Business surveys and repeated surveys: A simulation-based study

Keywords: repeated survey, business survey, time series model, Monte Carlo study.

1. Introduction

Most countries provide a business survey indicator that is repeated at regular time intervals in different sectors such as the manufacturing industry, construction, retail trade, services, and financial services. Business surveys make use of stratified sampling methods among companies so that big companies can appear regularly (see [1]). They are thus repeated surveys. The main question in repeated surveys is how to summarize the results, either using only the last survey or using some (weighted) average of the most recent surveys. Countries such as Belgium, France, and Turkey use seasonal adjustment procedures, like Census X-11, Census X-13ARIMA, or Tramo Seats, to provide the estimator. The European Commission (EC) centralizes the indicators for European countries and candidate countries using Dainties for the seasonal adjustment of the survey results. However, in the literature on repeated surveys, different approaches are suggested: a classical approach, a time series approach, and an approach based on signal extraction and the Kalman filter.

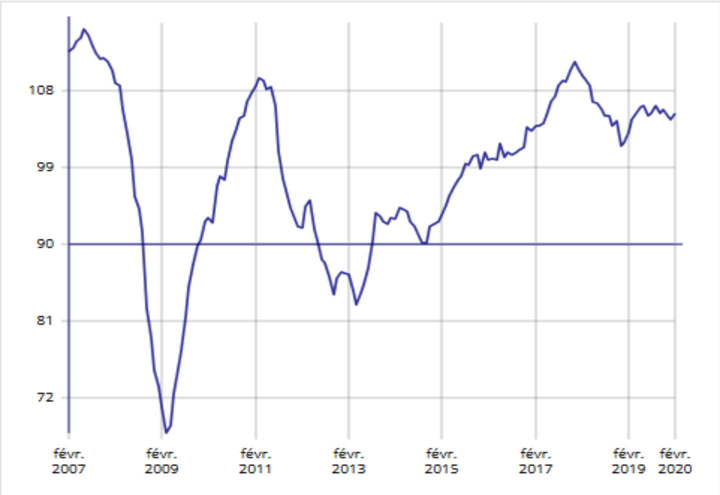

Figure 1. The business climate indicator - All sectors – Metropolitan France, February 25, 2020 (Source: https://www.insee.fr/fr/statistiques/3532408?sommaire=3530678)

In this study, we present the Monte Carlo simulation results of the seasonal adjustment approach and of the simplest time series approach suggested by the literature on repeated surveys. In the time series approach, we have retained the weighted average between the last value and a forecast. For this purpose, we consider the data of the business climate indicator for France for the period of January 2012 to February 2020.

2. Methods

2.1. The classical approach

Based on works by Jessen, Yates, Patterson, and others (see the survey paper [2]), the classical approach takes care of the overlapping of individual respondents in successive surveys. Moreover, it requires individual unit values to be available.

The classical approach does not exploit the time series aspects of repeated surveys. The papers [3] and [4] suggest that the value y_t obtained from the last survey about the parameter θ_t, the population value, is not necessarily the best one. Indeed, we can write $y_t = \theta_t + e_t$, where e_t is the survey sampling error. First, θ_t is not necessarily constant and, second, the e_t's are not necessarily independent, contrarily to what is assumed in the classical approach. From there, two approaches are possible, a time series approach and a signal extraction approach.

2.2. The time series approach

As pointed out by [5], past information obtained from the previous estimates is used to improve the current estimate. The improved survey estimator is a function defined on all samples taken between two constant time periods 1 and t, instead of one time period t. The time series approach is seen as an alternative to the classical approach.

One way is to model directly the data values, like [6]. One simple way is to estimate θ_t by some appropriate weighted average of the last observed values y_{t-k}, for $k = 0, 1, \ldots, K$, for some K and some weights. They propose the formula

$$\hat{\theta}_t = (1 - \pi) y_t + \pi \hat{y}_{t-1}(1), \qquad (1)$$

where π is based on the characteristics of the survey and $\hat{y}_{t-1}(1)$ is the forecast computed at time $t - 1$ for time t. Thus specifying and estimating an appropriate ARIMA model can then be used to obtain $\hat{y}_{t-1}(1)$. This is somewhat related to what is still used nowadays in some business surveys.

2.3. The approach based on signal extraction and the Kalman filter

Although suggested by several of the already cited authors, the paper [7] has explored the relationship between the θ_t and the e_t. The idea is to extract an estimator of θ_t from the y_{t-k}. This can be done by using a state-space model and the Kalman filter for calculating estimates of the state vector. See the survey by [8].

3. RESULTS

We have noticed in Section 3 that the techniques proposed in the literature of repeated surveys are not used in the business surveys conducted in the European Union (EU) and candidate countries. It should be interesting to compare the techniques they use with those proposed by the repeated survey literature. Following the paper of [5], who have produced a simulation study which suggests the improvement of (1) with respect to the last estimate, we have retained the two simplest techniques, the last value and the weighted average (WA) between the last value and a forecast, see (1).

