	97m06	97m9		98m01		
Trough		99m09	98m11		99m03	99m06		98m04
Peak		00m08	00m10		00m12	00m08	00m05	00m03
Trough	02m02	02m02	02m01	02m9	02m07	02m05	01m12	01m12
False signals		4		0		2		0
Missed turnings		0		4		0		2
MEAN		-4/14		-2/8		-2/10		-3/11
STDV		5.3		10.4		3.9		4

Notes: *DCI* is the abbreviation for the deviation cycle indicator of the manufacturing industry. *CLI* is the abbreviation for the composite leading indicator. The mean (MEAN) and standard deviation (STDV) of the differences in timing of turning points refer to the differences in months between the timing of the corresponding turning points of the *DCI* and the *CLI*. A false signal occurs if the *CLI* shows a cycle and the *DCI* does not. A missed turning point occurs if the *DCI* shows a cycle and the *CLI* does not. Due to data availability, the *CLI* starts at a later date and therefore no turning points corresponding to the ones of the *DCI* exists at the beginning of the sample.

* For Belgium, Germany and Italy, the most recent turning point is not reported and not taken into account in the calculations. As shown in Figure A1, a competitive turning point date might become the relevant one as the estimates of the *DCI* get revised with the arrival of new data observations.

References

- Agresti, A. M. and B. Mojon. 2001. "Some stylised facts on the euro area business cycle," ECB Working Paper series no. 95.
- Baxter, M. and R. King. 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *Review of Economics and Statistics* 81, pp. 575-93.
- Burns, A. F. and W. C. Mitchell. 1946. *Measuring Business Cycles*. New York: NBER.
- Berk, J. M. and J. A. Bikker. 1995. "International interdependence of business cycles in the manufacturing industry: the use of leading indicators for forecasting and analysis," *Journal of Forecasting* 14, pp. 1-23.
- Bikker, J.A. and L. De Haan. 1988. "Forecasting business cycles: a leading indicator for the Netherlands," De Nederlandsche Bank, Quarterly Bulletin no. 3, pp. 71-82.
- Bry, G. and C. Boschan. 1971. *Cyclical analysis of time series: selected procedures and computer programs*. New York: NBER.
- Canova, F. 1998. "Detrending and Business Cycle Facts," *Journal of Monetary Economics* 41, pp. 475-12.
- Conference Board. 2000. *Business cycle indicators handbook*. New York: The Conference Board.
- Christiano, L. J. and T. J. Fitzgerald. 2003. "The Band Pass Filter," *International Economic Review* 44:2, pp. 435-65.
- Diebold, F. X. and G. D. Rudebusch. 1990. "A Nonparametric Investigation of Duration Dependence in the American Business Cycle," *Journal of Political Economy* 98, pp. 596-616.
- Fase, M. M. G. and J. A. Bikker. 1985. "De datering van economische fluctuaties: proeve van een conjunctuurspiegel voor Nederland 1965-1984," *Maandschrift Economie* 49, pp. 299-332.
- Gomez, V. 2001. "The use of Butterworth filters for trend and cycle estimation in economic time series," *Journal of Business and Economic Statistics* 19, pp. 365-73.
- Granger, C. W. J. 1966. "The typical spectral shape of an economic variable," *Econometrica* 34:1, pp. 150-61.
- Haan, L. de and F. W. Vjlselaar. 1998. "De vernieuwde DNB-conjunctuurindicator," *Economisch Statistische Berichten* 4365, pp. 660-64.
- Hall, R. E.; M. Feldstein, J. Frankel, R. Gordon, C. Romer, D. Romer and V. Zarnowitz. 2003. *The NBER's recession dating procedure*. Business cycle dating committee, National Bureau of Economic Research.
- Hamilton, J. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57, pp. 357-84.
- Harding, D. and A. R. Pagan. 2001. "Extracting, analysing and using cyclical information," unpublished document, available at <http://econrsss.anu.edu.au/~arpagan/pdf/romeconf.pdf>.
- Harding, D. and A. R. Pagan. 2002. "Dissecting the cycle: a methodological investigation," *Journal of Monetary Economics* 49:2, pp. 365-81.
- Harding, D. and A. R. Pagan. 2005. "Synchronisation of Cycles," *Journal of Econometrics*, 127:1 pp. 59-79.
- Harvey, A. C. and A. Jaeger. 1993. "Detrending, stylized facts and the business cycle," *Journal of Applied Econometrics* 8, pp. 231-47.
- Harvey, A.C. and T. M. Trimbur. 2003. "General model-based filters for extracting cycles and trends in economic time series," *Review of Economics and Statistics* 85, pp. 244-55.
- Hodrick, R. J. and E. C. Prescott. 1997. "Post-War U.S. Business Cycles: A descriptive empirical investigation," *Journal of Money Credit and Banking* 29, pp. 1-16.
- Koopmans, T. C. 1947. "Measurement without Theory," *Review of Economics and Statistics* 29, pp. 161-79.
- Kranendonk, H.; J. Bonenkamp and J. Verbruggen. 2004. "A leading indicator for the Dutch economy: methodological and empirical revision of the CPB system," in *IFO survey data in business cycle and monetary policy analysis*, J.E. Sturm and T. Wollmershäuser, eds. Springer Verlag.
- Marcellino, M. 2006. "Leading indicators: What have we learned?" in *Handbook of Economic Forecasting*, G. Elliot, C. W. J. Granger and A. Timmerman, eds. North-Holland: Elsevier.
- Moore, G. H. and J. Shiskin. 1967. "Indicators of business expectations and contractions," NBER Occasional Study no. 103.

- OECD. 2002. "An update of the OECD composite leading indicators," downloadable on "<http://www.oecd.org/dataoecd/4/33/15994428.pdf>".
- Ohn, J.; L. W. Taylor and A. R. Pagan. 2004. "Testing for duration dependence in cycles in economic cycles," *Econometrics Journal* 7, pp. 528-49.
- Orphanides, A. and S. van Norden. 2002. "The unreliability of output-gap estimates in real time," *Review of economics and statistics* 84, pp. 569-83.
- Pedersen, T.M. 2001. "The Hodrick-Prescott filter, the Slutsky effect and the distortionary effect of filters," *Journal of Economic Dynamics and Control* 25, pp. 1082-101.
- Pollock, D. S. G. 2000. "Trend estimation and de-trending via rational square-wave filters," *Journal of Econometrics* 99, pp. 317-34.
- Priestley, M. B. 1981. *Spectral analysis and time series*. Academic Press.
- Ravn, M. R. and H. Uhlig. 2002. "On adjusting the HP-filter for the frequency of observations," *Review of economics and statistics* 84, pp. 371-76.
- Schleicher, C. 2004. "Kolmogorov-Wiener filter for finite time-series," Mimeo: University of British Columbia. Download on "http://www.iam.ubc.ca/theses/Schleicher/ChristophSchleicher_PhD_Thesis.pdf".
- Stock, J.H. and Watson, M.W., 1989, "New indexes of coincident and leading economic indicators", in *Macroeconomics Annual*, O. Blanchard and S. Fischer, eds. New York: NBER, MIT Press, pp. 351-94.
- Stock, J. H. and M. W. Watson. 1999. "Business cycle fluctuations in US macroeconomic time series," in B. Taylor and M. Woodford, eds. *Handbook of Macroeconomics*, Volume I.
- Stock, J. H. and M. W. Watson. 2003. "Forecasting output and inflation: the role of asset prices," *Journal of Economic Literature* 41, pp. 788-829.

Appendix

A technical note on testing for duration dependence

This appendix derives a regression based test statistic for duration. The state variable S_t is obtained by applying the dating rules (1) to the indicators of the deviation cycle. Moreover, let d_i be the duration of the i^{th} phase in number of months. Consider the random sample of n observations $(D_1, D_2, \dots, D_n)$ of durations from a discrete distribution with density function $f(d)$. Then the hazard rate function is defined as,

$$h(d) = \frac{f(d)}{1 - F(d)}$$

For a small Δ , $h(d)\Delta$ is the probability that the phase will terminate during the interval $(d, d + \Delta)$. If there is to be no duration dependence, then the termination probability must be constant, regardless of the duration of time spent in the phase and so does not depend on d , that is

$$H_0 : h(d) = \theta, \text{ for some } \theta > 0 \text{ and all } d > 0. \quad (\text{A.1})$$

In the discrete case, the null hypothesis of no duration dependence (A.1.) can only be satisfied by the geometric distribution. The geometric density of D with termination probability p reads as

$$P(D = d) = (1 - p)^d p \text{ for } 0 \leq d < \infty,$$

and its constant hazard function can be derived as:

$$h(d) = \frac{P(D = d)}{1 - P(D \leq d)} = \frac{(1 - p)^d p}{(1 - p)^d} = p.$$

Moreover, it holds in the geometric case that $E(D) = (1 - p)/p$, $Var(D) = (1 - p)/p^2$ and so

$$Var(D) - [E(D)]^2 - E(D) = 0.$$

So, the null hypothesis (A.1) of no duration dependence implies a testable moment condition.

However, in this study a minimum duration on the complete cycle of 18 months is imposed by the parameter settings of the CF-filter. So, effectively the distribution of D becomes a delayed geometric one $P(D^{delayed} = D - c)$, or equivalently :

$$P(D^{\text{delayed}} = k) = \begin{cases} 0 & \text{otherwise} \\ p(1-p)^{k-c} & \text{for } k = c, c+1, \dots \end{cases}$$

And it holds that

$$E(D^{\text{delayed}}) = \frac{(1-p)}{p} + c = E(D) + c, \quad \text{Var}(D^{\text{delayed}}) = \frac{(1-p)}{p} = \text{Var}(D)$$

and the testable moment restriction for the delayed geometric distribution reads as:

$$\text{var}(D) - [E(D)]^2 - (2c+1)E(D) - c(c+1) = 0. \quad (\text{A.2})$$

A regression based test statistic for duration dependence, which is of delayed geometrically distributed durations (A.2), can be constructed as a variable -addition test in a regression equation. The test is a modification of Ohn *et al.* (2004) to allow durations to be delayed in addition to be geometrically distributed.

Consider the following regression:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 S_{t-1} X_{t-1} + \text{error} \quad (\text{A.3})$$

where the number of consecutive months spent in a high-growth phase up and through time t is given by:

$$X_t = 1 + (1 - S_t(1 - S_{t-1}))(1 - S_{t-1}(1 - S_t))X_{t-1}.$$

It is possible to derive an analytic form for the estimate of β_2 which is identical to the moment restriction (A.2).

Consider:

$$\frac{1}{T} \sum_{t=1}^T S_{t-1} S_t = \bar{S} - \frac{n}{T}$$

where $\bar{S}$ is the sample mean of S_t , T is the sample size and n is the number of TP-phases. Moreover,

$$\begin{aligned} \overline{SX} &= \frac{1}{T} \sum_{t=1}^T S_t X_t = \frac{1}{T} \sum_{i=1}^n \frac{(d_i^{\text{delayed}} + 1) d_i^{\text{delayed}}}{2} \\ &= \frac{1}{T} \sum_{i=1}^n \frac{(d_i + 1 + c_m)(d_i + c_m)}{2} = \frac{n}{2T} \left[\bar{d}^2 + (2c_m + 1)\bar{d} + (c_m^2 + c_m) \right] \end{aligned}$$

where d_i is the duration of the i^{th} high-growth phase¹⁰ minus the minimum of c_m periods.

Finally,

$$\frac{1}{T} \sum_{t=1}^T S_t S_{t-1} X_{t-1} = \frac{1}{T} \sum_{t=1}^n \frac{(d_i - 1 + c_m)(d_i + c_m)}{2} = \overline{SX} - \bar{S}.$$

Now, the standard linear regression of $S_t - \bar{S}$ on $S_{t-1} - \bar{S}$ and $S_{t-1} X_{t-1} - \overline{SX}$, with Δ the determinant of the scaled (by $1/T$) cross product matrix of the two regressors, results in:

$$\begin{aligned} \Delta \hat{\beta}_2 &= \bar{S}(1 - \bar{S}) \left(\overline{SX} - \bar{S} + \frac{n}{T} c_m - \overline{SX} \bar{S} \right) - \overline{SX}(1 - \bar{S}) \left(\bar{S}(1 - \bar{S}) - \frac{n}{T} \right) \\ \Leftrightarrow \frac{T \Delta \hat{\beta}_2}{n(1 - \bar{S})} &= \overline{SX} + \bar{S} \left(c_m - \frac{T}{n} \bar{S} \right) \\ &= \frac{1}{T} \sum_{i=1}^n \frac{(d_i + 1 + c_m)(d_i + c_m)}{2} + \frac{n(\bar{d} + c_m)}{T} \left(c_m - \frac{n}{T} \frac{T}{n} (\bar{d} + c_m) \right) \\ \Leftrightarrow \frac{2T^2 \Delta \hat{\beta}_2}{n^2(1 - \bar{S})} &= \bar{d}^2 - 2\bar{d} - (2c_m - 1)\bar{d} - c_m(c_m - 1) \\ &= \text{var}(D) - [E(D)]^2 - (2c_m - 1)E(D) - c_m(c_m - 1). \end{aligned} \quad (\text{A.4})$$

Note that for $c_m := c + 1$ equations (A.2) and (A.4) are equivalent. The variable addition test statistic $H_0 : \beta_2 = 0$ for equation (A.3) is therefore also effectively testing for no delayed duration dependence. Because of the binary nature of the variables involved one needs to consider the heteroskedastic and autocorrelation consistent t -statistic.

¹⁰ The analysis can be performed for low-growth (PT-) phases as well by defining $S^{\text{low-growth phase}} = 1 - S^{\text{high-growth phase}}$.

Figure A1 International deviation cycles in the manufacturing industry and composite leading indicators

