



EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1334 / MAY 2011

**ON THE IMPORTANCE
OF SECTORAL AND
REGIONAL SHOCKS
FOR PRICE-SETTING**

by Guenter W. Beck,
Kirstin Hubrich and
Massimiliano Marcellino



EUROPEAN CENTRAL BANK

EUROSYSTEM



WORKING PAPER SERIES

NO 1334 / MAY 2011

ON THE IMPORTANCE OF SECTORAL AND REGIONAL SHOCKS FOR PRICE-SETTING¹

by Guenter W. Beck², Kirstin Hubrich³
and Massimiliano Marcellino⁴



In 2011 all ECB publications feature a motif taken from the €100 banknote.

NOTE: This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

This paper can be downloaded without charge from <http://www.ecb.europa.eu> or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=1813977.

- ¹ We thank Todd Clark, Timothy Cogley, Gabriel Fagan, Bartosz Mackowiak, Ben Malin, Serena Ng, Kristoffer Nimark, Giorgio Primiceri, Pierre-Daniel Sarte, James Stock, Rob Vigfusson, Mark Watson, Ken West, Alexander Wolman and participants at the NBER Summer Institute 2010, the European Economic Association Congress 2010, a workshop at the Federal Reserve Bank of St. Louis, and at seminars at the European Central Bank, the Federal Reserve Board as well as the Federal Reserve Banks of Kansas City and Philadelphia for helpful comments. The usual disclaimer applies.
- ² University of Siegen, Hölderlinstr. 3, 57076 Siegen, Germany and CFS; e-mail: guenter.beck@uni-siegen.de
- ³ Research Department, European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: kirstin.hubrich@ecb.europa.eu.
- ⁴ European University Institute, Via Roccazzini, 9, I-50014 Fiesole, Florenz, Italy; Bocconi University and CEPR; e-mail: Massimiliano.Marcellino@EUI.eu.

© European Central Bank, 2011

Address

Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address

Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone

+49 69 1344 0

Internet

<http://www.ecb.europa.eu>

Fax

+49 69 1344 6000

All rights reserved.

Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, <http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html>

ISSN 1725-2806 (online)

CONTENTS

Abstract	4
Non-technical summary	5
1 Introduction	8
2 Data and descriptive statistics	11
3 Econometric methodology: a new approach	14
3.1 The model	14
3.2 Estimation of a factor model for over-lapping data blocks	16
4 Monte-Carlo simulations	18
4.1 Base case	19
4.2 Additional experiments	21
5 Empirical results	22
5.1 Aggregate-sector decomposition	22
5.2 Aggregate-sector decomposition: our approach	23
5.3 Analysis of the regional component	26
6 Robustness analysis: month-on-month versus year-on-year changes	29
6 Conclusions	30
References	32
Tables	36
Appendices	43

Abstract

We use a novel disaggregate sectoral euro area data set with a regional breakdown to investigate price changes and suggest a new method to extract factors from overlapping data blocks. This allows us to separately estimate aggregate, sectoral, country-specific and regional components of price changes. We thereby provide an improved estimate of the sectoral factor in comparison with previous literature, which decomposes price changes into an aggregate and idiosyncratic component only, and interprets the latter as sectoral. We find that the sectoral component explains much less of the variation in sectoral regional inflation rates and exhibits much less volatility than previous findings for the US indicate. We further contribute to the literature on price setting by providing evidence that country- and region-specific factors play an important role in addition to the sector-specific factors. We conclude that sectoral price changes have a “geographical” dimension, that leads to new insights regarding the properties of sectoral price changes.

JEL Classification: E31, C38, D4, F4

Keywords: *Disaggregated prices, euro area regional and sectoral inflation, common factor models*

Non-technical Summary

A central element of a majority of contemporary macroeconomic models is the assumption of nominal rigidities in goods markets. The rationale for incorporating price stickiness into these models is that there exists strong empirical evidence in favor of stickiness in prices at an aggregate level. Moreover, the empirical fit of models usually improves considerably when nominal rigidities are allowed for. A standard assumption in DSGE models is Calvo pricing, where firms adjust prices according to staggered contracts (time-dependent pricing). Alternative assumptions include state-dependent pricing, menu costs, information frictions or rational inattention. The relatively broad consensus about the importance of stickiness in nominal goods prices that emerged, has been challenged in recent years, however. Newer studies that analyze the behavior of micro price data have come to somewhat puzzling results: They find that these prices are not only very volatile, i.e. the frequency of price changes is high, but also exhibit low persistence (see e.g. Bils and Klenow, 2004, and Alvarez et al., 2006), in contrast to the findings concerning the behavior of aggregate data.

One explanation to reconcile the evidence on disaggregate and aggregate prices, has been presented by Boivin, Giannoni and Mihov (2008) and Mackowiak, Mönch and Wiederholt (2009c). These papers argue that the differences in inflation persistence at the aggregate and disaggregate level may be due to different responses of aggregate and sectoral prices to macroeconomic and sector-specific shocks. Decomposing a broad set of disaggregate sectoral price data into an aggregate and an idiosyncratic or sectoral component these authors find that the aggregate component exhibits considerable persistence but contributes only little to changes in sectoral prices. The sectoral component on the other hand shows no persistence but is very volatile and explains most of the movements in sectoral prices. Thus, the seemingly contradictory evidence on the different behaviour of disaggregate and aggregate prices can be attributed to the fact that the former are mostly determined by very volatile sectoral shocks with low persistence whereas the latter are pre-dominantly influenced by highly persistent aggregate shocks with low volatility.

Mackowiak et al (2009c) relate their findings to three different models of price-setting and ask whether any of these models is capable to explain the observed patterns of sectoral price changes. They show that both the Calvo- and the sticky-information model are compatible with the observed pattern of sectoral price dynamics only for extreme parameter values and conclude that the rational-inattention model fits the observed behavior of sectoral prices best since it postulates that firms react more to volatile sector-specific shocks than to aggregate macroeconomic shocks.

In this paper, we use a novel disaggregate sectoral euro area data set with a regional breakdown and suggest a new method to extract factors from over-lapping data blocks. We also show that this method has good properties in small samples. Using both the new disaggregate data set and our new method allows us to separately estimate an aggregate, sectoral and idiosyncratic component of price changes, thereby extending previous literature that decomposes price changes into an aggregate and an idiosyncratic component only, where the latter is interpreted as the sector-specific component. Since the sector-specific component in previous analyses is computed as an idiosyncratic component, it captures by construction the effects of all factors that

influence sectoral inflation rates including the actual sector-specific component and other non-sector-specific factors. One of these additional factors can be measurement errors, as acknowledged in the literature. We argue that other important, non-sector-specific elements in the residual component result from aggregating geographic-specific factors across regions. If these non-sectoral aspects play an important quantitative role in explaining the idiosyncratic component of sectoral price changes, the behavior of the sectoral component analyzed in the previous literature might not correspond to the behavior of the actual sector-specific component but might result from combining the effects of very different elements and might only be loosely related to actual sectoral elements. If we proceed as in the previous literature based on our euro area data set, we obtain aggregate and sector-specific components that behave very similar to the ones obtained in previous papers for US data.

Employing our newly proposed method to extract factors from over-lapping data blocks, we decompose our novel data set of euro area regional sectoral inflation rates into an aggregate, a sector-specific, a country-specific, a country-sector specific and an idiosyncratic component. This decomposition might imply quite different properties of the different components of inflation, and might therefore lead to different conclusions regarding the validity of different pricing models as in previous literature. For instance, higher persistence of the sectoral component would provide more support for the Calvo and sticky information model, and relatively lower volatility might imply less support for the rational inattention model.

It turns out that the sectoral component based on our new decomposition exhibits much less volatility than previous findings for the US indicate and explains much less of the variation in the data. In particular, we find that the sectoral component explains on average only about 14% and the country-specific sectoral component only about 21% of the overall volatility in sectoral regional prices. This is substantially less than the 85-90% explained volatility by sector-specific shocks found in previous studies for sectoral prices and in our results if we apply previous methods to our data set.

However, in line with previous US results, we find in our new decomposition that the sector-specific component exhibits little persistence on average, although persistence varies substantially across sectors. Since we find overall a clear negative relationship between the persistence and the volatility of the inflation components, our results largely confirm previous findings by Mackowiak et al (2009c) that the rational inattention model provides a plausible explanation of observed changes in sectoral prices since firms pay more attention to inflation components the more volatile they are, and react to it faster. The substantially lower volatility of our estimated sectoral component in comparison with previous studies should be noted, though.

We also find that country- and region-specific factors play an important role in addition to the sector-specific factors. The region-specific component, excluding other factors such as measurement error, explains about 13% of the overall variation of inflation rates. We find that economic characteristics of regions, such as the growth rate of the respective region and its competitiveness structure, show a significant link to the variance in regional sectoral inflation rates that is due to region-specific shocks, underlining that regional shocks are indeed an important driving force behind inflation developments.

Overall, our results suggest that previous findings that sectoral shocks to prices (or what was interpreted as sectoral shocks) are a dominant source of changes in sectoral prices need to be reconsidered. Disaggregate forces do play an important role in price determination, but sectoral shocks are complemented by regional (and for the euro area country-specific) shocks. However, our results provide suggestive evidence in favor of the rational-inattention model and against the Calvo and sticky-information model. The rational-inattention model might be adequate to allow for region-specific shocks since from our empirical analysis they appear on average to have similar volatility as sectoral shocks, with comparable relatively high standard error, and low persistence.

1 Introduction

A central element of a majority of contemporary macroeconomic models is the assumption of nominal rigidities in goods markets. The rationale for incorporating price stickiness into these models is that there exists strong empirical evidence in favor of stickiness in prices at an aggregate level. Moreover, the empirical fit of models usually improves considerably when nominal rigidities are allowed for. A standard assumption in DSGE models is Calvo pricing, where firms adjust prices according to staggered contracts (time-dependent pricing). Alternative assumptions include state-dependent pricing, menu costs, information frictions or rational inattention. The relatively broad consensus about the importance of stickiness in nominal goods prices that emerged, has been challenged in recent years, however. Newer studies that analyze the behavior of micro price data have come to somewhat puzzling results: They find that these prices are not only very volatile, i.e. the frequency of price changes is high, but also exhibit low persistence¹, in stark contrast to the findings concerning the behavior of aggregate data.

To reconcile the evidence on disaggregate and aggregate prices, several explanations have been put forward. One strand of the literature argues that the apparent persistence of aggregate inflation may be the result of an aggregation bias which arises as the consequence of aggregating heterogeneous sectoral price series.² Other authors such as Cogley and Sargent (2005) or Clark (2006) argue that the observed aggregate persistence of prices may reflect a structural break in the mean of inflation during the sample. A third explanation presented in Boivin et al. (2009) states that the differences in inflation persistence at the aggregate and disaggregate level may be due to different responses of aggregate and sectoral prices to macroeconomic and sector-specific shocks. Decomposing a broad set of disaggregate sectoral price data into an aggregate and an idiosyncratic or sectoral component these authors find that the aggregate component exhibits considerable persistence but contributes only little to changes in sectoral prices. The sectoral component on the other hand shows no persistence but is very volatile and explains most of the movements in sectoral prices. Thus, the seemingly contradictory evidence on the different behavior of disaggregate and aggregate prices can be attributed to the fact that the former are mostly determined by very volatile sectoral shocks with low persistence whereas the latter are pre-dominantly influenced by highly persistent aggregate shocks with low volatility.

¹See, e.g., the papers by Bils and Klenow (2004) or Alvarez et al. (2006).

²See, e.g., Granger (1980), Pesaran and Smith (1995) and Imbs et al. (2005).

The results by Boivin et al. (2009) are confirmed in a recent study by Mackowiak et al. (2009). Similar to Boivin et al. (2009) these authors decompose a large set of disaggregate monthly U.S. sectoral consumer price data into an aggregate and a sectoral component. They find that the sectoral component not only explains the bulk of variations in sectoral prices but that this component also shows no sign of persistence. In a second step, these authors relate their findings to three different models of price-setting and ask whether any of these models is capable to explain the observed patterns of sectoral price changes. The three models that the authors consider are multi-sector versions of the Calvo (1983) model, the sticky-information model à la Mankiw and Reis (2002) and the rational-inattention model by Mackowiak and Wiederholt (2009). They show that both the Calvo- and the sticky-information model are compatible with the observed pattern of sectoral price dynamics only for extreme parameter values and conclude that the rational-inattention model fits the observed behavior of sectoral prices best since it postulates that firms react more to sector-specific shocks than to aggregate macroeconomic shocks.

A different view is taken in Carvalho and Lee (2010) who develop a multi-sector sticky-price DSGE model that can endogenously deliver differential responses of sectoral prices to aggregate and sectoral shocks. In their model, sectoral labor market segmentation and input-output linkages produce a pricing interaction which is called “non-uniform” because it takes the form of a strategic complementarity in price setting across sectors and that of a strategic substitutability within sectors. The authors show that this non-uniform price interaction allows the model to match a wide range of sectoral price facts documented in Boivin et al. (2008) without the need of extreme assumptions, amongst them the empirically well-documented differential response of sectoral prices to aggregate and sectoral shocks.

The sector-specific component in previous analyses is computed as an idiosyncratic component. Hence, it captures by construction the effects of all factors that influence sectoral inflation rates but are not common to all of them. It might therefore represent a mixture of the actual sector-specific component and other non-sector-specific factors. One of these additional factors can be measurement errors, as Boivin et al. (2009) acknowledge. Other important non-sector-specific elements in the residual component result from aggregating geographic-specific factors across regions. If these non-sectoral aspects play an important quantitative role in explaining the idiosyncratic component of sectoral price changes, the behavior of the sectoral component which Boivin et al. (2009)

and Mackowiak et al. (2009) analyze might not correspond to the behavior of the actual sector-specific component but might result from combining the effects of very different elements. In other words, what these authors identify as sectoral components (and shocks) could be only loosely related to actual sectoral elements.

To shed light on this important issue, in this paper, we use a novel disaggregate sectoral euro area data set with a regional breakdown and develop a new method to extract factors from over-lapping data blocks. We can therefore estimate aggregate, sectoral, country-specific and regional components of price changes. This finer decomposition can imply quite different properties of the different components of inflation, leading to different conclusions regarding the validity of different pricing models. For instance, higher persistence of the sectoral component would provide more support for the Calvo and sticky information model, and relatively lower volatility might imply less support for the rational inattention model.