Since the details on the survey methodology are vague, we have no way to estimate π. We have therefore chosen arbitrarily three values π = 0.2, 0.5, and 0.8. For the business survey techniques, instead of Dainties, Census X-11, and Census X-13ARIMA, for practical reasons, we have chosen Tramo-Seats, like in Turkey. Because a file is available, we started from the French data from January 2012 to February 2020 and denoted F_t, t = 1, ..., 98. The idea is to simulate N series (y_t, t = 1, ..., 98) so that $y_t = F_t + P_t + D_t$, where P_t is produced by a periodical seasonal component with period 12 months, basically a cosine function with amplitude A, and the random deviates D_t are generated by an autoregressive process with autocorrelation coefficient R on the basis of pseudo-random errors E_t that are normal with mean zero and standard deviation S. More precisely,

$$P_t = A\cos(2\pi(t-5)/12)\,, t = 1, \dots, 98,\ D_t = RD_{t-1} + E_t,\ t = 2, \dots, 98,$$

where $D_1 = E_1/(1-R^2)$ (so that the random deviates follow a stationary process) and the seasonal component is equal to 0 in February to avoid bias. For A, S, and R, we have used, respectively, A = 5, 10, 15, S = 5, 10, 15, and R = 0.2, 0.5, 0.8. We have generated N = 1000 series for each of the 27 cases, using the same stream of pseudo-random numbers to reduce variability. Simulation results are obtained by RJDemetra package in R and we have checked on a small number of series that RJDemetra gives the same results as JDemetra+.

Table 1. Results of rescaled MSE for the simulations for 1000 replications for different values of A, S, and R = 0.5. In each row, the smallest rescaled MSE is shown in italics.

A/S	**Forecast**	**RS π = 0.2**	**RS π = 0.5**	**RS π = 0.8**
5				
5	119.08	90.49	*81.28*	95.99
10	60.28	80.10	57.87	*53.40*
15	46.85	76.38	51.53	*43.21*
10				
5	109.39	96.37	*82.30*	90.97
10	87.01	89.85	*70.62*	73.18
15	52.91	81.20	55.64	*48.02*
15				
5	105.17	89.99	*76.10*	85.71
10	91.84	96.91	*77.40*	79.02
15	74.80	88.79	65.75	*64.06*

The results are expressed as a mean squared error (MSE) across the simulations for the value in February 2020 (105.4) for each A/S/R combination, and each of the five estimators: the forecast $\hat{y}_{t-1}(1)$, the elementary repeated survey (RS) weighted averages obtained by (1) for π = 0.2, 0.5, and 0.8, and finally the seasonally adjusted value, like used in many countries. Since we are interested mainly in comparison with respect to the latter, we have computed 100 times the ratio of the real value of the A/S/R simulation results to the seasonal adjustment data MSE values. Only the results for R = 0.5 are shown here due to a lack of space. Table 1 shows these rescaled MSE so that numbers below 100 show an improved estimator with respect to seasonal adjustment value estimator.

In Table 1, it is seen that the results for π = 0.5 give the smallest MSE values for 5 out of the 9 combinations of A/S/R. However, the minimum MSE values are obtained for π = 0.8 in the remaining cases. As a consequence, in our simulations, the simple RS estimators (1) for π = 0.5 or π = 0.8 is always better than the estimator provided by the seasonally adjusted value as it is used by EC and most of the countries being considered. In the results, the estimator provided by the forecast (which corresponds to an RS weighted average for π = 1) is not systematically better than the seasonally adjusted value estimator.

4. Conclusions

The simulation results obtained using Tramo-Seats with the RJDemetra package for R indicate that, for each of the 27 A/S/R combinations considered, the weighted averages for π = 0.5 and 0.8 beat, using MSE for February 2020, the Tramo-Seats seasonally adjusted value estimator. As a concluding remark, we suggest to consider RS methods as estimators for business surveys, not necessarily the simplest methods used here, but well more advanced methods. Since the simplest methods provide better results than the existing estimators, it is reasonable to suppose that methods using more information will improve these results. The reason why we did not apply the approach based on signal extraction and the Kalman filter is that the method will depend on the knowledge of how the units enter into the survey. That confidential information is only known by the organizers of the survey. This is, however, not a limitation of our study, since the simplest repeated survey methods beat the methods being used. More details on business surveys (see [1]), and on repeated surveys, and more results will be given in the full paper as well as an R script. We did not consider consumer surveys in this paper but it is possible that similar conclusions can be obtained.

References

[1] European Commission (2020), The Joint Harmonised EU Programme of Business and Consumer Surveys User Guide, https://ec.europa.eu/info/files/user-guide-joint-harmonised-eu-programme-business-and-consumer-surveys_en.

[2] D. A. Binder and M. A. Hidiroglou, Sampling in time, Handbook of Statistics 6, P. R. Krishnaiah and C. R. Rao, eds., Elsevier Science Publishers B.V. (1988), 187-211.

[3] B.J.N. Blight and A.J Scott, A stochastic model for repeated surveys**,** Journal of the Royal Statistical Society Series B 35 (1973), 61–68.

[4] A. J. Scott and T. M. F. Smith, Analysis of repeated surveys using time series models, Journal of the American Statistical Association 69 (1974), 674-678.

[5] M. A. Ismail, H. A Auda, and Y. A Elzafrany, On time series analysis for repeated surveys, Journal of Statistical Theory and Applications 17 (2018), 387-396.

[6] A. J. Scott, T. M. F. Smith, and R. G. Jones, The applications of time series methods to the analysis of repeated surveys, International Statistical Review 45 (1977), 13-28.

[7] W. Bell and S. Hillmer, Time series methods for survey estimation, Bureau of Census, Statistical Research Division Report Series (1987) CENSUS/SRD/RR-87/20.

[8] M. Feder, Time series analysis of repeated surveys: the state-space approach, Statistica Neerlandica 55 (2001), 182-199.