Institute for Employment Research

The Research Institute of the Federal Employment Agency

IAB-Discussion Paper 14/2013

Articles on labour market issues

Labour market forecasting

Is disaggregation useful?

Enzo Weber
Gerd Zika

ISSN 2195-2663

Labour market forecasting

Is disaggregation useful?

Enzo Weber (IAB, University of Regensburg)
Gerd Zika (IAB)

Mit der Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

The “IAB-Discussion Paper” is published by the research institute of the German Federal Employment Agency in order to intensify the dialogue with the scientific community. The prompt publication of the latest research results via the internet intends to stimulate criticism and to ensure research quality at an early stage before printing.

Contents

Figures

Abstract

Using the example of short-term forecasts for German employment figures, the article at hand examines the question whether the use of disaggregated information increases the forecast accuracy of the aggregate. For this purpose, the out-of-sample forecasts for the aggregated employment forecast are compared to and contrasted with forecasts based on a vector-autoregressive model, which includes not only the aggregate but also the numbers of gainfully employed people at the industry level. The Clark/West test is used in the model comparison. It becomes evident that disaggregation significantly improves the employment forecast. Moreover, fluctuation-window tests help identify the phases during which disaggregation increases forecast accuracy to the strongest extent.

Zusammenfassung

In dem vorliegenden Artikel wird am Beispiel der Kurzfristprognose für die deutschen Erwerbstätigenzahlen der Frage nachgegangen, ob die Nutzung von disaggregierten Informationen die Prognosegüte für das Aggregat erhöht. Hierzu werden die out-of-sample-Prognosen für die aggregierte Erwerbstätigenprognose mit den Prognosen auf Basis eines Vektor-autoregressiven Modells verglichen, in dem neben dem Aggregat auch die Erwerbstätigenzahlen auf Branchenebene einfließen. Für den Modellvergleich wird der Clark/West-Test verwendet. Es zeigt sich, dass Disaggregation die Erwerbstätigenprognose signifikant verbessert. Daneben werden mit Hilfe von Fluctuation-Window-Tests die Phasen identifiziert, in denen die Disaggregation die Prognosegüte am stärksten erhöht.

JEL classification: J23, C53

Keywords: forecast, disaggregation, employment

Acknowledgement: We are grateful to Alexander Vosseler and participants of the Regensburg-IAB Econometric Seminar for their comments.

1 Introduction

Since the global economic crisis of 2008/2009 at the very latest, the focus has shifted back to the accuracy of forecast methods (cf. e.g. Drechsel/Scheufele 2012, Hendry/Hubrich 2011, Angelini et al. 2011). Back then, no forecast model was able to predict the sharp decline of economic performance after the financial market crisis. Labour market developments, on the other hand, are still playing an inferior role in specialist forecast literature, even though the variety of labour market developments e.g. in Germany and Southern Europe is highly relevant.

One of the most important variables (if not the single most important one) is the number of gainfully employed people, i.e. all people who contribute to the gross domestic product by means of their work. An important question is whether the number of gainfully employed people should be predicted as an aggregate, or whether disaggregated information is more advantageous to the forecast. In long-term forecasts, for instance, the development of the employment figures is often initially estimated at the industry level based on input/output models and only aggregated in retrospect. With short-term forecasts, on the other hand, it is not clear whether disaggregation improves forecast accuracy in general. Pertinent literature concludes that this depends fundamentally on the properties of the variable to be forecast and its sub-aggregates. It would also be conceivable for the inclusion of sectoral information to improve forecast accuracy, especially in certain labour market development stages. In Germany, for example, the great recession of 2008/2009 led to significant dislocations at the industry level with respect to employment development. While jobs were abolished in the manufacturing industry, which suffered greatly from the decrease in export volumes, employment continued to grow uninterrupted in major parts of the service sector.

This paper now examines German data as to whether the use of disaggregated information improves forecast accuracy as compared to a direct aggregate forecast, and whether there are typical stages in labour market development where this improvement becomes evident. For this purpose, the out-of-sample forecasts for the aggregated employment forecast are compared to and contrasted with forecasts based on a vector-autoregressive model (VAR), which includes not only the aggregate but also the numbers of gainfully employed people at the industry level. Hendry and Hubrich (2011) show that this method is equivalent to summing up individual forecasts created at the industry level. Since this is a nested model comparison, we will use the Clark/West test (Clark/West 2007).

Furthermore, we will also examine the impact of different labour market development stages. To this end, we aim to examine a great variety of different stages. However, since this would make the underlying estimation period too short for a rolling-window database, we will also use a recursive-window database. We thus hope to reach a conclusive statement as to which model has its particular advantages in which stage of labour market development. The statistical review will be carried out with the help of fluctuation-window tests in the style of Giacomini/Rossi (2010).

It becomes clear that a forecast of the number of gainfully employed people is more accurate when disaggregated information is used. Moreover, it is possible to identify certain stages which cause this effect and some branches which prove to be especially relevant.

The paper at hand is structured as follows: Section 2 below explains the models used. Subsequently, the results are discussed in Section 3, and the corresponding conclusions are drawn in Section 4.

2 Models

We know from labour market theory that employment figures hinge on a variety of influencing values such as wage level, production volume, working hours, etc. But we do not feel that it makes sense to integrate all influencing values into a holistic model where our research question is concerned. Instead, the models should initially have as little complexity as possible. The first model comparison leaves out any other influencing values entirely; we are estimating an AR(p) process. Taking disaggregated data into account, and due to the mutual dependencies of gainfully employed people in different sectors, we are assuming a vector-autoregressive model (VAR(p) model):

$$y_t = A_1 \cdot y_{t-1} + A_2 \cdot y_{t-2} + \cdots + A_p \cdot y_{t-p} + \varepsilon_t$$

with

$$y_t \coloneqq \left(y_{g,t}, y_{1,t}, \cdots, y_{N-1,t}\right)' \qquad A_l := \begin{pmatrix} \alpha_{g,g}^l & \alpha_{1,g}^l & \cdots & \alpha_{N-1,g}^l \\ \alpha_{g,1}^l & \alpha_{1,1}^l & \cdots & \alpha_{N-1,1}^l \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{g,N-1}^l & \alpha_{1,N-1}^l & \cdots & \alpha_{N-1,N-1}^l \end{pmatrix}, l = 1, \cdots, p$$

$$\varepsilon_t \coloneqq \left(\varepsilon_{g,t}, \varepsilon_{1,t}, \cdots, \varepsilon_{N-1,t}\right)'$$

where $y_{g,t}$ is the number of gainfully employed people in the economy as a whole at a given time t and $y_{n,t}$ is the number of gainfully employed people in the industry n $(n = 1, \cdots, N-1)$ at a given time t. Even though we distinguish between N industries in total, only $N-1$ of these are considered in the model, as otherwise we would have perfect multicollinearity.

Hendry and Hubrich (2011) show that this method yields forecasts for the aggregate which are equivalent to summing up individual forecasts for all industries. It is therefore possible to treat a typical disaggregation problem in the framework of this model. When the A_l for all $\alpha_{n,g}^l$ and n $(n = 1, \cdots, N-1)$ are set to zero a priori in the matrices l $(l = 1, \cdots, p)$, the gainfully employed at the industry level $y_{n,t}$ are excluded from the model for the aggregate, which results once again in the AR(p) process for the overall employees $y_{g,t}$. The models are thus nested.

While we have up to now considered the completely aggregated case and the completely disaggregated case, it would also be possible to achieve the best forecasts at a medium degree of disaggregation. In order to examine that effect, it would not be practicable to repeat the model comparison with varying numbers of industries, because the results are influenced by the types of industry which are collected. Therefore, we will begin by identifying the principal components of the gainfully employed people in different sectors and then proceed to include these in the VAR model as additional variables.

