

ORIGINAL ARTICLE

Open Access



A new indicator for nowcasting employment subject to social security contributions in Germany

Christian Hutter*

Abstract

Contrary to the number of unemployed or vacancies, the number of employees subject to social security contributions (SSC) for Germany is published after a time lag of 2 months. Furthermore, there is a waiting period of 6 months until the values are not revised any more. This paper uses monthly data on the number of people subject to compulsory health insurance (CHI) as auxiliary variable to better nowcast SSC. Statistical evaluation tests using real-time data show that CHI significantly improves nowcast accuracy compared to purely autoregressive benchmark models. The mean squared prediction error for nowcasts of SSC can be reduced by approximately 20%. In addition, CHI outperforms alternative candidate variables such as unemployment, vacancies and industrial production.

Keywords: Nowcasting, Real-time data, Employees, Social security contributions, Compulsory health insurance

JEL Classification: C53, E24, E27, J21

1 Introduction

Obtaining information on important macroeconomic variables as early as possible is important for forecasters, especially if they conduct short-term predictions. Among the main consumers of these predictions are policymakers who need regularly updated and reliable data as basis for their decisions, particularly in times of economic turmoil. This is also true for the labour market, where the focus often is on employment and unemployment development. However, while the number of unemployed in Germany is published with almost no delay and not revised in later months, the number of employees subject to social security contributions (SSC) is not published until after a time lag of 2 months. In addition, one has to wait 6 months to know the final data that are not subject to regular revisions any more.

Thus, data without publication lag and/or uncertainty about the current and past employment development are highly valuable. Especially during times of economic

crisis such as the great recession of 2008/2009 or the Corona pandemic of 2020, questions of how fast and how severely the employment figures are affected become highly relevant.

This paper aims to close this gap by investigating an auxiliary variable that has been rather neglected so far: employees subject to compulsory health insurance (CHI). This variable is published by the Federal Ministry of Health on a monthly basis and with almost no delay. Furthermore, it covers 97% of employees subject to SSC so that both variables are closely linked. These are promising prerequisites for improving nowcasts of employees subject to SSC.

In order to assess the usefulness of the potential auxiliary variable, I conduct out-of-sample nowcast tests based on real-time vintages. Hence, the paper takes account of the real-time, revised nature of the data (Clark and McCracken 2009). Consequently, I only calculate nowcasts using information that would have been available at the time. Nowcasts based on the best purely autoregressive benchmark model are compared to those stemming from a model enhanced by current and past values of the auxiliary

*Correspondence: christian.hutter@iab.de
Institute for Employment Research (IAB), Nuremberg, Germany

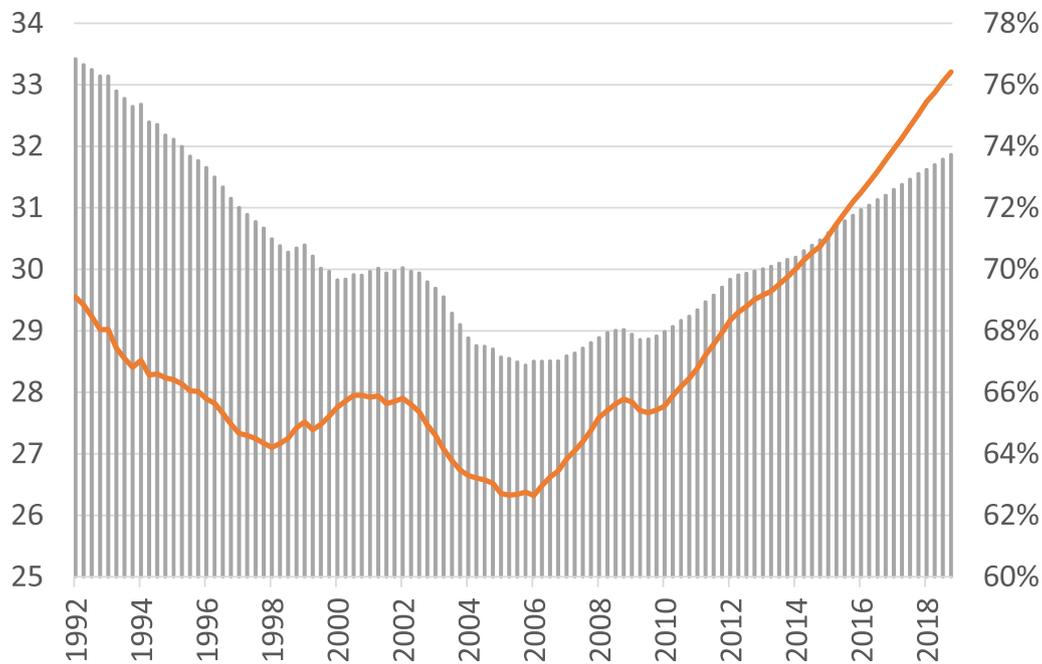


Fig. 1 Employees subject to social security contributions, 1992–2018. The graph shows the development of employees subject to social security contributions in million (orange line, left scale) and their share among all gainful workers (grey bars, right scale). Sources: Statistics Department of the Federal Employment Agency, Federal Statistical Office

variable. Statistical tests à la Clark and West (2007) that take into account the nested model environment show that employees subject to CHI indeed help to significantly outperform the purely autoregressive benchmark. Beyond statistical significance, the results are also economically relevant: The mean squared prediction error for both nowcast horizons (t_{-1} and t_0) can be reduced by approximately 20%. The value added of *CHI* is emphasized by the fact that it also outperforms other variables such as unemployment, vacancies or industrial production.