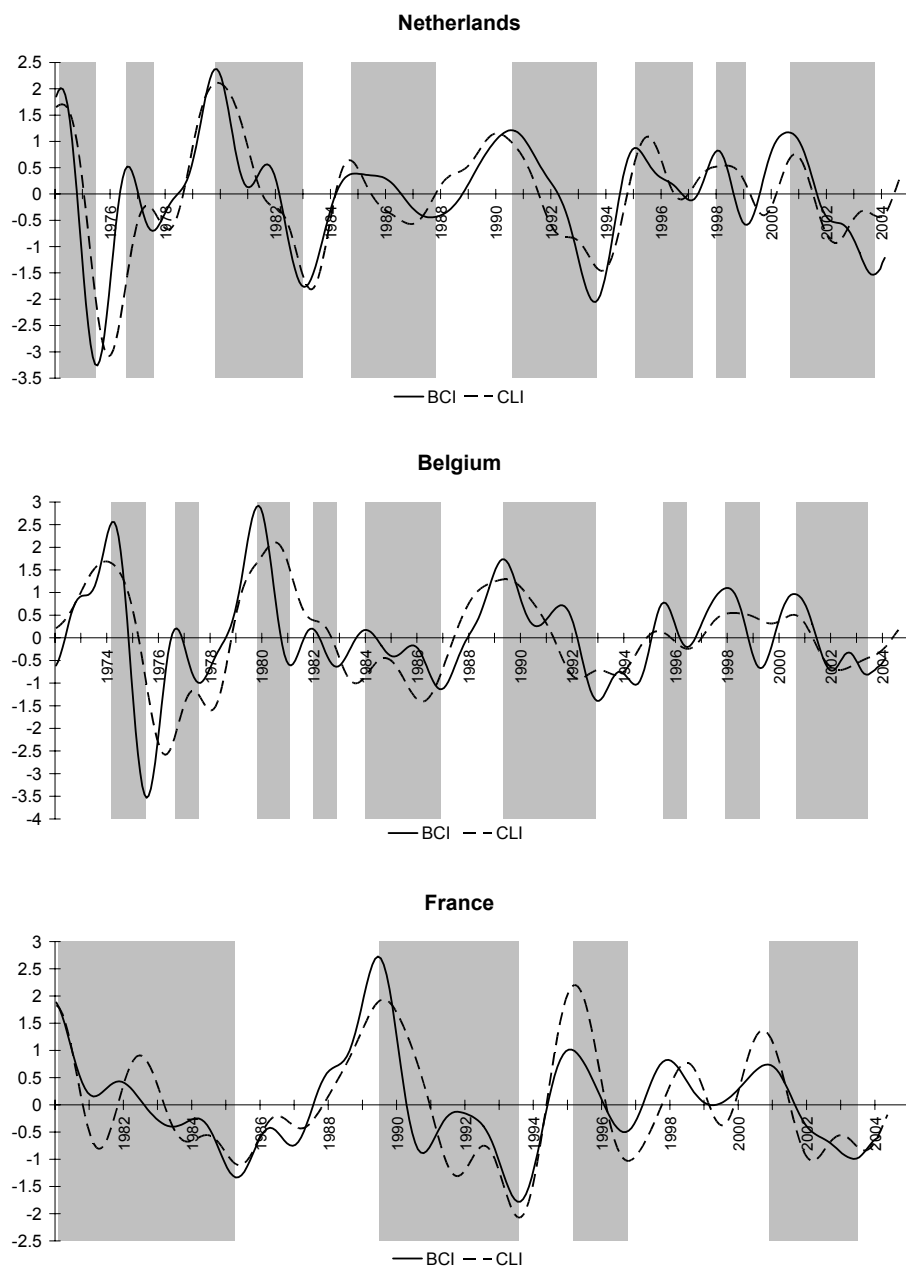


Figure A1 continued

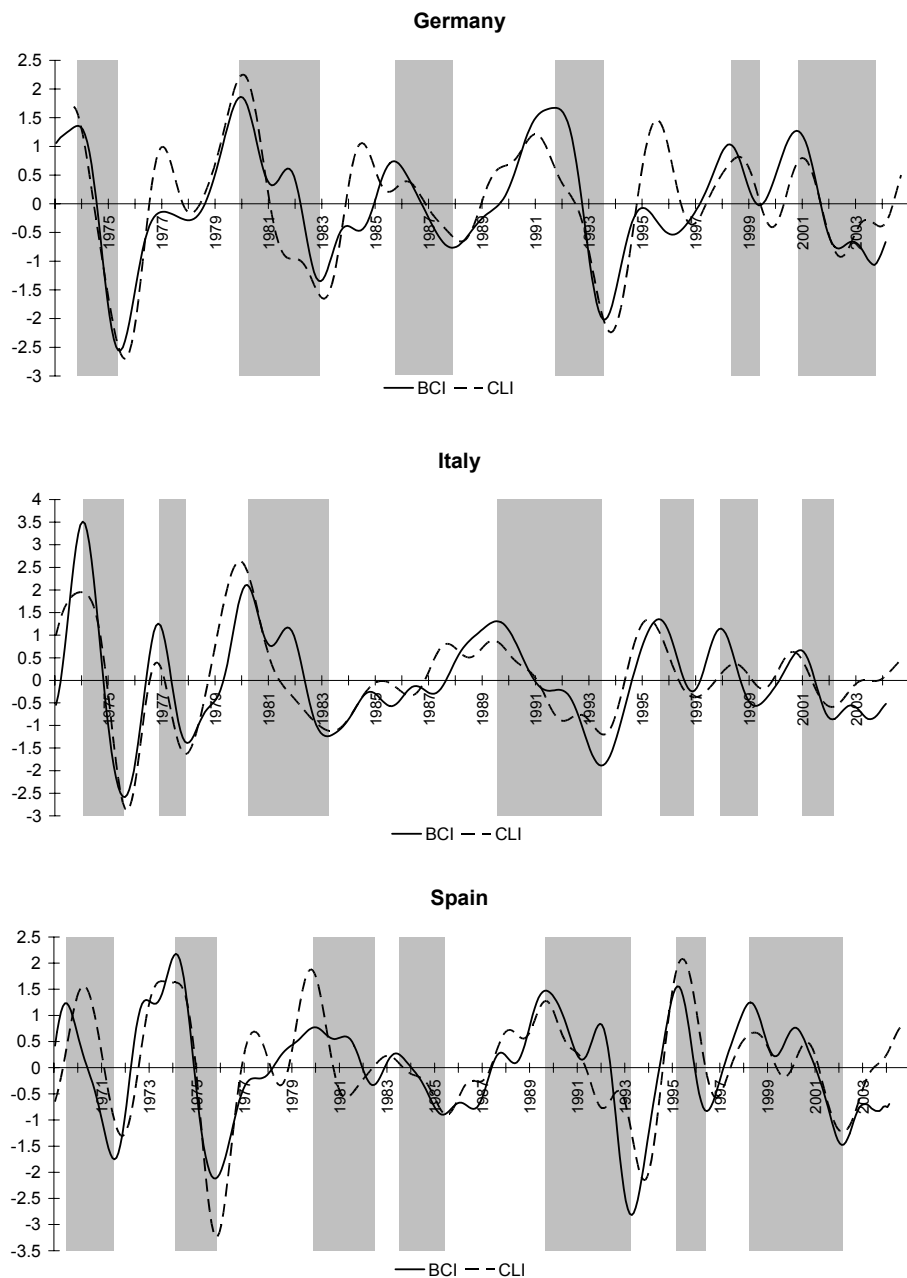


Figure A1 continued

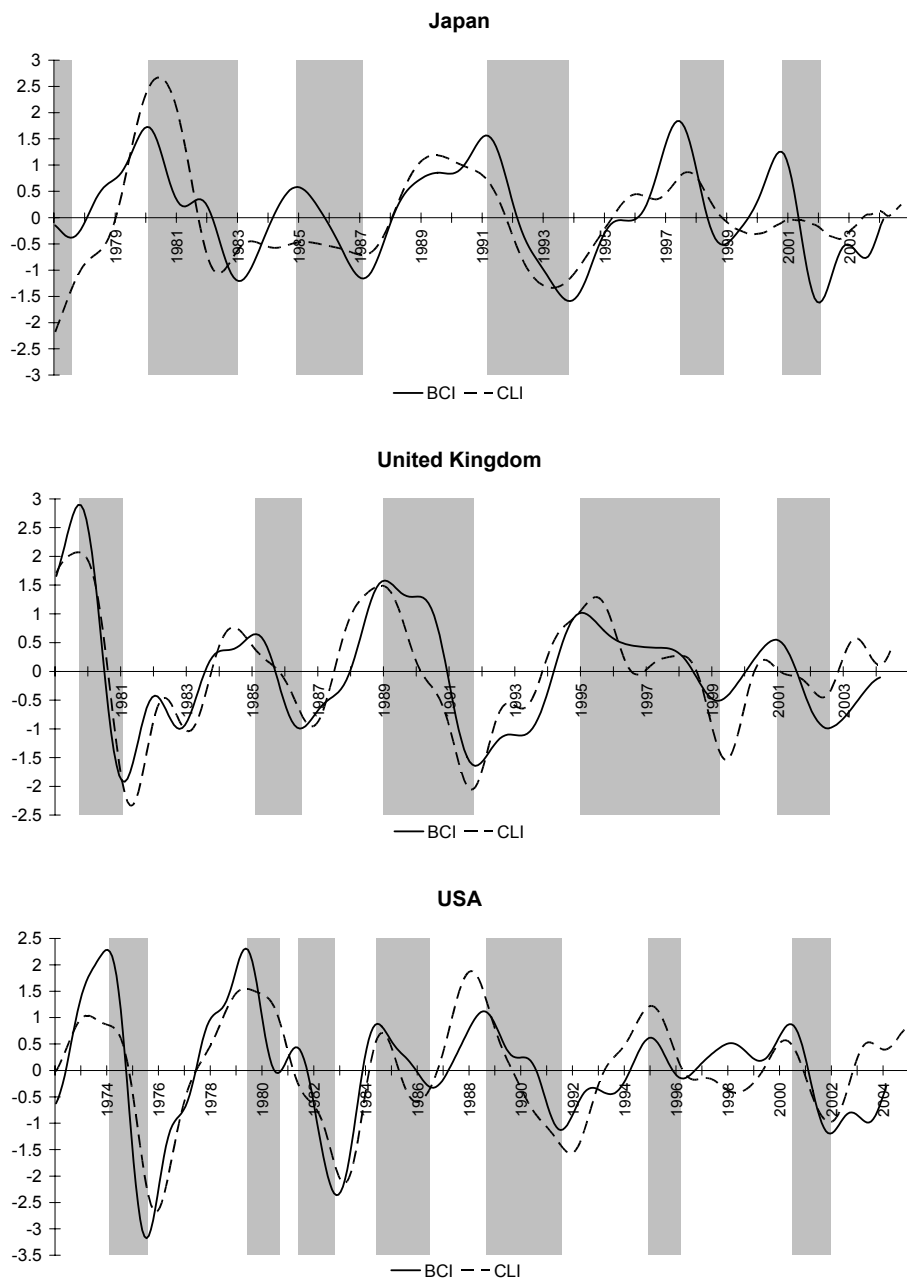


Table A1 Data source

Description	Code	Source*
The Netherlands		
Industrial production	IRAA04MB	CBS
Consumer price index	IRDE31MA	CBS
Short term interest rate	IKGB01MA	Prebon/Yamane IKGB11MA, from 1999 DNB-FM IKGB01MA
Long term interest rate	IKGE11MA	DS
IFO, expected business situation	IJDEFDMA	DS
Hourly wage industry	IRDK11MA	CBS
Storage final products	IJNLFAMA	DS
Expected business activity	IRAD41MA	CBS
Belgium		
Industrial production	IIBEAXMB	BIS
Consumer price index	IIBEBAMA	BIS
Short term interest rate	IKGB01MA	BIS IIBELBMA, from 1999 DNB-FM IKGB01MA
Long term interest rate	IIBELDMA	DS
IFO, expected business situation	IJDEFDMA	DS
Hourly wage industry	IJBFEFMA	DNB S&I*
Domestic sales prices	IIBEBCMA	BIS
Labour costs per unit product	IJBFEFMA	DNB S&I*
France		
Industrial production	IJFRFLMH	DS
Consumer confidence	IIFRAMMB	DS
Producer confidence	IIFRAQMB	DS
Equity price	IIFRLXMA	BIS
Labour costs per unit product	IJFRFFMA	DNB S&I IIFRAGKA*
Judgement of order arrivals	IJFRFBMA	DS *
Storage final products	IJFRFAMA	DS
Level order position	IJFRFCMA	DS
Germany		
Industrial production	IJDEFLMH	DNB S&I (ECB, BIS)
Consumer confidence	IIDEAMMB	DS
Short term interest rate	IKGB01MA	DNB-FM IKGB41MA, from 1999 DNB-FM IKGB01MA
IFO, expected business situation	IJDEFDMA	DS
Hourly wage industry	IIDEADMB	BIS
Judgement of order arrivals	IJDEFBMA	DS
Storage final products	IJDEFAMA	DS
Level order position	IJDEFCMA	DS

Notes: * Quarterly to Monthly

BIS = Bank of International Settlements, CBS = Statistics Netherlands,
 DNB-FM = De Nederlandsche Bank - division Financial Markets,
 DNB S&I = De Nederlandsche Bank - division Statistics & Information,
 DS = Datastream, ECB = European Central Bank

Table A1 **Data source (continued)**

Description	Code	Source*
Italy		
Industrial production	IJITFLMH	DNB S&I (ECB, BIS)
Consumer price index	IIITBAMA	DS
Short term interest rate	IKGB01MA	BIS IIITLBMA, from 1999 DNB-FM IKGB01MA
Long term interest rate	IIITLDMA	DS
IFO, expected business situation	IJDEFDMA	DS
Hourly wage industry	IIITADMA	BIS
Domestic sales prices	IIITBCMA	BIS
Producer confidence	IIITAQMB	DS
Prospects total economy	IJITFEMA	DS
Judgement of order arrivals	IJITFBMA	DS
Level order position	IJITFCMA	DS
World market prices commodities	IRDF01MA	HWWA - Institut für Wirtschaftsforschung
Spain		
Industrial production	IJESFLMH	DNB S&I (ECB, BIS)
M1	IJESFFMA	DNB S&I IIESHXMA + IIESTXME
IFO, expected business situation	IJDEFDMA	DS
Storage final products	IJESFAMA	DS
Level order position	IJESFCMA	DS
World market prices commodities	IRDF01MA	HWWA - Institut für Wirtschaftsforschung
Japan		
Industrial production	IIJPAXMB	BIS
Consumer price index	IIJPBAMA	BIS
Short term interest rate	IIJPLBMA	BIS
Long term interest rate	IIJPLDMA	BIS
Consumer confidence	IJJPFBMA	DNB S&I IIJPAMKB*
Hourly wage industry	IIJPADMA	BIS
Domestic sales prices	IIJPBCMA	BIS
Storage final products	IJJPFAMA	DS *
Consumption	IJJPCMA	DNB S&I IIJPCFKB*

Table A1 Data source (continued)

Description	Code	Source*
United Kingdom		
Industrial production	IIGBAXMB	DS
Consumer confidence	IIGBAMMB	DS
Producer confidence	IIGBAQMB	DS
Consumer price index	IIGBBAMA	DS
Short term interest rate	IIGBLBMA	Financial Times
Domestic sales prices	IIGBBCMA	BIS
Total wages	IJGBFFMA	OECD/QNA GBR.CQRSA.S1 Compensation of employees*
Expected future industrial production	IJGBFDMA	DS
Storage final products	IJGBFAMA	DS
Level order position	IJGBFCMA	DS
Prospects total economy	IJGBFEMA	DS
United States		
Industrial production	IIUSAXMB	BIS
World market prices commodities	IRDF01MA	HWWA - Institut für Wirtschaftsforschung
Storage final products	IJUSFAMA	DS
Hourly wage industry	IIUSADMB	BIS
Equity price (S&P corporate 500)	IIUSLXMA	BIS
Short term interest rate	IIUSLBMA	BIS
Long term interest rate	IIUSLDMA	BIS
Domestic sales prices	IIUSBCMA	BIS

Notes: * Quarterly to Monthly

BIS = Bank of International Settlements, *CBS* = Statistics Netherlands,

DNB-FM = De Nederlandsche Bank - division Financial Markets,

DNB S&I = De Nederlandsche Bank - division Statistics & Information,

DS = Datastream, *ECB* = European Central Bank

Comparing Probability Forecasts in Markov Regime Switching Business Cycle Models

Daniel R. Smith* and Allan Layton**

Abstract

We evaluate techniques for comparing the ability of Markov regime switching (MRS) models to fit underlying regimes of a series of interest. This is particularly important in the business cycle literature where one may be interested in determining whether using leading indicators to allow transition probabilities to vary improves the ability of MRS models to fit the NBER business cycle chronology. This is typically done using the quadratic probability score, or QPS (Diebold and Rudebusch, 1989). Although it is possible to statistically compare the QPS statistics for two MRS models using the Diebold and Mariano (1995) (DM) test statistic for comparing forecasts, we find using a Monte Carlo experiment that the DM statistic tends to under-reject (the null of "no difference in forecast accuracy") when comparing MRS models. This we believe is because of the strong non-normality of the forecast errors of such models. Furthermore, using simulation-based inference we demonstrate that leading indicators improve the fit of an MRS model of the US business cycle chronology by 24 percent, such improvement having a p -value of 0.001.

Key Words: Markov Regime Switching, Diebold and Mariano statistic, Quadratic Probability Score, Monte Carlo, Business Cycle

JEL Classification: C15, C53, E32

*

Simon Fraser University, Faculty of Business Administration (Burnaby, Canada); drsmith@sfu.ca

**

University of Southern Queensland, Faculty of Business (Toowoomba, Australia); layton@usq.edu.au

Résumé

Nous évaluons des techniques afin de comparer l'aptitude des régimes alternatifs de Markov à s'adapter aux régimes sous-jacents dans la série souhaitée. Cela est particulièrement important pour la littérature sur les cycles économiques où il peut y avoir un intérêt à déterminer si le fait d'utiliser des indicateurs avancés, afin d'autoriser la variation de probabilités de transition, améliore ou non la capacité des régimes alternatifs de Markov à s'adapter à la chronologie du cycle économique du Bureau National de Recherches Économiques. La technique la plus couramment utilisée est celle du QPS (quadratic probability score) (Diebold et Rudebusch (1989)). Bien qu'il soit possible de comparer statistiquement les probabilités QPS de deux régimes alternatifs de Markov en utilisant le test statistique de prévision comparative Diebold et Mariano (1995) (DM), nous nous rendons compte, grâce à une simulation de Monte-Carlo, que la statistique DM tend à rejeter l'hypothèse nulle "aucune différence dans l'exactitude de la prévision" lors de la comparaison des régimes alternatifs de Markov. Nous pensons que cela est dû à une non normalité significative des erreurs de prévisions de tels modèles. De plus, en utilisant des déductions fondées sur des simulations nous démontrons que les indicateurs avancés améliorent l'adéquation d'un régime alternatif de Markov à la chronologie du cycle économique US à hauteur de 24 %. Une telle amélioration ayant une p -value de 0.001.

1 Introduction

The Markov regime switching (MRS) model has become an important tool for modeling many financial and economic time series. It has been particularly prominent in the business cycle literature. In this literature it is very common for researchers to be interested in determining the extent to which their models of the business cycle reproduce broad characteristics of the real world business cycle they purport to represent, including phase amplitudes and durations. Of particular interest in the context of MRS models is the degree to which a particular model can reproduce the actual phase change chronology of the business cycle. In the US, the most widely accepted business cycle chronology is that compiled by the National Bureau of Economic Research (NBER). This is a monthly chronology which extends back to before the first world war.¹ The "actual" chronology of a country's business cycle is not directly observable and so many other countries do not have such a widely accepted set of dates for the peaks and troughs in the cycle but, for the US, it is reasonable to regard the NBER chronology as, in effect, amounting to the actual US business cycle chronology. The vast majority of US business cycle scholars have, for many years, regarded it as such, and have proceeded to test their models for their consistency with the NBER dates.

By allowing the transition probabilities in MRS to vary with leading indicators we can assess their usefulness in predicting business cycle phase changes. The most important metric for assessing the ability of various MRS models to "fit" the business cycle is the so-called quadratic probability score (QPS) introduced by Diebold and Rudebusch (1989). The QPS statistic is simply a mean square error measure comparing the probability of a regime (the prediction) with an indicator variable for the regime. Smaller QPS statistics identify models that fit the NBER chronology better. The magnitude of the QPS statistic has been used in the literature to compare different regime-switching business cycle models. This metric can be used to determine, for example, whether using leading indicators to allow the transition probabilities to vary through time can improve their fit to the actual business cycle chronology. Examples of researchers who have used QPS in the business cycle context include Chen (2002), Diebold and Rudebusch (1989), Filardo (1994), Kholodilin and Yao (2005), Lahiri and Wang (1994), Layton (1998), Chauvet (1999) and others. Unfortunately the sampling distribution of the QPS statistic is unknown, and formal statistical comparisons of the fit of different model-generated chronologies to the actual NBER business cycle chronology are rare. The only attempts at a statistical comparison of models using the QPS statistic of which we are aware is that by Chen (2002) and Kholodilin and Yao (2005) who use the asymptotic Diebold-Mariano (1995) test.

The Diebold and Mariano (1995) statistic (DM statistic) is an important general-purpose tool for comparing two different competing sets of model-based forecasts. Diebold and Mariano (1995) propose a test statistic which involves calculating an asymptotic Z-statistic using the variance of the loss-differential (in our context, the difference between the squared

¹ For the purposes of this paper only the chronology dating from the 1950s will be used. The chronology is available from the NBER website.

forecast errors). This can be applied in a straightforward fashion to the QPS statistic. In many applications the theoretical asymptotic standard normal distribution for the Z-statistic appears to be a very good approximation and the test performs satisfactorily. However, comparing QPS statistics in MRS models presents somewhat of a specific challenge to the DM approach because the distribution of the loss-differential is extremely non-normal. The forecast errors are usually very small while the system stays in a given regime but can be relatively large around the times of regime change, giving rise to an empirical distribution much more peaked than the normal distribution.

This strong non-normality is exacerbated when considering differences in the squared forecast errors. As a result the normal distribution is quite a poor approximation in even quite large samples. We demonstrate using a Monte Carlo experiment that in a wide range of models the DM statistic consistently under-rejects the null hypothesis even when the samples are as large as 500 or even 1000 observations. We therefore recommend caution when implementing the DM statistic to compare MRS models of the US business cycle.

Given the poor performance of the DM statistic in samples that are typical in many business cycle empirical analyses we turn to a purely empirical approach to formal statistical inference. We present an empirical application to modeling the business cycle using the experimental coincident index (XCI) of Stock and Watson. We find that allowing the transition probabilities to vary with their experimental leading index improves the model's statistical fit, as measured by log-likelihood, and its ability to fit the NBER chronology, as measured by its QPS statistic, by around 24 percent. This QPS improvement has an empirical p -value of 0.001 and is almost $1/20^{\text{th}}$ of the p -value of 0.018 of the DM test statistic using the theoretical asymptotic distribution of the standard normal.

The paper proceeds as follows. In Section 2 we present the MRS model of the business cycle and present an application to the US business cycle. We discuss the Diebold-Mariano (1995) statistic in Section 3. In Section 4 we evaluate the finite-sample properties of the DM statistic using a Monte Carlo experiment. Section 5 assesses the ability of leading indicators to improve the fit of MRS models to the NBER chronology and we conclude in Section 6.

2 Regime Switching Models

The Markov Regime Switching (MRS) model of Hamilton (1989) has become a very important tool in understanding the dynamics of both economic and financial time series. The conditional distribution of a time series variable, y_t , depends on the value of a discrete state latent variable S_t . We will present our analysis using a very simple two-state model where $S_t = \{1, 2\}$ and there are no autoregressive dynamics. The simplest MRS specification models the dynamics of S_t as a Markov process with constant transition probabilities:

$$\begin{aligned}
p(S_{t+1} = 1 | S_t = 1) &= p \\
p(S_{t+1} = 2 | S_t = 1) &= 1 - p \\
p(S_{t+1} = 1 | S_t = 2) &= 1 - q \\
p(S_{t+1} = 2 | S_t = 2) &= q
\end{aligned}$$

We model the conditional distribution of y_t given S_t as

$$f(y_t | S_t = i; \theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_t - \mu_i)^2}{2\sigma_i^2}\right)$$

where $\theta = (p, q, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)'$ is the parameter vector. We denote this model the constant transition probability (CTP) model. We estimate the parameters by maximum likelihood and infer the state of the economy using Hamilton's (1989) recursive filter. This filter produces a sequence of regime probabilities for each observation in the data.

We can extend the model to include time-varying transition probabilities (TVTP) which depend on some variable X_t . In MRS models of the business cycle it is common to define X_t using leading indicators, e.g., Diebold, Lee, and Weinbach (1994), Fillardo (1994) and Layton (1998), and, in a regime-switching model of interest rates, Gray (1996) defines X_t as the level of short-term interest rates. The transition probabilities are defined as:

$$p(S_{t+1} = 1 | S_t = 1, X_t) = p_{t+1} = (1 + \exp(-\beta_{01} - \beta_{11}X_t))^{-1}$$

$$p(S_{t+1} = 2 | S_t = 2, X_t) = q_{t+1} = (1 + \exp(-\beta_{02} - \beta_{12}X_t))^{-1}$$

using the logistic transformation to ensure that the transition probabilities lie between zero and one.