So the second model comparison will again compare the AR(p) process for the overall employees $y_{g,t}$ to a $\mathrm{VAR}_{\mathrm{PC}}$(p) model. However, this time the VAR(p) model includes the principal components rather than the disaggregated number of gainfully employed people. So one has:

$$y_t \coloneqq \left(y_{g,t}, pc_{1,t}, \cdots, pc_{C,t}\right)'$$

where $pc_{c,t}$ is the p-th principal component $(c = 1, \cdots, C)$ at a given time t. Employment is typically influenced by the development of economic activity. That is why we examine in the third model comparison which effects result when the price-adjusted gross value added z_t is included in the model as an additional, explanatory variable. Hence, the following VARX(p) model was examined:

$$y_t = A_1 * y_{t-1} + A_2 * y_{t-2} + \cdots + A_p * y_{t-L} + C_1 * z_t + C_2 * z_{t-1} + \cdots + C_{L+1} * z_{t-p} + \varepsilon_t$$

with

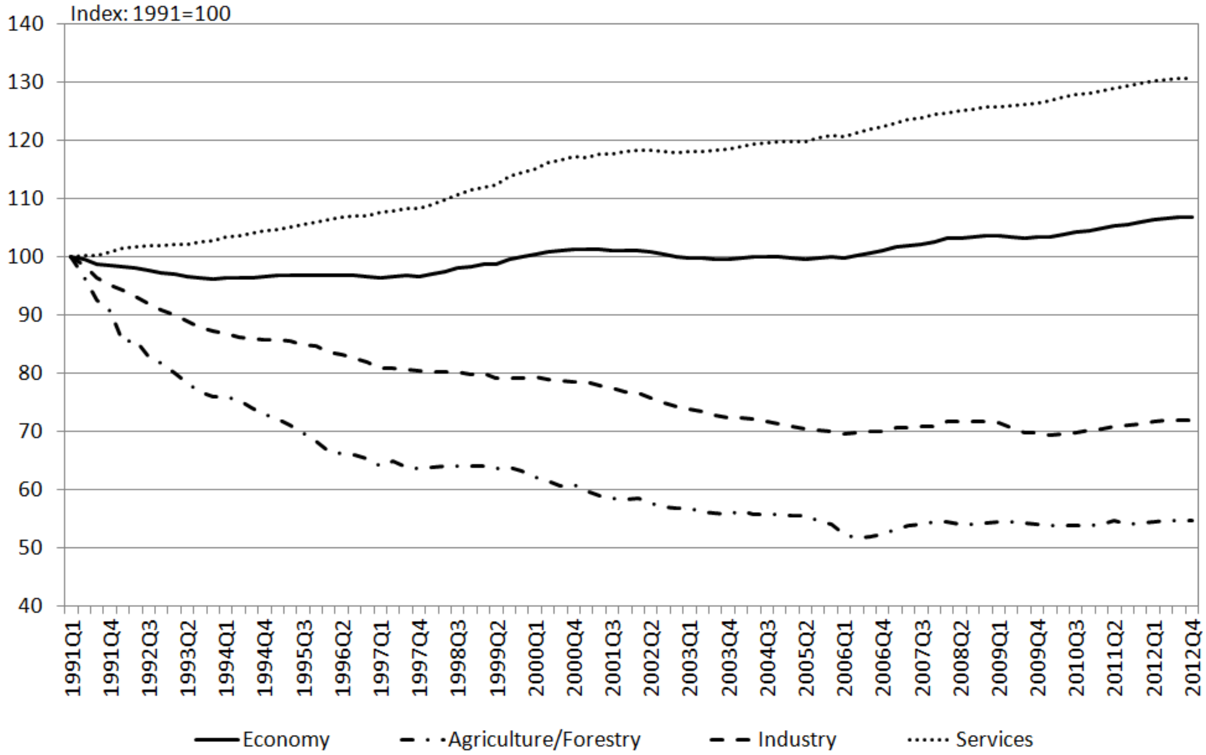
$$C_l \coloneqq \left(\gamma_g^l, \gamma_1^l, \cdots, \gamma_{N-1}^l\right)', l = 1, \cdots, p$$

3 Model Comparison

3.1 Database

The model comparisons are based on the quarterly statistics from the national accounts (Volkswirtschaftliche Gesamtrechnungen, VGR) of the German Federal Statistical Office. The data for the whole of Germany range from 1991 Q1 to 2012 Q4. The quarterly statistics of the VGR distinguishes between 10 industries in total. In the early 1990s, the economy as a whole experienced a decline in employment, which could be compensated during the upswing at the turn of the millennium (cf. Figure 1). Since the middle of the last decade, employment has been steadily increasing, with only a short-term interruption due to the financial crisis of 2008/2009.

Figure 1
Employees from 1991 Q1 to 2012 Q4 – index (1991 Q1 = 100)



Source: Federal Statistical Office; own calculations.

At the industry level, however, employment development varies greatly. For instance, the reduction of employment at the beginning of the 1990s only affected the manufacturing industry and agriculture. The services sector, on the other hand, registers a continuous growth in employment, even during the great recession, which affected the manufacturing industry most and caused it to reduce employment.

3.2 Procedure

Our first step is to specify the lag length for the models used. Since the available database is not very long, a maximum lag length of 4 quarters would seem the best choice for reasons of economy. A test concerning the optimal lag structure resulted in a length of p = 1 pursuant to both the Schwarz as well as the Hannan-Quinn information criteria. But not only these criteria speak in favour of a lag length of 1. Model comparisons with different lag lengths also showed that the forecast error is generally smallest at a lag length of 1.

The following model comparisons are thus carried out:

- AR(1) vs. VAR(1)
- AR(1) vs. $VAR_{PC}(1)$
- ARX(1) vs. VARX(1)

In order to determine the forecast errors, the available database is split into an estimation period and an evaluation period with $T = 88$ points in time. A recursive window is used to make sure that as many labour market cycles as possible can be considered in the evaluation period; however, this leads to a relatively short initial estimation period. We chose the period from 1997 to 2012 as our evaluation period, thus spanning $T_1 = 64$ points in time. So with a forecast horizon of $h = 1$ quarter, there are only 24 points in time (1991 Q1 up to and including 1996 Q4) available for the estimation. With a forecast horizon of $h = 6$ quarters, it would only be 19 points in time (1991 Q1 up to and including 1995 Q3). For every additional forecast, however, the estimation period is extended by one quarter.

Using a rolling window, on the other hand, results in a constant number of points in time during the estimation period for each forecast. They are merely postponed by one period. The advantage here is that the test results are not influenced by the duration of the estimation period. On the other hand, however, the estimation period should be long enough to achieve sound results. We chose a length of 40 points in time for the estimation period. Taking the forecast horizon h into account, the evaluation period thus lasts from 2002 Q2 to 2012 Q4 and includes $T_1 = 43$ points in time. Assuming $h = 1$, the estimation period of the first forecast is the period from 1992 Q2 up to and including 2002 Q1, and that of the last forecast is the period from 2002 Q4 up to and including 2012 Q3.

We are using iterative forecasts, i.e. for a preferred lag length of 1, we have:

$$\hat{y}_{h,t} = \hat{A}_1 \cdot \hat{y}_{h-1,t-1} \,, \qquad h = 1, \cdots, 6 \quad , \quad t \in [T - T_1 + 1, \cdots, T]$$

with $\hat{y}_{0,t} = y_t$

In those models which include an exogenous variable, information on z is required for the employment forecast. However, this is not available in out-of-sample forecasts. Therefore, an AR(1) process with coefficient ρ_1 is added to the existing model in order to allow predicting the exogenous variable. The forecasts of the overall employees therefore result in such models pursuant to:

$$\hat{y}_{h,t} = \hat{A}_1 \cdot \hat{y}_{h-1,t-1} + \hat{C}_1 \cdot \hat{z}_{h,t} + \hat{C}_2 \cdot \hat{z}_{h-1,t-1} \qquad h = 1, \cdots, 6 \;\; ; \;\; \hat{z}_{h,t} = \hat{\rho}_1 \cdot \hat{z}_{h-1,t-1}$$

and $t \in [T - T_1 + 1, \cdots, T]$

where again $\hat{y}_{0,t,} = y_t$ and $\hat{z}_{0,t} = z_t$

3.3 Clark/West Tests

Since all these model comparisons compare a VAR model to an AR process which can be directly derived from the VAR model by setting parameters to zero, we are examining nested models. We are therefore using the Clark/West test statistic to assess the accuracy of the estimations. Clark and West (2007) argue that the mean squared prediction error (MSPE) of the larger model is distorted upwards, because parameters must be estimated which should be set to zero under the null hypothesis of equal predictive accuracy of both models. The smaller reference model would be more efficient if it were not encumbered by the need to estimate the parameters of a redundant variable. Ergo, the usual Diebold-Mariano style tests (1995) are undersized and have poor power in nested models. The test statistic is:

$$CW_h = \sqrt{\frac{1}{\sigma_h^2 \cdot T_1}} \sum_{t=T-T_1+1}^{T} \left(\hat{\varepsilon}_{1,h,t}^2 - \left[\hat{\varepsilon}_{2,h,t}^2 - \left(\hat{y}_{1,h,t} - \hat{y}_{2,h,t}\right)^2\right]\right)$$

where $\hat{\varepsilon}_{1,h,t}^2 = \left(y_{1,h,t} - \hat{y}_{1,h,t}\right)^2$ is the MSPE for the h-step forecast of the more economical model, i.e. the AR(1) process, and $\hat{\varepsilon}_{2,h,t}^2 = \left(y_{2,h,t} - \hat{y}_{2,h,t}\right)^2$ is the MSPE from the larger model, i.e. the VAR(1). $\left(\hat{y}_{1,h,t} - \hat{y}_{2,h,t}\right)^2$ is the correction factor introduced by Clark/West.