The paper proceeds as follows: The subsequent data section introduces the main target variable and the potential for improving its nowcast accuracy. Furthermore, it presents the auxiliary variable and proposes a method how to deal with the revisions connected to it. Section 3 presents the nowcasting equations and the results of the evaluation tests. It also investigates the performance of alternative candidate indicators. The final section concludes.

2 Data

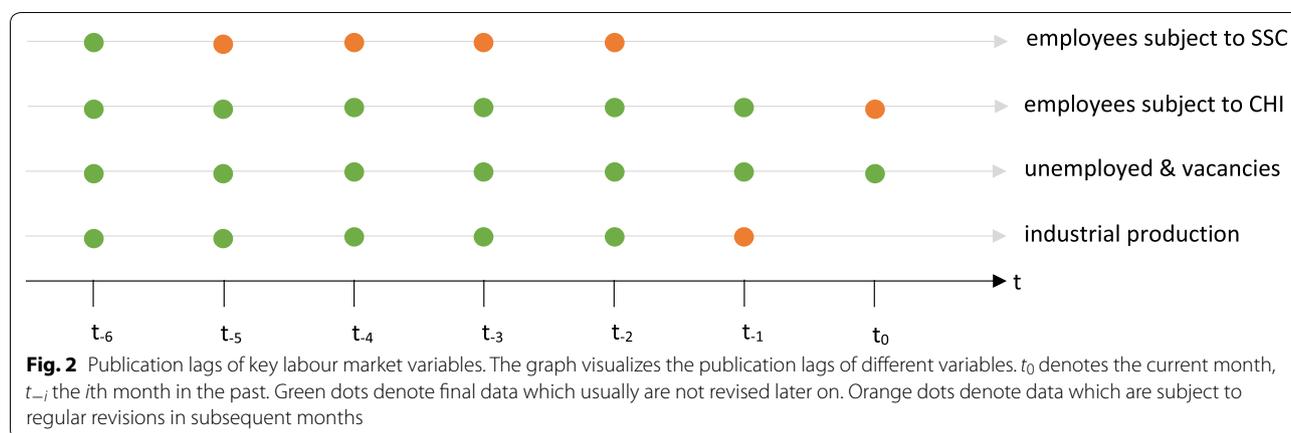
2.1 The target variable: employees subject to social security contributions

Employees subject to social security contributions (SSC) are by far the biggest group among all gainful workers in Germany. In 2018, 32.96 million people or 73% worked either full- or part-time with the duty to

pay SSC. Figure 1 shows the development of employees subject to SSC at a monthly frequency as published by the Federal Employment Agency (FEA). Their number has been rising at an above-average rate since 2005, only temporarily interrupted by the great recession, and their share among all gainful workers is now as high again as in the mid-1990s.

There is an ongoing discussion about the reasons for this strong development (e.g. Klinger and Weber (2020), Launov and Wälde (2016), Dustmann et al. (2014)). The main drivers are to be found among labour-market-specific candidate factors such as increasing matching efficiency, labour supply and job creation intensity as well as decreasing separation propensity (Hutter et al. 2019). However, beyond the question how strongly the labour market performs during good times, questions how fast and severely the employment figures are hit in times of economic turmoil attract particular attention. For assessing the development of the labour market, employees subject to SSC are among the most important variables: First, they cover the lion's share of all gainful workers, and second, they are published on a monthly basis. However, they suffer from a considerable publication lag, especially compared to other key variables of the German labour market such as unemployment or vacancies.

Figure 2 visualizes the availability and publication lags of a range of variables that are published on a monthly



basis. While the numbers of unemployed and vacancies are known with almost no delay¹ (i.e. in t_0), the number of employees subject to SSC suffers from a publication lag of 2 months. Furthermore, past values ranging from t_{-2} to t_{-5} are extrapolated values and hence regularly revised. One has to wait half a year (t_{-6}) until the published values can be considered as being final.

Table 1 shows the average extent of the monthly revisions required for past seasonally adjusted values of employees subject to SSC. They are calculated based on real-time data ranging from 2005m01 to 2018m12. As “true” values, I have taken the data of the 2019m06-vintage since this is the first month in which the original (i.e. not seasonally adjusted) 2018m12-values are final.² Logically, the values after a waiting period of 2 months are subject to the highest revisions in absolute value. They amount to an error of approximately 25,000 people on average. For the oldest extrapolated values (i.e. after a waiting period of 5 months), this error decreases to 17,000 people. The even older values that are not extrapolated any more are subject to the lowest revisions which stem from the usual learning effects inherent to seasonal adjustment.³ Consequently, after a waiting period of 6 or more months, the average errors for the seasonally adjusted time series are rather stable (approximately 11,000 to 12,000 people). Although the revisions are only 0.04% to 0.09% of the stock of employees, they are rather substantial when compared to the mean absolute monthly change, especially in case of the extrapolated values (up to 58%, see Table 1).

Table 1 Mean absolute monthly revisions for employees subject to SSC

Waiting period	Seasonally adjusted data			No s.a. MAR
	MAR	% (level)	% (dif)	
2 months (e)	25,120	0.09	57.84	24,229
3 months (e)	20,615	0.07	47.27	20,030
4 months (e)	19,332	0.07	44.34	19,433
5 months (e)	17,190	0.06	39.37	17,729
6 months (f)	11,616	0.04	26.30	0
7 months (f)	11,934	0.04	27.10	0
8 months (f)	11,759	0.04	26.78	0
9 months (f)	11,481	0.04	26.19	0
10 months (f)	11,255	0.04	25.56	0

The table shows the mean absolute revision (MAR) for employees subject to SSC. (e) and (f) denote extrapolated and final values, respectively. % (level): MAR in percent of the average level of the variable. % (dif): MAR in percent of the average absolute monthly difference of the variable. no s.a.: original (i.e. not seasonally-adjusted) data

This emphasizes that the FEA’s decision to wait 6 months before reporting final values is indeed justified.