2.1 The quadratic probability score

It is common when modeling the US business cycle using MRS to compare the regime probabilities with the NBER business cycle chronology. This is done both informally by comparing graphs of a binary regime indicator variable with the regime probabilities, or more formally using the QPS statistic:

$$QPS = \frac{1}{T} \sum_{t=1}^T (\Pr(S_{t+1} = 1 | I_t) - \lambda_{t+1})^2$$

where: $\Pr(S_{t+1} = 1 | I_t)$ is the ex ante forecast probability of a recession in the next period formed using information available up to the current period which is denoted by I_t , λ_{t+1} is a

dummy variable taking the value one when the NBER chronology indicates that the US economy was in recession in month $t+1$ (a recession is defined as any month between a peak (non-inclusive) and the subsequent trough (inclusive)). The metric originated in weather forecasting (Brier, 1950) and was introduced to the MRS literature by Diebold and Rudebusch (1989). It has since been applied to business cycles by Filardo (1994), Lahiri and Wang (1994), Layton (1998), Layton and Smith (2000), Chauvet (1998) and others. It can be used to compare, for example, the performance of constant and varying transition probability models. In particular, a model with a small QPS statistic might be preferred to another model with a larger QPS statistic as it more closely reproduces the NBER chronology.

2.2 Regime switching models of the US business cycle

We estimate a two-regime CTP model for the experimental coincident index (XCI) developed by Stock and Watson (1989). We also use their experimental leading index (XLI) to estimate a TVTP model. Our sample period is from March 1959 through December 2003, giving a sample size of 538 observations. We plot the data in Figure 1 in the Appendix. Their relationship to the underlying NBER chronology is clearly evident. The parameter estimates for both the CTP and TVTP models are reported in Table 1. Consider first the CTP results. Both regimes are quite persistent (as evidenced by the estimated transition parameters being close to one) and the average growth rate in the expansion state (state 0) is much higher than in the recession state. The variance of XCI is much higher when the economy is in recession. A very similar pattern emerges from the TVTP model which appears to be a superior explanation of the XCI data as evidenced by a likelihood ratio statistic of 18.86 for the null hypothesis of constant transition probabilities.² The forecast and smoothed regime probabilities are presented in Figure 2 for the CTP model, and in Figure 3 for the TVTP model. Both models appear to do a pretty good job of fitting the NBER chronology.

We also calculate the QPS statistics for each model using the NBER official chronology as the reference of the actual state of the US economy. These dates may be downloaded from the NBER website <http://www.nber.org/cycles/>. The TVTP model produces a 24% improvement in QPS over the CTP model, suggesting that allowing the transition parameters to vary according to XLI has quite substantially improved the power of the model to predict

² Under the null of no improvement in fit to the XCI data, this statistic is distributed as chi-squared with 2 degrees of freedom and so the calculated value of 18.86 equates to a p -value <0.0001 . We also calculated the robust Wald test for this same null hypothesis. The value of this robust test statistic is 14.5708 which again is significant at better than the 0.0001 significance level.

the NBER-determined dates of the US business cycle phase changes. Of course, whether this improvement is a statistically significant one needs to be formally determined.³

Table 1 Parameter estimates of the CTP and TVTP MRS models for XCI

Parameter	CTP Model		TVTP Model	
	Estimate	Robust SE	Estimate	Robust SE
p	0.9727	0.0104		
q	0.9185	0.0310		
μ_1	0.3422	0.0279	0.3484	0.0312
μ_2	-0.1310	0.0964	-0.1433	0.0997
σ_1^2	0.1459	0.0136	0.1448	0.0158
σ_2^2	0.6705	0.1761	0.6515	0.1861
β_{01}			2.4261	0.4311
β_{11}			0.3193	0.1339
β_{02}			2.8984	0.6691
β_{12}			-0.4013	0.1641
LL	-385.3328		-375.9028	% Change:
QPS-Ex Ante	0.1104		0.0844	-24%

Notes: LL stands for log likelihood and QPS stands for quadratic probability score. QPS-Ex Ante is the QPS statistic using ex ante probabilities of the state in period $t+1$ made in period t , $p(S_{t+1}=1|I_t)$. The TVTP model's QPS is 24% lower than that for the CTP model.

³ The model we use is very simple and does not allow for autoregressive dynamics in the conditional mean other than that which is generated by changes in regime. To explore this possibility we also fit both the CTP and TVTP models allowing for AR(1) dynamics. Consistent with results reported in Layton and Smith (2000) we find that adding AR(1) dynamics to the CTP model reduces the model fit as measured by QPS which increases to 0.1196. Interestingly, we find that the TVTP model with AR(1) dynamics actually improves on the basic TVTP model, producing a QPS statistic of only 0.0790. We thank an anonymous referee for this suggestion.

3 Comparing Forecast Accuracy: The Diebold-Mariano (1995) Statistic

Suppose we are interested in comparing two forecasts of some random variable w_t , which are denoted by $\hat{w}_{i,t}$ for $i=1,2$. Diebold and Mariano (1995, DM hereafter) construct an asymptotic Z-statistic for comparing these two forecasts. Define the forecast errors as $e_{i,t} = w_t - \hat{w}_{i,t}$. The DM test evaluates the forecasts in terms of an arbitrary loss metric $g_{i,t}$ which we will define here as the squared error $g_{i,t} = e_{i,t}^2$. The test statistic is then given by

$$Z = \frac{T^{-1} \sum_{t=1}^T (g_{1t} - g_{2t})}{\sqrt{S/T}}$$

where S is an estimator of the variance of the sample mean loss differential $T^{-1} \sum_{t=1}^T (g_{1t} - g_{2t})$. DM suggest estimating S using the spectral density of the loss differential $g_{1t} - g_{2t}$ at zero frequency, similar to the estimators presented in Newey and West (1987) and Andrews (1991).⁴ The numerator of the Z-statistic is the difference between the mean-squared prediction errors from each forecast.

We can use the DM statistic to compare different MRS models in terms of their ability to reproduce the NBER chronology. In the context of MRS models w_t is defined as the business cycle indicator variable λ_t and the two forecasts are the regime probability predictions estimated from the two models. In this context the numerator of the Z-statistic is nothing other than the difference between the QPS statistics from each model.

Under suitable regularity conditions we can rely on the law of large numbers to derive the asymptotic standard normal distribution for the Z-statistic. For most applications the normal distribution will be appropriate in even quite modest sample sizes. However, the MRS model presents a particular challenge for the DM test. Virtually all MRS models are characterized by quite persistent regimes often with probabilities of staying in a given regime from one period to the next of around 0.9. While the economy is in that regime the regime probability forecast errors are usually quite small. However, at periods around phase changes at either a business cycle peak or trough, the forecast errors can become relatively large.

We illustrate the extreme non-normality with our MRS model of the business cycle presented in the previous section. The problem arises because of the discrete nature of λ_t (only taking the values of zero or one) and the fact that most of the calculated loss differentials (the $(e_{1t}^2 - e_{2t}^2)$) are either very close to zero or quite large through the dataset. This is illustrated in Figure 4 in the Appendix where we plot the time-series of the forecast errors (e_{it}) for both the CTP and TVTP models, as well as the calculated loss differential for each observation in the dataset using the MRS models discussed in the previous section.

⁴ Following Newey and West (1987) we use an estimator based on the Bartlett Kernel and calculate the optimal number of lags to include following Andrews (1991).

Of course, non-normality in the underlying data is not a problem as we only require that the sample mean of the loss differential converges to a normal distribution asymptotically. Even in the MRS model the Z-statistic will converge to a standard normal distribution asymptotically. However, the pronounced non-normality in the loss-differential when comparing regime probability forecasts from constant and varying transition probability MRS models is such that the DM statistic converges to the normal distribution extremely slowly. In the next section we demonstrate in a Monte Carlo experiment that, for empirically relevant sample sizes (for monthly business cycle analyses), the asymptotic normal distribution is a poor approximation to the likely actual finite sample distribution of the calculated DM statistic.

4 A Monte Carlo Experiment

To understand how well the asymptotic distribution approximates the finite sample distribution in MRS models we undertake the following Monte Carlo experiment. For each of 1000 replications, indexed by i , we simulate a sequence of regimes ($S_t^{(i)}$), observations ($y_t^{(i)}$) and leading indicators ($X_t^{(i)}$) as follows:

1. Simulate T observations (T being 250, 500 and 1000) on a useless leading indicator, $X_t^{(i)}$, using a zero-mean AR(1) process with autocorrelation coefficient $\rho = \{0.5, 0.9\}$.
2. Simulate the state vector $S_t^{(i)}$ independently of X .⁵
3. Given the sequence of states $S_t^{(i)}$ we simulate the artificial coincident index series by drawing a standard normal variate $z_t^{(i)}$ as:

$$y_t^{(i)} = \mu_1 + (S_t^{(i)} - 1)(\mu_2 - \mu_1) + z_t^{(i)} \sqrt{\sigma_1^2 + (S_t^{(i)} - 1)(\sigma_2^2 - \sigma_1^2)}.$$

Parameter values are selected as in Table 2.

For each simulated triple, $y_t^{(i)}, X_t^{(i)}, S_t^{(i)}$, we then estimate both a CTP and a TVTP model. (Note that the CTP model is, of course, the correct specification). The CTP model is estimated using only $y_t^{(i)}$, whereas we use both $y_t^{(i)}$ and $X_t^{(i)}$ when estimating the TVTP model. We then use the actual $S_t^{(i)}$ and the estimated model-generated regime probabilities to calculate QPS^{CTP} and QPS^{TVTP} for each simulation, i , and calculate the DM statistic for that simulation. This is repeated for each of the 1000 replicates. We report the rejection frequencies using the two-sided test of the null hypothesis of equal predictive accuracy at both the 5 and 1 percent nominal sizes in Table 2 for a range of ten different model parameterizations which are reported at the head of each column.

⁵ Specifically, let $u_t^{(i)}$ be a draw from a continuous uniform random number defined on $[0, 1]$. We set $S_t^{(i)} = 1$ when $u_t^{(i)} \leq (1-q)/(2-p-q)$ and $S_t^{(i)} = 2$ otherwise. We then recursively simulate future states: if $S_{t-1}^{(i)} = 1$ then $S_t^{(i)} = 1$ if $u_t^{(i)} \leq p$ otherwise set $S_t^{(i)} = 2$ and if $S_{t-1}^{(i)} = 2$ then set $S_t^{(i)} = 2$ if $u_t^{(i)} \leq q$ otherwise set $S_t^{(i)} = 1$. Values for p and q are selected as in Table 2.

Table 2 Rejection frequencies from the Diebold-Mariano (1995) statistic on MRS models: Two Sided Test

Model	1	2	3	4	5	6	7	8	9	10	11	12
p	0.8	0.8	0.9	0.9	0.95	0.95	0.9	0.9	0.95	0.95	0.9	0.9
q	0.8	0.8	0.9	0.9	0.95	0.95	0.9	0.9	0.95	0.95	0.95	0.95
μ_1	2	1	2	1	2	1	2	1	2	1	2	1
μ_2	-2	-1	-2	-1	-2	-1	-2	-1	-2	-1	-2	-1
σ_1^2	1	1	1	1	1	1	1	1	1	1	1	1
σ_2^2	1	1	1	1	1	1	2	2	2	2	1	1

T	A	Panel A: $\rho=0.5$											
250	0.05	0.002	0.055	0.001	0.021	0.002	0.016	0.004	0.033	0.005	0.025	0.000	0.006
	0.01	0.000	0.013	0.000	0.002	0.000	0.002	0.000	0.005	0.000	0.002	0.000	0.001
500	0.05	0.001	0.021	0.004	0.012	0.007	0.016	0.007	0.028	0.013	0.018	0.000	0.000
	0.01	0.000	0.003	0.000	0.001	0.000	0.002	0.000	0.002	0.001	0.000	0.000	0.000
1000	0.05	0.000	0.022	0.006	0.014	0.011	0.017	0.012	0.025	0.019	0.018	0.000	0.002
	0.01	0.000	0.003	0.000	0.000	0.000	0.002	0.001	0.002	0.001	0.001	0.000	0.000

Panel B: $\rho=0.9$													
250	0.05	0.001	0.034	0.005	0.016	0.001	0.020	0.004	0.024	0.008	0.020	0.000	0.003
	0.01	0.000	0.008	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.001	0.000	0.002
500	0.05	0.004	0.017	0.004	0.017	0.009	0.012	0.006	0.018	0.016	0.017	0.000	0.002
	0.01	0.000	0.005	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.003	0.000	0.000
1000	0.05	0.001	0.023	0.004	0.012	0.012	0.012	0.007	0.018	0.015	0.017	0.000	0.000
	0.01	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.000	0.000

The most striking result to take from this table is the poor finite sample performance of the DM statistic. Using the appropriate critical value from the standard normal distribution, even when we have 500 observations (a sample size that is particularly relevant as we have 538 observations in our empirical model) and have a very persistent leading indicator ($\rho=0.9$) we only reject between 0.4 percent and 1.8 percent of samples when the nominal size is 5 percent. A similar pattern emerges even when using $T=1000$ observations where, for the same nominal size of 5 percent, the rejection frequency varies from 0.1 percent to 2.3 percent. In fact at the five percent level we typically reject in less than one percent of the samples though we sometimes reject around 2 to three percent of samples. The general lesson to take from this experiment is that the DM statistic rejects far less frequently than the nominal size would suggest. We therefore believe that one should be very cautious about accepting the null hypothesis of equal forecast accuracy that is indicated by a small calculated

DM statistic relative to the standard normal distribution. Rather, we recommend evaluating the statistical significance of any QPS improvement using Monte Carlo methods.⁶

5 Do Leading Indicators Help Fit the Business Cycle?

A leading indicator such as XLI may be regarded as helping to explain the business cycle if the regime probability forecasts improve when the transition probabilities are allowed to vary according to the leading indicator. In this section we statistically test the QPS improvement of our TVTP MRS model of the US business cycle.

Earlier, we reported that we found that allowing the transition probabilities to vary with XLI improves the QPS statistic over the CTP model by a seemingly impressive 24%. Furthermore, the DM statistic is -2.34 which has a p -value of 1.86 percent and is statistically significant at the 5 percent but not the 1 percent level. Our simulation experiments above suggest that this p -value may be understated as we uncovered a tendency for the DM statistic to under-reject when applied to regime shifting models. For this reason we undertake two Monte Carlo experiments to compute the finite-sample size of this DM statistic. We also evaluate the power of the tests to identify leading indicators that help improve the ability of the models to capture the business cycle chronology.

5.1 Assessing QPS improvements using simulations

In order to conduct the experiments the first step is to estimate a model for XLI using the actual data on XLI. The dynamics of XLI are modeled as an AR(2)-GARCH(1,1) model with Student- t innovations. We settled on this model after considering a range of AR-GARCH models and as indicated by the Ljung-Box specification tests, it does an admirable job of whitening the residuals (see Table 3 for parameter estimates).

We now present two different approaches to evaluating the statistical significance or p -value by Monte Carlo methods. In our first experiment we generate 1000 artificial useless "leading indicators" which have the same time-series dynamics as XLI but contain no information about the business cycle. We simulate sample paths using the estimated parameters from the univariate AR(2)-GARCH-T time-series model reported in Table 3.⁷ Although the artificial leading indicators and XLI display the same time-series dynamics, the artificial data are randomly generated independent of the sample path of XLI and therefore have no systematic relationship to the US business cycle.

⁶ A similar picture emerges when using a one-sided test, though the under-rejection is slightly less pronounced.

⁷ When simulating each of the 1000 artificial series of the leading indicator we initialize the first two values of the series at their unconditional mean, simulate 1538 subsequent observations, and then drop the initial 1000 to avoid any contamination from the starting values selected.

Table 3 Parameter estimates for AR(2)-GARCH model for XLI

	Parameter	Robust SE
ϕ_0	0.3393	0.0725
ϕ_1	1.1424	0.0477
ϕ_2	-0.2452	0.0500
σ^2	1.2366	0.3323
α	0.0454	0.0526
β	0.9381	0.0906
ν	0.1086	0.0349
LL	-809.8408	
LB(z , 12)	13.0500	0.3654
LB(z^2 , 12)	18.3271	0.1061

Notes: The conditional mean is modeled as $\mu_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2}$ and the conditional variance is modeled as $h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1}$ where $e_t = y_t - \mu_t$, and the last term captures asymmetry in the conditional volatility specification. We report the unconditional volatility σ^2 , which is given by $\sigma^2 = \omega / (1 - \alpha - \beta)$, rather than ω directly.

The standardized residual $z_t = e_t / h_t^{0.5}$ is a standardized T random variable with ν^{-1} degrees of freedom.

LB(z , 12) is the Box-Ljung portmanteau test for autocorrelation in the standardized residual with 12 lags, and LB(z^2 , 12) tests for serial correlation in the squared standardized residuals.

For each artificial leading indicator we re-estimate the TVTP MRS model's parameters (using the actual XCI data in each case) and calculate the corresponding QPS – using the NBER dates. Because we use the NBER dates in the calculation of QPS this simulation experiment will provide the empirical distribution of the QPS statistic improvements – in comparison to that obtained using the CTP model for XCI – under the null that the leading indicator is of no value in relation to the NBER business cycle chronology. This experiment helps us gauge the likelihood that the improvement in the QPS statistic that we observe in the real data could have arisen purely by chance.

The resulting empirical distribution of the QPS improvement is provided visually in the upper plot in Figure 5 in the Appendix (labeled "Actual NBER Dates"). It is clear from the figure that, under the null hypothesis that XLI is without useful informational content for NBER phase changes, the probability of observing an improvement as large as the actual observed improvement of 0.0261 (or 24%), which is indicated by the vertical line, is effectively zero with a p -value of 0.001. The evidence from the DM statistic is much less compelling as the p -value was only 1.8 percent.

As an aside, we also calculated the DM statistic on each of these 1000 artificial samples. It will be recalled that the asymptotic distribution of the DM statistic under the null of no

predictive improvement is the standard normal. In fact, the asymptotic standard normal distribution provides a very poor approximation to the actual finite sample distribution in this experiment. The Bera-Jarque normality test finds evidence against normality at the 0.0015 level. The DM test under-rejects quite severely as it rejects the null hypothesis at the 5 percent level in less than 1 percent of samples. And this is despite using quite a large sample of 538 observations.

We also compute the p -value using an experiment in which we simulate both the coincident and leading indicators. In this case for each of 1000 replications we generate the leading indicator $X_t^{(i)}$ as above using the parameter estimates in Table 3. We also generate the regimes $S_t^{(i)}$ using the same approach as in Section 3 but using the parameter estimates from the CTP MRS model reported in Table 1. Finally, rather than using the actual XCI series, here we generate an artificial coincident index, $y_t^{(i)}$, using the relevant CTP model parameter estimates from Table 1. For each simulated triple $y_t^{(i)}, X_t^{(i)}, S_t^{(i)}$ we analyze the model fit by estimating both the CTP model – which is the correctly specified model – and a TVTP model which uses a meaningless leading indicator. The CTP model is estimated using only $y_t^{(i)}$, whereas we use both $y_t^{(i)}$ and $X_t^{(i)}$ when estimating the spurious TVTP model. The QPS statistics for both the CTP and TVTP models and the DM Z-statistic are computed using $S_t^{(i)}$ rather than the NBER dates (see bottom graph in Figure 5).

We find that the DM statistic in this experiment exhibits the same size distortion reported previously. In particular, at the 5 percent level, we reject the null hypothesis in only 0.87 percent of samples. The Bera-Jarque statistic for the distribution of the DM statistic across the simulations is 33.79 which is strongly significant with a p -value of only 4.6×10^{-8} indicating that the finite sample distribution (with nonetheless over 500 sample observations) of the DM statistic is very different from the asymptotic normal. Furthermore, our simulations suggest that the calculated p -value of the DM statistic of -2.34 from the real data is actually 0.001 (only 1 of the 1000 simulations were larger) which is very different from the asymptotic p -value of 1.8 percent. This again highlights the size distortions in the DM statistic when applied to MRS business cycle models.

