

# Nowcasting Austrian Short Term Statistics

Markus Fröhlich<sup>1</sup>

Early estimates for Austrian short term indices were produced using multivariate time-series models. The article presents a simulation study with different models (vector error correction models, vector autoregressive models in levels – both with unadjusted and seasonally adjusted time-series) used for estimating total turnover, production, etc. In a preliminary step, before time-series were provided for nowcasting, the data had to undergo an editing process. In this case a time-series approach was selected for data-editing as well, because of the very specific structure of Austrian enterprises. For this task basically the seasonal adjustment program X13Arima-Seats was used for identifying and replacing outlying observations, imputation of missing values and generating univariate forecasts for every single time series.

*Key words:* Nowcast; cointegration; outlier adjustment; vector error correction models; (seasonal) unit-roots.

## 1. Introduction

Official statistics are used by a variety of public and private organizations to assess and monitor economic and social phenomena. The demand for timely statistics continues to increase as national and international decision-makers (governments, central banks, etc.) must rapidly adjust their policies in accordance with the most recent indicators to meet mandated requirements. However, there are often considerable time lags between releases of these official statistics. Attempts to produce more timely statistics include efforts to speed up the production process by reducing the transmission deadlines or by redesigning the sample (Statistics Finland 2004a), extended and increased imputation of missing data (Fröhlich et al. 2010), or employing econometric models (Ladiray and O'Brian 2003; Statistics Finland 2004b, or Bodo et al. 1991).

The European Union's monthly short-term business statistics (STS) represents one of the most important sources for the assessment of the European economy. Deadlines for data transmission for this statistical program have been fixed to 45 days (for the production and turnover variables) and to 60 days (for the number of employees and hours worked variables) after the end of the reference period, in accordance with the EU-Regulations European Commission (1998) and European Commission (2005). Small Member States are allowed to extend these deadlines by 15 days. Statistics Austria releases its preliminary

<sup>1</sup> Statistics Austria, Guglgasse 13, 1140 Vienna, Austria. Email: markus.froehlich@statistik.gv.at.

**Acknowledgments:** This article was prepared for the ICES-Meeting in Geneva in June 2016. The author wishes to thank Alexander Kowarik and Angelika Meraner for presenting the topic at two conferences and to the associate editor and three anonymous referees for their intensive engagement and their very thoughtful comments and suggestions which fundamentally improved the quality of this article.

STS data 55 days after the end of the reference month and its final STS data 85 days after the end of the reference month. It is desirable to reduce the lag time between collection and release for these measures in the near future. Accordingly, we investigated methods of producing early estimates for the Austrian STS that perform well both in times of regular (stable) economic development and in times of regime shifts. Specifically, our goal was to investigate methods of generating reliable preliminary STS estimates 30 days after the end of the reference period ( $t + 30$ ), effectively reducing the data transmission by 25 days.

To accomplish this, we needed to find solutions to two main issues: (i) “cleaning” (unprocessed) raw data at the micro level (enterprises or establishments) and (ii) imputing missing data on an aggregated NACE (Nomenclature statistique des Activités économiques dans la Communauté Européenne – Statistical Classification of Economic Activities in the European Community) level (total industry and construction – Sections C to F of NACE 2008). All calculations and analyzes described in this article were performed using well documented packages in R (R Core Team 2010).

To address the first issue, we concentrated on eliminating outlying observations and imputing missing values for past periods with univariate time series models. The purpose of this task was to create complete aggregate series from the reduced data. The data cleaning step was necessary because the production data editing procedures were not available at an earlier point in the data transmission cycle ( $t + 30$ ) due to limited human resources and limited access to administrative data used for validation in production. In our approach, every business unit was treated as an individual time series, and the imputation procedure accounts for series-level outliers and the trading day composition of each month. The outlier identification procedure focused on the microdata; possible outliers at the macro level were not considered. We identified different R packages for this task, with software choice depending on the characteristics of the individual time series; see Section 2 for a complete discussion. After developing the unit-level models, a univariate one-step ahead forecast for each variable was generated for each enterprise and variable.