To sum up, forecasters face the challenge that not only the future but also the near past of a key labour market variable is unknown or at least subject to substantial revisions. Of course, this challenge arises also with other variables such as GDP or productivity. Nonetheless, the fact that at any given month, the current employment figures are not known is rather unsatisfactory. In the following, I will show how to substantially improve nowcasts of these unknown employment figures.

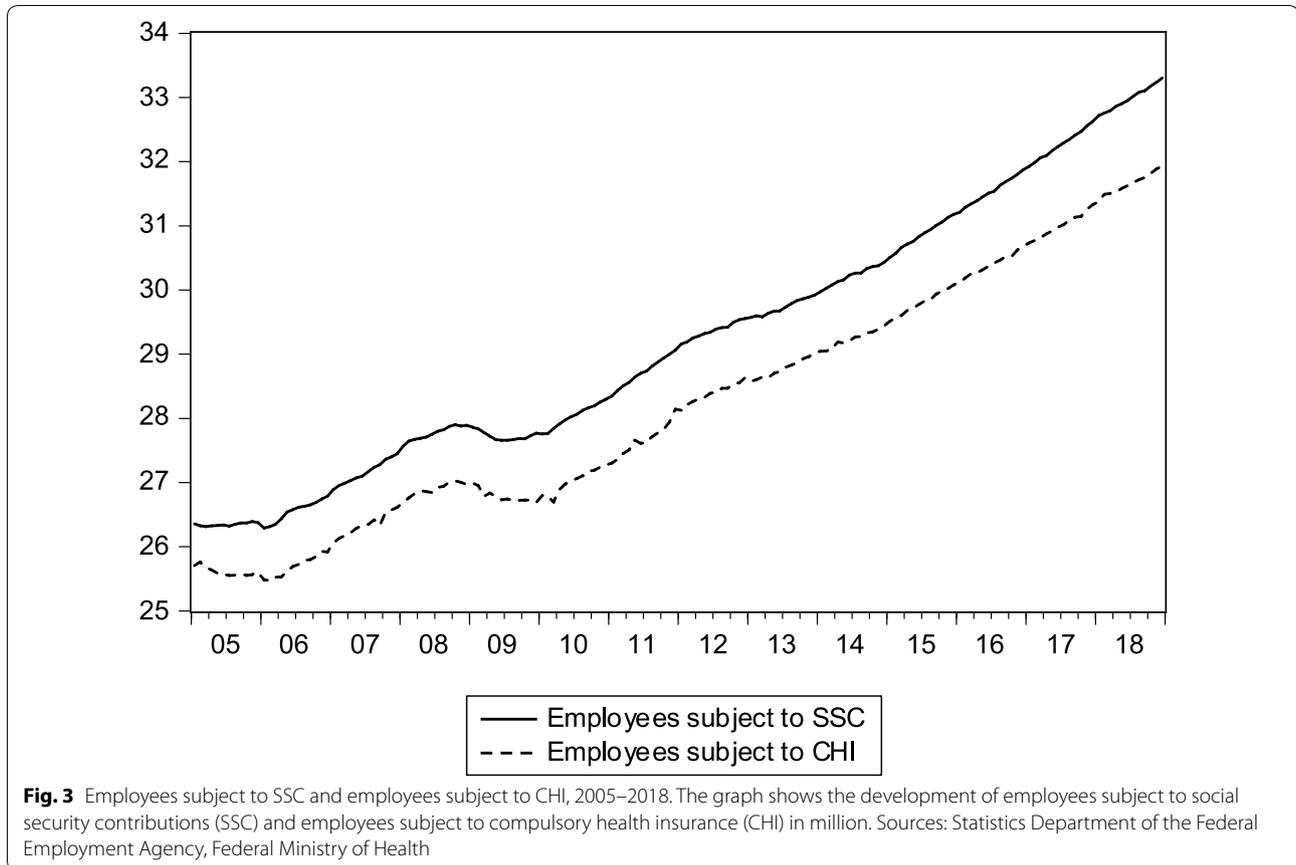
2.2 The auxiliary variable: employees subject to compulsory health insurance

The aim is to improve the knowledge about the development of employees subject to SSC of the (still unknown) near past and present. For this purpose, any auxiliary variable should fulfil several requirements. First, like the

¹ Usually, the data are published following a press conference at the end of the reporting month or the beginning of the subsequent month.

² There was a major revision in 2014m06. Consequently, before this major revision, the data of the 2014m05-vintage were considered as “true” values.

³ The original (not-seasonally-adjusted) value is not revised after a waiting period of 6 or more months, which explains why the MAR in Table 1 is 0 in these cases.



target variable it should be available on a monthly basis so that monthly updates could be performed. Second, it should be available without, or with only a minor, delay. Third, it should be subject to no or only to minor revisions. Fourth, it should economically be linked to the target variable in a sense that it is sufficiently plausible why both variables should comove.

Considering these requirements, there is a candidate variable that has been rather neglected so far: the number of employees subject to compulsory health insurance (CHI). They cover all employees subject to obligatory health insurance according to §25 SGB III (Sozialgesetzbuch III, Social Security Code III) as reported by the health insurance companies to the Federal Ministry of Health (FMH) in Germany. This variable is available at monthly frequency, with only minor delay.⁴ Furthermore, only the latest data point is subject to regular revisions in the subsequent month (compare Fig. 2), and I will show below that these revisions do not pose a major problem

since they have a pattern that can be forecasted very well. The most promising fact is that the vast majority of employees subject to CHI belongs—by definition—also to the group of employees subject to SSC.