3.4 Fluctuation-Window Tests

The Clark/West test, however, only tells us whether one model is superior to the other model across the entire evaluation period. But it is not possible to use this test to determine whether there are certain stages in labour market development where one or the other model is preferable. To this end, we are carrying out fluctuation-window tests in the style of Giacomini and Rossi (2010). For these tests, the initial evaluation period spanning $T_1 = 64$ and $T_1 = 43$ points in time, respectively, is split into 56 and 35 sub-evaluation periods, respectively, with a length of $T_2 = 8$ quarters. For each sub-period, the Clark/West test statistic is calculated on the basis of the recursive (rolling) windows, where the estimation period is prolonged (postponed) by one period for each sub-period:

$$CW_{t_2,h} = \sqrt{\frac{1}{\sigma_{t_2,h}^2 \cdot T_2}} \sum_{t=T-T_1+t_2}^{T-T_1+t_2+T_2} \left(\hat{\varepsilon}_{1,h,t}^2 - \left[\hat{\varepsilon}_{2,h,t}^2 - \left(\hat{y}_{1,h,t} - \hat{y}_{2,h,t}\right)^2\right]\right)$$

with $t \in [T - T_1 + 1, \cdots, T]$ and $t_2 \in [1, \cdots, T_1 - T_2]$

The critical values depend on the size relation of the estimation window and the evaluation window and are taken from Giacomini and Rossi (2010).

3.5 Test Results

The section below discusses the individual model comparisons. To this end, first the Clark/West tests with the different databases are analysed and after that, the results of the fluctuation-window tests.

3.5.1 AR(1) versus VAR(1) model

The first model comparison shows that the mean squared prediction error can be reduced when sectoral information is included (cf. Table 1). Regardless of the choice of evaluation period (recursive-window database or rolling-window database) and the length of the forecast period, disaggregation yields significantly more accurate predictive results. However, the benefits of sectoral information decrease slightly with a longer forecast horizon. Moreover, it becomes clear that a rolling-window database, i.e. a fixed length of the evaluation period (40 quarters), makes the differences stand out more clearly.

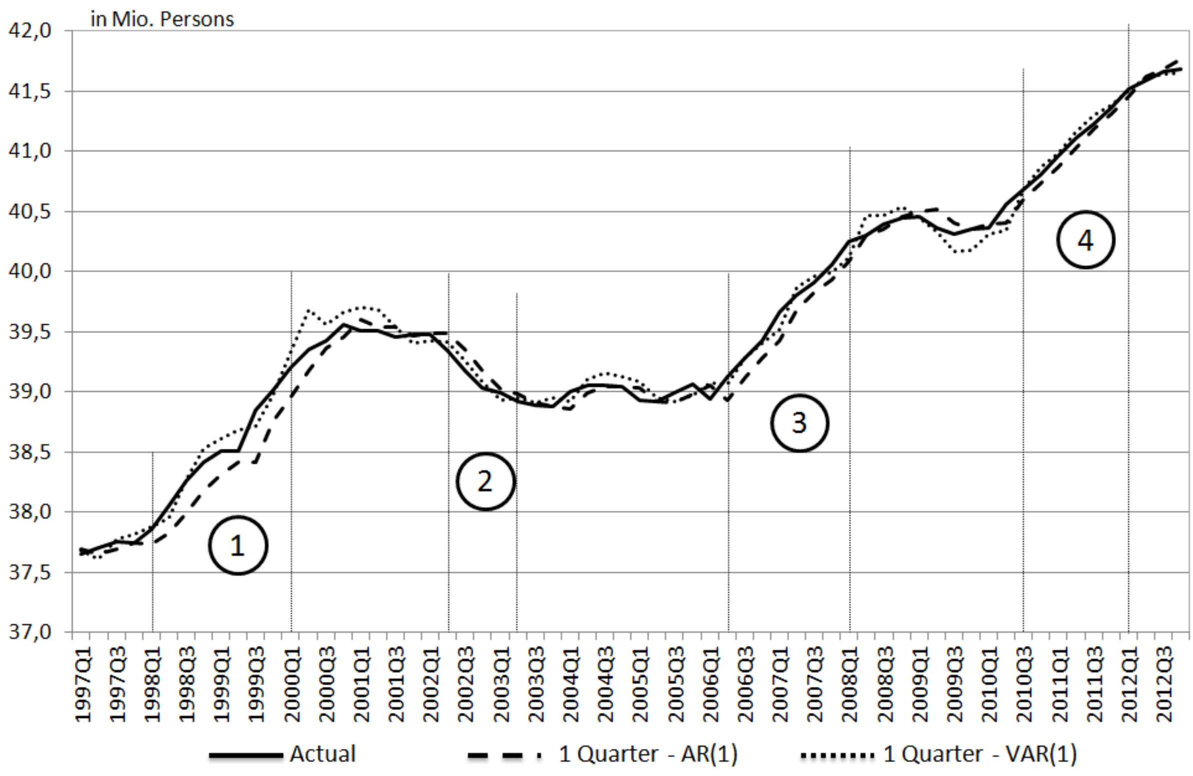
Table 1
AR(1)- vs. VAR(1) model – Clark/West test statistic (p-value)

Database	Quarters Forecast					
	1	**2**	**3**	**4**	**45**	**6**
Recursive Window	2,743 (0,003)	2,614 (0,004)	2,529 (0,006)	2,332 (0,010)	2,114 (0,017)	1,935 (0,027)
Rolling Window	3,366 (0,000)	3,472 (0,000)	3,361 (0,000)	3,123 (0,001)	2,697 (0,003)	2,318 (0,010)

Source: Own calculations.

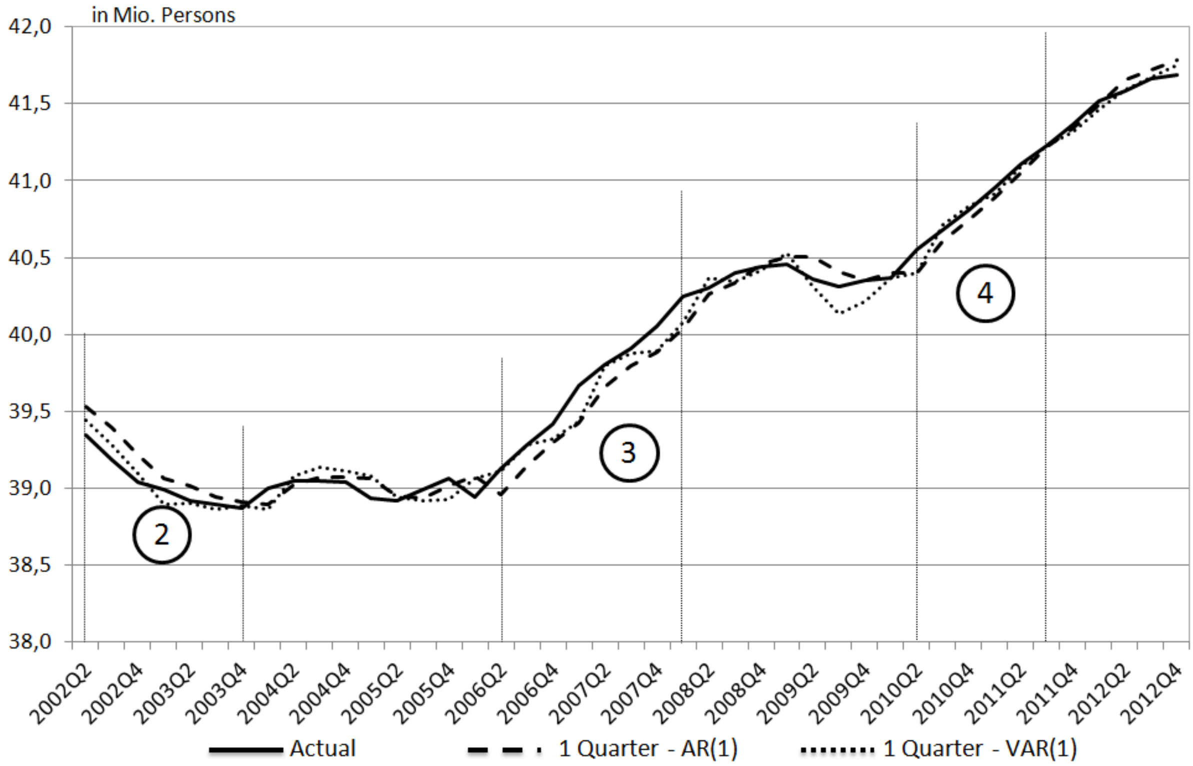
This, however, does not allow the conclusion that the sectoral forecast will *always* yield more accurate forecasts. This is further supported by the Figures 2a and 2b, where the actual values and the forecast values of the one-quarter forecast of both models are depicted for the recursive-window database (Figure 2a) and the rolling-window database (Figure 2b). Ultimately, we can identify 3 and 4 periods of time, respectively, where the dotted line is closer to the straight line, meaning that the use of disaggregated information allows for more accurate forecasts. What stands out is that during these periods, employment has been either increased or reduced continuously, whereas no clear statement can be made on the periods of time that contained the inflection points and saddle points.