It turns out that the sectoral component now exhibits much less volatility than previous findings for the US indicate, and explains much less of the variation in the data. However, in line with previous US results, we find that the sector-specific component exhibits little persistence on average, although persistence varies substantially across sectors.

We also find a clear negative relationship between the persistence and the volatility of the inflation components, confirming previous findings by Mackowiak et al. (2009) and supporting the rational inattention model as a plausible explanation of observed changes in sectoral prices, since firms pay more attention and react faster to more volatile inflation components.

Regarding the role of the geographic dimension, an important finding is that the country- and in particular region-specific factors play a major role as drivers of regional sectoral prices. We also find that regional economic characteristics, such as the growth rate of the respective region and its competitiveness structure, have a significant influence on explanatory power of the regional factors, indicating that what we have extracted is indeed truly regional.

Given the international dimension of our data set it is natural to relate our results to findings in the literature on international pricing. That literature shows that prices across two markets behave very different when there is a national border between the two markets or not. Two of our findings might be particularly relevant in the context of that literature: First, our country-specific and country-specific sectoral components together

explain about 30% of the variance in the data while the regional component only explains about 13%. This result is, e.g., consistent with recent findings by Gopinath et al. (2011), who show that international borders create a substantially larger discontinuity in price changes than state and provincial boundaries. Second, we find that labor markets do not play a role in explaining the importance of regional factors for price changes. This is in line with another finding in Gopinath et al. (2011), that relative cross-border retail prices are mainly driven by changes in relative wholesale costs and not by local non-traded costs such as nominal wages.³

Overall, our results indicate that region-specific shocks are important for inflation dynamics in addition to sector-specific shocks, which is plausible in the framework of the rational inattention model, since it is intuitive that consumers or producers are more attentive to region-specific shocks than to aggregate shocks. In that sense the rational inattention model does encompass the existence of a relevant regional component in addition to sectoral price setting. Moreover, in the euro area there remains an important role for country-specific factors as drivers of price movements, in line with findings in the literature on international pricing. Our paper therefore is related to two strands of the literature that have received recent renewed interest, the literature on sectoral shocks and price setting and the literature on international pricing.

The rest of the paper is organized as follows: In Section 2 we shortly describe our data and provide some stylized facts on the extent of differences in inflation rates across sectors and regions in the euro area. In Section 3 we introduce the econometric framework used to analyze the determinants of changes in regional sectoral prices. Section 4 provides a Monte Carlo assessment of the small-sample properties of our proposed factor estimation algorithm, which turn out to be very good. In Section 5 we present and discuss the empirical results. In Section 6 we assess the robustness of our findings. Finally, in Section 7 we summarize our main findings and conclude.

2 Data and descriptive statistics

To determine and characterize the factors driving changes in sectoral prices in the European Monetary Union (EMU), we collected a large set of regional European sectoral price

³It should be noted that the analysis by Gopinath et al. (2011) is based on retail prices in the US and Canada, while we investigate euro area CPI inflation data.



index data. More precisely, we compiled a data set that includes sectoral consumer price index (CPI) data from six EMU member countries (Austria (AU), Germany (DE), Finland (FI), Italy (IT), Portugal (PO) and Spain (ES)), and that comprises a total of 61 locations, covering about 60% of the euro area in terms of GDP. The regions are the same as in Beck et al. (2009), where they analyze an all items data set with a regional breakdown.⁴ For each region, in addition to the all-items inflation considered in Beck et al. (2009), we have the following sectors: 1. food and non-alcoholic beverages (food); 2. alcoholic beverages, tobacco and narcotics (alco); 3. clothing and footwear (clot); 4. housing, water, electricity, gas and other fuels (hous); 5. furnishings, household equipment and routine household maintenance (furn); 6. health (heal); 7. transport (tran); 8. communication (comm); 9. recreation and culture (recr); 10. education (educ); 11. restaurants and hotels (hote). Overall, the data set includes 730 series, spanning the period 1995(1) to 2004(10) on a monthly frequency, non-seasonally adjusted and in index form.⁵

The inflation rate in a given country c , region r and sector s at time t denoted by $\pi_{c,r,s,t}$, is computed as the month-on-month proportional change in the (log of the) respective sectoral price index, $p_{c,r,s,t}$, i.e.,

$$\pi_{c,r,s,t} = \ln(p_{c,r,s,t}) - \ln(p_{c,r,s,t-1}), \quad (1)$$

with $c = 1, \dots, C$, $r = 1, \dots, R_c$, $s = 1, \dots, S_r$, and $t = 1, \dots, T$, and where C denotes the number of countries in our dataset, R_c denotes the number of regions in country c and S_r denotes the number of sectoral series available for region r .

For our econometric analysis, the data are seasonally adjusted, standardized and series with clear signs of structural breaks or shifts in variance are dropped. Moreover, outliers larger than 4 standard deviations are replaced by averages of the adjacent observations. We have also dropped Austria, since sectoral data are only available at a regional level since 2001. The resulting “cleaned” data set contains 418 series.

Table 1 reports descriptive statistics for the (unstandardized) data series included in this cleaned data set. Results are presented for all data series (*Total sample, All sectoral*) and subsamples which include all series from a given country (*Data grouped by countries*) or a given sector (*Data grouped by sectors*). Moreover, results are reported for the

⁴An overview of the regions included in our sample and the short names used in this paper is given in Tables A and B of Appendix A.

⁵For the remaining euro area countries comparable regional data are not available or at least not for a similar time span.

regional aggregate price indices (*Total sample, All aggregate*). Several interesting features of the reported statistics are noteworthy. Specifically, when looking at the total sample, we can see that there exists considerable heterogeneity in mean inflation rates across series. Moreover, similar to findings of studies on sectoral inflation, we find that regional sectoral inflation rates are on average very volatile but exhibit little or no persistence.⁶ However, results are different when we look at aggregate regional inflation rates. The degree of persistence is considerably higher,⁷ whereas the volatility and the cross-sectional dispersion are significantly lower. The degree of commonality on the other hand seems to be larger.

The numbers in the second and third panels of Table 1 show that there are considerable differences in (long-run) average inflation rates both across countries (reaching from about 1.1% for German sectoral inflation rates to about 2.6% for both Spanish and Portuguese inflation rates) and sectors (reaching from about 1.3% for clothing to about 2.9% for hotel). Moreover, for all groups in these panels we can observe that the regional sectoral inflation rates are both very volatile and show little persistence. Interesting insights are provided by considering the deviation of the average correlation of the inflation rates within a group from the aggregate inflation rate of a group.⁸ This statistic can be seen as a proxy measure for the degree of comovement in a given group. The results show that the extent of comovement for sectoral regional inflation rates is clearly higher when the series are grouped either by countries or sectors relative to the case when all series are taken into account. This indicates that regional sectoral inflation rates could not only be driven by sector-specific factors but that also country-specific factors could matter.

Table 2 reports descriptive statistics when the series of our sample are grouped by country-specific sectors. The reported numbers show that there is considerable dispersion in long-run average inflation rates across sectors even within countries. Volatility is large across national sectors and is comparable in size. Persistence on the other hand is always very low. The correlation is even higher than for the country-specific sectoral groupings.

Two final questions deserve an answer. First, to which extent has the “cleaning” pro-

⁶Persistence here is measured as the sum of the estimated coefficients of an AR model with 13 lags, following Boivin et al. (2009).

⁷It must be noted though that the observed degree of persistence is still considerably lower than that found in many other studies. One reason for this finding is probably related to our data sample period (1996 - 2004) for which other studies such as Altissimo et al. (2006) or Mishkin (2007) also found a relatively small degree of persistence in aggregate inflation.

⁸The aggregate inflation rate of a group is computed as a weighted average of the series included in the group, see footnotes to Table 1 for details.

cess changed the general pattern of our data? Tables C and D of Appendix ?? report descriptive statistics for the raw data. They show that the pattern of the results for mean values, persistence and within-group correlations is similar to that of the cleaned dataset. As could be expected, the numbers for volatility are smaller in the cleaned data set, since outlying values are eliminated from the latter. Overall, we can conclude that the “cleaning” process required to make the data suited for the subsequent econometric analysis did not alter their information content.

Second, are the sectoral regional inflation rates in the cleaned dataset stationary or integrated? Beck et al. (2009) run formal unit root tests on the all-items regional inflation series, but they do not obtain a definitive answer, since the single equation tests do not reject non-stationarity in most cases while the panel tests systematically reject non-stationarity. Hence, they perform the analysis for both the levels and the first differences of inflation, finding qualitatively similar conclusions. Based on this result and on the fact that the average persistence measures reported in Table 2 are low, we focus on the levels of the inflation series.

In summary, the descriptive analysis of this Section, based on a new dataset for the euro area with both a regional and a sectoral breakdown, confirms previous findings that sectoral price changes are not only very volatile but also exhibit little persistence. Our results furthermore indicate that changes in sectoral prices seem to have a “geographical” dimension that has not been explored in the literature thus far.

3 Econometric methodology: A new approach

3.1 The model

To analyze the determinants of changes in sectoral prices previous studies have proposed to decompose $\pi_{c,r,s,t}$ as follows:⁹

$$\pi_{c,r,s,t} = \alpha_{c,r,s} f_t^a + u_{c,r,s,t} \quad (2)$$

where $\alpha_{c,r,s} f_t^a$ represents the aggregate component related to macroeconomic developments while $u_{c,r,s,t}$ is interpreted as the sector-specific component. Based on this decom-

⁹See, e.g., equation (2) of Boivin et al. (2009) or equation (1) of Mackowiak et al. (2009). Inflation rates are demeaned and their variances are normalized to one before estimation.

position, the statistical properties of both the aggregate and sector-specific components are then examined, and the relative contribution of each component to the overall volatility of $\pi_{c,r,s,t}$ is determined. Using this approach, previous studies have found that the aggregate component exhibits relatively low volatility but high persistence, while the sector-specific component displays high volatility and no persistence. Moreover, the latter is found to explain about 85-90% or more of the movements in $\pi_{c,r,s,t}$, and therefore sectoral inflation rates essentially behave like their sector-specific component.

One problematic aspect of this methodological approach is that the sector-specific component $u_{c,r,s,t}$ is computed as a residual variable, and therefore it captures the effects of all elements which influence sectoral inflation rates but are not common to all of them. In other words, a (possibly large) part of $u_{c,r,s,t}$ could be totally unrelated to sectoral movements.

The use of our regional sectoral inflation rates allows us to decompose the residual term $u_{c,r,s,t}$ further, and to explicitly extract a sectoral factor whose characteristics and relative importance in explaining variations in $\pi_{c,r,s,t}$ we can then analyze. More specifically, we decompose $u_{c,r,s,t}$ as follows:

$$u_{c,r,s,t} = \beta_{c,r,s}f_t^c + \gamma_{c,r,s}f_t^s + \delta_{c,r,s}f_t^{sc} + e_{c,r,s,t} \quad (3)$$

and therefore analyze the following model for $\pi_{c,r,s,t}$:

$$\pi_{c,r,s,t} = \alpha_{c,r,s}f_t^a + \beta_{c,r,s}f_t^c + \gamma_{c,r,s}f_t^s + \delta_{c,r,s}f_t^{sc} + e_{c,r,s,t}. \quad (4)$$

In this equation, f_t^a are k_a aggregate factors common to all of the units (e.g., related to monetary policy, raw material prices, or external developments), f_t^c are k_c country-specific factors that only affect variables in country c (e.g. fiscal policy or nation-wide labour market legislation), f_t^s are k_s sector-specific factors that only affect variables in sector s (e.g. tariffs decided at the European Union level on goods belonging to a specific sector or increases in the costs of inputs specific to a given sector), and f_t^{sc} are k_{sc} sector- and country-specific factors that only affect variables in sector s of country c (e.g. changes in value added taxes for goods in a specific sector or the implications of sectoral wage bargaining at the national level). $e_{c,r,s,t}$ denotes the remaining idiosyncratic component that includes measurement error and, importantly, a regional component as we will argue.¹⁰

¹⁰To motivate our empirical decomposition theoretically one could proceed analogously to Mackowiak et al. (2009). These authors model the price-setting decisions of monopolistic competitive firms and show that changes in sectoral prices are determined by aggregate and sectoral factors (in an additive fashion). If one uses their setting, assumes that regional goods markets are segmented and allows, e.g., for region-specific shocks to production conditions and/or wage-setting one can derive an expression which shows that changes in regional sectoral prices are determined as the sum of region-specific, sector-specific and aggregate factors.

The factors within each group are assumed to be orthonormal, and the factors across groups are assumed to be uncorrelated with each other. The factors are also assumed to be uncorrelated with the idiosyncratic term $e_{c,r,s,t}$, which has limited correlation across units and over time in order to satisfy the conditions in Stock and Watson (2002a) and Stock and Watson (2002b). Under the assumptions we have made, the model is identified, which makes the loadings and the factors estimable.

Additional details on the relationship between the previous and our new detailed decomposition can be found in Appendix A, while the following sections of the paper present an estimation procedure for the model discussed above, as well as some Monte Carlo experiments that investigate its small sample properties. Readers more interested in the economic analysis than in the technical details can skip the next sections and go directly to Section 5.

3.2 Estimation of a factor model for over-lapping data blocks

To estimate the different types of factors in (4), we extend the previous literature on extracting factors from non-overlapping data-blocks¹¹ to over-lapping data blocks. A parametric approach combined with Maximum-Likelihood estimation could be applied, see e.g. Koopman and Jungbacker (2008). However, given the complex structure of our estimation problem, with a very high number of factors to be estimated and uncertain correlation structure in the idiosyncratic components, a non-parametric procedure provides a more robust alternative. Hence, we develop a modified version of the non-parametric principal component based estimator of Stock and Watson (2002a) and Stock and Watson (2002b).

Starting with the aggregate factors f_t^a , which influence all variables under analysis, Stock and Watson's method can be directly applied. Therefore, the k_a estimated factors \hat{f}_t^a coincide with the first k_a principal components of $\pi_{c,r,s,t}$.