Figure 3 compares the development of seasonally adjusted employees subject to CHI (dotted line) with that of employees subject to SSC (solid line). It visualizes the pronounced comovement between both variables and emphasizes their particular close relationship since employees subject to CHI make up, on average, for almost 97% of employees subject to SSC.

This leads to the question why, despite its obvious advantages, employees subject to CHI have not attracted much attention as auxiliary variable for nowcasting employment subject to SSC. The answer to this question might lie in high revisions of employees subject to CHI. Table 2 shows that between 2007m6⁵ and 2018m12, the mean absolute revision of the t_0 -values (which are the only ones that get revised, compare Fig. 2) amounts to 88,740 people or 73.25% of the mean absolute monthly

⁴ The figures are published within a few days after the reporting month at the following website: <https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte/>.

⁵ 2007m5 is the last month with missing data for the t_0 -value of employees subject to CHI.

Table 2 Mean absolute monthly revisions for employees subject to CHI

Time range		2007m6 to 2018m12		2014m1 to 2018m12	
4-step AP	s.a.	MAR	% (dif)	MAR	% (dif)
no	No	88,740	73.25	99,958	88.95
	Yes	82,212	157.81	89,511	182.33
yes	No			24,001	21.36
	Yes			24,796	49.99

The table shows the mean absolute revision (MAR) for employees subject to CHI, with and without applying the 4-step adjustment procedure (AP) that exploits the patterns in the revisions as described in the main text. Since this adjustment procedure requires an initial estimation period before it can be applied, revisions are only calculated for the period ranging from 2014m1 to 2018m12. s.a.: seasonal adjustment. % (dif): MAR in percent of the average monthly difference of the variable

change of this variable. Even when considering seasonally adjusted data⁶, the situation does not improve, quite the contrary: The mean absolute revision of the t_0 -values remains very high (82,212 people) and amounts to 157.81% of the respective mean absolute monthly change. In addition, this issue is not less relevant when only the subperiod 2014m1 to 2018m12 is considered. Here, one can expect that the latest seasonally adjusted data point will be revised by an amount that is almost twice as large as the average absolute monthly change of the variable itself. This seems to contradict one central requirement stated above and explains the hitherto reluctance to give employees subject to CHI more attention.

However, a deeper look into the nature of the revisions reveals that they are subject to a strong seasonal pattern. Figure 4 zooms into the development of the revisions – defined as the difference between final values and originally reported values—during the recent 7 years. It shows that the November-, December- and January-values usually experience a downward-revision while in the other months, the values are revised upwards. The fact that the seasonal pattern is rather pronounced and stable gives rise to believe that controlling for it substantially reduces the uncertainty about the current employment development. Beyond the seasonal pattern, there seems to be a linear trend in the revisions, turning revisions that are negative or neutral on average more and more into positive ones. Indeed, a regression of the revisions on four autoregressive lags⁷, a constant, seasonal dummies and a

⁶ Contrary to employees subject to SSC, official seasonally adjusted data do not exist for employees subject to CHI. Therefore, the data are seasonally adjusted in own calculations using the (multiplicative) Census X-12 ARIMA procedure.

⁷ Statistical tests reveal that no more than 4 autoregressive lags are necessary to eliminate serial correlation in the residuals.

trend reveals that the deterministic terms are significant even at the 1% significance level.

The crucial point is that to the extent revisions are stable and predictable the uncertainty coming with them can be reduced. For this purpose, I propose pursuing the following 4-step adjustment procedure:

- Step 1: Calculate a monthly time series of the known revisions of employees subject to CHI. The revisions y are defined as the difference between the final (CHI_t^{final}) and originally reported (CHI_t^{first}) values.

$$y_t := CHI_t^{final} - CHI_t^{first} \quad (1)$$

- Step 2: Use an AR(4)-model with constant, seasonal dummies and linear trend to control for the deterministic patterns in y .⁸

$$y_t = c + \sum_{i=1}^4 \alpha_i y_{t-i} + \sum_{s=2}^{12} \beta_s dummy_s^{seas} + \gamma t + \epsilon_t \quad (2)$$

- Step 3: With the help of Eq. (2), make a forecast of the revision in t_0 .
- Step 4: Calculate the expected value of CHI_t^{final} for t_0 as the sum of CHI_t^{first} and the forecasted revision from step 3.

This procedure yields the adjusted final value of employees subject to CHI. Again, the revisions after applying the 4-step adjustment procedure can be computed and compared to those before the adjustment. Logically, Eq. (2) requires an initial estimation period (until 2013m12 in this case). As a consequence, a proper comparison with real-time data is conducted for the following out-of-sample period (2014m1 to 2018m12). Table 2 shows that exploiting the deterministic patterns substantially reduces the mean absolute revisions, both for original and for seasonally adjusted values. In the latter case, the mean absolute revision amounts to less than 25,000 people. Hence, the average size of the revisions can be reduced by 72%. Furthermore, although the revisions relate to t_0 -values, i.e. after no waiting period, their average size is comparable to those of the employees subject to SSC after a waiting period of 2 months (compare Table 1). This is a promising feature and emphasizes the potential of employees subject to CHI as indicator for the target variable.

⁸ As estimation method, I use weighted least squares in order to put more weight on recent observations. As weighting variable, I use t^2 . Robustness checks reveal that seasonal autoregressive (SAR) models do not outperform the method outlined here.