Figure 2a
Employees – Actual, AR(1)- and VAR(1)- 1-quarter forecasts – recursive window –



Source: Federal Statistical Office; own calculations.

Figure 2b
Employees – Actual, AR(1)- and VAR(1)- 1-quarter forecasts – rolling window –

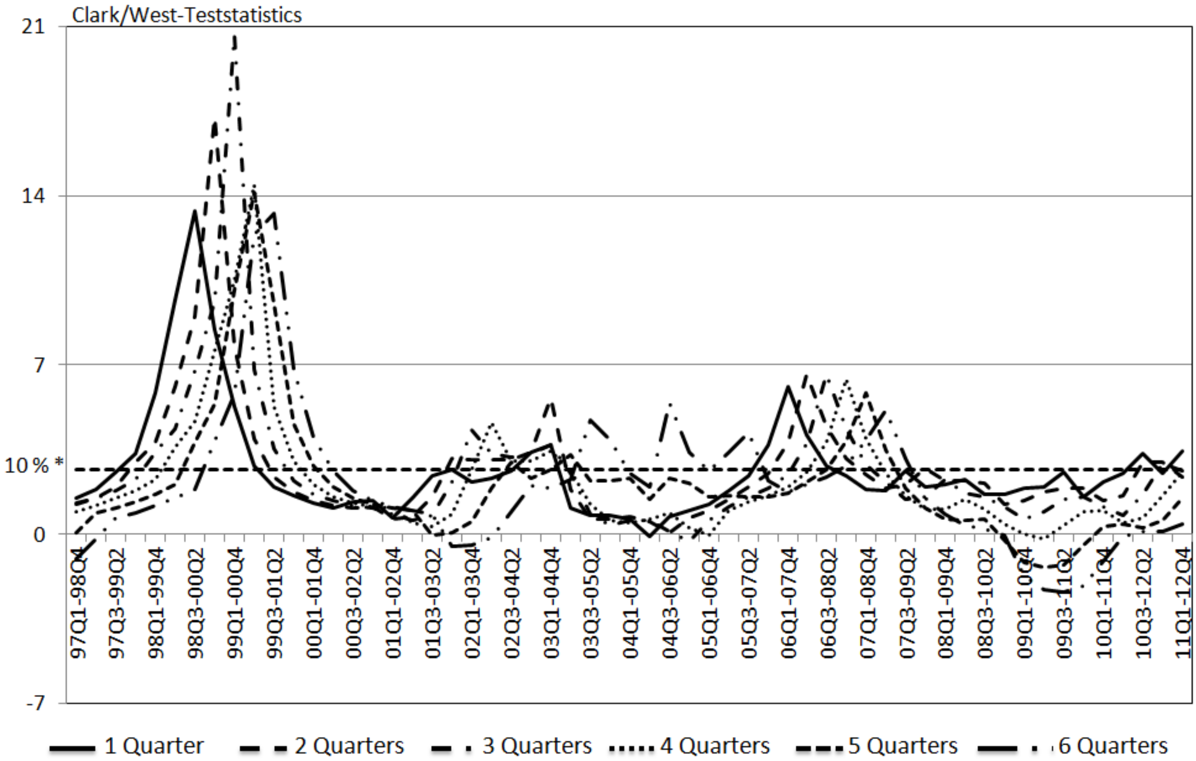


Source: Federal Statistical Office; own calculations.

It now remains to find out, with the help of the fluctuation-window test, whether disaggregation really yields significantly better forecasts during these previously identified periods of time. As explained before, this involves splitting the initial evaluation period into many short evaluation periods. Figures 3a and 3b show the test statistics of these short evaluation periods for the different forecast horizons. If our assumption regarding the identified periods is correct, these 4, respectively 3 periods in Figures 3a and 3b should appear as peaks.

For the recursive-window database, these peaks are clearly identifiable especially for the first and the third period (cf. Figure 3a, straight line). But the test statistics for the periods 2 and 4 are also still significant. In general, the Clark/West test statistics yield a positive value, with a few exceptions, across all evaluation periods, i.e. the forecasts with additional disaggregation tend to be more accurate.

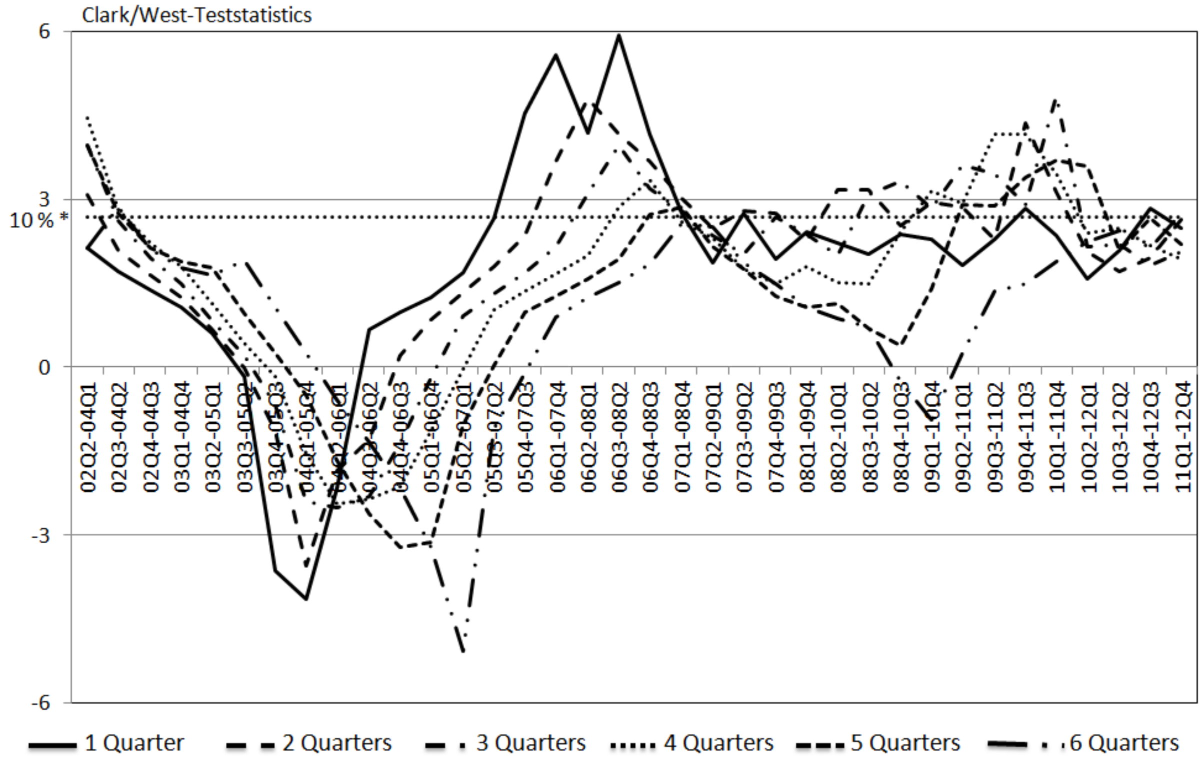
Figure 3a
Results of the fluctuation-window test for different forecast horizons – AR(1)- vs. VAR(1) model – recursive window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

The results of a rolling-window database (cf. Figure 3b) are a little more differentiated. The only unambiguous period here is period 3. For period 4, the test statistic is borderline significant, but not anymore so for period 2; however, significance is still achieved here for other forecast horizons. As explained above, the evaluation period in the rolling-window database does not begin until 2002 Q2, due to the estimation period spanning 40 quarters. It strikes the reader that the first peak is negative. On the other hand, the test statistics for forecasts after early 2006 are close to the critical value for most horizons.

Figure 3b
Results of the fluctuation-window test for different forecast horizons – AR(1)- vs. VAR(1) model – rolling window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

What stands out is the parallel course of the test statistics, each delayed by one quarter, for the different forecast horizons as well as the drifting apart at the current margin and/or in the past few years. Whether, and to which extent, the use of disaggregated information leads to an improvement of the aggregated forecast appears to depend on the point in time at which the forecasts are made and/or the stage one is in at that moment. Since the peaks for the individual quarters occur delayed by exactly one period, the corresponding forecasts are always based on the same data points.

Apart from the parallel course of the individual test statistics, it also stands out in Figures 2a and 2b that the difference between the results increases towards the end of the evaluation period, depending on the forecast horizon. When taking a recursive-window database as a basis and disaggregating at the current margin, forecast accuracy goes down as the forecast horizon increases. This effect cannot be observed on a rolling-window database. Here, the test statistics for all forecast horizons are all very close to the critical value, also at the current margin. It appears that the disaggregated model can adjust better to the current employment dynamics for longer-term forecasts when a rolling window is used.