5.2 Power

To understand the ability to detect improvements in model fit when the data are actually generated by varying transition probability models we simulate the leading indicators $X_t^{(i)}$ as above and, using these, we generate the regimes $S_t^{(i)}$ and coincident index $y_t^{(i)}$ from a TVTP model using the parameters reported in Table 1. Using these simulated data we generate Monte Carlo critical values of 0.0115 (1 percent) and 0.0202 (5 percent) which allow us to reject the null hypothesis of no improvement in 92.9 and 97.0 percent of samples. However, we only reject in 83.8 and 91.6 percent of samples using respectively the 1 and 5 percent critical value from the asymptotic distribution of the DM statistic. Both types of tests based on the QPS statistic have good power to detect varying transition probabilities implied in our model of business cycle dynamics. Yet the simulation-based critical values have higher power because of the poor size of the DM statistic using the asymptotic standard normal distribution.

6 Conclusions

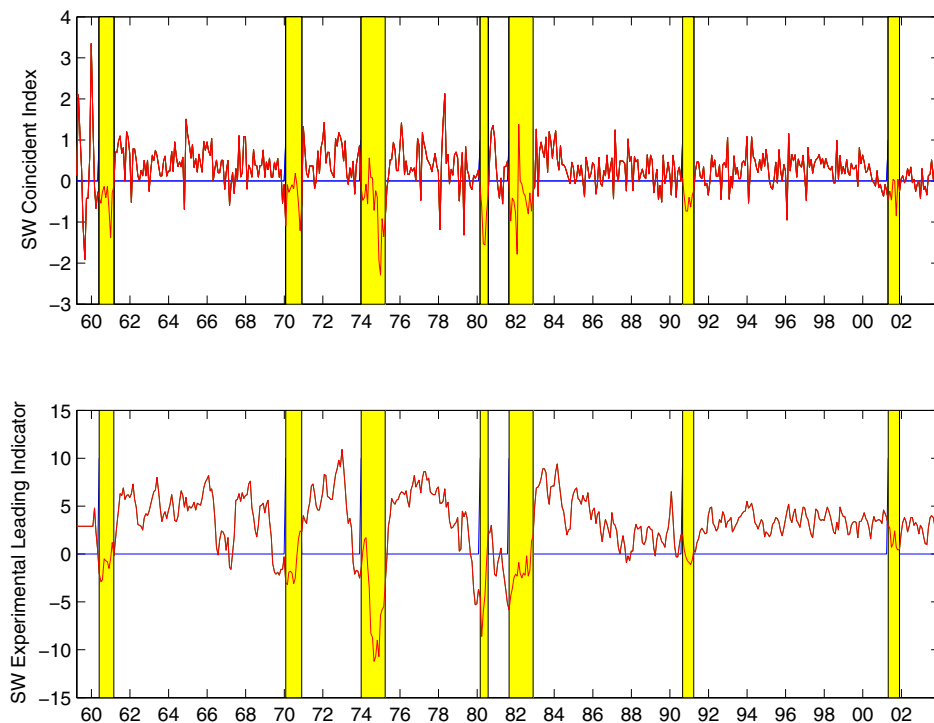
The use of the quadratic probability score to evaluate the extent to which a business cycle model replicates the official NBER business cycle chronology is now quite common in US empirical business cycle modeling. One important application of the QPS statistic is to measure the ability of leading indicators to improve the fit of regime switching business cycle models to the NBER chronology. The Diebold-Marino (DM) statistic is a powerful tool widely used to compare two different forecasts of some time-series variable. It is simple to calculate and does a satisfactory job in many empirical applications. However, we have demonstrated that the DM statistic is not well suited to evaluating changes in QPS statistics associated with competing business cycle models as the asymptotic normal distribution in this case is a very poor approximation to the true distribution. The DM statistic tends to under-reject even in samples of as many as 500 or 1000 observations. As an alternative, we present an empirical method to assess the statistical significance of the QPS statistic in the context of Markov regime switching (MRS) models of the business cycle. The method involves constructing the distribution of the QPS statistic by Monte Carlo simulations for the application at hand. Using this method we find that the experimental leading index of Stock and Watson provides statistically significantly improved forecasts of the NBER business cycle chronology for the US over and above what can be provided by Stock and Watson's experimental coincident index alone.

References

- Andrews, D.W.K. 1991. "Heterokedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, pp. 817-58.
- Brier G.W. 1950. "Verification of Forecasts Expressed in Terms of Probabilities," *Monthly Weather Review* 78, pp. 1-3.
- Campbell R. Harvey. 1989. "Forecasting Economic Growth with the Bond and Stock Markets," *Financial Analysts Journal* September/October, pp. 38-45.
- Chauvet, Marcelle. 1999. "Stock Market Fluctuations and the Business Cycle," Working paper, UC Riverside.
- Chen, Max. 2002. "The Use of the Composite Leading Index for Forecasting Business Cycle Turning Points," *Annual Meeting of the American Statistical Association, Joint Statistical Meetings - Business & Economic Statistics Section*, pp. 509-18.
- Diebold, Francis X. and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, pp. 253-63.
- Diebold, F.X. and R. Rudebusch. 1989. "Scoring the Leading Indicators," *Journal of Business* 62, pp. 369-402.
- Diebold, F.X.; J.-H. Lee and G. Weinbach. 1994. "Regime Switching with Time-Varying Transition Probabilities," in *Nonstationary Time Series Analysis and Cointegration*, C. Hargreaves, ed. Oxford: Oxford University Press, pp. 283-302. (Advanced Texts in *Econometrics*, C.W.J. Granger and G. Mizon, eds.).
- Filardo, Andrew J. 1994. "Business Cycle Phases and Their transitional Dynamics," *Journal of Business and Economic Statistics* 12, pp. 299-308.
- Filardo, Andrew J and Stephen F. Gordon. 1998. "Business Cycle Durations," *Journal of Econometrics* 85, pp. 99-123.
- Gray, Stephen. 1996. "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics* 42, pp. 27-62.
- Hamilton, James. 1989. "A New Approach to the Analysis of Time Series and the Business Cycle," *Econometrica* 57, pp. 357-84.
- Kholodilin, Konstantin A. and Vincent W. Yao. 2005. "Measuring and Predicting Turning Points Using a Dynamic Bi-Factor Model," *International Journal of Forecasting* 21, pp. 525-37.
- Lahiri, Kajal and Jiazhao G. Wang. 1994. "Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model," *Journal of Forecasting* 13, pp. 245-63.
- Layton, Allan P. 1998. "Further on the Influence of Leading Indicators on the Probability of US Business Cycle Phase Shifts," *International Journal of Forecasting* 14, pp. 63-70.
- Layton, Allan P. and Daniel R. Smith. 2000. "A Further Note on the Three Phases of the US Business Cycle," *Applied Economics* 32, pp. 1133-43.
- Newey, W. and K.D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, pp. 703-8.
- Stock, James H. and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Indicators," *NBER Macroeconomics Annual*, pp. 351-93.

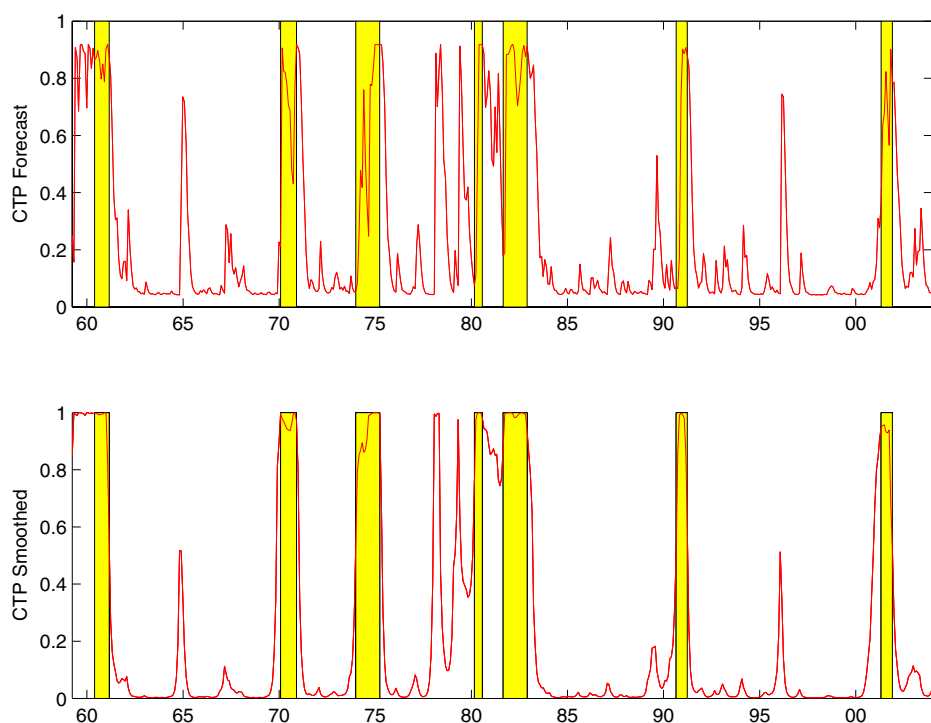
Appendix

Figure 1 Underlying data



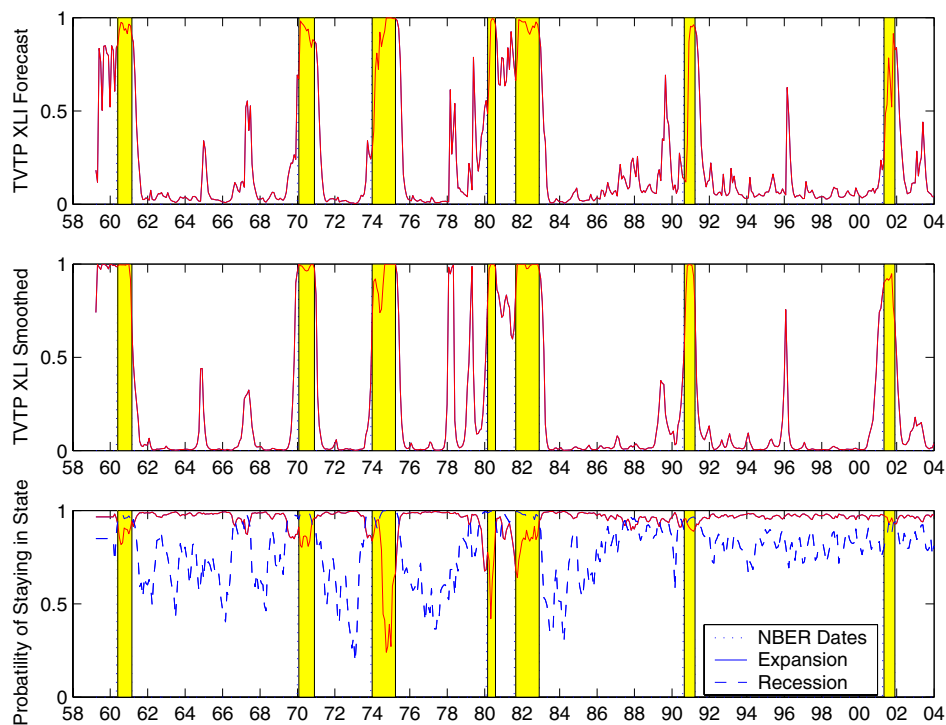
Notes: The shaded areas denote recessions as identified by the NBER.
 Upper panel: Month-on-month growth rates of the Stock-Watson experimental coincident index (XCI)
 Lower panel: Previous month's value of the experimental leading indicator (XLI).

Figure 2 **Forecast and smoothed probabilities of a recession from MRS model with constant transition probability (CTP)**



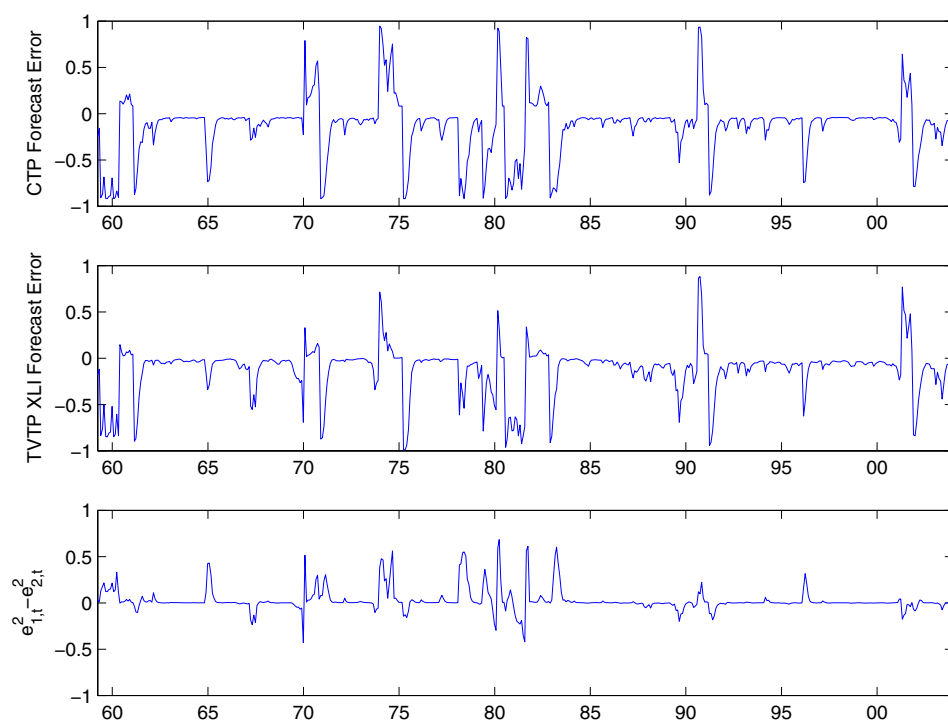
Note: The shaded areas denote recessions as identified by the NBER.

Figure 3 Forecast and smoothed probabilities of recessions from MRS model with time-varying transition probabilities (TVTP) using Stock and Watson's experimental leading indicator (XLI)



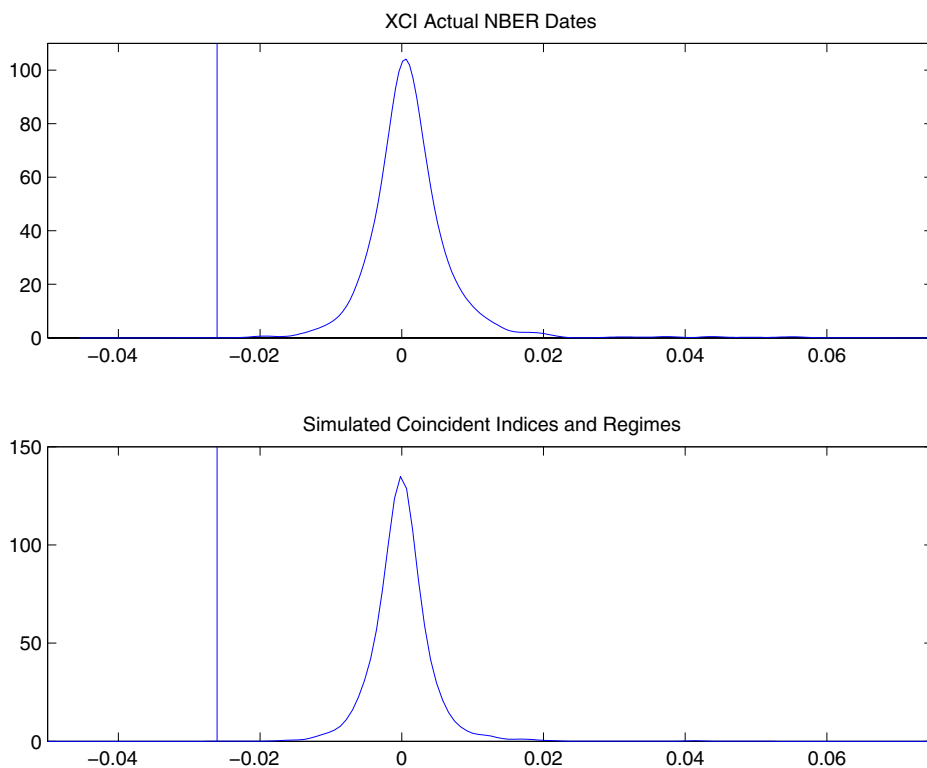
Note: The shaded areas denote recessions as identified by the NBER.

Figure 4 **Forecast error (difference between the recession dummy and the forecast probability of a recession next period)**



Notes: Upper panel: Constant transition probability (CTP) model
 Middle panel: Time-varying transition probability (TVTP) model using the experimental leading indicator (XLI)
 Lower panel: Difference between the two squared forecast errors

Figure 5 **Plot of QPS improvement from TVTP when models estimated for artificially generated series for XLI**



Notes: The figure provides a Kernel-based nonparametric estimate of the density of the difference between the QPS from a CTP and TVTP MRS models where the leading indicators have no informational content as far as XCI is concerned. The upper panel uses XCI and actual NBER dates with artificial leading indicators. The lower panel uses artificial leading indicators, regimes and coincident indices. The vertical lines represent the actual improvement we found using XLI.

Report

Improving Timeliness for Short-Term Economic Statistics

Richard McKenzie*

Abstract

Effective business cycle analysis, and indeed the monitoring of a countries economic performance from a policy perspective, requires access to timely high quality short-term economic statistics (STES). Consequently in recent years there has been a lot of pressure on national statistics organisations (NSOs) to better serve their users by improving the timeliness of release for their short-term economic indicators. In response to this demand, NSOs have focused on improving the efficiency and methodology of their statistical production processes. So this begs the question: where would one look to find comprehensive documentation on good practices used by NSOs to improve the timeliness of their short-term economic statistics? The answer is the *STES Timeliness Framework*, a structured collection of documentation on a range of *good practices* currently used by NSOs for improving timeliness, reducing costs or improving accuracy for short-term economic statistics. This resource is freely available in the form of an intuitive, user friendly website developed by the OECD Short-Term Economic Statistics Expert Group at www.oecd.org/std/research/timeliness. This report outlines the principles behind the development of this framework, explains its structure and reviews its current usage by statisticians.

Key Words: Timeliness, Reducing Cost, Short Term Economic Statistics

* OECD Statistics Directorate, Short-Term Economics Statistics Division; richard.mckenzie@oecd.org

Résumé

Une analyse efficace du cycle économique, et en fait le contrôle de la performance économique d'un pays en terme de politique nécessite un accès, en temps utile, à des statistiques économiques à court terme de grande qualité, par conséquent, ces dernières années il y a eu beaucoup de pression sur les Offices Nationaux de Statistiques afin qu'ils soient plus utiles à leurs utilisateurs en améliorant les délais de parution de leurs indicateurs économiques à court terme. En réponse à cette demande les Offices Nationaux de Statistiques se sont attachées à améliorer l'efficacité ainsi que la méthodologie de leurs processus de production statistique. Et cela pose la question : où chercherait on si l'on voulait trouver de la documentation complète sur les pratiques effectives utilisées par les Offices Nationaux de Statistiques afin d'améliorer les délais de leurs statistiques économiques à court terme ? La réponse est : dans le Timeliness framework des statistiques économiques à court terme (cadre de travail sur l'amélioration des délais), un ensemble structuré de documentation sur l'ensemble des pratiques effectives actuellement utilisées par les Offices Nationaux de Statistiques afin d'améliorer les délais, réduire les coûts ou augmenter la qualité des statistiques économiques à court terme. Cet outil est disponible librement sous la forme d'un site Internet intuitif et convivial développé par le groupe d'experts des statistiques économiques à court terme de l'OCDE au www.oecd.org/std/research/timeliness. Ce rapport expose les principes à l'origine du développement de ce cadre de travail, explique sa structure et révisé son utilisation actuelle par les statisticiens.