To address the second issue (task (ii)), we employed a multivariate time-series approach for predicting the following key-economic indicators: production, turnover, number of hours worked and number of employees. We refer to this as “nowcasting”, employing the Eurostat definition (Mazzi and Cannata 2016, 27) that is,

“A very early estimate produced for an economic variable of interest over the most recent reference period calculated on the basis of incomplete data using a statistical or econometric model different from the one used for regular estimates. Soft data should not play a predominant role in nowcasting models. Nowcasts may be produced during the very same reference period for which the data is produced.”

To increase the robustness of the forecast models, we sought alternative source data (auxiliary information) that could be incorporated into the forecast models to accurately predict the desired estimates. Unfortunately, we had no access to “soft” data, such as business expectation surveys performed by research institutes or other potentially useful auxiliary data. Instead, we used information collected from the early responding units to predict the missing data values (see Section 3). Subsection 3.1 presents the simulation designed to demonstrate the performance of different considered models.

Since 2013, the presented nowcasting methods have been adopted to produce early estimates (30 days after the end of the reference period) for selected STS variables at a highly aggregated level. However, the data cleaning approach described in Section 2 had not yet been implemented for these published statistics. Consequently, we use the heuristic procedure on the entire data set. Depending on the findings of this exercise, we want to extend the procedure to a more detailed NACE level and to other variables. For this to be implemented in production, the entire procedure – data-cleaning and imputation plus estimation of missing data with the multivariate model – should run to completion within one day.

## 2. Data Cleaning and Imputation: Treatment of Univariate Time Series

The Austrian STS sample is composed of 10,000 to 12,000 enterprises and establishments that report about 40 variables each on a monthly basis. In the production system, data editing and imputation (basically automatic editing, selective editing and interactive editing) is performed by subject matter experts who access numerous administrative sources, like tax reports and social security accounts. As mentioned in Section 1, these production system procedures and data were not available 30 days after the end of the reference month. Consequently, we developed separate data-cleaning procedures to replace dominating outliers in the individual series (in particular for the current period) and impute missing values for past periods.

The data cleaning algorithms described in this section yield outlier-adjusted and gap-filled time series for every single business unit in the STS population using a variety of R-packages. Our primary choice was the `x12`-package (Kowarik et al. 2014), which offers automatic time-series modelling, as well as the incorporation of trading day effects and control of the sensitivity of outlier identification. The latter feature was especially important as we focus on the correction of only the most obvious outliers to avoid “over-adjusting”. However, the required minimum length of the series limited the usage of `x12` for selected univariate time series in our data; the length of the series has to be at least 36 observations for monthly data and 20 observations for quarterly data, with at least one observation available for every month/quarter (see U.S. Census Bureau 2017). For series of length 25 to 36, we use the `tsclean` function in the `forecast`-package (Hyndman and Khandakar 2008). This function provides gap-filling functionality as well as outlier identification, but does not offer control of the sensitivity of the outlier identification (a serious drawback for our purposes). The function produces good results for most series. However, specific variables in our data set exhibit very specific patterns which are hardly processable with standard software tools. In the course of our research, we found that the procedure fails to identify outlying observations in the case of missing values preceding outliers due to the interpolation step, that precedes outlier identification. These effects are most pronounced in short time series (less than two years) and for series with little or no variability (e.g., number of employees).

Consequently, we developed a heuristic procedure to process the remaining series (of length 24 or shorter). In the latter case, outlier identification was restricted to the last two observations of each individual time series that is, the observations for the reference month and the month preceding the reference month. Treated observations for the remainder of

the series were already available from the edited production data. In all cases, identified outliers were eliminated from the data set and replacement values and forecasts were produced with the nonseasonal Holt-Winters algorithm described later in this section. The output of the data-cleaning step was an outlier-adjusted series with imputed missing historical values along with an additional univariate one-step ahead forecast.

For our approach, we constructed separate time series for every single business unit (enterprise/establishment) for every variable, resulting in 10,000 to 12,000 unique time series per variable. This data set was separated into three groups: (i) units with more than three years of observations, (ii) newly established units or units with less than 25 observations and (iii) any other units.