So it can be concluded that, regardless of how the database is used, disaggregation significantly increases forecast accuracy as compared to an aggregated forecast and that certain periods of time can be identified where this becomes especially evident.

3.5.2 AR(1) versus VARPC(1) model

In the second model comparison, the AR(1) process is compared to the $VAR_{PC}(1)$ model. In the case of the latter model, the ten available economic industries are first used to identify up to nine principal components, which are then included in the VAR model together with the aggregate variable.

The tests show that, with a recursive-window database, disaggregation always yields significantly better forecast accuracy, regardless of the number of principal components (cf. Table 2). It also becomes clear, however, that the test statistics lose significance as the forecast horizon is increased. Improvements occur up to a number of five principal components, but there is hardly any additional increase in significance after that.

On a rolling-window database, the results regarding the increasing forecast horizon are similar. There are, however, some deviations where the number of principal components is concerned. For example, there is no identifiable significant difference between the aggregated and the disaggregated forecast when 3 principal components are used for forecast horizons of more than 4 quarters and when 9 principal components are used for a forecast horizon of 2 or more quarters. Apart from that, all disaggregated forecasts are significantly more accurate in this case as well.

Since the disaggregated forecast is more accurate than the aggregated forecast to a highly significant degree when 5 principal components are used, regardless of the assumed database, we will only discuss the results for this case in the following. The results of the fluctuation-window test for the remaining cases are available on demand.

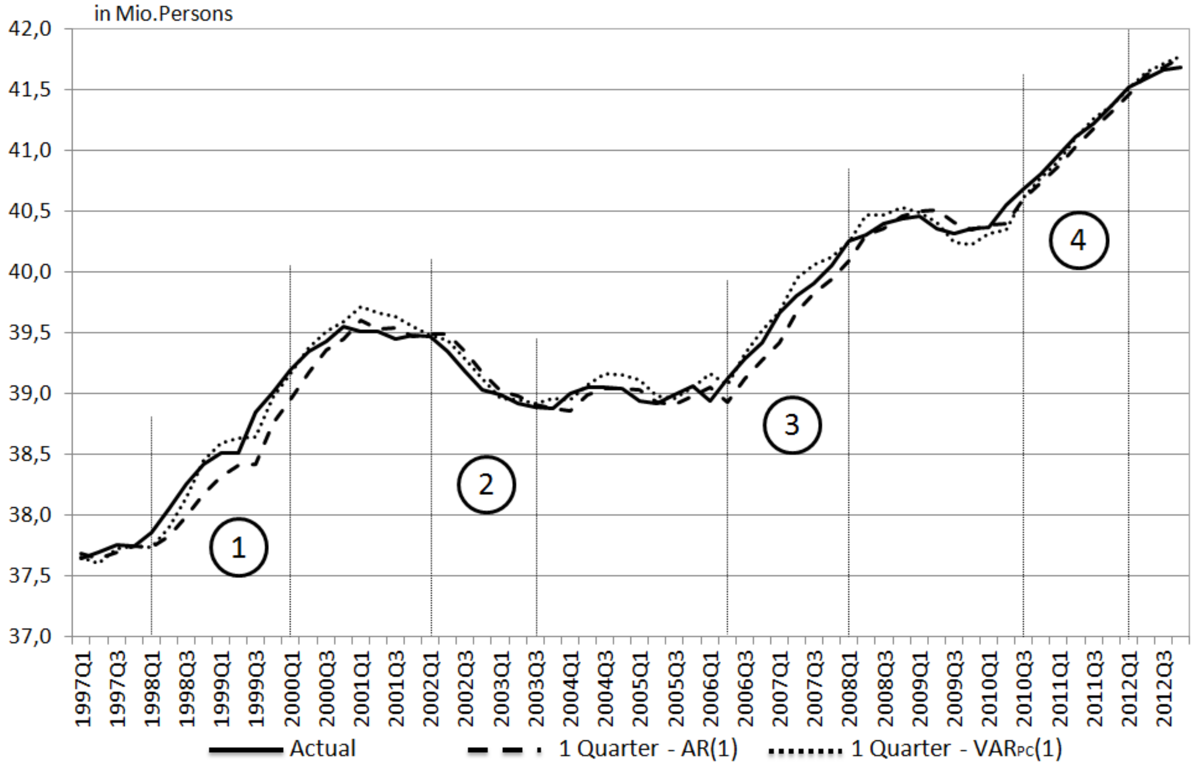
When using 5 principal components, it is possible, in turn, to identify 4, respectively 3, periods where the use of sectoral information increases forecast accuracy (cf. Figures 4a and 4b).

Table 2
AR(1) vs. VAR$_{PC}$(1) model – Clark/West test statistic (p-value)

Database	Principal Components	Quarters Forecast					
		1	2	3	4	45	6
Recursive Window	1	1,794 (0,036)	1,752 (0,040)	1,721 (0,043)	1,704 (0,044)	1,686 (0,046)	1,669 (0,048)
	2	1,755 (0,040)	1,664 (0,048)	1,587 (0,056)	1,514 (0,065)	1,429 (0,076)	1,337 (0,091)
	3	2,003 (0,023)	1,924 (0,027)	1,820 (0,034)	1,694 (0,045)	1,584 0,057)	1,494 (0,068)
	4	2,231 (0,013)	2,147 (0,016)	2,011 (0,022)	1,831 (0,034)	1,648 (0,05)	1,495 (0,067)
	5	2,804 (0,003)	2,687 (0,004)	2,515 (0,006)	2,314 (0,010)	2,105 (0,018)	1,937 (0,026)
	6	2,728 (0,003)	2,577 (0,005)	2,358 (0,009)	2,147 (0,016)	1,944 (0,026)	1,785 (0,037)
	7	2,924 (0,002)	2,732 (0,003)	2,497 (0,006)	2,268 (0,012)	2,028 (0,021)	1,836 (0,033)
	8	2,714 (0,003)	2,621 (0,004)	2,484 (0,006)	2,300 (0,011)	2,078 (0,019)	1,886 (0,030)
	9	2,743 (0,003)	2,614 (0,004)	2,530 (0,006)	2,332 (0,010)	2,114 (0,017)	1,935 (0,027)
Rolling Window	1	1,566 (0,059)	1,719 (0,043)	1,820 (0,034)	1,816 (0,035)	1,684 (0,046)	1,293 (0,098)
	2	2,286 (0,011)	2,443 (0,007)	2,508 (0,006)	2,424 (0,008)	2,263 (0,012)	1,847 (0,032)
	3	2,098 (0,018)	2,055 (0,020)	1,831 (0,034)	1,460 (0,072)	1,080 (0,140)	0,714 (0,238)
	4	2,898 (0,002)	2,800 (0,003)	2,517 (0,006)	2,118 (0,017)	1,707 (0,044)	1,365 (0,086)
	5	3,063 (0,001)	3,142 (0,001)	2,934 (0,002)	2,569 (0,005)	2,200 (0,014)	1,900 (0,029)
	6	2,891 (0,002)	2,981 (0,001)	2,824 (0,002)	2,594 (0,005)	2,368 (0,009)	2,134 (0,016)
	7	3,127 (0,001)	3,286 (0,001)	3,257 (0,001)	3,031 (0,001)	2,642 (0,004)	2,140 (0,016)
	8	3,214 (0,001)	3,387 (0,000)	3,320 (0,000)	3,077 (0,001)	2,688 (0,004)	2,302 (0,011)
	9	1,509 (0,066)	1,061 (0,144)	-1,054 (0,146)	-1,051 (0,147)	-1,052 (0,147)	1,052 (0,146)