¹¹ See e.g. Kose et al. (2003), Beck et al. (2009), Moench et al. (2009), Diebold et al. (2008) and Stock and Watson (2008).

Let us consider now the country-specific factors f_t^c . We might think of using as estimators the first k_c principal components of all variables for each country $c = 1, \dots, C$. However, these principal components would depend on f^a , and therefore the resulting estimators of f_t^c would be correlated with those of f_t^a , mixing aggregate and country information. To tackle this problem we could take the principal components of $\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a$ for each country, where the loadings $\widehat{\alpha}_{c,r,s}$ are obtained by OLS regressions of $\pi_{c,r,s,t}$ on the estimated factors \widehat{f}_t^a . The use of the estimated rather than true aggregate factors requires the total number of variables ($N = \sum_{c=1}^C \sum_{r=1}^{R_c} \sum_{s=1}^{S_r} 1$) to be large and to grow faster than the number of observations (T); in particular, it should be $\sqrt{T}/N \rightarrow 0$, see Bai and Ng (2002) for details. The use of the estimated rather than the true loadings is justified by the consistency of the OLS estimator when T diverges.

In order to estimate the sector-specific factors f_t^s , we could follow a similar procedure and use as estimators the first k_s principal components of $\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a$ for each sector. However, since some of the observations in $\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a$ are used to construct both the estimators of f_t^c and those of f_t^s , the resulting estimated factors would be correlated, in contrast with the assumption of no correlation between f_t^c and f_t^s . Therefore, we need an additional modification to estimate f_t^c and f_t^s .

Let us therefore consider the model

$$\begin{aligned} & \frac{1}{S_r} \sum_{s=1}^{S_r} \left(\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a \right) \stackrel{asympt}{=} \frac{1}{S_r} \sum_{s=1}^{S_r} \left(\pi_{c,r,s,t} - \alpha_{c,r,s} f_t^a \right) = \\ & = \left(\frac{1}{S_r} \sum_{s=1}^{S_r} \beta_{c,r,s} \right) f_t^c + \frac{1}{S_r} \sum_{s=1}^{S_r} \gamma_{c,r,s} f_t^s + \frac{1}{S_r} \sum_{s=1}^{S_r} \delta_{c,r,s} f_t^{sc} + \frac{1}{S_r} \sum_{s=1}^{S_r} e_{c,r,s,t}. \end{aligned}$$

If S_r is large, since the sector-specific factors f_t^s are orthogonal across sectors by assumption, the term $\frac{1}{S_r} \sum_{s=1}^{S_r} \gamma_{c,r,s} f_t^s$ vanishes. Hence, for each country, we suggest to estimate the country-specific factors as the first k_c principal components of the R_c ($c = 1, 2, \dots, C$) variables $\frac{1}{S_r} \sum_{s=1}^{S_r} \left(\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a \right)$, which are also no longer dependent on the sector specific factors when S_r is large. Then, for each sector, the sector specific factors can be estimated as the first k_s principal components of the $\sum_{c=1}^C \sum_{r=1}^{R_c} I(r_s)$ variables

$$\pi_{c,r,s,t} - \widehat{\alpha}_{c,r,s} \widehat{f}_t^a - \widehat{\beta}_{c,r,s} \widehat{f}_t^c. \quad ^{12}$$

¹² $I(r_s)$ represents a dummy variable equal to one if data for the considered sector s are available in region r and equal to zero if no data for sector s are available for region r .

This procedure requires the number of sectors S_r to be large. When this is not the case, an iterative method can produce better results. In the first step, f_t^c and f_t^s are estimated as indicated in the previous paragraph, which yields \hat{f}_t^{c1} and \hat{f}_t^{s1} . In the second step, the residuals $\pi_{c,r,s,t} - \hat{\alpha}_{c,r,s}\hat{f}_t^a - \hat{\gamma}_{c,r,s}\hat{f}_t^{s1}$ are computed, and their first k_c principal components are used to construct \hat{f}_t^{c2} . Notice that this is an alternative method to get rid of the correlation between \hat{f}_t^c and \hat{f}_t^s . In the third step, the residuals $x_{c,r,s,t} - \hat{\alpha}_{c,r,s}\hat{f}_t^a - \hat{\beta}_{c,r,s}\hat{f}_t^{c2}$ are computed, and their first k_s principal components are used to construct \hat{f}_t^{s2} . In the fourth step, the residuals $\pi_{c,r,s,t} - \hat{\alpha}_{c,r,s}\hat{f}_t^a - \hat{\gamma}_{c,r,s}\hat{f}_t^{s2}$ are computed, and their first k_c principal components are used to construct \hat{f}_t^{c3} . The procedure continues like this until successive estimates of the factors are sufficiently close. In particular, we suggest to stop the iterations when $\max_c \left\{ \max_t |\hat{f}_t^{c,i} - \hat{f}_t^{c,i-1}| \right\} < 0.001$ and $\max_s \left\{ \max_t |\hat{f}_t^{s,i} - \hat{f}_t^{s,i-1}| \right\} < 0.001$.

The final set of factors to be estimated are the country- and sector-specific factors f_t^{sc} . For each sector in a given country, we use as estimators the first k_{sc} principal components of the $\sum_{r=1}^{R_c} I(r_s)$ variables $\pi_{c,r,s,t} - \hat{\alpha}_{c,r,s}\hat{f}_t^a - \hat{\beta}_{c,r,s}\hat{f}_t^c - \hat{\gamma}_{c,r,s}\hat{f}_t^s$ (i.e., for a given country, the dataset is composed of a given sector for each region).

In the presentation so far, we have considered the number of factors as known. To relax this assumption, the various k_{iS} can be determined on the basis of a proper information criterion. We will follow the method proposed by Bai and Ng (2002) in our empirical analysis.

4 Monte-Carlo simulations

In this Section we assess the small sample performance of our factor estimation method in the presence of a block structure in the loading matrix, namely, when there are factors that only affect subgroups of the variables, as in the case of the country specific or sector specific factors. The first subsection presents the basic Monte Carlo design and associated results. The second subsection discusses results for a variety of modifications of the design. The final subsection compares the so far standard approach of decomposing sectoral inflation rates into an aggregate and a sectoral component with our proposal of further decomposing the latter in order to identify the truly sectoral elements.

4.1 Base case

We assume that the inflation rate of region r in country c and sector s is given by:

$$\pi_{r,c,s,t} = \alpha_{r,c,s}f_t^a + \beta_{r,c,s}f_t^c + \gamma_{r,c,s}f_t^s + e_{c,s,t}. \quad (5)$$

In the base case we suppose that there are 2 countries and 2 sectors, with 30 regions in each country. Therefore, (5) can be written in matrix notation as

$$X_t = \mathbf{A}\mathbf{F}_t + e_t$$

where X_t is of dimension 120×1 , where $N = 2 * 2 * 30 = 120$, while the \mathbf{A} matrix of loadings is 120×5 , the \mathbf{F}_t matrix containing the factors at time t is 5×1 (since there is one aggregate factor, two country factors and two sector factors), and the idiosyncratic errors are grouped in the 120×1 vector e_t .

Due to the specific factor structure, the loadings matrix \mathbf{A} is specified as follows

$$\mathbf{A} = \begin{bmatrix} \alpha_{1,1,1} & \beta_{1,1,1} & 0 & \gamma_{1,1,1} & 0 \\ \alpha_{2,1,1} & \beta_{2,1,1} & 0 & \gamma_{2,1,1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{30,1,1} & \beta_{30,1,1} & 0 & \gamma_{30,1,1} & 0 \\ \alpha_{1,1,2} & \beta_{1,1,2} & 0 & 0 & \gamma_{1,1,2} \\ \alpha_{2,1,2} & \beta_{2,1,2} & 0 & 0 & \gamma_{2,1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{30,1,2} & \beta_{30,1,2} & 0 & 0 & \gamma_{30,1,2} \\ \alpha_{1,2,1} & 0 & \beta_{1,2,1} & \gamma_{1,2,1} & 0 \\ \alpha_{2,2,1} & \vdots & \beta_{2,2,1} & \gamma_{2,2,1} & 0 \\ \vdots & 0 & \vdots & \vdots & \vdots \\ \alpha_{30,2,1} & 0 & \beta_{30,2,1} & \gamma_{30,2,1} & 0 \\ \alpha_{1,2,2} & 0 & \beta_{1,2,2} & 0 & \gamma_{1,2,2} \\ \alpha_{2,2,2} & \vdots & \beta_{2,2,2} & 0 & \gamma_{2,2,2} \\ \vdots & 0 & \vdots & \vdots & \vdots \\ \alpha_{30,2,2} & 0 & \beta_{30,2,2} & 0 & \gamma_{30,2,2} \end{bmatrix},$$

and each non-zero element of \mathbf{A} is drawn from a standard normal distribution. Each factor is instead generated as an $AR(1)$ process with persistence 0.8 and standard normal errors, and the factors are independent. The idiosyncratic errors are also independent and each of them is standard normal. The sample size is $T = 100$, and we run $R = 1000$ simulations.

We compare the performance of the standard principal component based factor estimators introduced by Stock and Watson (2002a, 2002b) and of our procedure proposed in Section 3. We consider three evaluation criteria. First, the correlation between the true and estimated factors. Second, the true and estimated persistence of the factors. Third, the true and estimated percentage of variance explained by the factors. These are three basic ingredients for the economic analysis that we conduct with our model, and it is therefore important to assess the reliability of our proposed estimation method with reference to them. Note also that since in this context each common aggregate, country and sectoral component of each variable is just equal to the factor multiplied by a constant, the results on the correlation and persistence of the factors translate directly to the components.

In Tables 3 - 5 we present the results for the three criteria. We report both the mean and selected percentiles of the empirical distribution of the criteria over the R replications. The latter information is important to assess the robustness of the estimation method.

Four main findings emerge. First, and obviously, the values for the aggregate factor are equal for the two estimation methods, and therefore we focus on the country and sector factors. Second, in terms of correlation with the true factors, Table 3 highlights that our estimation method provides much higher values than the unrestricted Stock and Watson approach, not only in terms of averages but also of all the percentiles of the distribution. The average correlation for our method is around 0.80, compared with about 0.40 for the unconstrained principal component estimator. Even more important, the 25th percentile is about 0.74 for us and 0.20 for the unconstrained estimator, so that there is a non-negligible percentage of cases where the latter yields estimated factors fairly different from the true ones. We obtain similar results for the sectoral factors. Third, in terms of estimated persistence, from Table 4 it emerges that the two methods are fairly similar for the country factors, but the values are higher and closer to the true values for the sectoral factors. The median values for the four country and sector factors are in the range 0.65-0.67 for our approach versus 0.47-0.77 for the unconstrained-principal-component approach. In both cases the values are slightly smaller than the theoretical value (0.8). Hence, in practice, it could be that the country and sectoral shocks are slightly more persistent than what turns out from the model estimation. Finally, Table 5 indicates that the standard approach

underestimates on average the explanatory power of the sectoral factors (about 10% versus a true value of 0.27%), and overestimates the role of the idiosyncratic components (37% versus a true value of about 18%). Our approach is biased in the same directions, but the extent of the problem is much smaller, 22% versus 27% for the sectoral factor and 24% versus 18% for the idiosyncratic error. Therefore, the sectoral component could be slightly more relevant than what results from the estimation of our empirical model. However, to support such a conclusion we need to verify that the results we have obtained are robust to modifications in the experimental design.

4.2 Additional experiments

The results reported so far are quite good, but we need to assess their robustness to a variety of changes in the experimental design. In particular, we consider a number of modifications of the data generating process which could all deteriorate the performance of the factor estimation methods, and in some cases could make it more difficult to distinguish between the common and the idiosyncratic component. These experiments include a reduction in the persistence of the factors, lower volatility of the factors, larger variance for each idiosyncratic error, a decrease in the number of regions in each country, a decrease in the temporal dimension, the use of a uniform rather than standard normal distribution to draw the non-zero elements of the loading matrix, and an increase in the number of countries and sectors from 2 to 3.

The results of all these experiments are summarized in Tables 8 - Table 10 and discussed in detail in the Not-for-publication Appendix B. Basically, the performance of the estimation method deteriorates as expected, but it remains quite good.

We have also carried out a bootstrap exercise where we use a data generating process similar to the estimated model in our empirical analysis, with 1 aggregate, 5 country and 9 sectoral factors. The results are largely similar to those reported so far, despite the complex structure of the model. The uncertainty of the estimates is higher, not surprisingly given the complexity of the model, but the correlation between the estimated and the true factors remains high, e.g. for the aggregate factor it is 0.95.¹³

Overall, the results of the set of experiments we have conducted highlight the importance of modifying the standard principal component factor estimator in the presence of a block structure for the matrix of loadings. Our approach substantially improves the correlation between the estimated

¹³Tables with results for this bootstrap experiment are available upon request.

and the true factors, as well as their estimated persistence and explanatory power, though the persistence remains slightly underestimated and the role of the idiosyncratic component slightly overestimated.

5 Empirical results

In this section we present the results from decomposing changes in regional sectoral prices into their determinants, as discussed in the previous Sections. We start with reporting the results for the standard approach that decomposes sectoral regional inflation rates into an aggregate and an idiosyncratic component only. Afterwards, the findings for our more disaggregate decomposition of sectoral price changes as shown in equation (4) are discussed. Finally, we investigate in more details the role and determinants of the regional component.