We used the R-package `x12`, for the (i) group of time series with more than 35 observations. `x12` is a wrapper function of X13-ARIMA-SEATS, provided by the U.S. Census Bureau (Findley et al. 1998). X13-ARIMA-SEATS is state-of-the art seasonal adjustment software, used in national statistical offices worldwide. The X13 seasonal adjustment procedure decomposes each time series into three separate components (a trend component, a seasonal and a residual component) and applies filters to the first two components to obtain a seasonally adjusted series. In addition, the program offers model selection capabilities, as well as numerous quality diagnostics. The R package `x12` offers the full functionality of the X13 seasonal adjustment and pre-adjustment procedures; it does not offer SEATS seasonal adjustment. This limitation does not affect us, since we only employ the X13 pre-adjustment procedures, specifically the regARIMA modelling, which is used for outlier identification and replacement (treatment), trading day effect adjustment, gap-filling, and forecasting.

A regARIMA model is defined by

$$y_t = \sum_i \beta_i x_{it} + z_t \quad (1)$$

where  $y_t$  is the time series to be modeled,  $x_{it}$  are regression effects, such as outliers, trading days, and so on,  $\beta_i$  are the regression coefficients and  $z_t$  are the regression errors that are assumed to follow a (seasonal) ARIMA model. According to the U.S. Census Bureau (2017), a general multiplicative seasonal ARIMA model (SARIMA) for a time series  $z_t$  can be written as:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \theta(B)\Theta(B^s)a_t \quad (2)$$

where  $B$  is the backshift operator ( $Bz_t = z_{t-1}$ ),  $s$  is the seasonal period,  $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$  is the nonseasonal autoregressive (AR) operator of order  $p$ ,  $\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})$  is the seasonal AR operator of order  $P$ ,  $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$  is the nonseasonal moving average (MA) operator of order  $q$ ,  $\Theta(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})$  is the seasonal MA(Q) operator and the  $a_t$  are IID with mean zero and variance  $\sigma^2$  (white noise).  $(1-B)^d$  and  $(1-B^s)^D$  imply nonseasonal and seasonal differencing of order  $d$  and  $D$  respectively. For more information on ARIMA and seasonal ARIMA models, see Box et al. (2008).

In certain series, a trading day adjustment is necessary to account for the lack of equidistance of observation periods. The problem is most evident for flow series, that is, series where each observation is the sum of a variable over a specific period. The trading

day effect can be divided into a length-of-month effect (a seasonal effect) and an effect which is induced by the different number of working (trading) days in the periods under comparison. For example, the amount of sales in a specific month may not be directly comparable to the value obtained from the same month a year ago because each month has a different number of working days. In our case, the trading day regressor is not considered important for outlier identification since we concentrated on identification and treatment of very extreme values, which cannot be accounted for by a trading day effect. However, the trading day regressor was included in the univariate forecast models.

The automatic outlier identification procedure in  $\times 12$  is a stepwise regression approach based on the work of [Chang et al. \(1988\)](#). In the default procedure, regressors are fit for each point of the series, and their corresponding t-statistics are compared against specified critical values. The program offers tests for additive outliers, level shifts, temporary changes, ramps, and seasonal outliers. Outliers can be identified automatically, but users can fix specific outliers by themselves.

We used the automatic outlier identification of additive outliers. Level shifts, temporary changes, and ramp outliers identify and “correct” level shifts in the series. However, except for the last two months in the series, the unit level data had been validated by subject matter experts. In short, the identification of outliers for this data set needed particular attention, because validated original data values should not be modified. Additionally, additive outliers in the very end of the series could provide evidence of a regime-shift and therefore should be retained without treatment in the regARIMA model. Moreover, “unusual” values for variables like turnover and production are easily conceivable. Examples include the construction sector, where holdings are created to cooperate in specific projects or expensive goods, like aircraft, which are demanded erratically. We determined the critical value for outlier identification heuristically by increasing the t-value of the regressand depending on the variable. Detected outliers were substituted by the replacement values, which were calculated within the pre-adjustment process of  $\times 12$ . The regARIMA model, estimated according to Equation 1, was used to generate (univariate) forecasts of the series. Ultimately, about 80% of all series were treated with  $\times 12$ .