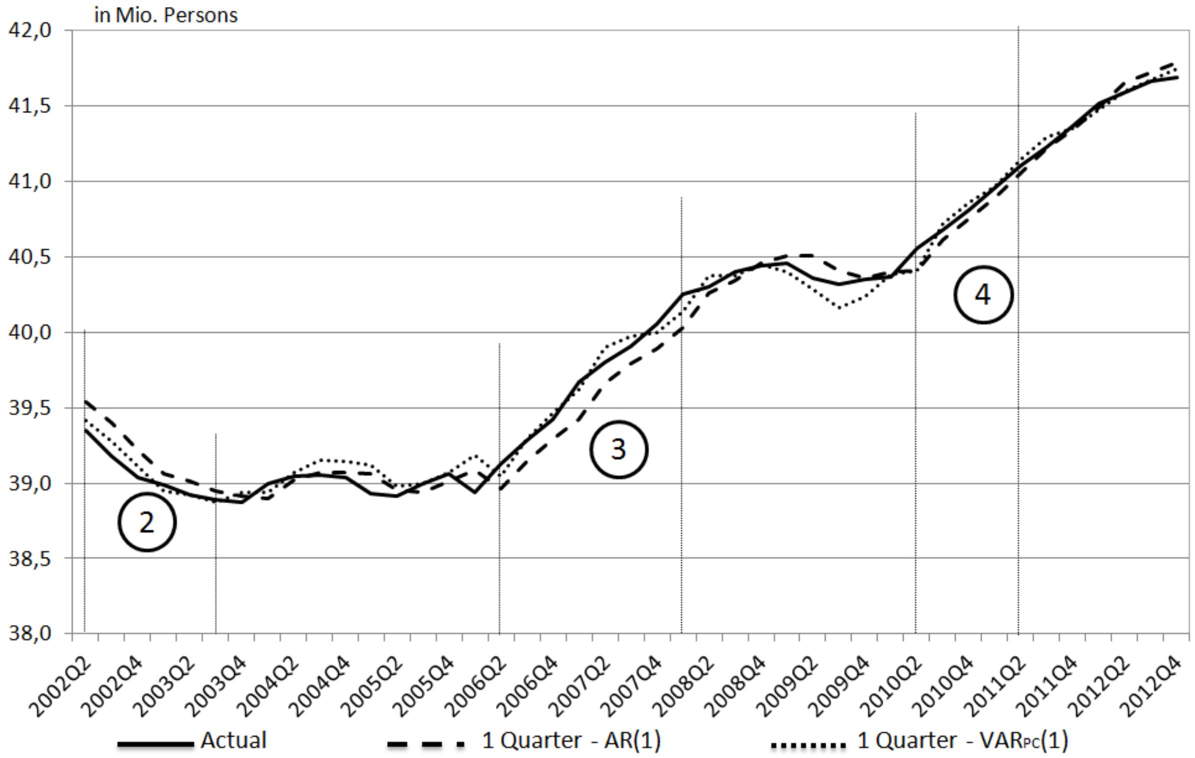
Source: Own calculations.

Figure 4a
Employees – Actual, AR(1)- and VAR$_{5PC}$(1)- 1-quarter forecasts; 5 PCs – recursive window –



Source: Federal Statistical Office; own calculations.

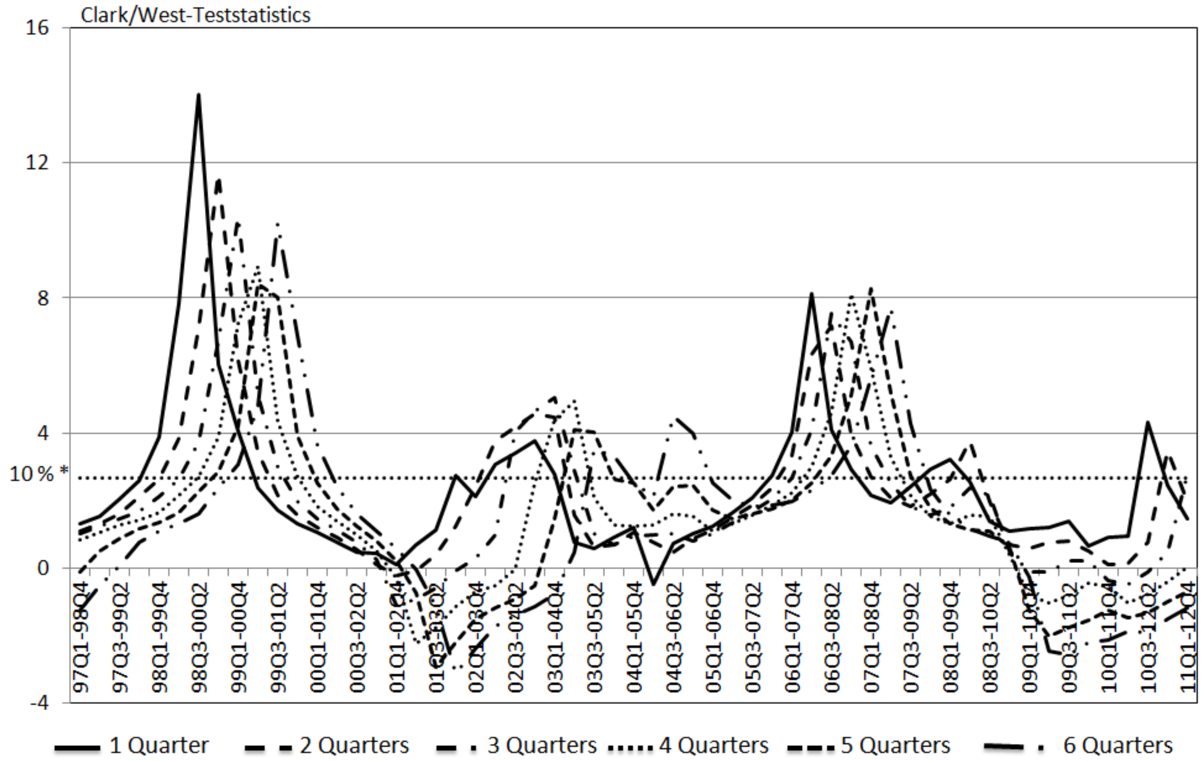
Figure 4b
Employees – Actual, AR(1)- and VAR$_{5PC}$(1)- 1-quarter forecasts; 5 PCs – rolling window –



Source: Federal Statistical Office; own calculations.

These periods of time can also be found as peaks in the behaviour of the individual test statistics of the fluctuation-window tests (cf. Figures 5a and 5b). In the case of the recursive-window database, the peaks for the periods 2 and 4 are even a little more pronounced when compared to the previous model comparison. When seen across the entire evaluation period, the behaviour of the test statistics strongly resembles their behaviour without principal components. Similar results could also be achieved using 8 or 9 principal components. With any other number of principal components, the test statistics would deteriorate.

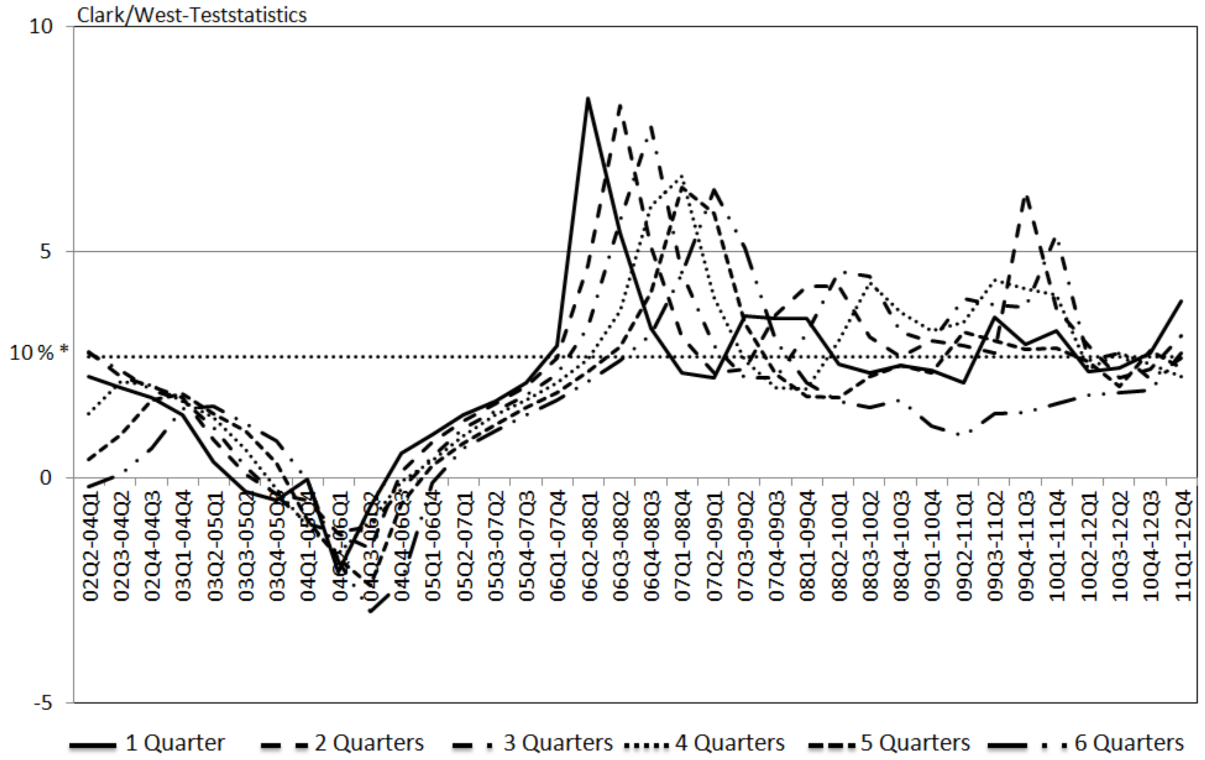
Figure 5a
Results of the fluctuation-window test for different forecast horizons – AR(1)- vs. VAR$_{5PC}$(1) model – recursive window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

When using a rolling-window database, the peaks also show a tendency to be more pronounced due to the identification of principal components. Moreover, another period (2007 Q3 to 2009 Q4) can be identified for the 1-quarter forecast where disaggregation is significantly better. On the whole, test statistic's behaviour when using 5 principal components once again resembles that from the first model comparison. However, when using 5 principal components, all disaggregated forecasts (with the exception of the 6-quarter forecast) since early 2006 lead to significantly better forecast accuracy.