5.1 Aggregate-sector decomposition

The first two columns of Table 6 report results for the case where changes in sectoral regional prices are decomposed into an aggregate and an idiosyncratic component only. Thus, in this case we proceed analogously, e.g., to Boivin et al. (2009) and Mackowiak et al. (2009)¹⁴ and first extract the aggregate component from the inflation rates and then treat the residuals from regressing actual price changes on the estimated aggregate factor, denoted by $u_{c,r,s,t}$, as the sector-specific component. Since the Bai and Ng (2002) criterion indicates $k_a = 1$, the reported results are based on a model with one area-wide factor only. The characteristics of the so obtained aggregate and sector-specific components are very similar to those obtained by the above mentioned studies. We find, e.g., that the sector-specific component is on average more than four times more volatile than the aggregate component. For the median volatility the difference in volatility is even larger (by a factor of almost six). The persistence numbers show that the sector-specific component exhibits basically no persistence (the mean persistence parameter takes a value of -0.050, the median value is 0.071), whereas the aggregate component displays considerably more

¹⁴See also Mackowiak and Smets (2009) for an analysis of inflation in the euro area, and Foerster et al. (2008) for an analysis of industrial production using related decompositions.

persistence (mean/median persistence value of about 0.3).¹⁵ Concerning the relative importance of the aggregate and the sector-specific component for explaining changes in sectoral prices, our results also confirm previous findings. The numbers in the first two columns of the last panel of Table 6 show that the aggregate component explains only very little of observed changes in sectoral prices (only about 8%), whereas the idiosyncratic component $u_{c,r,s,t}$ explains the remaining 92% and therefore is the dominant determinant of sectoral regional inflation.

Overall, the findings for the decomposition of sectoral regional inflation rates into an aggregate and an idiosyncratic component suggest that the extremely low persistence in sectoral regional inflation rates documented in Table 1 is due to the fact that sectoral regional inflation rates are almost exclusively driven by the nonpersistent idiosyncratic component, interpreted as sectoral component in previous studies.

Moreover, our results are in line with the conclusions drawn by Mackowiak et al. (2009) with respect to the plausibility of leading price-setting models. As in their paper, our idiosyncratic component (interpreted as the sectoral component in this simple decomposition) basically behaves like a white-noise process and our aggregate component exhibits some autocorrelation. These findings imply that sectoral indices immediately respond to sectoral shocks but only gradually to aggregate shocks. As Mackowiak et al. (2009) show, such a quick response to a sectoral shock is not compatible with the Calvo and the sticky-information price-setting models for plausible parameter values, but it is compatible with the rational-inattention model. Compatibility with the rational-inattention model is due to the fact that the idiosyncratic component, interpreted as the sectoral component, is considerably more volatile than the aggregate component.

5.2 Aggregate-sector decomposition: Our approach

As discussed in the Section 3, the results we got on the role of the sectoral component might no longer hold if the idiosyncratic component $u_{c,r,s,t}$ in equation (4) does not only represent the sector-specific element but is a mixture of different factors. In fact, since $u_{c,r,s,t}$ is obtained by “cleaning” the sectoral regional inflation rates from the aggregate

¹⁵Our numbers for the persistence of the aggregate component are substantially smaller than those reported, e.g., by Boivin et al. (2009). One major reason for this difference is that our data sample is different. If we restrict the data by Boivin et al. (2009) to a sample period comparable to ours, we obtain a significant drop in the persistence of the aggregate component. Evidence in favor of a substantial drop in the persistence of U.S. inflation in recent years is discussed in Mishkin (2007), for the euro area an analogous discussion is contained, e.g., in Altissimo et al. (2006).

component, it captures the effects of any factors that influence the respective sectoral prices and are not common to all sectoral prices. Potential determinants of changes in sectoral prices that are not common, but also not sector-specific, can be idiosyncratic measurement errors or what we call geography-related factors. The latter include country-specific factors, such as national fiscal policy or nation-wide labour market legislation, or country-sector-specific factors, such as changes in value added taxes for goods in a specific sector in a given country. As a consequence, the properties of the true sector-specific component might be considerably different from the characteristics obtained for $u_{c,r,s,t}$, which has previously been referred to as the sector-specific component.

To disentangle the impact of the various determinants of $u_{c,r,s,t}$, we decompose it into a country-specific (C), a sector-specific (S), a country-specific sectoral (CS) and an idiosyncratic component (Idios.), as discussed in Section 3. The results for this decomposition are reported in columns three to six of Table 6. Since the Bai and Ng (2002) criteria systematically indicate $k_i = 1$, the reported results are based on a model with one area-wide, one country, one sector, and one country-sector factor.

The figures show that we must modify our above drawn conclusions concerning the behavior and the relative importance of the sector-specific component for explaining changes in sectoral regional price changes and the conclusions from previous results in the literature using the simple decomposition of sectoral price changes into a macroeconomic and a sector-specific component. Whereas we confirm previous findings that the sectoral component is on average more volatile than the aggregate component, the volatility of the sectoral component is considerably smaller than in other papers, less than 0.4. Moreover, even though our estimated sectoral component still displays very low persistence on average, the difference in persistence relative to the aggregate component has become considerably smaller compared to previous findings. The median persistence of the sectoral component is now about 0.15 in comparison to a median persistence of about 0.3 for the aggregate component.

The numbers for the sector-specific component in the last panel of Table 6 show that the sector-specific component explains on average only about 15% of the overall variance in regional sectoral price changes. The number increases to about 35% when adding the contribution of the country-specific sectoral factor. However, even in this case it is still far below the 92% found using the previous decomposition.

To sum up, our results for the sector-specific component differ significantly from previous findings in important dimensions. The relatively low volatility together with the small proportion of overall variance explained by the sector-specific component suggests that the sector-specific component is not the main driving force explaining movements and characteristics of sectoral regional price changes.

The question then arises which of the remaining elements in the idiosyncratic component $u_{c,r,s,t}$ is the major driving force behind changes in sectoral prices? The numbers in the third panel of Table 6 show that it is the region-specific idiosyncratic component $e_{c,r,s,t}$ that by far explains most of the overall variation in sectoral prices (about 47% on average). Given its relatively high volatility and its low (on average negative) persistence, we can conclude that $e_{c,r,s,t}$ is indeed the variable that predominantly determines the behavior of sectoral price changes. From an economic point of view, the idiosyncratic component can basically capture two effects: First, it can reflect measurement errors and secondly, it can reflect the reaction of price-setters to local conditions.¹⁶ We will come back to this issue in the next subsection.

Another noteworthy feature of our decomposition results concerns the behavior and the role of the country-specific factors. The third panel of Table 6 shows that the country-specific factors explain almost as much of overall volatility in sectoral prices as the pure sector-specific factors do. Moreover, on average they appear to be as volatile as the sector-specific components. However, they are considerably more persistent than either the sector-specific and even the aggregate components. To understand this result it is instructive to consider the potential factors underlying the country-specific components. As we argued in Section 3, we think that national fiscal policies and nation-wide labour market legislation are potential causes for the existence of country-specific factors.

In terms of the economic implications of our empirical findings, we think they provide suggestive evidence in favor of the rational-inattention model by Mackowiak and Wiederholt (2009). Apart from the country-specific component, there is a clear negative relationship between the persistence of a factor and its volatility, as suggested by this model: The more volatile a factor is, the more attention firms pay to it, and the faster they react to it. Moreover, the relevance of the country-specific factors as drivers of price movements is in line with findings in the literature on international pricing, see e.g. Gopinath et al. (2011).

¹⁶When examining the factors driving regional output fluctuations in the U.S. Clark (1998) also found that regional factors play a very important role in addition to industry composition.

Summarizing, the results of this subsection suggest that the sectoral component exhibits much less volatility than previous findings for the US indicate, and explains much less of the variation in the data. Country factors and, even more, region-specific factors play an important role in addition to the sector-specific factors. The existence of a relevant country- and region-specific component has important implications for previously obtained results in the literature. However, our results still support the rational-inattention model by Mackowiak and Wiederholt (2009), while they are rather in disagreement with the Calvo- and the sticky-information models.

5.3 Analysis of the regional component

We have seen that the region-specific idiosyncratic component $e_{c,r,s,t}$ is the variable that predominantly determines the behavior of sectoral price changes. We have argued that the idiosyncratic component can basically capture two effects: First, it can reflect measurement errors and, secondly, it can be related to the reaction of price-setters to local conditions. We now analyze this issue further. In doing so, we first try to identify the factors common to all sectoral inflation rates of a given region, then examine their relative importance and major time-series characteristics, and finally relate them to local economic conditions as suggested by economic theory.

5.3.1 How important is the regional component?

The region-specific idiosyncratic component $e_{c,r,s,t}$ identified above is obtained by “cleaning” changes in regional sectoral inflation rates by a variety of different factors. Nevertheless, it still represents a composition of at least three different factors, namely a factor common to all sectoral prices of the given region, a region-sector-specific factor and a “truly” idiosyncratic component including measurement error. We will now estimate the first of these three components and analyze its relative importance and its major time-series characteristics. To this end, we decompose $e_{c,r,s,t}$ as follows:

$$e_{c,r,s,t} = \delta_{c,r,s} f_t^{rc} + \varepsilon_{c,r,s,t}. \quad (6)$$

In this equation, f_t^{rc} are k_{rc} region-specific factors that only affect variables in region r of country c . $\varepsilon_{c,r,s,t}$ denotes the remaining idiosyncratic component. We have outlined the assumptions under which such a model is identified and the loadings and factors can be estimated in Section 3. The same assumptions are made here. To estimate the regional

factors we apply again the non-parametric principal component based estimator of Stock and Watson (2002a) and Stock and Watson (2002b), i.e. the k_{rc} factors \hat{f}_t^{rc} are estimated as the first k_{rc} principal components of the estimated $e_{c,r,s,t}$. To determine the number of regional factors we use also in this context the method proposed by Bai and Ng (2002), which suggests $k_{rc} = 1$ for all regions.

The results for the decomposition are reported in the last two columns of Table 6. The reported values indicate that the regional component explains on average a substantial part of the residual component $e_{c,r,s,t}$, about 25% (which corresponds to about 13% of the variability in the sectoral inflation rates), though there is substantial heterogeneity across regions. The volatility of the regional component is considerably smaller than that of $e_{c,r,s,t}$ and is comparable in size to that of the aggregate component. As $e_{c,r,s,t}$, it does not exhibit any persistence. The decomposition results also suggest that the idiosyncratic component cleaned for the region-specific influences still explains most of the changes in regional sectoral inflation rates (around 35%). It is the most volatile component and does not exhibit persistence.

5.3.2 Determinants of the regional component

In the last section we have seen that the regional component explains a substantial proportion of changes in regional sectoral prices (on average about 13%). We now want to examine potential economic determinants of the importance of the regional factors. In the following we therefore examine the influence of a variety of region-specific variables such as differences in regional economic structures, differences in regional economic developments, differences in market sizes and differences in the competitive structures of the regional economies on the relative importance of the regional factors. Unfortunately, the extent to which economic variables are available at a regional level in the euro area is limited. In particular, the frequency at which these data are collected is very low. As a consequence, we are not able to examine the question under consideration in a panel context but are only able to perform a purely cross-sectional analysis. As our dependent variable we choose the average variance explained by the regional factors for each region. To approximate the other economic influences we choose the following variables: (i) size of the service sector (SERV_r, to approximate differences in the regional economic structures), (ii) average GDP growth (DGDP_r, to approximate differences in regional economic (business cycle) developments), (iii) (log) of population (LPOP_r, to approx-

imate differences in market sizes), (iv) number of business units per regions (DENS_r, to approximate differences in the competitive structures of the regional economies). The data source is Eurostat's Regio database with annual data frequency on those variables.

The estimation results are as follows (White HAC standard errors are given in parenthesis):

$$\begin{aligned} Var_t^r &= \underbrace{0.195}_{(0.071)} + \underbrace{0.100}_{(0.054)} SERV_r - \underbrace{0.096}_{(0.028)} DGDP_r - \underbrace{0.007}_{(0.004)} LPOP_r + \underbrace{0.004}_{(0.001)} DENS_r + \underbrace{\varepsilon_r}_{(0.026)} \\ R_{adj}^2 &= 0.395 \end{aligned}$$

where Var_t^r represents the squared loadings from the regional factors, i.e. the average variance explained, that reflects the importance of the regional factors.

These results show that the average variance explained by the regional factors we have estimated can be related to plausible regional economic characteristics. All coefficients of the variables we included are statistically significant, at least at a 10% significance level. GDP growth and market density are highly statistically significant at a 1% significance level.

We find that market density exhibits a positive sign, indicating that higher local competition leads to a relatively higher importance of region-specific factors for local price setting. Furthermore, a larger size of the service sector in a region corresponds to a relatively higher importance of local factors, confirming the importance of asymmetric shocks due to sectoral specialization in different regions. Since services include a large part of non-tradable goods, it is reasonable for it to explain regional differences in variability of inflation.¹⁷ It is also reasonable that different business cycle developments as measured by different growth rates explain regional differences in inflation dynamics. If more dynamic markets are relatively higher integrated to other markets then this should be reflected in a relatively lower importance of regional factors. We also find that in relatively larger markets the relative importance of regional factors tends to be smaller.

We also find that labor markets do not play a role in explaining the importance of regional factors for price changes. This can be seen as consistent with the finding in

¹⁷Beck et al. (2009) did a similar cross-sectional regression of the level of all items regional inflation rates and also found that a number of reasonable regional economic characteristics including inflation in non-wage input factors as well as an indicator of competitive structure had a significant impact.

Gopinath et al. (2011) for the US and Canada that relative cross-border retail prices are mainly driven by changes in relative wholesale costs and thus local non-traded costs such as nominal wages do not seem to play an important role.¹⁸

An additional argument for the regional component in sectoral inflation to be important is that substitutability within regions is larger than across regions within a sector. For instance, if restaurant prices increase, substitution of consumption is likely to increase local supermarkets sales rather than restaurant revenues in other regions. This kind of substitution effect depends on the demand elasticity of the respective sector and region.¹⁹

Overall, our results indicate that economic characteristics of regions show a significant link to the variance in regional sectoral inflation rates that is due to region-specific shocks, underlining that regional shocks are indeed an important driving force behind inflation developments.

6 Robustness analysis: Month-on-month versus year-on-year changes

In this section we consider whether our results can be affected by the presence of a weak factor structure and the use of month on month rather than year on year inflation.

A potential problem for the reliability of the empirical results concerns the very low proportion of variance explained by the aggregate factor, about 8%. While this result is also found by other authors who analyze the behavior of sectoral prices²⁰, it casts some doubts on the appropriateness of the performed factor analysis. Indeed, Onatski (2006) and Kapetanios and Marcellino (2006) show that when the factor structure is weak (i.e., the fraction of variance explained by the first principal component is very small), the

¹⁸The results in Gopinath et al. (2011) indicate, however, that unconditionally a substantial fraction of the movements in cross border prices is accounted for by relative movements in retail markups. However, conditionally on the nominal exchange rate they find that the variation in the retail price gap at the border is almost entirely driven by variation in wholesale costs, not by variation in markups; for related evidence, see Eichenbaum et al. (2008). The importance of mark-ups might also be related to the degree of competition in a region that we find to be significant for the importance of regional factors.