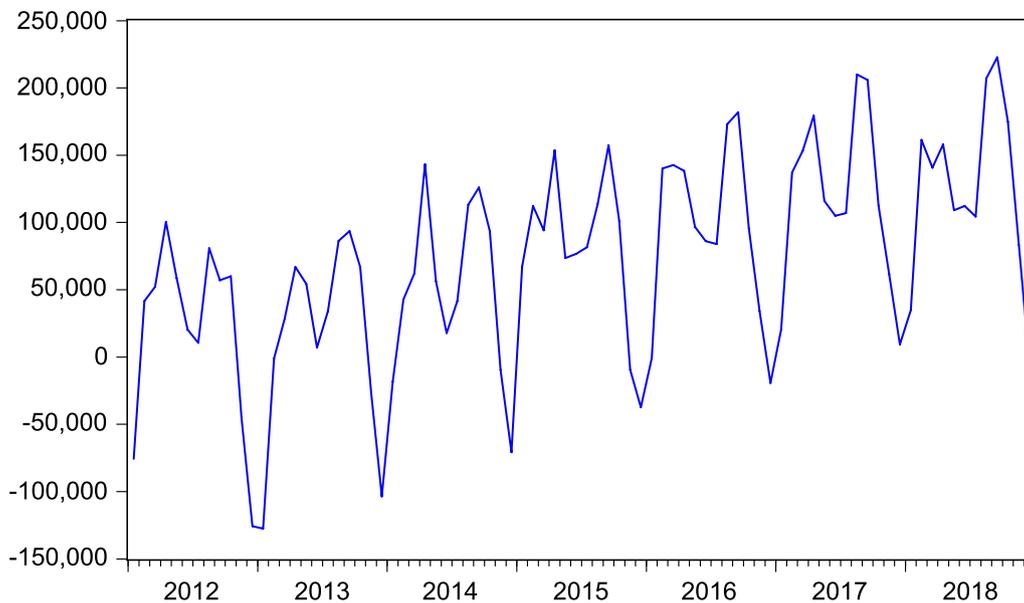


Fig. 4 Revisions of the t_0 -values of employees subject to CHI, 2012–2018. The graph shows the development of the revisions, measured as the difference between final values and originally reported values of employees subject to compulsory health insurance (CHI). Sources: Federal Ministry of Health, own calculations

3 Nowcast evaluation

3.1 Evaluation procedure

In the following, the two missing months of employees subject to SSC, i.e. the t_{-1} - and t_0 -values, are nowcasted with the help of current and past *CHI*-values available until t . The nowcasts are then compared to those stemming from a purely autoregressive benchmark model. Throughout this section, I use seasonally adjusted data.

In order to quantify the utility of employees subject to CHI for improving the knowledge about the current development of employees subject to SSC, I conduct a real-time nowcast exercise. This means that inference is based on statistical tests of out-of-sample performance. For this purpose, the observation period is split into an initial estimation period (2008m1–2013m12)⁹ and an evaluation period (2014m1–2018m12). In the former, the two competing estimation models are set on a solid base, while in the latter, nowcasts are conducted that can be compared to the realised true values of the target variable. Importantly, throughout the paper I only take information that would have been available at the time

the respective nowcasts are made. Since the variable *SSC* is subject to regular revisions, a candid setting requires the usage of real-time data. Like in Sect. 2.1, the vintage of 2019m06 is the source of the true values to which the nowcasts are compared.

The following paragraph discusses the choice of the underlying parsimonious benchmark model. Throughout this paper, I use weighted least squares as estimation method in order to put more weight on recent observations and allow the model to react more flexibly to potential structural changes.¹⁰ One could think of models relying solely on the own past such as AR(p)-models or random walk (RW). In their GDP growth application, Clark and West (2007) use an AR(1) with constant as benchmark model, Clark and McCracken (2015) use models with just a constant in order to predict stock returns. Sometimes AR models of higher order, determined by in-sample information criteria such as AIC or SC, are used. In any case, a candid evaluation should involve a thorough search for a good benchmark model, so that the bar for improvement is not set too low. The model with the best out-of sample performance was found in a systematic search as follows: First, all autoregressive models up to order 12 were investigated. In addition, due to the high persistence of *SSC*, also the respective models in differences (including a RW with

⁹ Just like in Sect. 2.2, also here I use data after 2007m6. The estimation of models with lagged variables requires a set of predetermined values, which is why the initial estimation period does not start until 2008m1. The split point was chosen ex-ante alongside full calendar years so that initial estimation period and evaluation period are approximately equally long. For a discussion on how to choose the split point, see Hansen and Timmermann (2012), for instance.

¹⁰ As weighting variable, I use t^2 , i.e. the squared trend.

drift) were included in the search. Once the best performing lag order was found, I investigated again systematically whether there is a subset model (i.e. a model that leaves out certain lags) that can improve upon the full specification. The resulting estimation specification of the best purely autoregressive benchmark model reads as follows:

$$\Delta SSC_{t-2} = c + \sum_{i=1}^6 \alpha_i \Delta SSC_{t-2-i} + \epsilon_{t-2} \quad i \neq 4 \quad (3)$$

It regresses the difference of employees subject to SSC on a constant and six autoregressive lags (without the fourth lag).¹¹ Note that the most recent value available in t is the one after 2 months waiting period. This model also substantially outperforms the so-called naïve nowcast, i.e. extrapolating the last known value of the target variable into the future.