Figure 5b
Results of the fluctuation-window test for different forecast horizons – AR(1)- vs. VAR$_{5PC}$(1) model – rolling window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

By identifying principal components, an improvement can be achieved for both rolling-window and recursive-window databases. The number of components is also relevant.

3.5.3 ARX(1) versus VARX (1) model

Now we will examine the behaviour of forecast accuracy when an additional exogenous variable is included in the model. To this end, we include an explanatory variable in the model which is used in employment forecast by default: price- and season-adjusted gross value added. The simple AR(1) process thus becomes an ARX(1) process, and the VAR(1) model becomes a VARX(1) model. It becomes clear that here, too, disaggregation has a significantly smaller forecast error than the aggregated forecast. Interestingly enough, however, the differences between the disaggregated and the aggregated forecast become a little smaller due to the inclusion of the gross value added (cf. Table 3). Apparently, the inclusion of the gross value added can compensate for part of the lacking disaggregated information.

While the p-values also deteriorate when a rolling-window database is used, the disaggregated forecasts here are still significantly more accurate, even if the forecast horizon is 6 quarters.

Table 3
ARX(1)- vs. VARX(1) model – Clark/West test statistic (p-value)

Database	Quarters Forecast					
	1	**2**	**3**	**4**	**45**	**6**
Recursive Window	2,881 (0,002)	2,740 (0,003)	2,596 (0,005)	2,167 (0,015)	1,564 (0,059)	1,142 (0,127)
Rolling Window	2,323 (0,010)	2,549 (0,005)	2,239 (0,013)	1,827 (0,034)	2,079 (0,019)	1,758 (0,039)

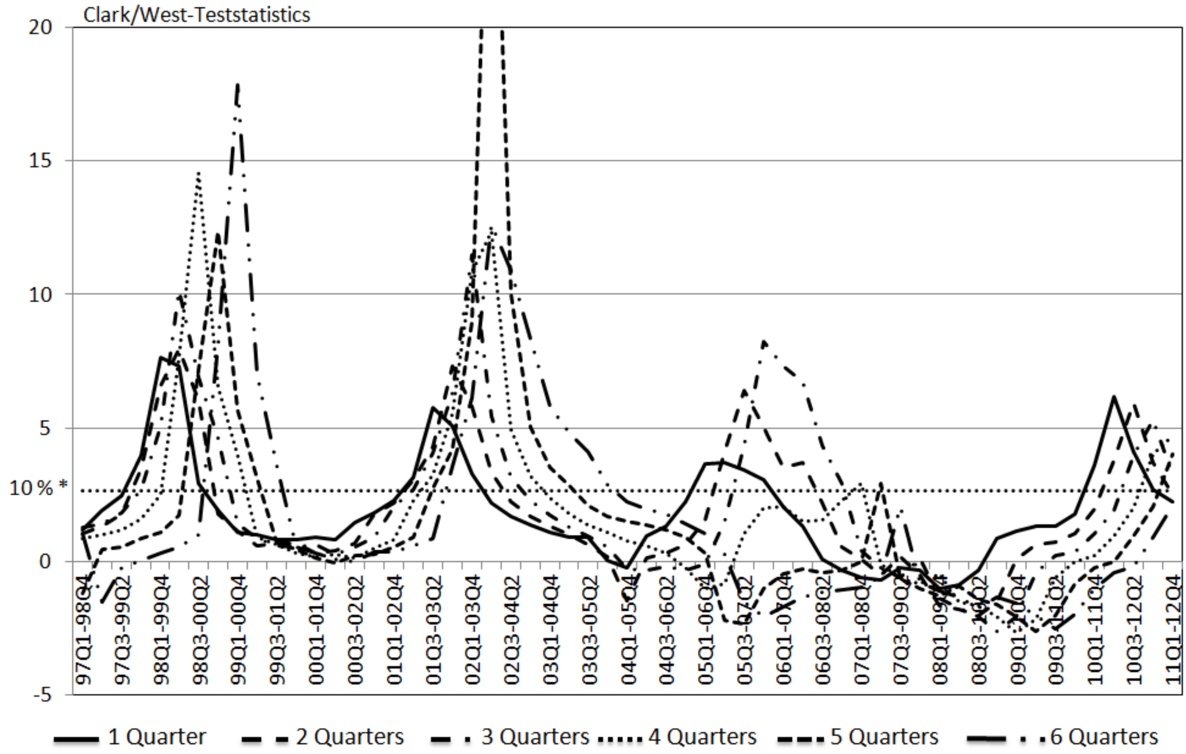
Source: Own calculations.

The behaviour of the test statistics from the fluctuation-window tests is also interesting. While the peaks can still be identified on a recursive-window database, they are much more pronounced for the first two periods in the forecasts with longer forecast horizons. This means that disaggregation can bring additional insights especially with longer forecast horizons. Interestingly enough, this does not apply to the third period, while the current margin still runs its usual course. It appears that about the year 2005, perhaps due to the radical labour market reforms (Hartz), the strong connections at the industry level between economic development (gross value added) and employment development - which accounted for the high forecast accuracy during the first two periods - were temporarily reversed or displaced.

Test statistics that included forecasts for the year 2005 also showed the greatest discrepancies when a rolling-window database was used (cf. Figure 6b). But the same is true for test statistics which concern the crisis years of 2009 and 2010.

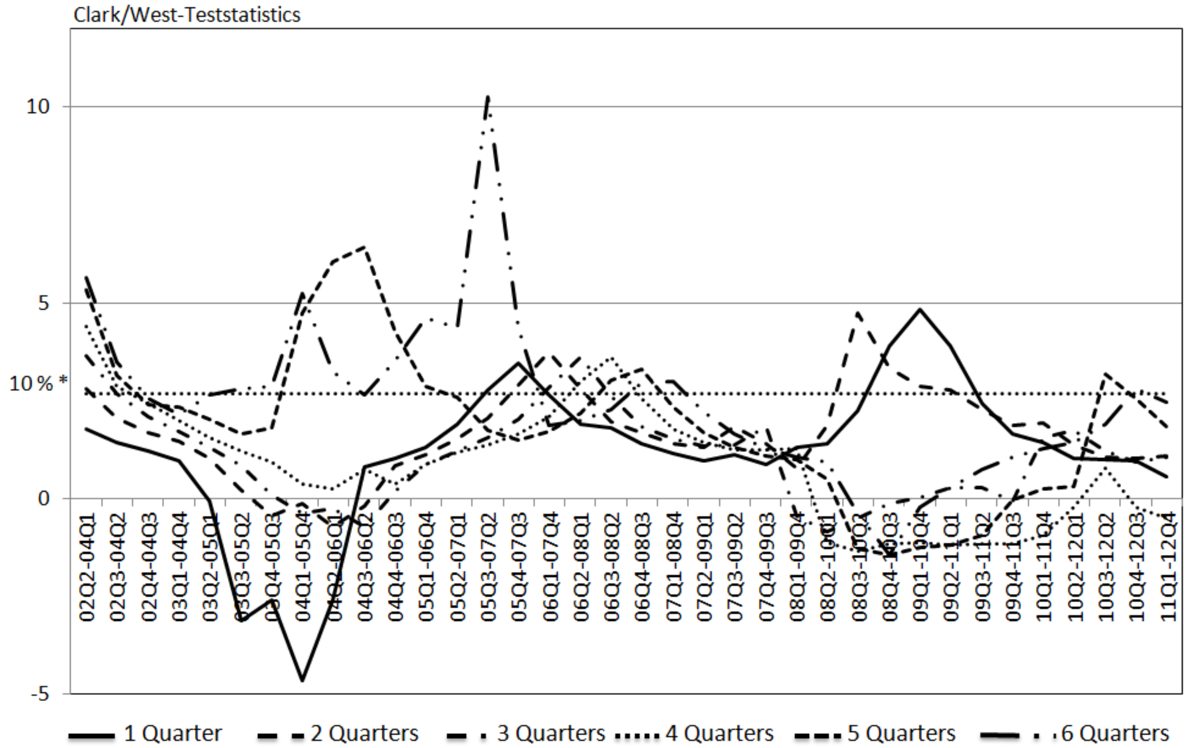
It can therefore be stated that when economic performance is considered, the use of sectoral information also improves forecast accuracy in general. However, it is no longer possible to clearly identify the periods where disaggregation is always advantageous. This result should not come as a surprise, considering the higher complexity of the model.