¹⁹The regression discontinuity approach underlying the analysis by Gopinath et al. (2011) assumes that equilibrium prices depend on many local factors such as the elasticity of substitution across stores, or demographic characteristics, all of which impact the effective transaction costs for a household and can vary with location.

²⁰Mackowiak et al. (2009), e.g., report that the first common component explains about 7% of the overall variation in their data, Boivin et al. (2009) find that the first five principal components of their data sample explain only about 15% of overall variation.

principal component based estimator of the factor is no longer consistent. Intuitively, there is too little commonality to separate what is common from what is idiosyncratic.

We therefore redo our analysis using year-on-year changes in sectoral prices. This transformation averages out some of the idiosyncratic variation in the month-on-month series, thus strengthening the factor structure. The choice of this transformation has two additional positive side aspects. First, the year-on-year inflation rate is the key variable for monetary policy and, secondly, the twelve difference operator is also useful to remove seasonality from the price level series.²¹

The results are reported in Table 7. The Bai and Ng (2002) criteria still select one factor of each type, but the aggregate component now explains about 22% of the overall variation in sectoral price changes. Due to smoothing of the year-on-year transformation we get of course very different results in terms of volatility and, particularly, persistence of the components. All series are now substantially more persistent. However, the major result concerning the relative importance of the sectoral component for explaining changes in sectoral prices are mostly confirmed. The sectoral and country-specific sectoral component on average explain again only about 35% of the overall variation in price changes, as in the case for month-on-month inflation. Moreover, the sectoral component is only slightly more volatile than the aggregate component, and its persistence is smaller than that of the aggregate component but only to a relatively small degree. Overall, our qualitative results for month-on-month inflation are confirmed.

7 Conclusions

In this paper we use a novel and large set of euro area regional sectoral price data to analyze the importance and major characteristics of the determinants of price changes. We also propose a new method to extract factors from over-lapping data blocks. The use of the novel disaggregate sectoral euro area data set with a regional breakdown combined with our new factor estimation method allow us to separately estimate aggregate, sectoral, country-specific as well as regional components of price changes. Hence, we provide an improved estimate of the sectoral component of price changes, thereby extending previous

²¹However, twelve differencing could introduce a moving average component into the error term of models where the year on year inflation rate is the dependent variable, when the true dependent variable is the month on month inflation rate. In our context we do not find this problem, since standard tests for no correlation of the residuals of the models that we will present do not reject the null hypothesis in most cases.

literature that decomposes price changes into an aggregate and an idiosyncratic component only, where the latter is interpreted as the sector-specific component (e.g. Boivin et al. (2009) and Mackowiak et al. (2009)). We investigate whether our decomposition provides different results and interpretation than the simple decomposition into aggregate and idiosyncratic components, in particular regarding the importance and properties of the sectoral component. A further contribution of our paper is to investigate the importance of regional factors for price setting and discuss potential implications for the plausibility of price setting models in that context.

Our analysis is therefore related to two different strands of the literature with a rather loose connection so far: The literature on macroeconomic price setting models (e.g. Calvo (1983), Mankiw and Reis (2002) and Mackowiak and Wiederholt (2009)) and the role of sectoral shocks on the one hand, and the literature on international pricing and regional pricing competition on the other hand (e.g. Gopinath et al. (2011)).

Regarding the sectoral component of price changes, we find that it explains on average only about 14% of the overall volatility in sectoral regional prices, with the country-specific sectoral component contributing an additional 21%. This is substantially less than the 85-90% values reported in previous studies on sectoral prices. Moreover, our estimated sectoral component exhibits much less volatility than previous findings for the US indicate. On the other hand, previous results on the low persistence of sector-specific shocks are confirmed by our analysis. Overall, we find a clear negative relationship between the persistence and the volatility of the inflation components.

Investigating the question whether local factors do play a role for price changes, we find that country- and region-specific factors play an important role in addition to the sector-specific factors. Country-specific factors explain about 10% of overall volatility in sectoral-regional prices and exhibit a substantial degree of persistence. The region-specific component, excluding other factors such as measurement error, explains about 13% of the overall variation of inflation rates, so that regional shocks are indeed an important driving force behind inflation developments.

Overall, our results suggest that previous findings that show that sectoral shocks to prices (or what was interpreted as sectoral shocks) are a dominant source of changes in sectoral prices need to be reconsidered. Disaggregate forces do play an important role in price determination, but sectoral shocks are complemented by regional (and for the euro area country-specific) shocks. However, in line with Mackowiak et al. (2009), our

results provide suggestive evidence in favor of the rational-inattention model and against the Calvo and sticky-information model. The rational-inattention model might also be adequate to allow for region-specific shocks that from our empirical analysis appear on average to have similar volatility as sectoral shocks, with comparable relatively high standard error, and low persistence. This would be consistent with the idea that price setters for consumer prices devote similar attention to regional shocks and to sector-specific shocks. Moreover, the remaining important role for country-specific factors as drivers of price movements is in line with the finding in the literature on international pricing that international borders create a substantially larger discontinuity in price changes than state and provincial boundaries (Gopinath et al. (2011)).

Finally, the results from our analysis also suggest that further research is needed on the importance of the "geographical" or "regional" dimension in other countries, including the US. Recent US studies investigating regional differences other than for CPI inflation include e.g. Clark (1998), Hamilton and Owyang (2009), Ng and Moench (2009) and Stock and Watson (2008) on housing. It is intuitive that price setters for consumer prices are attentive to regional shocks, and that they are probably more attentive to regional than to aggregate shocks. An interesting question to address is whether informational frictions explain whether consumers can obtain and process information on regional developments more easily than on sectoral developments. A further interesting issue to investigate in the context of the present study are the real effects of monetary policy shocks in the presence of additional heterogeneity due to regional shocks. Given the current disagreement in the literature on the real effects of monetary policy in the presence of sectoral heterogeneity we leave that to future research.

References

- Altissimo, F., Ehrmann, M., and Smets, F. R. (2006). Inflation persistence and price-setting behavior in the euro area - a summary of the IPN evidence. ECB Occasional Paper No. 46.
- Alvarez, L. J., Dhyne, E., Hoeberichts, M., Kwapil, C., Le Bihan, H., Lnnemann, P., Martins, F., Sabbatini, R., Stahl, H., Vermeulen, P., and Vilmunen, J. (2006). Sticky prices in the euro area: A summary of new micro-evidence. *Journal of the European Economic Association*, 4(2-3):575–584.

- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Beck, G. W., Hubrich, K., and Marcellino, M. (2009). Regional inflation dynamics within and across euro area countries and a comparison with the US. *Economic Policy*, 57:141–184.
- Bils, M. and Klenow, P. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112:987–985.
- Boivin, J., Giannoni, M., and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated u.s. data. *The American Economic Review*, 99(1):350–384.
- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12:383–398.
- Carvalho, C. and Lee, J. W. (2010). Sectoral price facts in a sticky-price model. Manuscript. Federal Reserve Bank of New York.
- Clark, T. E. (1998). Employment fluctuations in U.S. regions and industries: The roles of national, region-specific and industry-specific shocks. *Journal of Labor Economics*, 16(1):202–229.
- Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics*, 21:563–587.
- Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the Post WWII US. *Review of Economic Dynamics*, 8:262–302.
- Diebold, F. X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics*, 146:351–363.
- Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2008). Reference prices and nominal rigidities. NBER Working Paper Series No.13829.
- Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2008). Sectoral vs. aggregate shocks: A structural factor analysis of industrial production. Federal Reserve Bank of Richmond, Working Paper No. 08-07.

- Gopinath, G., Gourinchas, P.-O., Hsieh, C.-T., and Li, N. (2011). International prices, costs and markup differences. *American Economic Review*, forthcoming.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238.
- Hamilton, J. D. and Owyang, M. (2009). The propagation of regional recessions. Federal Reserve Bank of St. Louis Working Paper No. 2009-013A.
- Imbs, J., Mumtaz, H., Ravn, M. O., and Rey, H. (2005). PPP strikes back: Aggregation and the real exchange rate. *Quarterly Journal of Economics*, 120:1–44.
- Kapetanios, G. and Marcellino, M. (2006). Factor-GMM estimation with large sets of possibly weak instruments. Queen Mary University of London Working Paper No. 577.
- Koopman, S. J. and Jungbacker, B. (2008). Likelihood-based analysis for dynamic factor models. Manuscript. VU University Amsterdam.
- Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International business cycles: World, region and country-specific factors. *The American Economic Review*, 93(4):1216–1239.
- Mackowiak, B., Moench, E., and Wiederholt, M. (2009). Sectoral price data and models of price setting. *Journal of Monetary Economics*, 56:78–99.
- Mackowiak, B. and Smets, F. (2009). Implications of microeconomic price data for macroeconomic models. Manuscript, Humboldt University Berlin.
- Mackowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*. forthcoming.
- Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics*, 117:1295–1328.
- Mishkin, F. S. (2007). Inflation dynamics. NBER Working Papers No. 13147.
- Moench, E., Ng, S., and Potter, S. (2009). Dynamic hierarchical factor models. Manuscript, Federal Reserve Bank of New York.

- Ng, S. and Moench, E. (2009). A factor analysis of consumption and housing market dynamics in the U.S. and the regions. Manuscript, Federal Reserve Bank of New York.
- Onatski, A. (2006). Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak. Manuscript, Columbia University.
- Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics*, 68:79–113.
- Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.
- Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indices. *Journal of Business and Economic Statistics*, 20(2):1167–1179.
- Stock, J. H. and Watson, M. W. (2008). The evolution of national and regional factors in U.S. housing construction. Manuscript, Princeton University.

Tables

Table 1: Descriptive statistics: Total sample, sectoral and country data

Sample	Nobs	Level		Volatility		Persistence		Disp	Corr(x_i, x)
		Mean	Std	Mean	Std	Mean	Std		
Total sample									
All sectoral inflation	418	2.057	1.063	4.768	2.436	0.005	0.537	5.245	0.154
All items inflation	61	2.228	0.636	2.249	0.528	0.233	0.260	2.020	0.482
Data grouped by countries									
DE	77	1.090	1.056	5.222	2.637	-0.162	0.624	5.150	0.323
ES	120	2.630	0.818	4.538	1.840	0.136	0.337	4.511	0.501
FI	43	1.455	0.911	5.992	3.498	-0.326	0.965	6.223	0.503
IT	124	2.085	0.700	3.774	1.591	0.070	0.426	3.881	0.229
PO	54	2.577	1.148	5.941	2.865	0.067	0.338	6.367	0.212
Data grouped by sectors									
alco	30	1.835	1.130	4.246	2.146	0.066	0.387	4.042	0.400
clot	35	1.269	1.567	5.398	4.781	-0.329	1.120	6.580	0.241
food	60	1.909	0.829	4.945	1.369	0.309	0.309	4.362	0.518
furn	56	1.495	0.798	2.574	1.314	-0.011	0.376	2.811	0.286
heal	27	2.557	0.674	3.767	1.006	0.023	0.339	3.376	0.441
hote	53	2.938	1.081	4.548	1.641	-0.087	0.720	4.518	0.239
hous	58	2.539	0.612	4.803	1.556	0.114	0.239	4.101	0.512
recr	57	1.448	0.729	5.967	2.318	-0.078	0.472	5.311	0.338
tran	42	2.558	0.623	6.534	2.302	-0.106	0.248	5.145	0.578

Notes:

- 1) Table 1 reports descriptive statistics for the data series included in our cleaned data set. Results are reported for all regional sectoral data series (“All sectoral inflation”) and subsamples which include all series from a given country (country data) or a given sector (sectoral data). In addition the first line also reports results for all aggregate price indices (“All items inflation”). Monthly inflation rates are multiplied by 1200.
- 2) The reported statistics include the mean and the standard deviation (std) of the time-series means of all inflation series included in a given group (level), the mean and the standard deviation (std) of the time-series standard deviation of all inflation series included in a given group (volatility), the mean and the standard deviation (std) of the persistence measures of all inflation series included in a given group, the time-series mean of the cross-sectional dispersion of all inflation series included in a given group and the mean correlation of all inflation series included in a given group with the group aggregate inflation rate.
- 3) The measure for persistence is based on the sum of the estimated coefficients of an AR model with 13 lags.
- 4) The group aggregate inflation rate is computed as a weighted average of the series included in the group. Regions are weighted by their relative economic size, sectors are weighted based on their euro area HICP weight in 2000.