1 Introduction

Improving the timeliness of short-term economic statistics (STES) is a topic which has attracted enormous interest within the international statistical community in recent times. For example, it has been a key focus for the European Union (EU) with work being undertaken on a wide range of initiatives to achieve the timeliness targets set out in the EMU action plan¹. It is also an important element of the International Monetary Fund's (IMF) SDDS².

To support these initiatives, there has been numerous meetings and conferences organised to facilitate information sharing; attempts to identify and distribute best practices; the formation of methodological expert groups to provide recommendations; etc. Indeed these types of activities date back to a 1972 meeting of the United Nations Statistical Commission, and the twenty third Plenary Session of the Conference of European Statisticians in 1975 (Ryten, 1997).

Despite the importance of the topic, and the large amount of work undertaken in the international arena, a comprehensive framework to assist Statisticians in making informed choices on the best methods to employ within different stages of the statistical production process to improve timeliness had never been developed. Consequently a key outcome of the June 2003 meeting of the OECD Short-Term Economic Statistics Expert Group (STESEG) was the formation of a multi-country task force to establish such a framework as a tool to assist NSOs to improve the timeliness of their short-term economic statistics, with due regards to minimizing any potential adverse affects on accuracy.

This paper outlines the process used by the STESEG task force for developing the Short-Term Economic Statistics Timeliness Framework, describing its purpose, how it is intended to be used and updated over time, and summarises recent data on its usage recorded by OECD web trends statistics. The framework was first released on the OECD website in September 2004 at www.oecd.org/std/research/timeliness as a structured collection of documentation on a range of *good practices* currently used by NSOs for improving timeliness, reducing costs or improving accuracy³ in the production of STES, for those indicators produced from business surveys and/or administrative records of businesses. The framework underwent its first annual update in September 2005 with the addition of new documentation reflecting developments in this field over the period May 2004 – July 2005.

¹ The European Monetary Union (EMU) action plan outlines a strategy for improvement of the timeliness and quality of European short-term economic statistics.

² The IMF's Special Data Dissemination Standard (SDDS) specifies requirements for the coverage, periodicity and timeliness of principal economic indicators, and for the dissemination of an advance release calendar.

³ The STES Timeliness Framework only covers practices for improving accuracy if it is clear that the method to improve accuracy could be traded off to either improve timeliness and / or reduce costs of production.

2 Development of the STES Timeliness Framework

A number of issues were taken into account when developing the *STES Timeliness Framework*. Firstly the framework was intended to promote good practices for improving the timeliness of short-term economic statistics but in order to do this related issues such as accuracy and cost needed to be considered. It was intended for the framework to cover all aspects of the statistical production process but the interrelated nature of these processes needed to be taken into account. Perhaps most importantly, for the framework to be a useful tool to NSOs it needed to be practically focused and applicable to all countries who we know operate under a variety of different circumstances. This section of the paper describes in detail how the above aspects, and many others, were taken into account in developing the *STES Timeliness Framework*.

2.1 *Timeliness as a quality indicator and its relationship to cost and accuracy*

Timeliness is an important indicator of quality for STES, and is recognised as such in statistical quality frameworks such as those developed by the OECD (2003) and Statistics Canada (2002). Other relevant quality indicators (OECD, 2003) are accuracy, relevance, accessibility, coherence, interpretability and credibility. It has often been suggested (Ryten, 1997; Gnoss, 2003) that significant improvements in timeliness cannot be achieved without a loss in accuracy. Oberg (2002) argues this need not be the case, citing examples where some countries appear to be both more timely and more accurate due to their efficiency of operational procedures. This assertion is supported by Ahnert (2001) and Walkner (2001) who found that countries with more timely first releases of STES did not necessarily have larger revisions.

It is obvious that to improve timeliness in STES without a significant loss in accuracy some change to operational procedures or methodological techniques is required. Indeed timeliness can be considered a characteristic of the whole statistical production process (SPC Expert Group, 2002) for STES. Therefore, a key purpose in designing the *STES Timeliness Framework* was for it to be a tool to assist Statisticians in making choices on methods to use within each stage of the statistical production process, which enables the timeliest estimates to be produced with minimal or no losses in accuracy. Alternatively, methods which reduce survey costs for the same timeliness and accuracy could be seen as relevant given this may provide the opportunity to spend more resources directed at improving timeliness through modifying other aspects of the statistical production process.

The issue of user requirements is also particularly relevant, as estimates produced to an earlier time frame and released only at aggregate levels could serve different user needs in comparison to more detailed statistics produced at a later time frame (Oberg, 2002). Of course minimising revisions between different releases for the same period remains an important issue. Consequently the *STES Timeliness Framework* aimed to cover aspects such

as producing preliminary estimates from sub-samples but also how analysis of revisions can lead to the improvement of preliminary estimation methodologies.

2.2 *Aspects of the statistical production process to be covered*

As discussed in section 2.1 it was deemed necessary that the *STES Timeliness Framework* should cover all aspects of the statistical production process. This required determining the processes involved in the production of short-term economic statistics and grouping them in a logical manner. It was also important that the detailed production processes referenced could be related to the issue of timeliness either directly or indirectly through the relationship between timeliness, cost and accuracy.

The starting reference for listing and grouping aspects of the statistical production process came from the "Checklist for Sample Surveys – Basic Survey Design course" (ABS, 1996). This checklist was organised into a two dimensional framework and refined extensively based on contributions from members of the task force and feedback from the 2003 and 2004 STESEG meetings where development of the *STES Timeliness Framework* was discussed. The final categories included in the framework are shown below in Table 1.

Having established a structure for the framework, the major task was to collect relevant documentation on good practices for improving timeliness, reducing cost or improving accuracy⁴ related to the processes included in the second column of the table. Of course the actual categories included in the *STES Timeliness Framework* were refined over the course of its development in conjunction with the collection of relevant documentation, through the processes described in section 3. Also, because different aspects of the statistical production process are interrelated this needed to be taken into account in the presentation of *good practices* for the various aspects of the statistical production process. That is, a *good practice* may encompass activities related to several of the aspects which appear in the framework reference table. This interaction between elements of the statistical production process was taken into account in the final presentation of the *STES Timeliness Framework* through the use of primary and secondary categories of papers for each separate activity as described in Section 4 (third paragraph).

⁴ The *STES Timeliness Framework* only covers practices for improving accuracy if it is clear that the method to improve accuracy could be traded off to either improve timeliness and / or reduce costs of production.

Table 1 **Final structure of the STES Timeliness Framework reference table**

Broad statistical process	Operational methods related to improving timeliness, reducing costs or improving accuracy for short-term economic statistics
1 Frame selection and administrative data use	1.1 Direct Use of Administrative Data for Short-Term Statistics 1.2 Maintaining Survey Frames from Administrative Data Sources
2 Questionnaire design	2.1 Flexible Questionnaire and Form Design 2.2 Minimising Data Items Collected
3 Sample design and selection	3.1 Efficient Sample Designs 3.2 Sample Designs for Sub Samples 3.3 Effective Business Selection Methods
4 Reference period & due dates	4.1 Choice of Reference Periods 4.2 Due Date for Businesses Providing Data
5 Data collection and validation (editing)	5.1 Efficient Data Collection Methods Using Technology 5.2 Follow Up of Non Respondents 5.3 Automated Data Validation (Editing) Methods 5.4 Selective (or Significance) Editing 5.5 Effectively Combining Technology and Staff Resources 5.6 Relationships with Respondents
6 Estimation	6.1 Data Imputation Methods 6.2 Preliminary Estimates from Sub Samples 6.3 Estimation Based on Lower Response Rates 6.4 Preliminary Estimates Based on Statistical Models
7 Evaluation & dissemination	7.1 Assessment of User Requirements 7.2 Analysis of Revisions 7.3 Quality Assessment of Timeliness Improvements 7.4 Use of Benchmarking Techniques 7.5 Efficient Dissemination Processes

2.3 *The need to provide a range of good practices for improving timeliness*

A common theme of recommendations by various past working groups on the issue of improving timeliness for STES has been to identify and agree on best practice, and benchmark current procedures against agreed best practice (e.g. SPC Expert Group, 2002). However, in developing the *STES Timeliness Framework* the STESEG task force considered this ideal to be very difficult to achieve for a number of reasons, including:

- what is best practice and who decides it? Would best practice for one country be best practice for another?
- would best practice apply to specific indicators, a group of indicators, or all STES (e.g. best practice procedures for the production of the producer price index could differ substantially to best practice procedures for the production of retail trade statistics);
- user requirements may differ between countries;
- differences in legislation between countries may mean that application of certain processes are not possible or unlikely to be successful (e.g. whether surveys are voluntary or compulsory, the existence of legislative requirements to use administrative data to reduce response burden, etc.);
- differences in economic structures between countries may mean that application of certain processes are not possible (e.g. the impact of predominantly monthly payrolls in Europe compared to predominantly weekly payrolls in the U.S);
- many countries may not have the resources or authority (i.e. depending on their institutional arrangements) to completely reorganise their production processes.

Consequently the task force agreed that the identification of best practice for improving timeliness for the entire statistical production process for STES was not possible due to the different exogenous influences facing the producers of STES in different countries. However, it seemed feasible that identifying a range of *good practices* for particular aspects of the statistical production process which may differ due to different initial conditions prevailing within the countries they were applied to could form an information resource applicable to a much wider group of countries.

Therefore a key purpose in developing the *STES Timeliness Framework* was to provide NSOs with a range of options for improving timeliness within different areas of the statistical production process to enable them to assess which options best suit their particular circumstances. For example, an NSO may have to consider such issues as: their capacity to fund change; the legislative circumstances within which they operate; which options are likely to have the maximum benefit on timeliness at lowest cost or impact on accuracy; which options best fit with their existing production processes and external circumstances; which options best meet the priorities of their users; etc..

The ideal goal of the framework is to provide an opportunity for NSOs to closely examine good practices used within other countries across the statistical production process and enable them to identify various techniques that could be applicable within their own statistical environment. This should allow NSOs to more readily evaluate different options for implementing new methods to improve timeliness or reduce the production costs of their STES. As such, the task force had to ensure that documentation on good practices was collected from a wide range of countries to be diverse enough to be of use to all NSOs that operate under a variety of different circumstances. The fact that documentation in the most recent update of the framework is sourced from 22 different countries and international organisations should go a long way to achieving this goal.

2.4 Ensuring a practical focus

A key objective in developing the *STES Timeliness Framework* was for it to have a practical focus. Therefore documentation referenced in the framework generally describes techniques already implemented within statistical organisations which enable STES to be produced in a very timely manner, or have enabled recent improvements in timeliness or reduced costs of production to be achieved. In addition, papers based on sound research with associated implementation plans have also been included. However purely theoretical papers or papers presenting future broad strategies were not within scope of the framework. Ideally, each paper referenced in the framework should give the reader ideas on possible changes or improvements they could make to one or several aspects of the statistical production process that could improve timeliness and / or reduce costs for the STES they produce.

3 Collection and Assessment of Documentation for Inclusion in the Framework

There were two main approaches used by the task force to collect documentation for inclusion in the framework. The first consisted of a literature search of statistical journals, international conferences and workshop proceedings which have covered issues relating to STES. The second involved a formal request to the STESEG representative of each OECD country to identify and coordinate the collection of relevant existing documentation from within their organisation. This documentation was then assessed by the task force to determine if it was suitable for inclusion in the framework.

3.1 Process of assessing documentation for inclusion in the framework

The assessment process consisted of reviewing documents provided by NSOs or those found through the literature search. Relevant documents submitted by NSOs could consist of internal documentation on methods or final project proposals for the implementation of new

methods, as well as other relevant papers describing processes, provided they satisfied either criteria 1, 2 or 3 below and criteria 4:

1. The document describes a process which enables STES to be produced in a very timely manner, or has enabled recent improvements in timeliness or reduced costs of production to be achieved.
2. The document explains a new method for improving timeliness which has been thoroughly evaluated and clear steps for implementation within the statistical production process are provided.
3. The document describes a method to improve accuracy or reduce costs which has an apparent relationship to timeliness such that it could be deemed to indirectly satisfy either criteria 1 or 2.
4. The techniques described in the document were regarded as methodological sound by the task force such that they could be classified as *good practice*.

In addition, the framework was restricted to cover processes for the important short-term economic statistics listed in Table 2, which are produced from business surveys and/or administrative records of businesses. A more detailed description of the definition of the short-term economic statistics included in this table is given in Appendix 3.

Table 2 Range of short-term economic statistics covered by the STES Timeliness Framework

Indicator	Common periodicity
Industrial production index	Monthly
Manufacturing output / sales	Monthly/quarterly
Construction production / output / sales	Monthly/quarterly
Retail trade, current price and volume	Monthly
Turnover for other service industries	Quarterly
Unemployment rate (not from household labour force survey)	Monthly
Total employment (not from household labour force survey)	Monthly/quarterly
Job vacancy rate	Monthly/quarterly
Hourly / weekly / monthly wages	Monthly / Quarterly
Labour costs / prices	Quarterly

3.2 *Papers sourced from journals / conferences*

Noting the assessment criteria outlined in Section 3.1, a list of potentially relevant statistical journals and statistical conference / workshop proceedings were identified. Over 1200 papers were reviewed for inclusion in the framework from 9 different statistical journals and the proceedings from 22 statistical conferences / workshops (16 in the initial development of the framework for the September 2004 release and an addition 6 for the September 2005 update). Of the papers which related to the issue of reducing costs or improving the timeliness of STES (i.e. a much smaller subset of papers), the content was often theoretical or did not have clear operational applicability. However some sources did contain highly relevant material and 50 papers from 16 different sources were chosen for inclusion in the framework (31 papers from 11 sources in the initial September 2004 release and a further 19 papers from 5 new sources in the September 2005 update). All journals and conference proceedings that were reviewed are listed in Appendix 1, together with the number of papers included in the framework from each source.

3.3 *Papers provided by national statistical organisations*

An official request was made to NSOs asking them to identify and provide relevant existing documentation for possible inclusion in the STES Timeliness Framework for both its initial development in 2004 and subsequent update in 2005. A detailed set of guidelines to aid NSOs in determining whether documentation was appropriate to send to the task force to be considered for inclusion in the framework was provided⁵. These requests were generally well responded to and many NSOs went to considerable effort to provide documentation which was appreciated and acknowledged by the task force. In summary, the request was made to 27 OECD Member countries of which 25 undertook some research to identify relevant documentation. For the initial request in 2004 a total of 129 papers from 21 countries were submitted to the task force to assess for possible inclusion in the framework. Of these, 46 papers from 18 different Member countries⁶ were deemed suitable and were included in the initial version of the STES Timeliness Framework. For the 2005 update 33 papers from 14 organisations were submitted for assessment of which 13 were included in the September 2005 update of the framework.

There was naturally some overlap between the papers submitted from NSOs and those obtained from the literature search of statistics journals and conferences. This was somewhat reassuring in regards to the consistency of the processes applied. In total, 22 different countries or international organisations have contributed documentation to the *STES Timeliness Framework*, and these contributions by country / organization can be identified at http://www.oecd.org/document/7/0,2340,en_2649_201185_33633799_1_1_1_1,00.html.

⁵ see Appendix 2 or <http://www.oecd.org/dataoecd/13/6/33630498.pdf>

⁶ Papers were accepted in national language provided there was an abstract in English. Some papers provided in national language were then translated to English by Eurostat. As a result, over 95% of the papers in the current version of the framework are in English.

4 Accessing and Using the *STES Timeliness Framework*

The *STES Timeliness Framework* can be accessed on the OECD website at: www.oecd.org/std/research/timeliness. The framework is presented within the structure shown in Table 3 below, which outlines the process for producing short-term economic statistics within broad categories. The headings in column 2 are statistical processes within which a range of techniques can be applied to improve timeliness or reduce costs. A user then clicks on a heading to access more detailed information on these techniques based on the research undertaken in developing the framework.

Each of the web pages accessed through the column 2 headings contain a brief summary of the main techniques used to improve timeliness or reduce costs through this statistical process, followed by links to detailed papers. Thus the framework demonstrates the opportunities that exist for improving timeliness or reducing costs within each stage of the statistical production process. All detailed papers contain abstracts or introductions briefly summarising the content of the paper, either within the document or in a separate web page which appears before accessing the document.

Because many of the documents included within the framework provide information on more than one aspect of the statistical production process as defined by the column 2 headings, a particular paper can be referenced under several headings. When the papers were assessed for inclusion in the framework, task force members noted the column 2 heading (i.e. statistical process) that the paper most related to, and other column 2 headings it also contained information on. Consequently a user will see two categories of papers within each column 2 heading web page which reflects this distinction; these are the *primary papers* and *secondary papers* headings. This aspect of the framework reflects the fact that statistical production processes are generally integrated within an overall strategy. Papers which covered an extremely varied range of practices were also included under heading *8.1 Papers Covering Several Methods to Improve Timeliness*. In addition the heading *8.2 All Papers Sorted by Country* provides access to all documents included in the framework listed by country of origin.

The chosen structure for presenting the framework was agreed by the taskforce after careful consideration and several revisions to draft proposals. Its principle aims are to provide summary information on a range of topics at a glance, with related detailed information being clearly identifiable and easy to access. Therefore it could be considered as a general model for the presentation of good statistical practices on the internet, which may be applicable to other statistical working groups undertaking similar initiatives.

Table 3 **The STES Timeliness Framework as it appears on the internet**

Broad statistical process	Links to documentation on proven operational methods to improve timeliness or reduce costs for short-term economic statistics
1 Frame selection and administrative data use	1.1 Direct Use of Administrative Data for Short-Term Statistics 1.2 Maintaining Survey Frames from Administrative Data Sources
2 Questionnaire design	2.1 Flexible Questionnaire and Form Design 2.2 Minimising Data Items Collected
3 Sample design and selection	3.1 Efficient Sample Designs 3.2 Sample Designs for Sub Samples 3.3 Effective Business Selection Methods
4 Reference period & due dates	4.1 Choice of Reference Periods 4.2 Due Date for Businesses Providing Data
5 Data collection and validation (editing)	5.1 Efficient Data Collection Methods Using Technology 5.2 Follow Up of Non Respondents 5.3 Automated Data Validation (Editing) Methods 5.4 Selective (or Significance) Editing 5.5 Effectively Combining Technology and Staff Resources 5.6 Relationships with Respondents
6 Estimation	6.1 Data Imputation Methods 6.2 Preliminary Estimates from Sub Samples 6.3 Estimation Based on Lower Response Rates 6.4 Preliminary Estimates Based on Statistical Models
7 Evaluation & dissemination	7.1 Assessment of User Requirements 7.2 Analysis of Revisions 7.3 Quality Assessment of Timeliness Improvements 7.4 Use of Benchmarking Techniques 7.5 Efficient Dissemination Processes
8 Overview	8.1 Papers Covering Several Methods to Improve Timeliness 8.2 All Papers Sorted by Country

5 Usefulness of the STES Timeliness Framework to NSOs

As the contents of the *STES Timeliness Framework* have been sourced from 22 countries and covers issues for the entire statistical production process it should be diverse enough to be of use to all NSOs that operate under a variety of different circumstances. Both the framework and the process of submitting documentation to be assessed for inclusion in the framework provides NSOs with the opportunity to:

- gain wider recognition for their work;
- access information about other countries practices extremely relevant to their daily work which would otherwise be very difficult to obtain, as this type of information is rarely published in journals etc.;
- closely examine good practices in countries most likely to be similar to their own;
- quickly evaluate different options for implementing new methods to improve the timeliness or reduce the costs of production for their short-term economic statistics with greater certainty of the likely outcomes.

The STES Timeliness Framework is also a useful tool for OECD Non-Member countries currently developing their statistical systems, and as an aid for use within bilateral programs at country level (i.e. where a country provides aid to another on developing the statistical capacity).