Figure 6a
Results of the fluctuation-window test for different forecast horizons – ARX(1)- vs. VARX(1) model – recursive window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

Figure 6b
Results of the fluctuation-window test for different forecast horizons – ARX(1)- vs. VARX(1) model – rolling window –



* Critical value of the 10 percent significance level.
Source: Own calculations.

3.5.4 Which branches should be considered?

Based on the test results, we have shown that the number of workers can be predicted significantly better by using disaggregated information. But so far, we don't know which industries are particularly important for forecasting the aggregate. This question will be explored in this section.

To investigate this question we consider the factor loadings of the principal components. The factor loadings indicate for each principal component the weight of the different sectors in a linear combination. Thus each sector would have a factor loading of one if all branches had the same weight. For reasons of clarity, we only consider the longest estimation sample, which provides the most observations for determining the loadings.

The test results in Section 3.5.2 show that five principal components achieve the most significant results for the recursive window, whereas further components bring no additional improvements. Therefore, the loadings of the first five principal components are relevant. In Table 4, the branches industry including 'Energy' (Branch No. 2), 'Construction' (Branch No. 3), 'Trade, Hotels, Transport' (Branch No. 4), 'Business-related Services' (Branch No. 8) and the branch 'Public services, education, health' (Branch No. 9) show large factor loadings (e.g, greater than two) in at least one of these principal components. In the variable aggregate employment each branch gets a natural loading of one. Thus, we can conclude that larger loadings indicate an important role of the relevant sectors over and above their natural role in the aggregate.[1]

Table 4
Factor Loadings of the Principal Components

Principal Compo-nents	Branches									
	1	2	3	4	5	6	7	8	9	10
1	0,197	2,152	0,702	2,311	0,296	0,319	0,099	0,995	2,230	0,698
2	-0,442	-2,891	-1,100	0,445	0,206	-0,087	0,183	2,672	1,282	0,693
3	-0,288	-2,099	2,916	0,663	-0,193	0,469	0,147	-1,820	1,122	0,284
4	-0,427	-0,489	-2,257	1,553	0,879	0,774	0,390	-2,126	-0,023	1,082
5	-0,426	-0,192	1,163	2,008	0,738	0,379	0,522	1,138	-3,033	0,402

Source: Own calculations.

In a further step, we consider the test statistics in case the simple AR(1) is compared with a VAR(1)-model which includes the five "unimportant" respectively the "important" sectors. This comparison ultimately confirms our above considerations,

[1] For better comparability, the mean of the absolute values of the loadings of each principal component is normalized to one.

since the test statistics in the second case are clearly larger and more significant than in the first case (see Table 5).

Table 5
AR(1) vs. VAR(1)-Model with 5 Branches – Clark/West Test Statistic (p-value)

Database	Branches	Quarters Forecast					
		1	2	3	4	45	6
Recursive Window	**1, 5, 6, 7, 10**	1,636 (0,051)	1,559 (0,060)	1,468 (0,071)	1,426 (0,077)	1,410 (0,079)	1,399 (0,081)
Rolling Window	**2, 3, 4, 8, 9**	2,732 (0,003)	2,615 (0,004)	2,444 (0,007)	2,241 (0,013)	2,039 (0,021)	1,876 (0,030)

Source: Own calculations.

The choice of the sectors seems quite plausible. Manufacturing and construction are known to react early and strongly to the business cycle. By the same token, trade and transport are often considered as forerunners of the economy, just as business-related services such as temporary agency employment.

4 Conclusion

Using the example of short-term forecasts for German employment figures, this paper examined the question whether the use of disaggregated information increases the forecast accuracy of the aggregate. For this purpose, out-of-sample forecasts for the aggregated employment forecast were compared to and contrasted with forecasts based on a vector-autoregressive model, which includes not only the aggregate but also the numbers of gainfully employed people at the industry level.

As shown by means of the individual statistical out-of-sample test, a forecast of the number of gainfully employed people is more accurate when disaggregated information is used. The mean squared prediction error is significantly smaller when the employment forecast is influenced by the development of the number of gainfully employed people at the industry level. It does not matter whether the forecasts are based on a fixed-length database (rolling-window database) or whether all available data points are used (recursive-window database), but disaggregation was a little more significant in the former case. By varying the degree of disaggregation, further improvements can be achieved, but these are no longer decisive. Hence, regardless of the number of principal components, disaggregation resulted in significantly more accurate forecasts in most cases. Further analysis showed that especially the branches ‘Industry including Energy, ‘Construction’, ‘Domestic Trade, Accommodation, Transport’, ‘Business-related Services’ and ‘Public Services, Education, Health’ drive this result.

Further analysis has shown that disaggregation is especially important when the number of gainfully employed people is increasing or decreasing. In stagnation phases, on the other hand, disaggregation is not necessary for the forecasting process; a simple AR process does not significantly deteriorate forecast accuracy.

Furthermore, the advantage of disaggregation cannot be compensated for by using additional information, i.e. including an exogenous variable in the model. While the significant difference between the aggregated and the disaggregated forecast is reduced, it still exists. Nevertheless, the optimum number of principal components becomes more important.

Our paper ultimately shows that industry-based disaggregation mostly improves the accuracy of the forecast of aggregated employment significantly. In labour market situations where the use of disaggregated information does not lead to any improvement, there should at least not be any deterioration, either. The forecast value of industry information for the German gross domestic product is also recognised by Drechsel/Scheufele (2012). So there may also be room for improvement in practice.

References

Angelini, E.; Camba-Mendez, G.; Giannone, D. et al. (2011): Short-term forecasts of euro area GDP growth. Econometrics Journal, Vol. 14, 20–24.

Clark, T.E.; West, K.D. (2007): Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Diebold, F.X.; Mariano, R.S. (1995): Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3/1995), 253–263.

Drechsel K.; Scheufele R. (2012): Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment. Swiss National Bank Working Papers 16.

Hendry, D.F.; Hubrich, K. (2011): Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate. Journal of Business & Economic Statistics 29 (2), 216–227.

Giacomini, R.; Rossi, B. (2010). Model Comparisons in Unstable Environments. Journal of Applied Econometrics 25 (4), 595–620.

Recently published

No.	Author(s)	Title	Date
27/2012	Stops, M.	Job matching across occupational labour markets	11/12
28/2012	Klinger, S. Weber, W.	Decomposing Beveridge curve dynamics by correlated unobserved components	12/12
29/2012	Osiander, Ch.	Determinanten der Weiterbildungsbereitschaft gering qualifizierter Arbeitsloser	12/12
1/2013	Fuchs, J. Weber, E.	A new look at the discouragement and the added worker hypotheses: Applying a trend-cycle decomposition to unemployment published in: Applied Economics Letters, Vol. 20, No. 15 (2013), p. 1374-1378	1/13
2/2013	Nordmeier, D. Weber, E.	Patterns of unemployment dynamics in Germany	4/13
3/2013	Zabel, C.	Effects of participating in skill training and workfare on employment entries for lone mothers receiving means-tested benefits in Germany	4/13
4/2013	Stephani, J.	Does it matter where you work? Employer characteristics and the wage growth of low-wage workers and higher-wage workers	5/13
5/2013	Moczall, A.	Subsidies for substitutes? New evidence on deadweight loss and substitution effects of a wage subsidy for hard-to-place job-seekers	5/13
6/2013	Schmillen, A. Umkehrer, M.	The scars of youth: Effects of early-career unemployment on future unemployment experiences	5/13
7/2013	Mönnig, A. Zika, G. Maier, T.	Trade and qualification: Linking qualification needs to Germany's export flows	6/13
8/2013	Alm, B. Engel, D. Weyh, A.	Einkommenseffekte von Betriebswechslern: Neue Befunde für Ostdeutschland	6/13
9/2013	Pauser, J.	Capital mobility, imperfect labour markets, and the provision of public goods	8/13
10/2013	Bauer, A.	Mismatch unemployment: Evidence from Germany 2000-2010	8/13
11/2013	Werner, D.	New insights into the development of regional unemployment disparities	8/13
12/2013	Eggs, J.	Unemployment benefit II, unemployment and health	9/13
13/2013	Vallizadeh, E. Muysken, J. Ziesemer, Th.	Migration, unemployment, and skill downgrading: A specific-factors approach	9/13

As per: 2013-09-17
For a full list, consult the IAB website

http://www.iab.de/de/publikationen/discussionpaper.aspx

Imprint

IAB-Discussion Paper 14/2013

Editorial address
Institute for Employment Research
of the Federal Employment Agency
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Jutta Sebald

Website
http://www.iab.de

Download of this Discussion Paper
http://doku.iab.de/discussionpapers/2013/dp1413.pdf

ISSN 2195-2663

For further inquiries contact the author:

Enzo Weber
Phone +49.911.179 7643
E-mail enzo.weber@iab.de

Gerd Zika
Phone +49.911.179 3072
E-mail gerd.zika@iab.de