Table 2: Descriptive statistics: Country-sector-specific data

Sample	Nobs	Level		Volatility		Persistence		Disp	Corr(x_i, x)
		Mean	Std	Mean	Std	Mean	Std		
Germany									
alco	6	0.702	0.352	2.891	1.266	-0.411	0.436	2.144	0.522
clot	7	0.205	2.382	3.470	1.200	0.121	0.432	3.849	0.460
food	12	0.672	0.381	5.362	0.712	0.236	0.067	2.641	0.873
furn	9	0.282	0.283	1.631	0.375	0.232	0.275	1.381	0.551
hote	10	1.415	0.278	6.505	1.698	-0.898	1.151	3.820	0.644
hous	11	1.617	0.350	2.886	0.598	0.199	0.501	1.867	0.726
recr	10	0.566	0.261	8.119	0.359	-0.621	0.500	3.395	0.744
tran	12	2.507	0.109	8.619	0.697	-0.230	0.203	3.470	0.909
Spain									
alco	15	2.170	0.809	4.634	1.397	0.191	0.272	3.192	0.693
food	17	2.729	0.235	4.217	0.417	0.592	0.073	2.366	0.811
furn	17	2.045	0.584	2.731	1.023	0.018	0.293	2.451	0.535
heal	18	2.262	0.528	3.567	0.922	-0.089	0.283	2.795	0.591
hote	17	4.011	0.318	3.912	1.102	0.144	0.365	3.235	0.565
hous	18	3.021	0.509	5.069	1.176	0.018	0.282	3.410	0.713
recr	18	2.146	0.574	7.497	1.795	0.104	0.265	3.193	0.921
Finland									
alco	4	0.546	0.212	1.406	0.181	0.229	0.247	0.483	0.873
clot	5	-0.171	0.597	12.783	1.882	-2.461	1.542	7.978	0.747
food	5	1.406	0.420	6.168	0.996	-0.091	0.347	3.124	0.865
furn	5	0.905	0.199	4.241	0.967	-0.438	0.465	3.452	0.581
heal	4	2.904	0.232	4.877	0.448	-0.092	0.176	2.695	0.758
hote	5	2.302	0.110	3.455	0.751	-0.013	0.126	2.165	0.762
hous	5	1.982	0.213	4.970	1.284	0.196	0.107	2.506	0.901
recr	5	1.694	0.135	4.726	1.299	0.084	0.288	2.742	0.752
tran	5	1.638	0.162	10.161	1.422	-0.188	0.179	4.206	0.910
Italy									
clot	18	2.248	0.645	2.149	0.588	0.082	0.527	2.018	0.344
food	19	1.832	0.367	4.069	0.664	0.396	0.196	3.081	0.660
furn	18	1.449	0.454	1.897	0.519	-0.104	0.414	1.638	0.385
hote	14	2.680	0.670	3.892	0.953	0.073	0.546	3.621	0.380
hous	19	2.675	0.368	5.831	1.273	0.107	0.214	3.888	0.696
recr	17	1.303	0.438	3.518	1.265	-0.052	0.567	2.771	0.550
tran	19	2.455	0.360	4.878	0.698	-0.031	0.502	2.898	0.763
Portugal									
alco	5	3.250	0.597	6.977	1.953	0.131	0.339	5.200	0.543
clot	5	0.676	0.955	12.407	1.876	-0.306	0.136	9.877	0.558
food	7	2.608	0.504	7.499	1.382	-0.203	0.147	5.710	0.655
furn	7	2.509	0.403	3.959	2.032	0.149	0.189	3.655	0.484
heal	5	3.340	0.668	3.598	1.168	0.518	0.095	2.857	0.566
hote	7	3.476	0.899	5.389	1.757	0.141	0.342	4.968	0.329
hous	5	2.872	0.379	3.986	1.572	0.224	0.110	2.580	0.723
recr	7	1.093	0.624	5.792	1.553	0.050	0.492	4.589	0.506
tran	6	3.755	0.228	4.587	0.675	-0.028	0.295	2.877	0.729

Notes:

1) Table 2 reports descriptive statistics for the data series included in our cleaned data set. Results are reported for sectoral data of each included country. See the notes of Table 1 for further details on the computation of the statistics.

Table 3: Monte Carlo simulation results: Correlations with true factors (Base case)

Factor	Mean	Quantile				
		0.025	0.25	0.5	0.75	0.975
Unconstrained principal components						
Aggregate	0.884	0.276	0.879	0.949	0.975	0.994
Country 1	0.471	0.025	0.250	0.486	0.686	0.900
Country 2	0.420	0.032	0.207	0.406	0.614	0.898
Sector 1	0.362	0.013	0.154	0.333	0.543	0.841
Sector 2	0.329	0.018	0.159	0.308	0.474	0.778
Constrained principal components						
Aggregate	0.884	0.276	0.879	0.949	0.975	0.994
Country 1	0.816	0.347	0.763	0.866	0.930	0.982
Country 2	0.811	0.320	0.752	0.861	0.932	0.985
Sector 1	0.808	0.327	0.744	0.860	0.933	0.985
Sector 2	0.815	0.328	0.748	0.866	0.934	0.982

Notes:

- 1) Table 3 reports selected Monte Carlo simulation results for the “Base case” as described in Subsection 4.1. The reported statistics provide a comparison of the performance of the standard principal component based factor estimators introduced by Stock and Watson (2002a, 2002b, upper panel, denoted by “Unconstrained principal components”) and of our procedure (lower panel, denoted by “Constrained principal components”) based on the correlation coefficients between the true factors and the estimated factors.
- 2) The specifications for the base case are: $C = 2$; $S = 2$; $\rho = 0.8$; $\sigma_f = 1$; $\sigma_{id} = 1$; $T = 100$; $R = 1000$ (Notation: C : Number of countries; S : Number of sectors; ρ : Persistence of factors; σ_f : Standard deviation of shocks to factors; σ_{id} : Standard deviation of shocks to idiosyncratic component; T : Number of observations; R : Number of replications in Monte Carlo experiment.)
- 3) Reported statistics are the mean of the empirical distribution of computed correlation coefficients over the R replications, and selected percentiles of this distribution.

Table 4: Monte Carlo simulation results: Estimated persistence of true and estimated factors (Base case)

Factor	Mean	Quantile				
		0.025	0.25	0.5	0.75	0.975
Results for true factors						
Aggregate	0.771	0.495	0.710	0.793	0.854	0.934
Country 1	0.769	0.489	0.714	0.789	0.850	0.929
Country 2	0.771	0.474	0.712	0.795	0.853	0.928
Sector 1	0.770	0.479	0.713	0.790	0.851	0.930
Sector 2	0.766	0.473	0.705	0.788	0.850	0.925
Results for unconstrained principal components						
Aggregate	0.767	0.506	0.710	0.786	0.842	0.925
Country 1	0.736	0.454	0.672	0.759	0.822	0.903
Country 2	0.667	0.289	0.590	0.698	0.771	0.879
Sector 1	0.575	0.122	0.487	0.607	0.702	0.832
Sector 2	0.437	-0.165	0.314	0.481	0.611	0.778
Results for constrained principal components						
Aggregate	0.767	0.506	0.710	0.786	0.842	0.925
Country 1	0.646	0.241	0.566	0.677	0.766	0.881
Country 2	0.651	0.252	0.572	0.680	0.765	0.880
Sector 1	0.635	0.224	0.552	0.664	0.755	0.876
Sector 2	0.630	0.257	0.551	0.657	0.744	0.858

Notes:

- 1) Table 4 reports selected Monte Carlo simulation results for the “Base case” as described in Subsection 4.1. The reported statistics provide a comparison of the performance of the standard principal component based factor estimators introduced by Stock and Watson (2002a, 2002b, upper panel, denoted by “Results for unconstrained principal components”) and of our procedure (lower panel, denoted by “Results for constrained principal components”) based on the estimated persistence of true and estimated factors.
- 2) The specifications for the base case are: $C = 2$; $S = 2$; $\rho = 0.8$; $\sigma_f = 1$; $\sigma_{id} = 1$; $T = 100$; $R = 1000$ (Notation: C : Number of countries; S : Number of sectors; ρ : Persistence of factors; σ_f : Standard deviation of shocks to factors; σ_{id} : Standard deviation of shocks to idiosyncratic component; T : Number of observations; R : Number of replications in Monte Carlo experiment.)
- 3) Reported statistics are the mean of the empirical distribution of estimated persistence coefficients over the R replications, and selected percentiles of this distribution. The measure for persistence is based on the sum of the estimated coefficients of an AR model with 13 lags.

Table 5: Monte Carlo simulation results: Percentages of variances explained (Base case)

Factor	Mean	Quantile				
		0.025	0.25	0.5	0.75	0.975
Results for true factors						
Aggregate	0.272	0.226	0.256	0.271	0.288	0.318
Country	0.273	0.230	0.257	0.272	0.288	0.318
Sector	0.273	0.230	0.257	0.273	0.288	0.317
Idiosyncratic	0.183	0.156	0.173	0.183	0.192	0.211
Results for unconstrained principal components						
Aggregate	0.304	0.234	0.273	0.301	0.330	0.396
Country	0.229	0.202	0.220	0.229	0.238	0.254
Sector	0.096	0.082	0.090	0.095	0.101	0.111
Idiosyncratic	0.372	0.278	0.341	0.373	0.405	0.460
Results for constrained principal components						
Aggregate	0.304	0.234	0.273	0.301	0.330	0.396
Country	0.230	0.151	0.203	0.230	0.257	0.306
Sector	0.222	0.159	0.200	0.222	0.243	0.284
Idiosyncratic	0.245	0.183	0.216	0.239	0.267	0.337

Notes:

1) Table 5 reports selected Monte Carlo simulation results for the “Base case” as described in Subsection 4.1. The reported statistics provide a comparison of the performance of the standard principal component based factor estimators introduced by Stock and Watson (2002a, 2002b, upper panel, denoted by “Results for unconstrained principal components”) and of our procedure (lower panel, denoted by “Results for constrained principal components”) based on the percentages of variance explained by the true and the estimated factors.

2) The specifications for the base case are: $C = 2$; $S = 2$; $\rho = 0.8$; $\sigma_f = 1$; $\sigma_{id} = 1$; $T = 100$; $R = 1000$ (Notation: C : Number of countries; S : Number of sectors; ρ : Persistence of factors; σ_f : Standard deviation of shocks to factors; σ_{id} : Standard deviation of shocks to idiosyncratic component; T : Number of observations; R : Number of replications in Monte Carlo experiment.)

3) Reported statistics are the mean of the empirical distribution of variances explained by the respective factors over the R replications, and selected percentiles of this distribution.

Table 6: Volatility, persistence and relative importance of estimated factors: Month-on-month changes

	Aggr	$u_{c,r,s,t}$	C	S	CS	Idios.	R	Idios. (R)
Volatility								
Mean	0.216	0.959	0.280	0.294	0.398	0.664	0.243	0.562
Median	0.173	0.985	0.263	0.229	0.397	0.671	0.117	0.581
Min	0.000	0.769	0.001	0.001	0.001	0.256	0.000	0.078
Max	0.639	1.000	0.768	0.839	0.896	0.996	0.948	0.959
Std	0.174	0.057	0.160	0.232	0.218	0.183	0.264	0.171
Persistence								
Mean	0.294	-0.025	0.570	0.084	-0.088	-0.341	-0.220	-0.329
Median	0.294	0.071	0.708	0.149	-0.017	-0.214	-0.166	-0.236
Min	0.294	-3.254	0.309	-0.565	-1.871	-3.614	-2.620	-2.970
Max	0.294	0.863	0.710	0.440	0.684	0.818	0.816	0.909
Std	0.000	0.486	0.167	0.260	0.510	0.584	0.528	0.551
Variance explained								
Mean	0.077	0.923	0.104	0.140	0.206	0.474	0.128	0.346
Median	0.030	0.970	0.069	0.053	0.158	0.451	0.014	0.337
Min	0.000	0.591	0.000	0.000	0.000	0.065	0.000	0.006
Max	0.409	1.000	0.589	0.703	0.803	0.993	0.899	0.919
Std	0.104	0.104	0.106	0.179	0.184	0.238	0.220	0.187

Notes:

- 1) Table 6 reports summary statistics for the aggregate (Aggr), the residual component $u_{c,r,s,t}$, the country-specific (C), the sector-specific (S) country-sector-specific (CS) common components, and the idiosyncratic component (Idios.). Moreover, results are reported for the case when the idiosyncratic component is decomposed further into a region-specific common component (R) and a region-specific idiosyncratic component (Idios. (R)). Inflation rates are computed as month-on-month proportional changes. Common components are computed as the product $\lambda_{c,r,s} f_t^x$ where $\lambda_{c,r,s}$ denotes the region-sector-specific loading of a series and f_t^x (with $x \in aw, c, s, cs$ denotes factor x). The decomposition of a time series is done according to equation (5).
- 2) Statistics are computed for the volatility and the persistence of the common components. The volatility of a time series is measured by the standard deviation of the series. The measure for persistence is based on the estimation of an AR processes with 13 lags.
- 3) The proportion of variance explained by a factor is computed as the product $\lambda_{c,r,s}^2 var(f_t^x)$ where $\lambda_{c,r,s}$ denotes the region-sector-specific loading of a series and f_t^x (with $x \in aw, c, s, cs$ denotes factor x). The decomposition of a time series is done according to equation (5).
- 4) The reported statistics include the mean value (mean), the median value (median), the minimum value (min), the maximum value (max) and the cross-sectional standard deviation (std) of the respective variables.

Table 7: Volatility, persistence and relative importance of estimated factors: Year-on-year changes

	Aggr	$u_{c,r,s,t}$	C	S	CS	Idios.
Volatility						
Mean	0.402	0.871	0.384	0.311	0.409	0.439
Median	0.387	0.922	0.381	0.293	0.396	0.429
Min	0.000	0.373	0.003	0.002	0.000	0.010
Max	0.928	1.000	0.862	0.799	0.876	0.956
Std	0.244	0.140	0.232	0.193	0.219	0.190
Persistence						
Mean	0.980	0.845	0.916	0.705	0.830	0.589
Median	0.980	0.860	0.933	0.675	0.859	0.652
Min	0.980	0.060	0.825	0.610	-0.016	-0.857
Max	0.980	1.614	0.949	0.862	1.098	1.017
Std	0.000	0.133	0.045	0.085	0.151	0.260
Variance explained						
Mean	0.221	0.779	0.201	0.134	0.215	0.229
Median	0.150	0.850	0.145	0.086	0.157	0.184
Min	0.000	0.139	0.000	0.000	0.000	0.000
Max	0.861	1.000	0.743	0.638	0.767	0.914
Std	0.217	0.217	0.189	0.137	0.190	0.183

Notes:

1) Table 7 reports summary statistics for the aggregate (Aggr), the residual component $u_{c,r,s,t}$, the country-specific (C), the sector-specific (S) country-sector-specific (CS) common components, the idiosyncratic component (Idios.). Inflation rates are computed as year-on-year proportional changes. Common components are computed as the product $\lambda_{c,r,s} f_t^x$ where $\lambda_{c,r,s}$ denotes the region-sector-specific loading of a series and f_t^x (with $x \in aw, c, s, cs$ denotes factor x). The decomposition of a time series is done according to equation (5). 2) Statistics are computed for the volatility and the persistence of the common components. The volatility of a time series is measured by the standard deviation of the series. The measure for persistence is based on the estimation of an AR processes with 13 lags.