For the alternative model, the auxiliary variable (*CHI*) is used on top of autoregressive terms. After applying the 4-step adjustment procedure described in Sect. 2.2, the most recent value of *CHI* available in t is the current one, i.e. t_0 . Again, the best specification was searched in the same systematic way as described above. I found that the following model performs best:

$$\begin{aligned} \Delta SSC_{t-2} = & c + \sum_{i=1}^6 \alpha_i \Delta SSC_{t-2-i} \\ & + \sum_{j=0}^9 \beta_j CHI_{t-j} + \epsilon_{t-2}, \quad i \neq 4, j \neq 6 \end{aligned} \quad (4)$$

For obtaining the set of nowcasts, I use a recursive scheme, i.e. the size of the estimation period grows with each iteration. For instance, the first nowcast of the t_{-1} -value of *SSC* is conducted for 2014m1 after having estimated Eq. (3) with *SSC*-data until 2013m12 (and *CHI* data until 2014m2). Then, the 2014m1-values for *SSC* (and 2014m3-values for *CHI*) are added to the estimation period and the second nowcast of the t_{-1} -value is conducted for 2014m2. This approach is repeated until the last nowcast of the t_{-1} -value is conducted for 2018m12.

The benchmark model in Eq. (3) is nested in the model of Eq. (4), i.e. the former can be obtained by setting all β -parameters of the latter to zero. This is of crucial importance in tests of equal predictive accuracy. Clark and West (2007) argue that the mean squared prediction

Table 3 Nowcast evaluation for employees subject to SSC

Horizon	t_{-1}	t_0
Benchmark model		
MAPE	30,101	32,109
MSPE (bn)	1.392	1.772
RMSPE	37,315	42,091
Enhanced model		
MAPE	25,329	27,365
MSPE (bn)	1.119	1.407
RMSPE	33,459	37,512
Model comparison		
Δ_{adj} (bn)	0.536	0.969
CW-statistic	3.05***	2.99***

The table compares the nowcasting performance of the benchmark model (Eq. (3)) and the enhanced model (Eq. (4)) for two horizons: t_{-1} and t_0 . MAPE: mean absolute prediction error. MSPE: mean squared prediction error. RMSPE: root mean squared prediction error. Δ_{adj} denotes the adjusted difference between the MSPEs of the two models. Unit of MSPE and Δ_{adj} : billion. CW-statistic is the value of the test statistic following Clark and West (2007). *** means the null hypothesis of equal predictive accuracy is rejected at the 1% significance level

error (MSPE) of the larger model is upward-biased due to additional noise stemming from the need to estimate the parameters which—under the null hypothesis of equal predictive performance—(1) are zero in population¹² and which (2) are correctly set to zero in the parsimonious model. In a sense, the smaller benchmark model is more efficient and hence benefits from not carrying the burden of estimating the parameters of redundant variables to zero. Consequently, usual tests in the style of Diebold and Mariano (1995) are undersized and have poor power in a nested model environment. Therefore, I implement the nested-model test described in Clark and West (2007) (CW test in the following), applying a one-sided test for equal predictive accuracy with the alternative hypothesis being worse nowcast performance of the nesting model.

3.2 Results

I estimate Eqs. (3) and (4), i.e. the benchmark and the enhanced model, as described in the previous subsection. Specification tests show that in both models, the chosen lag length of 6 months is able to eliminate serial correlation in the residuals. All Q -statistics from a test of no residual autocorrelation (Ljung and Box 1978) are insignificant with very high p -values.

The resulting mean absolute prediction errors (MAPEs) for both models and both nowcast horizons are shown in the first two blocks of Table 3. For nowcasts 1 month ahead (i.e. the t_{-1} -values), the MAPE can be reduced by

¹¹ Importantly, the AR(6) without lag 4 not only performs best out-of-sample during the evaluation period starting in 2014m1 but also is the preferred choice according to the in-sample Akaike information criterion with data ranging until 2013m12.

¹² For a discussion of the difference between a null hypothesis of equal accuracy in the *population* vs. *finite sample*, see e.g. Clark and McCracken (2015, 2013).

16% from 30,101 to 25,329 people. For the 2-months-ahead nowcasts (i.e. the t_0 -values), the MAPE-reduction amounts to 15%.

Despite this clear result in terms of the MAPE, there could be a risk that the inclusion of the new auxiliary variable sporadically leads to more extreme deviations than the purely autoregressive benchmark model. This would be particularly unpleasant at the current edge, with very high uncertainty about future employment development in times of economic crisis. Therefore, Table 3 also shows the mean squared prediction error (MSPE) which penalizes extreme deviations more severely and hence provides an alternative way of quantifying nowcast errors.¹³ For nowcasts 1 month ahead (t_{-1}), the MSPE can be reduced by 20% from 1.392 bn to 1.119 bn. For the 2-months-ahead nowcasts (t_0), the MSPE-reduction amounts to 21%. This emphasizes that the theoretical risk of more extreme deviations does not materialize, quite the contrary: The model enhanced by *CHI* is especially successful in reducing large nowcast errors.

The third block of Table 3 provides the statistical inference. As described above, the main difference between tests à la Diebold and Mariano (1995) and Clark and West (2007) is that in the latter, the difference between the MSPEs stemming from the benchmark model and the enhanced model is adjusted due to its downward bias in case of nested model environments. Hence, Table 3 reports Δ_{adj} , i.e. the adjusted difference. The resulting CW-test-statistics amount to 3.05 and 2.99, respectively, which means that the null hypothesis of equal predictive accuracy can be rejected even at the 1% significance level.

To sum up, the nowcast evaluation conducted in this subsection shows that the usage of *CHI* indeed significantly improves the assessment of current employment development. Importantly, the MAPE when nowcasting the t_0 -values can be reduced to a size that is even lower than the MAPE of the t_{-1} -values stemming from purely autoregressive benchmark models. Hence, additional months of information about employment development can be gained with the help of *CHI* while at the same time keeping additional uncertainty under control.