5.1 Usage of the STES Timeliness Framework

The usage of the *STES Timeliness Framework* can be partially assessed through OECD web trends statistics (see below). In addition the framework home page provides the opportunity for people / institutions to register as an official user to ensure they are informed of all future updates to the framework. An interesting statistic is that 12 registrations from institutions of non OECD member countries have been received which indicates that it is indeed a useful tool for countries currently developing their statistical systems.

Since its official launch in September 2004, the *STES Timeliness Framework* has proved to be very popular as measured by the OECD web trends statistics. Not only was access to the framework home page high in the launch month of September 2004 which recorded 862 visits, but it has continued to be one of the most popular OECD statistics web pages with the framework's home page averaging over 330 visits per month. This popularity no doubt reflects both the user friendly interface and extensive range of information included in the *STES Timeliness Framework*.

A power point training seminar has also been developed which can be made available to NSOs who wish to give a demonstration of the *STES Timeliness Framework* to their staff. So far five NSOs have requested this training package which can be obtained by emailing richard.mckenzie@oecd.org.

6 Future Updates

The OECD Short-Term Economic Statistics Expert Group has a mandate to continue updating the *STES Timeliness Framework* on an annual basis provided there is sufficient evidence of its ongoing use. As such the framework is promoted with statistical networks in several different ways to ensure as wide an audience as possible is aware of its existence.

References

- Ahnert, H. 2001. "Revision analysis for selected Euro area and EU country short-term indicators, Short-term statistics: improving timeliness and cooperation," Eurostat, proceedings from 2001 conference.
- Australian Bureau of Statistics. 1996. "Checklist for Sample Surveys, Basic Survey Design course notes," Commonwealth of Australia, Appendix 1.
- Gnoss, R. 2003. "Economic Statistics, Accuracy, Timeliness & Relevance," ISI – Eurostat – BEA conference proceedings, Government Printing Office, Washington pp. 48-50.
- McKenzie, R., 2003. "Development of a framework to assist in improving timeliness for short-term economic statistics," OECD, Paris, 2nd annual STESEG meeting, June.
- McKenzie, R., 2004. "The Short-Term Economic Statistics Timeliness Framework," OECD, Paris, 3rd Annual STESEG meeting, June.
- OECD. 2003. *OECD Quality Framework*. Paris.
- Oberg, S. 2002. "Quality and timeliness of Statistics; is it really a trade-off?" DGINS conference, Paper 3.2, 88th.
- Ryten, J. 1997. "Timeliness and Reliability: A Necessary Trade-off. Economic Statistics, Accuracy, Timeliness & Relevance," ISI – Eurostat – BEA conference proceedings, Government Printing Office, Washington.
- SPC Expert Group. 2002. "Sampling for timely European indicators – Final report," Item 9 of the 47th meeting of the Statistical Programme Committee.
- Statistics Canada. 2002. *Statistics Canada's Quality Assurance Framework*. Ottawa.
- Walkner, C. 2001. "Why US business indicators are faster available than those from the EU – some reflections and a comparison of practices, Short-term statistics: improving timeliness and cooperation," Eurostat, proceedings from 2001 conference.

Appendix 1

Journals and statistical conference proceedings reviewed

Journals reviewed

1. Journal of Official Statistics, 1992 – 2005. 5 papers included.
2. Survey of Current Business, 1994 – 2004. No relevant papers.
3. ASA, Journal of Business & Economic Statistics, 2001 – 2004. Mostly theoretical, nothing relevant to timeliness.
4. ILO, Bulletin of Labour Statistics. Focus is on concepts and methodology, nothing relevant to timeliness.
5. Statistics Canada, Survey Methodology, 1997 – 2003. Mostly theoretical, nothing relevant to timeliness.
6. Research in Official Statistics, Eurostat, 1999 – 2002. Mostly theoretical, nothing relevant to timeliness.
7. Review of Economics and Statistics, 2002 – 2003. Mostly econometric user based and theoretical.
8. Journal of forecasting, 2001 – 2003. Not related to the production of official statistics.
9. Agency Methodology Reports from BLS, Census Bureau & BEA from 2000 – 2004. Only relevant papers had already been submitted by these organizations for the framework.

Conferences / workshops / websites reviewed

1. OECD STESEG 2003. 8 papers included
2. UNSC, Work Session on Statistical Data Editing, 2003. 5 papers included.
3. International Conference of Establishment Surveys 2 (ICES2), 2000. 5 papers included.
4. Eurostat 2001. Short-term statistics: improving timeliness and cooperation. 4 papers included
5. OECD STESEG 2002. 2 papers included.
6. EU / US benchmark study on infa-annual statistics, 2001. 2 papers included.
7. ISI 2003. 1 paper included
8. Quality in Official Statistics, Stockholm 2001. 1 paper included.
9. International Conference on Improving Surveys, August 2002. 1 paper included

10. OECD/Eurostat Workshop on benchmarking techniques, 2004. A link to the proceedings is referenced which contains several papers and presentations on benchmarking techniques.
11. CES Plenary sessions, 2003. Substantive topics of the conference (e.g. confidentiality, globalisation) were not specifically relevant to timeliness.
12. International Conference on Questionnaire Development, Evaluation, and Testing Methods, 2002. Some papers were considered but were thought not to be relevant enough.
13. Baltic-Nordic Conference on Survey Sampling, Sweden 2002. Most papers of a technical nature.
14. IMF SDDS / DQAF. Reports against set of criteria which only briefly cover timeliness or related issues. Focus is on compliance with a specified target standard rather than listing of methods used to achieve this.
15. International Association of Official Statisticians conferences. Includes ISI 2001 which was reviewed but no suitable papers found, other recent conferences have been focused on specific topics not strongly related enough to specific issues for timeliness, e.g. New Economy, Information Society Statistics etc.
16. Federal Committee on Statistical Methodology Conferences (2001 & 2003: 2 new papers added, some others already included)
17. European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 May 2004. 2 papers included.
18. Data Quality for International Organisations, Wiesbaden, Germany, 27-28 May 2004. Contents were not relevant as they did not relate closely enough to the production of national statistics.
19. 55th Session of the International Statistical Institute (ISI), Sydney, Australia, 5-12 April 2005. 5 papers were included, these were also provided in the NSO submissions.
20. OECD / Eurostat Workshop on Frontiers in Benchmarking Techniques and Their Application to Official Statistics, Luxembourg, 7-8 April 2005. 1 paper included and reference made to proceedings.
21. UNECE, Work Session of Statistical Data Editing, Ottawa, Canada, 16-18 May 2005. Very relevant conference from which 7 papers were included. There was also significant overlap with papers provided by NSOs.
22. UNECE Seminar on Improved Data Reporting, June 13 2005. 3 papers included.

Appendix 2

Guidelines provided to NSOs for the submission of documentation

The Short-Term Economic Statistics (STES) *Timeliness Framework* is a structured collection of documentation on a range of *good practices* currently used by national statistical organisations for improving timeliness, reducing costs or improving accuracy⁷ in the production of STES (see Appendix 3 for the list of in scope STES).

The documentation contained within the current version of the *STES Timeliness Framework*, available on the OECD website at: www.oecd.org/std/research/timeliness, is sourced from 18 OECD countries. The submission of new documentation is now invited for inclusion in forthcoming updates of the framework. Contributing to the *STES Timeliness Framework* provides national statistical organisations with the opportunity to:

- gain wider recognition for their work;
- access information about other countries practices extremely relevant to their daily work which would otherwise be very difficult to obtain, as this type of information is rarely published in journals etc.;
- closely examine good practises in countries most likely to be similar to their own;
- quickly evaluate different options for implementing new methods to improve the timeliness of their short-term economic statistics with greater certainty of the likely outcomes.

Guidelines for the submission of documentation

Documentation included in the *STES Timeliness Framework* must have a practical focus. In general, documentation should describe techniques already implemented within statistical organisations which enable STES to be produced in a very timely manner, or have enabled recent improvements in timeliness or reduced costs of production to be achieved. For example, relevant documents may include internal documentation on methods or final project proposals for the implementation of new methods, as well as other relevant papers describing processes.

The current structure of the *STES Timeliness Framework* clearly shows the statistical processes for which documentation is sought (www.oecd.org/std/research/timeliness), with each link providing further information on the content relating to the short titles used in the main reference table. Documents can also cover several of the statistical processes included

⁷ The STES Timeliness Framework only covers practices for improving accuracy if it is clear that the method to improve accuracy could be traded off to either improve timeliness and / or reduce costs of production.

in the framework within the one paper. Documents can be provided in national language, but must include a short abstract in English, and also have a contact name.

The submission of documentation for assessment for inclusion in the framework, and any relevant queries, should be sent to richard.mckenzie@oecd.org

Users can also register to receive all future updates to the framework by emailing std.timeliness@oecd.org

Appendix 3

Short-term economic statistics in scope of the STES Timeliness Framework

The *STES Timeliness Framework* only intends to cover processes for the important short-term economic statistics listed in the table below, which are produced from business surveys and/or administrative records of businesses.

Indicator	Common periodicity
Industrial production index	Monthly
Manufacturing output / sales	Monthly / Quarterly
Construction production / output / sales	Monthly / Quarterly
Retail trade	
Current price	Monthly
Volume	Monthly
Turnover for other service industries	Quarterly
Unemployment rate (not from household labour force survey)	Monthly
Total employment (not from household labour force survey)	Monthly / Quarterly
Job vacancy rate	Monthly / Quarterly
Hourly / weekly / monthly wages	Monthly / Quarterly
Labour costs / prices	Quarterly

Extended definition of important short-term economic statistics covered by the framework

The text below gives a more detailed definition of each of the 10 important short-term economic statistics listed in the above table, for which methods of improving timeliness are covered in the *STES Timeliness Framework*.

Industrial production index

The majority of countries produce indices of industrial production. The data may come from a variety of business survey and or administrative sources. Generally the scope of these indices concerns production in industry excluding construction. This would therefore cover production in the Mining & Quarrying, Manufacturing, and Electricity Gas and Water supply industries (ISIC industries C, D & E), which is measured in a variety of ways across countries (i.e. in terms of the variables collected, with the common goal of measuring production).

Manufacturing output / sales

This indicator relates to the output of the Manufacturing industry, measured through business surveys of the Manufacturing industry or appropriate administrative data. The types of variables measured can differ between countries, but common types are: sales of goods manufactured; total turnover; stocks; new orders. Estimates are usually in the form on an index or as national currency levels.

Construction production / output / sales

This indicator relates to the output of the Construction industry. This may commonly be measured through business surveys of the Construction industry or appropriate administrative data. The variables measured can differ significantly between countries but common types include: building permits issued; value of work completed; work in place; work started. Note also that some variables may be measured in either monetary terms or physical quantities (e.g. number of buildings or surface area).

Retail trade: Current price; Volume

Almost all countries produce monthly estimates of sales / turnover in the Retail Trade industry, derived from business surveys of the Retail Trade industry or appropriate administrative data. *Current price* relates to the total value of sales (in index form or national currency level), whereas *volume* relates to data at constant prices – usually obtained by deflating the total value of sales estimate by relevant parts of the consumer price index.

Turnover for other service industries

This indicator relates to the output of other service industries such as ISIC industries G, H, I & K (i.e. Wholesale Trade; Hotels and Restaurants; Transport, Storage and Communication; Real Estate, Renting and Business Activities). This may commonly be measured through business surveys of these industries or appropriate administrative data. The variable generally collected is total turnover or sales.

Unemployment rate (not from household labour force survey)

This indicator relates to unemployment rate estimates derived from business surveys and / or administrative data. This indicator may exist only for a limited number of countries.

Total employment (not from household labour force survey)

This indicator relates to employment estimates derived from business surveys and / or administrative data. Many countries produce these estimates to provide reliable employment estimates by industry.

Job vacancy rate

This indicator relates to estimates of job vacancies / job vacancy rates / unfilled vacancies from business surveys and / or administrative data. This indicator may not exist for all countries.

Hourly / weekly / monthly wages

The majority of countries produce estimates of wage rates (i.e. hourly rates, weekly averages, monthly averages). These estimates usually come from specific business surveys or appropriate sources of administrative data. Data is usually estimated as an index or in national currency and for most countries will cover the majority of industries.

Labour costs / prices

Some countries have developed specific business surveys to measure a labour cost index / labour price index / employment cost index. The concept is to measure the total cost to employers for employing staff in the form of an index. Many countries compile this indicator as an extension to their estimate of wage rates, by combining this with other appropriate survey or administrative data.

Report

Cyclical Fluctuations in CIS Economies

Péter Benczúr*, Emin Muradov** and Attila Rátfai***

Abstract

This paper documents a number of stylized facts of quarterly frequency cyclical fluctuations in a specific group of developing economies, previously belonging to the same country organization, the former Soviet Union. We find that in these countries (1) fluctuations are in general less persistent than elsewhere; (2) private consumption is extremely volatile; (3) net exports are procyclical and persistent in commodity exporter countries; (4) government consumption is a very important, dominantly procyclical determinant of output; (5) Belarus, Russia, Ukraine, and to a smaller degree, Kazakhstan and Moldova are surprisingly similar in the behavior of their GDP components, industrial production and certain nominal variables; (6) there is mixed evidence regarding the dominance of supply versus demand shocks.

Key Words: Business Cycle Facts, Historical Origin, Commonwealth of Independent States

JEL Classification: E32

* Research Department, Magyar Nemzeti Bank and Department of Economics, Central European University; benczurp@mnb.hu

** Department of Economics, University of Texas at Austin; emin@eco.utexas.edu

*** Department of Economics, Central European University; ratfaia@ceu.hu

We thank CEU and GDN for financial support. The data used in this study are available at http://www.personal.ceu.hu/departs/personal/Attila_Ratfai/data/bmr_webpage.htm.

The views expressed herein are ours, and have not been endorsed by MNB.

Résumé

Le présent article documente un certain nombre de faits stylisés de fluctuations cycliques trimestrielles d'un groupe particulier d'économies en développement, qui appartenaient auparavant à la même organisation étatique, l'ancienne Union Soviétique. Nous trouvons que dans ces pays (1) les fluctuations sont en général moins persistantes que dans n'importe quel autre pays ; (2) la consommation privée est extrêmement instable ; (3) les exportations nettes suivent de très près le cycle de l'économie dans les pays exportateurs de produits. ; (4) la consommation du gouvernement est un déterminant de la production très important qui suit de manière dominante le cycle de l'économie ; (5) La Biélorussie, la Fédération de Russie, l'Ukraine, et à un moindre degré le Kazakhstan et la Moldavie sont étonnamment semblables quant au comportement des composantes de leur produit intérieur brut (PIB), la production industrielle ainsi que certaines variables nominales ; (6) La prédominance de l'offre par rapport aux chocs de la demande reste une évidence discutée.

1 Introduction

Are business cycles alike, particularly in countries with significant similarity in their recent political and economic history? Or, is it rather the interaction of potentially heterogeneous current shocks, and their propagation through the economy that are fundamental in shaping patterns of business fluctuations? To shed some light on this set of questions, this paper documents facts of business fluctuations in a group of developing economies, the ex-communist countries of the former Soviet Union. The countries, currently belonging to the Commonwealth of Independent States (CIS) we examine include Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russia and Ukraine.¹ The macroeconomic variables we focus on are GDP, private consumption, investment, government consumption, net exports, industrial production, employment, real wages, productivity, money, net capital flows, nominal interest rate, CPI, inflation and real effective exchange rate. The paper takes the perspective of modern business cycle analysis in characterizing economic fluctuations. The focus is on cross-country similarities and differences in the volatility, co-movement and persistence in the cyclical component of macroeconomic aggregates.

Until the fall of Communism, for about three-fourth of a century, the republics of the former Soviet Union shared many institutional features in organizing their political and economic systems.² After the early 1990s, however, these countries started to follow their own path out of the shadow of their communist past, preserving certain similarities, but showing increasingly more pronounced differences. On the one hand, for instance, Armenia and Georgia in the Caucasus are countries with important similarities in geography, economic development, size and openness. On the other hand, being rich in raw and/or processed commodities, large in population and territory, but relatively closed in international trade in goods and financial assets, Kazakhstan, Russia and Ukraine could also be grouped together. Finally, lacking basic commodities, being of smaller size and closely tied to, even somewhat dependent on their large neighbors, Belarus and Moldova are also countries with comparable characteristics.

While emerging markets are particularly intriguing objects of study in this context, research on characterizing their economic fluctuations is scant. Existing studies typically focus on the experience of certain individual countries. Kydland and Zarazaga (1997), for instance, examine the cyclical properties of the Argentine economy between 1970 and 1990. They document a number of interesting cyclical properties, many of them apparently pointing to the importance of supply side explanations of fluctuations. Kydland and Zarazaga find that output volatility is significantly higher in Argentina than in the United States. Private consumption expenditures are markedly more volatile than output. Net exports are

¹ Other ex-Soviet republics such as Azerbaijan, Tajikistan, Turkmenistan and Uzbekistan are excluded from the current study for the paucity of appropriate data. As they followed quite different path in development, the Baltic countries of Estonia, Latvia and Lithuania are also omitted.

² Kornai (1992) is a thorough overview and analysis of the socialist system.

countercyclical and government consumption is procyclical, though lagging the cycle. In terms of their country coverage, the most comprehensive studies are Agénor *et al.* (2000) and Benczúr and Rátfai (2005). Using industrial production as a proxy for output, Agénor *et al.* examine cyclical fluctuations in a diverse group of middle-income economies. Among other results, they document mostly countercyclical government expenditures, procyclical real wages, positive but weak contemporaneous correlation between money and output and no robust correlation between merchandise trade movements and output. Benczúr and Rátfai focus on the economies of Central and Eastern Europe (CEE). They find that consumption is highly volatile, and government spending and net capital flows are procyclical. Net exports are countercyclical, and real wages are dominantly procyclical. Overall, results in these papers highlight the importance of supply-side shocks in driving business cycles.

The rest of the paper is organized as follows. The next section briefly describes the data set and the empirical approach used in this paper. Section 3 presents the results, while Section 4 concludes.

2 Data and Methodology

As shown in Table 1 and described further in the Appendix, the quarterly frequency data for the countries studied here come from different sources, mainly from the International Financial Statistics of the IMF, local statistical offices and central banks. The series in general consist of 40 observations from 1995:1 to 2004:4. The sample is almost balanced, though not fully so. Some employment and industrial production data are entirely missing, and a few other series start only after 1995. Nonetheless, since highly unstable economic and political conditions make the quality of pre-1995 data questionable, in the relatively rare occasions when they are available, these observations are systematically omitted.