3) The proportion of variance explained by a factor is computed as the product $\lambda_{c,r,s}^2 var(f_t^x)$ where $\lambda_{c,r,s}$ denotes the region-sector-specific loading of a series and f_t^x (with $x \in aw, c, s, cs$ denotes factor x). The decomposition of a time series is done according to equation (5). 4) The reported statistics include the mean value (mean), the median value (median), the minimum value (min), the maximum value (max) and the cross-sectional standard deviation (std) of the respective variables.

A Not-for-publication Appendix: Computation of the aggregate sectoral price index

Other studies of disaggregate sectoral data such as Boivin et al. (2009) or Mackowiak et al. (2009) use national (or in the case of the EMU euro-area wide) rather than regional data. To see the relationship between our approach and that taken in these previous studies it is instructive to remember that national/euro-area wide data are obtained by aggregating regional price data. The weights which are used in this aggregation process normally correspond to the expenditure shares of the respective regions in total expenditure. The national/euro-area wide sectoral price index can therefore be thought of as computed as follows:

$$\pi_{s,t} = \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \pi_{c,r,s,t}, \quad (7)$$

where θ_c denotes the expenditure share of country c and $\theta_{c,r}$ denotes the expenditure share of region r of country c . As shown below, this term can be written as:

$$\pi_{s,t} = \bar{\alpha}_s^{rc} f_t^a + \bar{\gamma}_s^{rc} f_t^s + \sum_{c=1}^C \theta_c \bar{\beta}_{c,s}^r f_t^c + \sum_{c=1}^C \theta_c \bar{\delta}_{c,s}^r f_t^{sc} + \bar{e}_s^{rc}, \quad (8)$$

where a bar above a variable / parameter denotes the weighted average of this variable and the upper indices r or c indicate whether the average is taken across regions of a country or countries.²² Comparing this term with equation (2) we can see that the sector-specific component of Boivin et al. (2009) or Mackowiak et al. (2009), denoted by $u_{s,t}$ corresponds to the following expression:

$$u_{s,t} = \bar{\gamma}_s^{rc} f_t^s + \sum_{c=1}^C \theta_c \bar{\beta}_{c,s}^r f_t^c + \sum_{c=1}^C \theta_c \bar{\delta}_{c,s}^r f_t^{sc} + \bar{e}_s^{rc}. \quad (9)$$

This expression clearly illustrates that the time series properties of u_s , i.e. the “sectoral component” in previous studies, crucially depend on the time series properties of the country-specific, the country-sector-specific and the region-specific components.

²²Weights used in computing averages correspond to the respective expenditure shares. The upper index rc indicates that averages are first taken across regions of a country and then across countries.

To conclude, we derive the equation

$$\pi_{s,t} = \bar{\alpha}_s^{rc} f_t^a + \bar{\gamma}_s^{rc} f_t^s + \sum_{c=1}^C \theta_c \bar{\beta}_{c,s}^r f_t^c + \sum_{c=1}^C \theta_c \bar{\delta}_{c,s}^r f_t^{sc} + \bar{e}_s^{rc}, \quad (10)$$

Using equation (4), we obtain:

$$\pi_{s,t} = \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} [\alpha_{c,r,s} f_t^a + \beta_{c,r,s} f_t^c + \gamma_{c,r,s} f_t^s + \delta_{c,r,s} f_t^{sc} + e_{c,r,s,t}]. \quad (11)$$

This term can be rearranged as follows:

$$\begin{aligned} \pi_{s,t} &= \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \alpha_{c,r,s} f_t^a + \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \beta_{c,r,s} f_t^c + \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \gamma_{c,r,s} f_t^s \\ &\quad + \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \delta_{c,r,s} f_t^{sc} + \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} e_{c,r,s,t} = \\ &= f_t^a \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \alpha_{c,r,s} + \sum_{c=1}^C \theta_c f_t^c \sum_{r=1}^{R_c} \theta_{c,r} \beta_{c,r,s} + f_t^s \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} \gamma_{c,r,s} \\ &\quad + \sum_{c=1}^C \theta_c f_t^{sc} \sum_{r=1}^{R_c} \theta_{c,r} \delta_{c,r,s} + \sum_{c=1}^C \theta_c \sum_{r=1}^{R_c} \theta_{c,r} e_{c,r,s,t}. \end{aligned} \quad (12)$$

Since the parameters θ_c represent expenditure shares of a given state/country in total national/euro area wide expenditures we have $\sum_{c=1}^C \theta_c = 1$. Similarly, we obtain for the parameters $\theta_{c,r}$: $\sum_{r=1}^{R_c} \theta_{c,r} = 1$. Denoting the weighted average of a variable/parameter x across countries/regions as \bar{x}^c / \bar{x}^r we can rewrite equation (13) as follows:²³

$$\pi_{s,t} = \bar{\alpha}_s^{rc} f_t^a + \bar{\gamma}_s^{rc} f_t^s + \sum_{c=1}^C \theta_c \bar{\beta}_{c,s}^r f_t^c + \sum_{c=1}^C \theta_c \bar{\delta}_{c,s}^r f_t^{sc} + \bar{e}_s^{rc}. \quad (13)$$

²³Upper index rc indicates that averages are first taken across regions and then across countries.

B Not-for-publication Appendix: Additional Monte Carlo experiments

The results reported so far are quite good, but we need to assess their robustness to a variety of changes in the experimental design. In particular, we consider the following modifications, which could all deteriorate the performance of the factor estimation methods.

First, a reduction in the persistence of the factors from 0.8 to 0.4 and 0.1. Lower persistence decreases the overall variance of the factors and makes them dynamically more similar to the idiosyncratic errors. Both features can be expected to complicate the factor estimation. Second, lower volatility of the factors, the variance of the errors in the AR(1) model for the factors passes from 1 to 0.1, which decreases their overall explanatory power. Third, larger variance for each idiosyncratic error, from 1 to 10, which reduces the relative explanatory power of the factors. Fourth, a decrease in the number of regions in each country from 30 to 15, so that less information is available. Fifth, a decrease in the temporal dimension from 100 to 50, which should lower the precision in the estimation of the loadings and of the factor persistence. Finally, each non-zero element of the \mathbf{A} matrix of loadings is extracted from a uniform rather than standard normal distribution, which centers the loadings on 0.5 rather than 0.

The results of all these experiments are summarized in Tables 8 and 9. We only report the average values for each criterion over the R replications and different experimental designs in order to save space.²⁴

The main findings are rather in line with the theoretical expectations and can be summarized as follows. First, decreasing the persistence of the factors lowers their explanatory power but the correlation between true and estimated factors is barely affected, while naturally the estimated persistence decreases. Second, decreasing the size of the shocks to the factors does decrease their relative explanatory power, but the findings on the correlation between true and estimated factors and the estimated persistence of the latter are quite robust. The effects of a larger variance for the idiosyncratic errors are very similar. Thirdly, decreasing the number of regions lowers the correlation between estimated and true factors, leaves the results on the estimated persistence and variance explained basically unchanged though. A significant drop in all three evaluation criteria is found when

²⁴Tables with quantiles are available upon request.

the number of time series observations is decreased from 100 to 50. Finally, generating the loadings from a uniform distribution seems to only affect the explanatory power of the different components, with a more marked overestimation of the role of the aggregate and idiosyncratic components.

As an additional experiment, we evaluate the consequences of an increase in the number of countries and sectors from 2 to 3. This augments the total number of factors from 5 to 7 (one aggregate, 3 country and 3 sector factors), and hence makes estimation more complex.

The results reported in Table 10 suggest that there are no major differences with respect to the two countries - two sectors case of Tables 3 - 5. In particular, the correlation between the true and estimated factors remains high, and substantially higher for our estimation method than for the unconstrained principal components; the persistence of the factors is underestimated, in particular for the sectoral factors; and the role of the aggregate and idiosyncratic components is slightly overestimated by our method, while the standard approach overestimates the importance of the country component.

Overall, the results of the set of experiments we have conducted highlight the importance of modifying the standard principal component factor estimator in the presence of a block structure for the matrix of loadings. Our approach substantially improves the correlation between the estimated and the true factors, as well as their estimated persistence and explanatory power, though the persistence remains slightly underestimated and the role of the idiosyncratic component slightly overestimated.

Table 8: Monte Carlo simulation results: Additional experiments for two-countries-two-sectors setting

Correlation coefficients and estimated persistence						
Experiment	Statistic	Aggregate	Country1	Country2	Sector1	Sector2
$\rho = 0.4$	Corr - UPC	0.945	0.461	0.424	0.386	0.366
	Corr - CPC	0.945	0.893	0.892	0.888	0.892
	Pers - true	0.361	0.357	0.357	0.356	0.345
	Pers - UPC	0.281	0.282	0.218	0.172	0.120
	Pers - CPC	0.281	0.219	0.213	0.210	0.192
$\rho = 0.1$	Corr - UPC	0.952	0.445	0.420	0.385	0.372
	Corr - CPC	0.952	0.903	0.901	0.898	0.903
	Pers - true	0.052	0.051	0.046	0.043	0.029
	Pers - UPC	-0.083	-0.097	-0.125	-0.122	-0.135
	Pers - CPC	-0.083	-0.099	-0.116	-0.108	-0.137
$\sigma_f = \sqrt{0.1}$	Corr - UPC	0.835	0.442	0.396	0.342	0.297
	Corr - CPC	0.835	0.767	0.766	0.761	0.768
	Pers - true	0.771	0.769	0.771	0.770	0.766
	Pers - UPC	0.761	0.709	0.625	0.522	0.361
	Pers - CPC	0.761	0.593	0.595	0.595	0.596
$\sigma_{id} = \sqrt{10}$	Corr - UPC	0.835	0.442	0.396	0.342	0.297
	Corr - CPC	0.835	0.767	0.766	0.761	0.768
	Pers - true	0.771	0.769	0.771	0.770	0.766
	Pers - UPC	0.761	0.709	0.625	0.522	0.361
	Pers - CPC	0.761	0.593	0.595	0.595	0.596
$R = 15$	Corr - UPC	0.864	0.446	0.404	0.357	0.305
	Corr - CPC	0.864	0.759	0.758	0.758	0.773
	Pers - true	0.769	0.771	0.768	0.773	0.773
	Pers - UPC	0.761	0.736	0.664	0.574	0.442
	Pers - CPC	0.761	0.643	0.646	0.635	0.631
$T = 50$	Corr - UPC	0.826	0.461	0.396	0.339	0.278
	Corr - CPC	0.826	0.717	0.714	0.722	0.730
	Pers - true	0.730	0.720	0.718	0.719	0.747
	Pers - UPC	0.686	0.609	0.419	0.175	-0.059
	Pers - CPC	0.686	0.397	0.419	0.360	0.399
<i>Uniform</i>	Corr - UPC	0.729	0.459	0.400	0.348	0.347
	Corr - CPC	0.729	0.789	0.795	0.800	0.804
	Pers - true	0.769	0.770	0.769	0.772	0.774
	Pers - UPC	0.721	0.733	0.618	0.606	0.466
	Pers - CPC	0.721	0.659	0.664	0.652	0.658

Notes:

1) Table 8 reports robustness Monte Carlo simulation results (mean values) for the two-countries-two-sectors setting as described in Subsection 4.2 (UPC: Unconstrained-principal-component approach; CPC: Constrained-principal-component approach).

Table 9: Monte Carlo simulation results: Additional experiments for two-countries-two-sectors setting

Variance explained					
Experiment	Statistic	Aggregate	Country	Sector	Idios.
$\rho = 0.4$	Var - true	0.228	0.229	0.230	0.313
	Var - UPC	0.245	0.218	0.102	0.435
	Var - CPC	0.245	0.215	0.207	0.333
$\rho = 0.1$	Var - true	0.218	0.219	0.219	0.343
	Var - UPC	0.234	0.215	0.102	0.448
	Var - CPC	0.234	0.208	0.202	0.355
$\sigma_f = \sqrt{0.1}$	Var - true	0.130	0.131	0.131	0.608
	Var - UPC	0.155	0.181	0.103	0.561
	Var - CPC	0.155	0.115	0.114	0.616
$\sigma_{id} = \sqrt{10}$	Var - true	0.130	0.131	0.131	0.608
	Var - UPC	0.155	0.181	0.103	0.561
	Var - CPC	0.155	0.115	0.114	0.616
$R = 15$	Var - true	0.273	0.271	0.273	0.183
	Var - UPC	0.311	0.231	0.096	0.362
	Var - CPC	0.311	0.222	0.217	0.250
$T = 50$	Var - true	0.273	0.274	0.272	0.182
	Var - UPC	0.322	0.231	0.094	0.353
	Var - CPC	0.322	0.209	0.204	0.265
<i>Uniform</i>	Var - true	0.229	0.228	0.229	0.314
	Var - UPC	0.375	0.312	0.118	0.195
	Var - CPC	0.375	0.130	0.127	0.368

Notes:

1) Table 9 reports robustness Monte Carlo simulation results (mean values) for the two-countries-two-sectors setting as described in Subsection 4.2 (UPC: Unconstrained-principal-component approach; CPC: Constrained-principal-component approach).

Table 10: Monte Carlo simulation results: Three-countries-three-sectors setting

Factor	Correlation		Estimated persistence			Factor	Variance explained		
	UPC	CPC	true	UPC	CPC		true	UPC	CPC
Aggregate	0.965	0.965	0.775	0.758	0.758	Aggregate	0.273	0.293	0.293
Country 1	0.424	0.878	0.769	0.788	0.651	Country	0.273	0.279	0.240
Country 2	0.368	0.879	0.767	0.721	0.663	Sector	0.272	0.071	0.218
Country 3	0.330	0.884	0.771	0.644	0.665	Idios.	0.182	0.358	0.250
Sector 1	0.293	0.854	0.769	0.535	0.614				
Sector 2	0.259	0.851	0.772	0.436	0.619				
Sector 3	0.225	0.849	0.771	0.278	0.613				

Notes:

1) Table 10 reports robustness Monte Carlo simulation results for the three-countries-three-sectors setting as described in Subsection 4.2. The reported statistics provide a comparison of the performance of the standard principal component based factor estimators introduced by Stock and Watson (2002a, 2002b, upper panel, denoted by “UPC”) and of our procedure (lower panel, denoted by “CPC”).