Since a crucial point of this paper is to provide more timely and precise employment information especially at the current edge, it must be ensured that the good overall performance of the enhanced model also holds towards the end of the sample. In order to check robustness of the results, I conduct the evaluation for several subperiods. To keep the focus on the current edge, I gradually remove observations from the beginning of the evaluation period.

Table 4 Nowcast evaluation for different subperiods

Subperiod	2015m1–2018m12		2016m1–2018m12	
	t_{-1}	t_0	t_{-1}	t_0
Benchmark model				
MAPE	30,025	32,398	24,886	28,037
MSPE (bn)	1.435	1.806	0.946	1.240
RMSPE	37,880	42,492	30,755	35,207
Enhanced model				
MAPE	25,106	28,633	19,236	23,695
MSPE (bn)	1.146	1.520	0.659	1.014
RMSPE	33,859	38,991	25,669	31,840
Model comparison				
$\Delta_{adj}(bn)$	0.536	0.887	0.562	0.931
CW-statistic	2.65***	2.81***	2.51***	2.49***

See Table 3

Table 4 shows the evaluation for the subperiods 2015m1 to 2018m12 and 2016m1 to 2018m12.

The results emphasize the reliability of *CHI* as auxiliary variable. In both subperiods, the prediction errors can be reduced by economically relevant amounts. While the MSPE-reductions are a bit smaller in the first subperiod, they amount to 30% (t_{-1}) and 18% (t_0) during the last three years of the sample. Noteworthy, despite the fact that fewer observations make it more difficult to reject the null hypothesis of equal predictive accuracy, the results still are statistically significant even at the 1% level. To sum up, yet again a potential risk does not materialize, quite the contrary: *CHI* proves to be a very well-performing auxiliary variable, especially so towards the current edge of the data.

3.3 Alternative indicator variables

The previous subsection showed the value of *CHI* for nowcasting *SSC* compared to the best purely autoregressive benchmark model. While a comparison to autoregressive benchmark models is good standard in the forecasting literature (see e.g. Lehmann and Weyh (2016) for a study on forecasting employment in a range of European countries), a comparison to other potential candidate variables sheds more light to the real value added of an indicator (see, for instance, Lehmann and Wohlrabe (2017) or Hutter and Weber (2015) for studies investigating employment or unemployment as target variable, respectively). In the following, I investigate potential alternative candidate variables. For being considered, they must fulfil certain requirements. First, they should be published on a monthly frequency. Second, real time vintages must be available in order to account for the real-time, revised nature of the data and ensure

¹³ Table 3 also shows—for informational purposes—the RMSPE as the square root of the MSPE.

Table 5 Nowcast evaluation: alternative leading indicators

Horizon	t_{-1}	t_0
Unemployment		
MAPE	29,963	33,008
MSPE (bn)	1.366	1.680
RMSPE	36,956	40,988
Vacancies		
MAPE	30,056	33,002
MSPE (bn)	1.319	1.624
RMSPE	36,313	40,304
Industrial production		
MAPE	29,650	31,687
MSPE (bn)	1.346	1.603
RMSPE	36,687	40,032

The table compares the nowcasting performance of alternative candidate variables

MAPE mean absolute prediction error, MSPE: mean squared prediction error, RMSPE root mean squared prediction error

a fair comparison throughout this paper (Clark and McCracken 2009). And third, the time series must be long enough to cover the estimation and evaluation periods described in Sect. 3.1. The alternative variables are the number of unemployed (U) and vacancies (V) and the index of industrial production (IP). Again, all variables are seasonally adjusted. While real time vintages of U and V stem from the statistics department of the FEA, the source of IP is the real time database of the German Central Bank.¹⁴ The chosen candidate variables have in common that they can be considered “hard” data that signal the *current* state of the economy or labour market (just as CHI does). This fits the purpose of this paper that focuses on nowcasting rather than forecasting.

For each alternative variable, a systematic search for the model with the best out-of-sample performance was conducted in the same way as described in Sect. 3.1. Importantly, just as it was the case with CHI , the most recent information available was included in the respective nowcasting equations. This means that unemployment and vacancies were allowed to enter with no delay, while for the index of industrial production the publication lag of 1 month was respected. The resulting best specifications comprise (in addition to the autoregressive part) U_t , V_t and V_{t-1} , and IP_{t-1} , respectively.

Table 5 shows that all variables have similar nowcasting performance as the purely autoregressive benchmark model. Importantly, no alternative candidate can outperform CHI . Compared to the best indicator (industrial

production), using CHI as auxiliary variable still reduces the $MAPE$ by 14 to 15% and the $MSPE$ by 12 to 17%, depending on the nowcast horizon.

Potential reasons for this result are that compared to CHI , industrial production suffers from delayed publication and only covers one sector of the economy. By contrast, unemployment and vacancies at first glance seem to be ideal candidate variables due to timely publication. However, the labour market development during the last fifteen years has shown that the pronounced employment upswing was driven to a substantial amount by sources beyond unemployment. For instance, a steady inflow from out of the labour force through migration or increasing participation of elderly and women has proven highly relevant. As a consequence, there were periods in which unemployment development did not mirror the strong employment upswing. Similarly, an increasingly tight labour market caused problems for firms to fill their vacancies (Klinger and Weber 2020), resulting in obvious problems for nowcasting employment with the help of unemployment and vacancies.

4 Conclusion

This paper closes the gap of current employment statistics in Germany, for which there is a publication lag of 2 months. It investigates in how far employees subject to compulsory health insurance (CHI), a variable that has been rather neglected so far, can serve as auxiliary variable to better nowcast the two missing months of employees subject to social security contributions (SSC).

A closer look at CHI emphasizes its ability to serve as indicator for the target variable. It is available on a monthly basis and with almost no delay. Furthermore, it covers 97% of SSC so that both variables are closely linked.