In the spirit of Lucas (1977) and Kydland and Prescott (1990), we think of fluctuations as deviation cycles. In practice, to generate the cyclical component of variables, we make a number of data transformations. First, constant price values of GDP and its components are deflated by the CPI to obtain constant price measures of these variables.³ Productivity figures are obtained as the ratio of constant price GDP to total employment. The real wage is defined as the ratio of nominal wages to the CPI. Inflation is the rate of change in the CPI. Increases in the exchange rate mean depreciation of the currency. With the exceptions of net exports, net capital flows, inflation and nominal interest rates, the variables are transformed into log form. Net exports and capital flows are expressed as their fraction of GDP, while inflation and nominal interest rates are left in log difference form. Then, all series are de-seasonalized

³ GDP deflator series are in general unavailable in these countries. We experimented with deflating current price investment, export and import figures by the PPI. As the results were qualitatively identical, we do not report them here.

using the X-11 procedure. Finally, all seasonally adjusted series are de-trended using the Hodrick-Prescott (HP) filter with the standard parameter for quarterly data, 1600.⁴

Table 1 Data Sources

	ARM	BLR	GEO	KAZ	KGZ	MDA	RUS	UKR
GDP	IFS	IFS	SO	IFS	SO	SO	SO	SO
Industrial Production	SO	N/A	SO	CB	SO	N/A	WIIW	SO -03:4
Consumption	IFS	IFS	SO	IFS	SO	SO	SO	SO
Investment	IFS	IFS	SO	IFS	SO	SO	SO	SO
Government Consumption	IFS	IFS	SO	IFS	SO	SO	SO	SO
Exports	IFS	IFS	SO	IFS	IFS	SO	SO	CB
Imports	IFS	IFS	SO	IFS	IFS	SO	SO	CB
Nominal Wage	SO	IFS and SO	SO	IFS	SO	ILO 97:1-	WIIW	CB 96:1-
Total Employment	N/A	N/A	SO	IFS	SO	ILO 99:1-03:4	WIIW	SO 99:1-
M1	IFS	IFS	IFS	CB	CB	IFS	IFS	EIU
CPI	IFS	IFS	SO	CB	SO	IFS	IFS	CB
Net Capital Flows	IFS	IFS 96:1-	IFS 97:1-	IFS	IFS	IFS	IFS	IFS
Nominal Interest Rate	IFS	IFS	IFS	EIU 97:1-	IFS 96:1-	IFS 95:4-	IFS	IFS
Real Effective Exchange Rate	IFS	CB 96:1-	CB	CB	SO	IFS	IFS	IFS
Sample Period	95:1-04:4	95:1-04:4	96:1-04:4	95:1-04:4	95:1-04:4	95:1-04:4	95:1-04:4	95:1-04:4
i) Unless otherwise indicated, the sample period is determined by the availability of current price GDP data in a particular country, as shown in the last row of the table. ii) Data sources and abbreviations: International Financial Statistics (IFS) of the IMF, local Central Banks (CB), local Statistical Offices (SO), the Economic Intelligence Unit (EIU), the International Labor Organization (ILO) and the Vienna Institute for International Economic Studies database (WIIW). N/A indicates missing or inadequately short series. ARM=Armenia, BLR=Belarus, GEO=Georgia, KAZ=Kazakhstan, KGZ=Kyrgyzstan, MDA=Moldova, RUS=Russia, UKR=Ukraine								

⁴ All of our main results are robust to using standard alternative approaches to filtering.

3 Results

The main results are summarized in Tables 2a through 2c. We focus on three sets of statistics characterizing volatility, cyclicity and persistence in the cyclical component of aggregate data. Absolute volatility is defined as the standard deviation of the series, while relative volatility is measured as the ratio of the standard deviation of the series and standard deviation of constant price GDP. For GDP, net exports and net capital flows, we report on absolute volatilities, while for other variables on relative volatilities. Variables are defined to be procyclical if they move in the same direction, countercyclical if they move in the opposite direction, and acyclical if they are uncorrelated with output. We report figures of "Cyclicity" showing the highest correlation in absolute value between the variable and constant price GDP, along with the figure for "Phase Shift" indicating the number of periods by which the variable at its highest correlation with GDP leads, lags or coincidental with output. The critical value for the correlation coefficient being different from zero is approximated by $2/\sqrt{T}$, where T is the number of observations. The degree of persistence is captured by the first-order autoregressive coefficient in the series.

Production and Expenditures (see Table 2a)

GDP volatility is fairly high in CIS countries. With the exception of Kyrgyzstan, the first autocorrelation coefficient is significant. When looking at higher order autocorrelations (not reported here), it is Moldova with four, Russia with two and Ukraine with also two significant coefficients. This indicates that GDP fluctuations are in general less persistent than in industrial or CEE countries. It is also interesting to note that Belarus, Moldova, Russia and Ukraine show both particularly high volatility *and* persistence.

Industrial production, with the exception of Armenia, is strongly procyclical and coincidental. Ukraine and Russia are notable again, showing the lowest volatility and the highest persistence. As a whole, industrial production is more persistent in larger countries such as Kazakhstan, Russia and Ukraine.

Private consumption, similarly to other emerging market economies, is extremely volatile in CIS countries, far exceeding the volatility of GDP itself, with the exception of Russia. Apart from Russia and Georgia, its persistence is low even when compared to CEE countries. With the exception of Kazakhstan and Kyrgyzstan showing no cyclicity, private consumption is procyclical and dominantly coincidental. Belarus, Russia and Ukraine have the highest degree of procyclicity and the lowest relative volatility as well.

Investment is procyclical and, with the exception of Kyrgyzstan, coincidental, and its relative volatility is in line with the international experience. All countries show relatively low degree of persistence in investment. Interestingly, consumption and investment exhibit very similar pattern of persistence in each country. Armenia, Kazakhstan, Kyrgyzstan and Moldova have the least persistent and most volatile investment pattern, while Belarus, Russia and Ukraine again show the lowest relative volatility and highest degree of procyclicity.

Table 2a **Production and expenditures**

	ARM	BLR	GEO	KAZ	KGZ	MDA	RUS	UKR
GDP								
Absolute Volatility	3.33	6.43	4.23	5.36	5.94	6.06	7.76	4.65
Persistence*	0.43	0.54	0.40	0.33	-0.02	0.68	0.80	0.64
Industrial Production								
Relative Volatility	2.04		1.52	2.65	1.82		0.57	0.78
Cyclical	-0.33		0.60	0.59	0.46		0.88	0.40
Phase Shift	-3 [§]		0	+1	0		+1	0
Persistence*	0.45		0.17	0.67	0.34		0.77	0.80
Private Consumption								
Relative Volatility	1.38	1.06	1.53	1.17	2.71	1.29	0.70	1.01
Cyclical	0.52	0.77	0.45	0.24	-0.29	0.52	0.74	0.70
Phase Shift	0	0	-4 [§]	0	+2 [§]	0	+1	0
Persistence*	-0.02	0.41	0.62	0.26	0.08	0.13	0.80	0.41
Investment								
Relative Volatility	3.70	1.58	2.22	3.06	6.33	4.31	1.99	1.70
Cyclical	0.40	0.48	0.49	0.38	0.36	0.44	0.86	0.79
Phase Shift	0	0	0	0	-3 [§]	-1	0	0
Persistence*	0.13	0.54	0.57	0.27	-0.08	0.11	0.71	0.37
Government Consumption								
Relative Volatility	2.40	1.30	5.05	2.00	2.27	3.50	1.54	1.68
Cyclical	-0.37	0.88	0.55	0.42	0.43	0.73	0.75	0.42
Phase Shift	-4	0	-1	0	-3	0	+2	0
Persistence*	0.29	0.57	0.37	0.41	-0.23	0.69	0.78	0.59
Net Exports								
Relative Volatility	3.27	2.56	3.10	4.96	7.25	5.58	5.15	3.44
Cyclical	-0.38	-0.22	-0.46	0.49	0.45	-0.48	0.66	0.31
Phase Shift	+3	+1	+1	0	0	-2	0 [§]	+4
Persistence*	0.31	0.04	0.53	0.57	0.40	0.27	0.81	0.24

Notes: All cyclical component series are at the quarterly frequency, de-seasonalized and de-trended by the Hodrick-Prescott filter. *Absolute Volatility* is the standard deviation of the variable. *Relative Volatility* is the ratio of the standard deviation of the variable and that of real GDP. *Cyclical* is the highest correlation in absolute value between the variable and real GDP. The critical value for the correlation coefficient is approximated by $2/\sqrt{T}$, where T is the number of observations. Negative figures for *Phase Shift* indicate lead, while positive ones lag to real GDP. *Persistence* is the AR(1) coefficient.
 ARM=Armenia, BLR=Belarus, GEO=Georgia, KAZ=Kazakhstan, KGZ=Kyrgyzstan, MDA=Moldova, RUS=Russia, UKR=Ukraine

[§] Unclear cyclical pattern.

Government expenditures, not surprisingly in countries descending from central planning and pervasive government influence, show pronounced patterns in CIS countries. With the exception of Armenia, they are massively procyclical and coincidental, even more so than in CEE countries. They also tend to be the most procyclical and most persistent component of GDP. Government expenditures are highly volatile; with the exception of Kyrgyzstan, their relative volatility is comparable to that of investment. Finally, Belarus, Russia and Ukraine exhibit the least volatile, most persistent and most procyclical government expenditures. Moldova shows high persistence and procyclicality, but its government expenditures are also highly volatile.

Net exports, in contrast to international experience, are procyclical in Kazakhstan, Kyrgyzstan. In Russia, the highest correlation coefficient is negative, but Russia also has a comparably high positive coefficient too. These countries are major commodity exporters (oil for Kazakhstan and Russia, natural gas for Kyrgyzstan and Russia), thus a positive comovement between net exports and GDP may point to important terms of trade shocks, triggering export-led expansions. Indeed, Agénor *et al.* (2000) also find that in some emerging countries there exist a positive correlation between industrial production and net exports, and those countries tend to be important commodity exporters, such as Morocco and Nigeria. Final, the commodity exporting countries also tend to show non-negligible persistence in net exports.

Labor Market (see Table 2b)

Employment is not particularly volatile, and with the exception of Ukraine, shows a high degree of persistence. It also tends to be procyclical, with the exceptions of Ukraine, Kyrgyzstan (acyclical) and Moldova (countercyclical). The only country with significant positive lags is Kazakhstan, pointing towards a negligible role of labor hoarding considerations.

Real wages are dominantly procyclical and persistent, underlying the importance of supply side shocks in fluctuations. Armenia and Kyrgyzstan stand out as exceptions in both aspects. It is noteworthy that Belarus, Moldova, Russia and Ukraine show the highest degree of real wage persistence. Kazakhstan exhibits the lowest relative volatility and it is also a country with one of the highest degree of wage controls; while Belarus exhibits the highest volatility and the largest increase in wage controls (see Table 3).

Productivity in CIS economies shows very similar patterns to other emerging market countries. Its relative volatility is about one, slightly higher than in countries of CEE. Productivity is strongly procyclical in the region, more so than in CEE economies. Apart from Kyrgyzstan, it also shows a medium degree of persistence, and with the exception of Moldova, it is synchronous.

Table 2b Labor Market

	ARM	BLR	GEO	KAZ	KGZ	MDA	RUS	UKR
Employment								
Relative Volatility			1.03	0.84	0.38	0.56	0.11	0.26
Cyclical			0.36	0.43	-0.14	-0.58 ^a	0.49	-0.33 ^a
Phase Shift			0	+3	-2	+4	-2	0
Persistence			0.75	0.87	0.84	0.51 ^a	0.79	0.04 ^a
Real Wages								
Relative Volatility	1.63	3.50	1.49	0.54	1.38	1.46	1.59	1.81
Cyclical	-0.48	0.56	0.61	0.50	-0.41	0.81	0.75	0.30
Phase Shift	+4	+1	+1	0	0	+1	+1	-3
Persistence	0.29	0.92	0.63	0.56	0.25	0.85	0.88	0.93
Productivity								
Relative Volatility			1.15	1.29	1.05	1.15	0.96	1.24
Cyclical			0.55	0.76	0.93	0.79 ^a	0.99	0.78 ^a
Phase Shift			0	0	0	+3	0	0
Persistence			0.42	0.47	0.06	0.73 ^a	0.78	0.55 ^a

Notes: See Table 2a.

^a Based on 20-24 observations. The significance levels are [0.41,0.45]

Financial and Monetary Variables (see Table 2c)

Money (M1) shows fundamentally homogeneous patterns. Its relative volatility is surprisingly small. It is dominantly procyclical; the two exceptions are Belarus and Ukraine, both being marginally countercyclical. The persistence in the money stock is also notable.

Net capital flows do not exhibit any clear patterns. Overall, they tend to be procyclical or acyclical, with absolute volatilities comparable to CEE countries, and with the exception of Russia, showing no persistence.

Nominal interest rates are in general persistent, though less so in Kyrgyzstan. They tend to be more volatile in countries with pronounced inflationary experience (Belarus, Russia, Ukraine and Armenia). There is no clear pattern in their cyclical.

The *Consumer Price Index* is countercyclical in five cases and procyclical in three, pointing to a mixed role of supply versus demand type shocks. In many cases there are significant positive and negative correlation coefficients as well. Similarly to interest rates, inflation and the real effective exchange rate, the price level shows the highest volatility in Belarus, Russia, Ukraine, and to some degree, Armenia. Georgia and Kazakhstan exhibit a low CPI volatility and a high degree of price controls, while Belarus – just like in the case of

real wages – features the highest CPI and inflation volatility, and the largest increase in price controls (see Table 3).

Table 2c **Financial and Monetary Variables**

	ARM	BLR	GEO	KAZ	KGZ	MDA	RUS	UKR
Money								
Relative Volatility	2.53	2.35	1.86	2.70	1.27	1.65	1.34	1.81
Cyclical	0.47	-0.33	0.62	0.58	0.39	0.65	0.49	-0.31
Phase Shift	-1	+3	-1	+2	-3 ^{\$}	-4	-1	+2
Persistence	0.62	0.92	0.65	0.82	0.75	0.77	0.72	0.62
Net Capital Flows								
Relative Volatility	6.50	3.16	3.44	4.32	7.17	5.98	4.73	5.27
Cyclical	0.34	-0.23	0.33	0.36	0.58	0.27	0.53	0.33
Phase Shift	+3	+3	+1	+3 ^{\$}	-4 ^{\$}	-3 ^{\$}	0	-4 ^{\$}
Persistence	0.38	-0.05	0.07	-0.10	0.26	0.32	0.58	-0.20
Nominal Interest Rate								
Relative Volatility	0.76	1.23	0.21	0.11	0.30	0.11	1.48	0.79
Cyclical	-0.32	-0.33	0.41	0.39	-0.29	-0.76	-0.17	0.38
Phase Shift	-1	+2	-4	-3 ^{\$}	+1	-1	+2	+3
Persistence	0.78	0.75	0.51	0.61	0.37	0.84	0.87	0.54
CPI								
Relative Volatility	1.37	4.20	0.63	0.97	0.94	1.28	1.62	2.25
Cyclical	0.33	-0.34	-0.56	-0.45	0.41	-0.74	0.81	-0.66
Phase Shift	+4	+1 ^{\$}	0	-2 ^{\$}	-3	+1	-4 ^{\$}	0
Persistence	0.78	0.94	0.65	0.91	0.83	0.89	0.83	0.89
Inflation								
Relative Volatility	0.77	1.37	0.53	0.45	0.55	0.59	0.88	0.85
Cyclical	-0.36	-0.78	-0.42	0.49	-0.33	-0.46	-0.67	0.56
Phase Shift	0	0	0 ^{\$}	-1	0	-2	0	+1
Persistence	0.10	0.62	0.28	0.39	-0.01	0.38	0.66	0.71
REER								
Relative Volatility	2.13	2.10	1.46	1.24	1.00	0.88	2.11	1.36
Cyclical	-0.48	0.77	-0.63	0.60	-0.48	-0.21	0.68	-0.61
Phase Shift	-4	0	0	+2	+4	-4	+2 ^{\$}	+2
Persistence	0.65	0.67	0.75	0.71	0.47	0.67	0.83	0.73
<i>Notes: See Table 2a.</i>								

Inflation is mostly countercyclical, with the exception of Kazakhstan and Ukraine, both of them being procyclical. This can again indicate a mixture of supply and demand shocks, with some dominance by the first type. Inflation series are also close to coincidental. Interestingly, the cyclical nature of the price level and inflation differs in five countries. Finally, Belarus, Russia and Ukraine stand out by showing the most volatile *and* persistent inflation experience.

The *real effective exchange rate* shows completely mixed patterns in cyclicity, with three countries being procyclical, three countercyclical and two acyclical. Interestingly, real exchange rates are less volatile in CIS countries than in other emerging market economies.

Table 3 Institutional Characteristics: 1996-2004

		ARM	BLR	GEO	KAZ	KGZ	MDA	RUS	UKR	USSR*
Political System										
Political rights	Mean	4.67	6.33	3.67	6.00	5.33	2.67	4.67	3.67	4.00
	Trend	0	+	-	0	+	0	+	+	-
Civil liberties	Mean	4.00	6.00	4.00	5.00	4.67	4.00	4.67	3.67	4.00
	Trend	0	0	0	0	+	0	+	-	-
Status	Mean	PF	NF	PF	NF	PF	PF	PF	PF	PF
	Trend	0	+	-	+	++	0	++	0	-
Economic freedom										
Total	Mean	3.21	3.89	3.66	3.80	3.57	3.28	3.67	3.75	
	Trend	--	+	-	-	-	-	-	-	
Fiscal burden	Mean	2.77	4.05	2.26	3.25	3.43	3.30	3.53	3.97	
	Trend	-	-	0	+	-	--	--	-	
Government Intervention	Mean	4.00	2.83	3.50	2.75	3.25	3.50	3.67	4.00	
	Trend	--	++	--	-	--	--	--	-	
Wages and Prices	Mean	3.33	4.00	3.67	3.50	3.00	3.00	3.00	3.00	
	Trend	-	++	-	-	0	0	0	0	

Notes: A higher score indicates a less liberalized system, thus a positive trend corresponds to a decline in the degree of liberalization. *Mean* is the average for years 1996, 2000 and 2006. ++ and -- denote large changes. *PF* stands for partly free, while *NF* means not free. Political system data is from the Freedom House (www.freedomhouse.org), while economic freedom indices are from the Heritage Foundation (<http://www.heritage.org/research/features/index/downloads.cfm>).

ARM=Armenia, BLR=Belarus, GEO=Georgia, KAZ=Kazakhstan, KGZ=Kyrgyzstan, MDA=Moldova, RUS=Russia, UKR=Ukraine

* values from 1992

Overall, the results show that fluctuations are less persistent in most CIS countries than elsewhere, thus one may think of them less as “business cycles”. Moldova, Russia and Ukraine are three important exceptions.

Belarus, Russia, Ukraine, and to a smaller degree, Kazakhstan and Moldova show a surprising similarity, especially in GDP components (relative volatility, cyclical and persistence), industrial production (relative volatility and persistence) and the behaviour of certain nominal variables such as prices and interest rates. These countries seem to have managed to maintain a close connection among various elements of output, in terms of procyclicality, persistence and relative volatilities. The appearance of inflationary episodes, the volatility of interest rates and prices, and the volatility and persistence of inflation might be the consequence of a more managed real side in these economies. Interestingly, these countries also share many relevant characteristics: they tend to be larger (Kazakhstan, Ukraine, Russia), more centralized (Belarus, Kazakhstan, Russia and Ukraine)⁵ and linked to each other (Belarus, Russia, Ukraine, and to some degree, Moldova), richer (Belarus, Kazakhstan, Russia, and Ukraine), and exporting more commodities (Kazakhstan, Russia, and Ukraine).

Three GDP components exhibit a particularly pronounced cyclical pattern. Like in many emerging economies, private consumption is extremely volatile, more than GDP.⁶ Net exports tend to be procyclical and relatively persistent in countries where commodity exports play a central role. Lastly, government consumption appears to be a very important determinant of output, being very volatile and the most procyclical and persistent component of GDP overall.

There is mixed evidence regarding the dominance of supply (technology) versus demand (like government expenditures) shocks. The first type is consistent with procyclical real wages (the exceptions are Armenia and Kazakhstan), procyclical productivity (no exceptions), countercyclical prices (the exceptions are Armenia, Kyrgyzstan and partly Russia), and countercyclical inflation (the exceptions are Kazakhstan and Ukraine). This picture is, however, balanced by the dominant role government expenditures seem to play in these economies.

4 Concluding Remarks

Do countries' past traits in basic political and economic institutions leave their imprint on the nature of business cycles? To answer this question, this study documents stylized facts of business cycle fluctuations in a group of CIS countries. Despite the obvious limitations on data, several interesting results appear to emerge from the analysis.

⁵ See Table 3 summarizing information on political system, total economic freedom and fiscal burden, and partly on government intervention indicators in these countries.

⁶ This is documented and discussed by Benczúr and Rátfai (2005) and Aguiar and Gopinath (2004).