2) The specifications for the Monte Carlo simulations are: $C = 3$; $S = 3$; $\rho = 0.8$; $\sigma_f = 1$; $\sigma_{id} = 1$; $T = 100$; $R = 1000$ (Notation: C : Number of countries; S : Number of sectors; ρ : Persistence of factors; σ_f : Standard deviation of shocks to factors; σ_{id} : Standard deviation of shocks to idiosyncratic component; T : Number of observations; R : Number of replications in Monte Carlo experiment.)

3) Reported statistics are the mean of the empirical distribution of computed correlation coefficients between the true and estimated factors (columns 2 and 3), the estimated persistence of the true and estimated factors (columns 4 to 6), and the percentages of variance explained by the true and estimated factors (columns 8 to 10) over the R replications.

Table 11: Monte Carlo simulation results: Volatility, persistence and relative importance of principal components

	Aggr	u	C	S	Id	Aggr	u	C	S	Id.
	$\rho = 0.8; \sigma_f = 1; \sigma_{id} = 1; T = 100$					$\rho = 0.4; \sigma_f = 1; \sigma_{id} = 5; T = 100$				
Volatility										
Mean	0.483	0.816	0.407	0.399	0.469	0.312	0.925	0.291	0.288	0.780
(Std)	0.039	0.029	0.043	0.034	0.039	0.023	0.010	0.022	0.019	0.016
Std	0.264	0.172	0.250	0.248	0.150	0.199	0.084	0.192	0.190	0.114
(Std)	0.014	0.020	0.018	0.015	0.015	0.012	0.010	0.012	0.012	0.007
Persistence										
Mean	0.763	0.485	0.644	0.627	0.026	0.280	-0.084	0.189	0.178	-0.229
(Std)	0.115	0.094	0.128	0.124	0.125	0.287	0.079	0.234	0.216	0.046
Std	0.000	0.259	0.086	0.086	0.460	0.000	0.419	0.169	0.170	0.466
(Std)	0.000	0.051	0.075	0.073	0.056	0.000	0.042	0.142	0.139	0.040
Variance explained										
Mean	0.304	0.696	0.230	0.222	0.245	0.138	0.862	0.122	0.119	0.621
(Std)	0.041	0.041	0.040	0.032	0.039	0.018	0.018	0.016	0.014	0.024
Std	0.255	0.255	0.225	0.220	0.160	0.145	0.145	0.134	0.131	0.175
(Std)	0.022	0.022	0.027	0.023	0.023	0.015	0.015	0.015	0.014	0.010

Notes:

1) Table 11 reports results for the summary statistics for the aggregate (Aggr), the residual component u , the country-specific (C), the sector-specific (S) and the idiosyncratic component (Id.) for selected Monte Carlo simulation exercises as described in Section 4.

2) Results are reported for two different model specifications concerning the persistence of the factors (0.8 vs. 0.4) and the innovation variance σ_{id} (1 vs. 5). The following specifications are common across the two considered setups: $C = 2$; $S = 2$; $\sigma_f = 1$; $T = 100$; $R = 1000$ (Notation: C : Number of countries; S : Number of sectors; ρ : Persistence of factors; σ_f : Standard deviation of shocks to factors; σ_{id} : Standard deviation of shocks to idiosyncratic component; T : Number of observations; R : Number of replications in Monte Carlo experiment.) 3) Statistics are computed for the volatility and the persistence of the common components extracted from the simulated data. The volatility of a time series is measured by the standard deviation of the series. The measure for persistence is based on the estimation of an AR processes with 13 lags.

4) The proportion of variance explained by a factor is computed as the product $\lambda_{c,r,s}^2 var(f_t^x)$ where $\lambda_{c,r,s}$ denotes the region-sector-specific loading of a series and f_t^x (with $x \in aw, c, s$ denotes factor x). The decomposition of a time series is done according to equation (5). 5) The reported statistics include the average (mean) and standard deviation (std) of the mean values and the cross-sectional standard deviations of the respective variables obtained across 1000 simulation replications.

C Not-for-publication Appendix Tables on Data and Descriptive Statistics

Table A: Countries and Regions Included in our Study

Germany (12 NUTS-I Regions)
Regions: Baden-Württemberg, Bayern, Berlin, Brandenburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Saarland, Sachsen, Sachsen-Anhalt, Thüringen
Data Source: Statistical offices of the individual German states
Austria (9 NUTS II Regions)
Regions: Burgenland, Kärnten, Niederösterreich, Oberösterreich, Salzburg, Steiermark, Tirol, Vorarlberg, Wien
Data Source: Statistics Austria
Finland (5 NUTS-II Regions)
Regions: Ita-Suomi, Etela-Suomi, Lansu-Suomi, Pohjois-Suomi, Aland
Data Source: Statistics Finland
Italy (20 Major Cities of NUTS-II Regions)
Regions: Ancona, Aosta, Bari, Bologna, Cagliari, Campobasso, Firenze, Genova, L'Aquila, Milano, Napoli, Palermo, Perugia, Potenza, Reggio Calabria, Roma, Torino, Trento, Trieste, Venezia
Data Source: Istituto Nazionale di Statistica (ISTAT)
Spain (18 NUTS-II Regions)
Regions: Andalucia, Aragon, Principado de Asturias, Baleares, Canarias, Caabria, Castilla y Leon, Castilla La Mancha, Cataluna, Ceuta y Melilla, Extremadura, Galicia, Comunidad Madrid, Cummunidad Murcia, Navarra, Pais Vasco, La Rioja, Comunidad Valenicana
Data Source: Instituto Nacional de Estadistica (INE)
Portugal (7 NUTS-II Regions)
Regions: Acores, Algarve, Altenejo, Centro, Lisbon, Madeira, Norte
Data Source: Instituto Nacional de Estadistica (INE)

Table B: Country/Region/Variable Short Names

Full	Short	Full	Short	Full	Short Name
Countries					
Austria	AU	Germany	DE	Finland	FI
Italy	IT	Spain	ES	Portugal	PO
Regions					
Cast. la Mancha	alba	Marche	anco		
Extremadura	bada	Baden-Württemb.	bade	Cataluna	barc
Puglia	bari	Bayern	baye	Berlin	berl
Emilia-Romagna	bolo	Brandenburg	bran	Burgenland	burg
Sardegna	cagl	Molise	camp	Ceuta e Melilla	ceut
Norte	coim	Algarve	evor	Centro	faro
Toscana	fire	Lisboa	func	Liguria	geno
Ita-Suomi	hels	Hessen	hess	Etela-Suomi	joen
Krnten	kaer	Lansi-Suomi	kokk	Galicia	laco
Canarias	lapa	Abruzzo	laqu	Alentejo	lib
La Rioja	logr	Madrid	madr	Mecklenburg-Vorp.	meck
Milano	mila	Murcia	murc	Campania	napo
Niedersachsen	nied	Niedersterreich	nied	Nordrhein-Westf.	nord
Obersterreich	ober	Pohjois-Suomi	oulu	Asturias	ovie
Sicilia	pale	Baleares	palm	Navarra	pamp
Umbria	peru	Reg.Aut.d.Acores	pont	Reg.Aut.d.Madreira	port
Calabria	regg	Lazio	roma	Sachsen-Anhalt	saan
Saarland	saar	Sachsen	sach	Salzburg	salz
Pais Vasco	sans	Cantabria	sant	Aragon	sara
Andalucia	sevi	Steiermark	stei	Aland	tamp
Thringen	thue	Tirol	tiro	Piemonte	tori
Trento	tren	Friuli-Venezia	trie	Valencia	vale
Castilla Leon	vall	Veneto	vene	Vorarlberg	vora
Wien	wien				

Table C: Descriptive statistics: Total sample, sectoral and country data

Sample	Nobs	Level		Volatility		Persistence		Disp	Corr(x_i, x)
		Mean	Std	Mean	Std	Mean	Std		
Total sample									
All	730	2.146	1.798	9.809	9.175	-0.142	0.826	11.989	0.117
Data grouped by countries									
DE	142	1.291	1.952	11.315	7.885	-0.263	0.870	10.444	0.223
ES	216	2.788	1.494	10.619	9.850	-0.248	0.939	10.956	0.173
FI	60	1.491	1.395	12.550	12.081	-0.312	1.402	13.791	0.273
IT	228	2.016	1.483	6.379	3.151	0.083	0.412	6.227	0.200
PO	84	2.762	2.380	12.553	14.210	-0.151	0.593	15.571	0.160
Data grouped by sectors									
alco	60	3.654	1.458	11.768	2.375	-0.019	0.567	7.194	0.410
clot	61	1.726	1.533	24.104	21.219	-1.166	1.867	21.511	0.600
food	61	1.890	0.839	7.081	2.441	0.342	0.504	6.276	0.496
furn	61	1.580	0.831	3.633	1.335	-0.060	0.442	3.533	0.357
heal	61	2.685	0.944	9.650	8.620	-0.059	0.359	6.427	0.247
hote	61	2.922	1.112	8.444	5.520	-0.393	1.112	8.104	0.276
hous	61	2.575	0.705	5.754	1.802	0.135	0.237	4.808	0.491
recr	61	1.574	0.735	12.443	6.497	-0.505	0.596	10.750	0.362
tran	61	2.721	0.646	7.732	1.934	0.003	0.292	5.523	0.658

Notes:

- 1) Table C reports descriptive statistics for the data series included in our raw data set. Results are reported for all data series (total sample) and subsamples which include all series from a given country (country data) or a given sector (sectoral data).
- 2) The reported statistics include the mean and the standard deviation (std) of the time-series means of all inflation series included in a given group (level), the mean and the standard deviation (std) of the time-series standard deviation of all inflation series included in a given group (volatility), the mean and the standard deviation (std) of the persistence measures of all inflation series included in a given group, the time-series mean of the cross-sectional dispersion of all inflation series included in a given group and the mean correlation of all inflation series included in a given group with the group aggregate inflation rate.
- 3) The measure for persistence is based on the sum of the estimated coefficients of an AR model with 13 lags.
- 4) The group aggregate inflation rate is computed as a weighted average of the series included in the group. Regions are weighted by their relative economic size, sectors are weighted based on their euro area HICP weight in 2000.

Table D: Descriptive statistics: Country-sector-specific data

Sample	Nobs	Level		Volatility		Persistence		Disp	Corr(x_i, x)
		Mean	Std	Mean	Std	Mean	Std		
Germany									
alco	11	2.835	0.103	10.705	0.309	-0.646	0.316	2.183	0.977
clot	12	0.351	1.766	7.704	4.372	0.076	0.750	5.645	0.750
food	12	0.603	0.391	8.846	0.979	0.323	0.054	3.384	0.923
furn	12	0.286	0.508	1.956	0.515	0.208	0.458	1.623	0.467
hote	12	1.371	0.508	18.505	4.090	-1.984	1.640	5.793	0.934
hous	12	1.616	0.319	3.756	0.685	0.178	0.207	2.110	0.742
recr	12	0.569	0.286	20.735	2.220	-1.070	0.468	4.219	0.937
tran	12	2.508	0.110	8.620	0.699	-0.231	0.203	3.470	0.910
Spain									
alco	18	4.981	0.714	13.112	2.290	-0.127	0.154	4.874	0.883
food	18	2.700	0.250	6.247	0.861	0.514	0.085	3.466	0.797
furn	18	2.093	0.594	3.872	0.911	-0.088	0.401	2.769	0.706
heal	18	2.232	0.590	4.337	0.837	-0.082	0.308	3.022	0.679
hote	18	4.042	0.368	5.947	1.543	-0.159	0.398	4.037	0.719
hous	18	3.150	0.454	5.401	1.695	0.072	0.176	3.726	0.715
recr	18	2.214	0.594	16.316	3.836	-0.617	0.510	4.876	0.972
Finland									
alco	5	-0.208	0.089	16.424	0.922	1.486	0.501	0.354	1.000
clot	5	0.700	0.591	49.104	4.782	-4.021	2.505	11.120	0.974
food	5	1.370	0.407	8.995	1.541	-0.014	0.503	3.698	0.917
furn	5	0.923	0.215	4.921	0.559	-0.551	0.441	3.786	0.638
heal	5	3.039	0.288	7.540	0.669	-0.432	0.353	3.462	0.837
hote	5	2.284	0.126	5.159	0.945	-0.014	0.209	2.739	0.816
hous	5	1.988	0.217	5.438	0.706	0.211	0.117	1.843	0.931
recr	5	1.764	0.135	7.475	0.793	-0.221	0.358	3.289	0.847
tran	5	1.635	0.153	10.466	0.789	-0.183	0.171	4.034	0.922
Italy									
clot	19	2.324	0.620	3.698	0.850	-0.034	0.664	2.541	0.642
food	19	1.826	0.368	4.927	0.834	0.439	0.168	3.572	0.681
furn	19	1.828	0.344	3.617	0.960	-0.123	0.429	2.341	0.596
hote	19	2.652	0.699	5.592	1.589	0.092	0.398	4.983	0.368
hous	19	2.718	0.420	7.509	0.955	0.182	0.203	4.268	0.714
recr	19	1.663	0.274	6.605	3.142	-0.140	0.428	4.412	0.568
tran	19	2.431	0.360	6.058	0.846	-0.021	0.260	3.365	0.779
Portugal									
alco	7	4.027	0.494	8.941	1.448	-0.169	0.379	5.791	0.641
clot	7	-0.011	1.034	50.375	19.804	-1.396	0.674	26.121	0.824
food	7	2.558	0.239	10.679	3.302	-0.075	0.248	8.406	0.667
furn	7	2.273	0.394	5.020	1.612	0.071	0.238	3.980	0.605
heal	7	3.532	0.522	5.580	2.706	0.480	0.245	3.998	0.559
hote	7	3.890	0.384	7.709	1.618	0.146	0.238	6.030	0.458
hous	7	2.839	0.677	6.227	2.130	0.041	0.475	3.792	0.709
recr	7	1.279	0.777	7.666	1.914	-0.439	0.782	5.973	0.467
tran	7	3.910	0.208	6.506	2.448	-0.133	0.381	4.146	0.719

Notes:

1) Table D reports descriptive statistics for the data series included in our raw data set. Results are reported for sectoral data of each included country. See the notes of Table 1 for further details on the computation of the statistics.