I document that the auxiliary variable is subject to high revisions. These revisions might be the main reason why CHI has not attracted more attention so far. However, I show that they are very stable and predictable and hence pose no major obstacle to a good nowcasting performance. In this context, the paper presents an easy-to-implement 4-step approach of how to deal with the CHI -revisions.

In an out-of-sample setting with real-time data, I conduct a nowcast evaluation exercise in a nested model environment following Clark and West (2007). Nowcasts of SSC based on the best purely autoregressive benchmark model are compared to the same model enhanced by current and lagged values of CHI . The results show that the model including the auxiliary variable significantly outperforms the benchmark model. Beyond statistical significance, the reduction of the mean squared prediction error is also substantial, with approximately

¹⁴ See https://www.bundesbank.de/dynamic/action/en/statistics/time-series-databases/time-series-databases/745590/real-time-data?statisticType=BBK_RTD&treeId=113205500

20% for both nowcast horizons. Furthermore, *CHI* also outperforms alternative candidate variables such as unemployment, vacancies and industrial production.

The results of this paper lay the foundations for a various set of beneficial applications. First, closing the gap of the publication lag of 2 months through nowcasting is useful in itself since it increases the information set available at any given time. Consequently, better knowing the current development of a variable should help improving genuine forecasts, too. Therefore, *CHI* could serve as auxiliary variable especially for short term forecasts of employment.

Furthermore, the results could be useful for other projections. For instance, the usage of labour represents a typical input for GDP nowcasts. Thus, with more timely and precise employment information, national accounts statistics could be improved. In addition, current representative survey results can benefit. E.g. in the German job vacancy survey (Bossler et al. 2019), SSC is used to project the (macro) vacancy data from the (micro) survey results, so that the precision of the publications can be increased. The same holds true for other leading indicators relying on employment information. As one example, the IAB labour market barometer (Hutter and Weber 2015; Hutter et al. 2016) uses current employment information for a weighting mechanism optimising the forecasting performance.

Acknowledgements

I am grateful to Dominik Ammon, Stefan Gründer, Sonja Podlech, Enzo Weber, and two anonymous referees for helpful suggestions and valuable input.

Authors' contributions

I am the sole contributor to the writing of this article. The author read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The real-time data used in this article stem from the Federal Ministry of Health (in case of the *CHI* data), from the Statistics Department of the Federal Employment Agency (in case of the *SSC* data), and from the German Central Bank (in case of industrial production). They can be downloaded from the respective websites. Alternatively, since they are aggregate data and not covered by data protection legislation, I can provide them upon request. The same holds for the evaluation codes necessary to replicate the results.

Competing interests

I declare that there are no competing interests, neither financial nor non-financial.

Received: 9 April 2020 Accepted: 19 June 2020

Published online: 30 June 2020

References

- Bossler, M., H. Gartner, A. Kubis, B. Küfner, Rothe, T.: The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, Waves 2000 to 2016 and subsequent quarters 2006 to 2017. FDZ-Datenreport, 03/2019 (en), (2019)
- Clark, T.E., McCracken, M.W.: Tests of Equal Predictive Ability with Real-Time Data. *J. Busi. Econ. Stat.* (2009)
- Clark, T.E., McCracken, M.W.: Advances in Forecast Evaluation. In Elliott, G., Granger, C., Timmermann, A. (eds.), *Handbook of Economic Forecasting*, 1st edition, Vol 2. Elsevier, (2013)
- Clark, T.E., McCracken, M.W.: Nested Fforecast model comparisons: a new approach to testing equal accuracy. *J. Econ.* **186**(2009–050A), 160–177 (2015)
- Clark, T.E., West, K.D.: Approximately normal tests for equal predictive accuracy in nested models. *J. Econ.* **138**, 291–311 (2007)
- Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. *J. Bus. Econ. Stat.* **13**(3/1995), 253–263 (1995)
- Dustmann, C., Fitzenberger, B., Schönberg, U., Spitz-Oener, A.: From sick man of Europe to economic superstar: Germany's resurgent economy. *J. Econ. Persp.* **28**, 167–188 (2014)
- Hansen, P.R., Timmermann, A.: Choice of Sample Split in Out-of-Sample Forecast Evaluation. *Economics Working Papers ECO2012/10*, European University Institute, (2012)
- Hutter, C., Weber, E.: Constructing a new leading indicator for unemployment from a survey among German employment agencies. *Appl. Econ.* **47**, 3540–3558 (2015)
- Hutter, C., Weber, E., Schmidt, K., Delfs, S.: Frühindikator für Arbeitslosigkeit und Beschäftigung: IAB-Arbeitsmarktbarometer 2.0. IAB-Kurzbericht, 09/2016, (2016)
- Hutter, C., Klinger, S., Trenkler, C., Weber, E.: Which Factors are behind Germany's Labour Market Upswing? IAB-Discussion Paper 20/2019, Nürnberg, (2019)
- Klinger, S., Weber, E.: GDP-employment decoupling in Germany. *Struct. Change Econ. Dynam.* **52**, 82–98 (2020)
- Launov, A., Wälde, K.: The employment effect of reforming a public employment agency. *Eur. Econ. Rev.* **84**, 140–164 (2016)
- Lehmann, R., Weyh, A.: Forecasting employment in Europe: are survey results helpful? *J. Bus. Cycle Res.* **12**, 81–117 (2016)
- Lehmann, R., Wohlrabe, K.: Experts, firms, consumers or even hard data? Forecasting employment in Germany. *Appl. Econ. Lett.* **24**, 279–283 (2017)
- Ljung, G.M., Box, G.E.P.: On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303 (1978)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)