Most importantly, fluctuations are more volatile and less persistent in these countries than elsewhere. Moreover, with very few exceptions, each variable considered showed a high degree of similarity across the countries, confirming that their common past and comparable institutional evolution do have an influence on their economic performance. A particularly interesting case is that of government consumption, which is strongly procyclical, volatile, coincidental, and persistent, reflecting pervasive past and current government influence.

Though these countries completely share their past, there are noticeable differences in their current characteristics, like their size, income level, degree of liberalization, or the role of natural resources in their exports. Belarus, Russia, Ukraine, and somewhat less, Kazakhstan and Moldova have much in common in these aspects, and that leaves a strong mark on their cyclical patterns as well.

References

- Agénor, Pierre-Richard; C. John McDermott and Eswar S. Prasad. 2000. "Macroeconomic Fluctuations in Developing Countries: Some Stylized Facts," *The World Bank Economic Review*, pp. 251-85.
- Aguilar, Mark and Gita Gopinath. 2004. "Emerging Market Business Cycles: The Cycle is the Trend," *NBER Working Paper* no. 10734.
- Benczúr, Péter and Attila Rátfai. 2005. "Economic Fluctuations in Central and Eastern Europe. The Facts", *CEPR Discussion Paper* no. 4846.
- Kornai, János. 1992. *The Socialist System. The Political Economy of Communism*. Princeton and Oxford: Princeton University Press and Oxford University Press.
- Kydland, Finn E. and Edward C. Prescott. 1990. "Business Cycles: Real Facts and Monetary Myth," *Federal Reserve Bank of Minneapolis Quarterly Review*, pp. 3-18.
- Kydland, Finn E. and Carlos Zarazaga. 1997. "Is the Business Cycle of Argentina Different?" *Federal Reserve Bank of Dallas Economic Review*, pp. 21-36.
- Lucas, Robert E. Jr. 1977. "Understanding Business Cycles", *Carnegie-Rochester Conference Series on Public Policy*, pp. 7-29.

Appendix: Data Sources

GDP. For Georgia, Kyrgyzstan, Moldova, Russia and Ukraine, the current price GDP data are from the local Statistical Office. For Armenia, Belarus and Kazakhstan, the data are the IFS 99B series.

Industrial production. For Armenia, Georgia, Kyrgyzstan and Ukraine, the data are from the local Statistical Office. For Kazakhstan, the series are from the local Central Bank. For Russia, the data are from the WIIW.

Private consumption. For Georgia, Kyrgyzstan, Moldova, Russia and Ukraine, the current price private consumption data are from the local Statistical Office. For Armenia, Belarus and Kazakhstan, the data are the IFS 96F series, which includes Non-Profit Institutions Serving Households.

Investment. For Georgia, Kyrgyzstan, Moldova, Russia and Ukraine, the current price gross fixed capital formation data are from the local Statistical Office. For Armenia, Belarus and Kazakhstan, the data are the IFS 93E series.

Government consumption. For Georgia, Kyrgyzstan, Moldova, Russia and Ukraine, the current price private consumption data are from the local Statistical Office. For Armenia, Belarus and Kazakhstan, the data are the IFS 91F series.

Exports/Imports. For Georgia, Moldova and Russia, the data are from the local Statistical Office. For Ukraine, the data are from the local Central Bank. For Armenia, Belarus and Kazakhstan, the data are the IFS 90C and 98C series, for Kyrgyzstan, the IFS 70 and 71 series. If in dollars, the figures are converted into local currency at the corresponding exchange rate. Net exports are defined as the difference between exports and imports.

Total employment. The figures for Kazakhstan are the index from the IFS 67 series, For Russia, the data are from the WIIW database; while for Georgia, Kyrgyzstan and Ukraine, they are taken from the local Statistical Office.

Wage. The nominal wage series for Kazakhstan are from the IFS 65 series. For Ukraine, the data are from the local Central Bank. For Armenia, Georgia and Kyrgyzstan, the data are from the local Statistical Office. For Russia, the data are from the WIIW database. For Moldova, the data are from the ILO. For Belarus, the series are obtained by linking the average monthly wage figures of the local Statistical Office with the quarter-to-quarter rate of change information in the IFS.

Money. The money stock is measured by M1. For Ukraine, the series on M1 are obtained from the Economic Intelligence Unit. For Kazakhstan and Kyrgyzstan, the data are obtained from the local Central Bank. For all the other countries, M1 is the IFS 34 series.

Net capital flow. Net capital flows are obtained as the sum of IFS 78bcd (capital account) and IFS 78bjd (financial account) series measured in US dollars. The series are converted to local currency using the corresponding exchange rate.

Nominal interest rate. For Kazakhstan, the data on lending rates are from the Economic Intelligence Unit. For the rest of the countries, the data are from the IFS 60 series.

Consumer Price Index. The CPI figures for Kazakhstan and Ukraine are obtained from the local Central Bank, while for Georgia and Kyrgyzstan they are from the local Statistical Office. For the rest of the countries, the data are from the IFS 64 series. Inflation is defined as the quarterly change in log prices.

Real effective exchange rate. The series for Belarus, Georgia, Kazakhstan and Kyrgyzstan are from the local Central Bank. For Armenia, Moldova, Russia and Ukraine, they are from the IFS.

OECD and CIRET News – 2007/1

OECD Short-term Economic Statistics	138
Short Review of the 28 th CIRET Conference	143
Conferences and Workshops	145

OECD Short-term Economic Statistics

OECD Workshop on Business and Consumer Tendency Surveys, Rome, 19 September 2006

The main objective of the Workshop was to present work undertaken at national and international levels following the joint European Commission (EC) – OECD Workshop on International Development of Business and Consumer Tendency Surveys held in Brussels in November 2005. The agenda for the Rome Workshop was also designed to complement both the discussion on more analytical issues during the 2006 CIRET Conference and work undertaken by the European Commission. Finally, the Workshop was envisaged as an intermediate process leading to more in-depth studies on some of the areas of future work identified in November 2005 leading up to the next joint EC – OECD Workshop scheduled to held in Brussels in the third quarter of 2007.

The 2006 OECD Workshop comprised four substantive sessions, namely the presentation of:

- the final version of the enhanced opinion survey international guidelines and recommendations endorsed at the EC – OECD Workshop in Brussels in November 2005;
- recent OECD work on standardised confidence indicators and zone aggregates, and the compilation / dissemination of composite leading indicators for large emerging non-member economies (Brazil, China, India, Indonesia, Russian Federation and South Africa) and new recent OECD Member countries;
- a number of recent national developments covering a range of areas such as data collection, use of internet surveys, response analysis, etc;
- OECD proposals for proposed supplementary consumer opinion survey questions to collect information from the general public on their current knowledge of the economic and social situation of their country and their sources for such information.

This article summarises the main meeting conclusions and outlines proposals for future work.

Enhanced international guidelines for BTS / COS

The development of recommendations by two task forces comprising national institutes, the OECD and EC entailed a thorough review of current national practice and available literature. The draft standards developed by the task forces were discussed at the November 2005 joint workshop and the version presented in Rome were embodied in the meeting report

with minor modifications to incorporate comments subsequently received from national institutes. In discussion at the Rome Workshop there was continued strong support for the enhanced recommendations and no further changes were suggested.

The enhanced guidelines will be given wide dissemination by the OECD and the European Commission over the coming months, for example, in the OECD's business and consumer opinion survey portal, in future editions of the OECD's BTS Handbook and the Commission's revised (1997) publication outlining the EU Programme of Business and Consumer Surveys. The OECD will use the guidelines to facilitate the benchmarking and comparison of current national practice. This work will be an extension of current activities to improve the quality and consistency (in terms of issues covered) of metadata disseminated with opinion survey data in the OECD's monthly *Main Economic Indicators* (MEI) database.

Future development work and research associated with the enhanced guidelines will now focus on a number of specific issues / aspects raised by the recommendations. Such work may be undertaken by new task forces or by an individual national statistical institute. The two main areas of work associated with the enhanced guidelines over the next twelve months therefore comprise:

- Further analyses by individual national institutes survey non-response in the context of the key characteristics
- The creation of a new task force to prepare a draft manual / handbook for the development and conduct of internet surveys also for presentation and discussion in 2007. The initial content of the handbook would cover a limited number of key issues and additional issues could be included in subsequent versions.

Another area of potential future work over the next twelve months discussed briefly at the Rome Workshop concerned recent work at the national level involving the development of innovative approaches to analysing and presenting survey output to users. This work parallels similar work by NSOs / central banks over recent years in the area of quantitative statistics to give more value-added to the basic data they collect and disseminate in order to reach new audiences and provide an enhanced service to existing users.

Recent OECD work on composite leading indicators, zone aggregates and standardised confidence indicators

Composite leading indicators and zone aggregates

This session presented recent OECD work for discussion on the compilation and dissemination of new composite leading indicators (CLIs) for large emerging non-member economies (Brazil, China, India, Indonesia, Russian Federation and South Africa) and recent OECD Member countries, and the compilation of standardised confidence indicators. Associated with the new CLIs was the potential provided for the compilation of new zone aggregates to facilitate comparisons between different regions of the globe.

The expansion of the OECD system of CLIs to include the new CLIs for the six recently new OECD member countries has implications for the calculation of the *OECD total area* and *OECD Europe area* aggregates. In addition, the inclusion of the new CLIs for all of the above twelve countries opens the possibility to calculate new area aggregates such as *Major Asian economies* and *Eastern Europe* (including or excluding the Russian Federation), and a *World proxy* to give information on the overall global development. The importance of such new regional or area aggregates is of course very much dependent on the existence of different cyclical patterns between these new aggregates and the established ones. However, the calculation of a *World proxy* aggregate is important in itself in so far that it will represent global development better than the *OECD total area* aggregate. Detailed results on the development of the new CLIs are presented in a paper available at the OECD-website.¹

Standardised confidence indicators

The second area of recent OECD work presented in Rome related to the development of selected regional standardised confidence indicators for the manufacturing survey and consumers. The development of these indicators built on earlier work presented at the 2005 Brussels meeting, though with further enhancements to the indicator compilation method, the inclusion of additional component data and changes in some of the zone definitions. The new zones presented comprised: Asia Major 5 (China, India, Indonesia, Japan and Korea); G7; Euro area; OECD total; and OECD plus non-member economies (Brazil, China, India, Indonesia, Russian Federation and South Africa). The new confidence indicators will be included in a new subject table in the OECD's MEI in the near future.

Following the OECD presentations on CLIs and standardised confidence indicators there was discussion on their use in cross-country comparisons and world cycle anticipation. A specific question raised was what weighting variable to use for calculation of indicators for regional or area totals. Further work by the OECD will entail:

- Development of a cyclical indicators area in the OECD's main data dissemination platform (OECD.stat) where all data related to the analysis of economic cycles will be made available. This area will include CLIs for OECD Member countries and non-member economies, business tendency and consumer survey indicators and the new confidence indicators.
- Investigation of the possibility to calculate regional or zone confidence indicators for other sectors (construction, retail trade and services) including a non-manufacturing sector (aggregates are currently only available for the United States and Korea).
- Exploration of the use of alternative weights to reflect the relative importance of the countries in the aggregation of confidence indicators to regional or zone aggregates such as the share of exports and imports of a country in total world trade.

¹ <http://www.oecd.org/dataoecd/35/22/36414874.pdf>

Recent national developments

In this session four presentations were given by national institutes on recent work at the country level, namely:

- The results of a world-wide survey to collect information from national institutes (total of 32) on current internet business tendency survey practices. The survey covered sector coverage, comparisons of response rates or different collection media (paper, interview, internet), practical issues (such as formats used (email / web questionnaires), and national assessments of the viability of internet surveys from the perspectives of both the respondents and the survey institute (Ifo Institute, Germany).
- An overview of French experience in the use of internet surveys for the collection of survey data covering the wholesale, industry and services sectors (with future intentions to collect data via this medium for construction, investment and retail trade). In particular, the INSEE presentation covered issues / procedures related to ensuring security, provision of information feedback to the respondent, questionnaire design, data validation and procedures to maximise response (INSEE, France).
- Long-term biases in consumer sentiment using micro data to analyse people's perceptions of the economic situation (ISAE, Italy).
- The recent introduction by ISAE of two new questions in their manufacturing survey to collect information concerning the cyclical behaviour of trade credit, the aim being to identify possible relationships between such credit and the business cycle (ISAE, Italy).

Proposed COS supplementary questions on general public's knowledge of current national economic and social situation

The aim of this session was to present an OECD proposal to delegates at the Rome Workshop for a new international survey on what people in the general community know about key economic and social phenomena in their countries. The results of such a survey would be presented and discussed at the second World Forum on "Statistics, Knowledge and Policy" to be held in Istanbul in June 2007.

It is envisaged that the survey could be undertaken in co-operation with research institutes that already run opinion surveys on businesses and households. The survey could be undertaken through the inclusion of not more than ten supplementary questions to existing questionnaires (in particular, to those used for consumer opinion surveys) on what people know about key economic and social indicators (GDP, inflation, unemployment, etc.). Questions could range from knowledge about the growth rate of GDP, the level and trends in unemployment, inflation, tax rate, public deficit, life expectancy, greenhouse gas emissions, etc.

Questions would of course have to be properly defined and set up and this will require considerable thought, benefiting from the experience of research institutes running consumer opinion surveys, who would be free to add questions for national purposes.

The results would be transmitted to the OECD, who would prepare a special publication showing comparative results, while national data could be used by research institutes for their own purposes.

Summary of future work identified at Rome Workshop

The priority areas for future work identified at the Rome Workshop for the next joint OECD - EC opinion survey meeting scheduled in the third quarter of 2007 in Brussels are summarised below:

- further analyses of the response characteristics of surveys conducted by national institutes;
- compilation of a manual / handbook on the design and implementation of internet based surveys. It is proposed that a task force be set up to draft text for consideration / discussion at the next joint meeting on a range of technical and methodological issues;
- the creation of a Reflection Group to formulate proposals for the supplementary questions for consumer opinion surveys referred to above;
- presentations by national institutes on the development of tools, etc., for the presentation of opinion survey data. This work parallels similar work by NSOs / central banks over recent years in the area of quantitative statistics to give more value-added to the basic data they collect and disseminate in order to reach new audiences and provide an enhanced service to existing users;

Other possible topics for the joint meeting in 2007 comprise applications:

- with investment, construction and retail surveys highlighting the benefits and usefulness of such surveys, e.g. work done by Statistics Sweden or Ifo Institute, Germany with their respective investment surveys;
- using more advanced statistical and econometric techniques (e.g. the Markov-switching model developed by DG ECFIN to forecast turning points in euro-area industrial production).

Enquiries about any aspect of the Workshop should be referred to Denis Ward (denis.ward@oecd.org) or Ronny Nilsson (Ronny.Nilsson@oecd.org).

Short Review of the 28th CIRET Conference

The 28th CIRET Conference was held in Rome, Italy, September 20–23, 2006.

The conference was hosted by the (ISAE) and co-hosted by the University «La Sapienza», Faculty of Statistics, and the Bank of Italy.

In the name of all participants, chairpersons, speakers and authors of conference papers, the CIRET Office would like to thank the local organiser for the great efforts and for the fruitful cooperation in organising the conference. We also want to thank for the support which several sponsors accorded to the local conference organizer. In addition, the Bank of Italy was so kind to host the OECD Workshop on Business and Consumer Tendency Surveys held on September 19.

In the opening session of the conference, *Prof. Marco Lippi* gave a lecture on “Real Time Indicators of Aggregate Economic Activity: Theoretical Issues and Recent Experience with Eurocoin”.

Under the overall heading “Cyclical Indicators and Economic Policy Decisions”, the sessions were organised according to the following topics:²

BTS - Analysis

Six sessions chaired by Franz-Josef Klein, Riccardo Cristadoro, Sergio de Nardis and András Vértés

BTS - Forecasting

Session chaired by Enrico Giovannini

BTS - Quantification Methods

Two sessions chaired by Marco Lippi and Tommaso Proietti

BTS - Methods

Two sessions chaired by Bianca Maria Martelli

BTS - Internationalisation and Competition

Session chaired by Hélène Erkel-Rousse

BTS - Long-run Cyclical Fluctuations and Structural Breaks

Session chaired by Spyros Arvanitis

² *BTS*: Business Tendency Surveys (including Investment Surveys); *CS*: Consumer Survey; *CLI*: Composite and Leading Indicators; *Analysis*: Use of Survey Data or Cyclical Indicators for Micro- or Macroeconomic Analysis

BTS - Uncertainty and Expectations Evaluation

Two sessions chaired by Philippe Scherrer

CS - Analysis

Session chaired by Günter Poser

CS - Expectations Processes

Two sessions chaired by Giuseppe Parigi

CS - New Methods

Session chaired by Richard Curtin

CLI - Analysis

Six sessions chaired by Jürgen Kröger, Ronny Nilsson, Gernot Nerb,
Maria Drozdowicz-Biec and Winfried Stier

CLI - Methods

Two sessions chaired by Jan-Egbert Sturm

Innovative Activity

Two sessions chaired by Karl Heinrich Oppenländer

LMA - Labour Market Analysis

Session chaired by Ullrich Heilemann

Two sessions were held about the *Special Topic on Regional and Local Surveys* chaired by Antonio G. Calafati and Paolo Carnazza. The CIRET Office thanks the Institute Tagliacarne for the generous hosting of these sessions and for the evening program in the marvellous gardens of the institute.

In the closing session, *Prof. Lucrezia Reichlin* (ECB) gave a lecture on “The Value of Survey Data for Nowcasting Economic Activity”.

The CIRET Website (<http://www.ciret.org/conferences/rome2006>) provides additional information about the conference. The download of papers, the list of participants as well as a picture gallery are available to CIRET members, to authors and to participants of the conference.

Conferences and Workshops*

Conferences

- **Forecasting Summit 2007, September 17-19, 2007**
Boston, MA, USA
<http://www.forecasting-summit.com>
- **Survey Data in Economics – Methodology and Applications, October 19-20, 2007**
Munich, Germany
<http://www.ifo.de>

Workshops

- **Workshop on Real-Time Data Analysis, July 23-24, 2007**
Zurich, Switzerland
<http://www.kof.ethz.ch>

* This page gives information about forthcoming events. Please send early notice to 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

Three issues of the journal are published in each volume.

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 double blind referee process.

Under the heading "Report" additional contributions, including topics such as sample design, data collection and validation, dissemination and evaluation of survey results, are published. Reports are reviewed by members of the Editorial Board.

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).

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16

PRINTED IN FRANCE

(33 2007 01 1 P) ISSN 1729-3618 – No. 55729 2007

Journal of Business Cycle Measurement and Analysis

The *Journal of Business Cycle Measurement and Analysis* is jointly published by the OECD and the Centre for International Research on Economic Tendency Surveys (CIRET) to promote the exchange of knowledge and information on theoretical and operational aspects of economic cycle research, involving both measurement and analysis.

Contents of this issue:

Consumer Sentiment Surveys: Worldwide Review and Assessment <i>Richard Curtin</i>	7
Deviation Cycles in Manufacturing: Business Cycle Measurement and Leading Indicators <i>Ard H.J. den Reijer</i>	43
Comparing Probability Forecasts in Markov Regime Switching Business Cycle Models <i>Daniel R. Smith and Allan Layton</i>	79
Report Improving Timeliness for Short-Term Economic Statistics <i>Richard McKenzie</i>	99
Report Cyclical Fluctuations in CIS Economies <i>Péter Benczúr, Emin Muradov and Attila Rátfai</i>	121
OECD and CIRET News – 2007/1	137

Subscribers to this printed periodical are entitled to free online access.

If you do not yet have online access via your institution's network, contact your librarian or, if you subscribe personally, send an email to

SourceOECD@oecd.org

2007
Volume 3, No. 1

OECDpublishing
www.oecd.org/publishing

ISSN 1729-3618
33 2007 01 1 P