For the second group (ii) of time series (newly established units and units with less than 25 non-missing observations), a heuristic approach was selected to identify outlying observations. If the series consisted of one or two observations, we determined a maximum value for each variable that subject matter experts considered too high for a newly established enterprise. In these cases, we had to investigate whether the enterprise had just changed its organisational level (and therefore changed its identifier) or if the value had to be classified as an outlier. When two observations were available, we considered the difference, assigning the larger of the values as an outlier if the difference exceeded the same maximum value.

In case of three to 25 observations, outliers were identified according to the following rules:

- $y_n > m_{0,5} + 4 * r_{1..(n-1)}$  for  $m_{0,5} > 0$  and  $r_{1..(n-1)}/m_{0,5} > 0.1$ ,
- $y_n > m_{0,5} + 50 * r_{1..(n-1)}$  for  $m_{0,5} > 0$  and  $r_{1..(n-1)}/m_{0,5} < 0.1$  and
- $\frac{y_n}{m_{0,5}} < 0.05$  for  $m_{0,5} > \min$ ,

where  $m_{0.5}$  is the median,  $r_{1..(n-1)}$  the range of variable  $y$  for observations 1 to  $(n-1)$  and  $y_n$  is the current observation for variable  $y$ ,  $min$  was a predetermined minimal value per variable. If the median was zero, the arithmetic mean was substituted instead. The first two rules were used for positive outliers. The second rule was intended for series with very low variability, such as the number of employees, which could legitimately change considerably in relative values for small establishments (for example, two versus three employees). Finally, the third rule is intended for detecting negative outliers, specifically zeroes which could indicate item nonresponse in our specific data set. All outlying observations identified here were eliminated from the data set (i.e., the enterprise was regarded as a nonreporting unit instead of an early reporting unit).

The heuristic approach in this step was combined with visual inspection of the series with the highest absolute values for each variable. The results of this heuristic procedure were considered satisfactory, in the sense that extreme values that have not been identified in this class did not influence the aggregate of total industry or construction.

Forecasts for the time series in this group (less than 25 observations) were generated via the nonseasonal Holt-Winters procedure. The (linear) prediction equation can be written as (Stier 2001)

$$\hat{x}_t(m) = L_t + mT_t$$

where  $m$  is the prediction horizon,  $T_t$  is the trend component for observation  $t$  and  $L_t$  is the level of the series for observation  $t$ . These components are calculated recursively according to:

$$L_t = \alpha x_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)(T_{t-1})$$

A starting value has to be determined for each component. The parameters  $\alpha$  and  $\beta$  can be estimated by minimizing the ex-post forecasting error for various combinations of the parameters between zero and one (for more details, see Stier 2001).

Finally, the third group of time series – short and “difficult” series (i.e., series with more than 24 observations plus series with more than 35 observations which could not be processed in x12) were adjusted with the function `tsclean`. This procedure uses Friedman’s super smoother (Friedman 1984b and Friedman 1984a) to calculate a smoothed version of the time series. Subsequently, the deviations of the original series to the smoothed series (called “residuals”) are calculated. Outliers are determined as those residuals, which are beyond the first and third quartile plus/minus three times the interquartile-range of the distribution of the “residuals”. For seasonal series with more than two cycles of observations, `tsclean` uses a periodic seasonal and trend decomposition using loess (STL); see Cleveland et al. (1990). After the decomposition, the same outlier identification is used as for nonseasonal series. As already mentioned above, the sensitivity of the outlier identification cannot be selected as an argument in the function. However, the sensitivity can be changed by testing different adaptations of the `tsclean` function directly. After outlier detection and treatment, missing values were interpolated and one-step ahead forecasts were produced.

Figures 1 and 2 illustrate the performance of the data-cleaning process for total turnover. The first figure shows how much the untreated first reported values for total turnover differ from final values. The comparison was made for the period of January through December 2016 for every NACE section separately. The horizontal bars (in grey) represent the range of maximum and minimum absolute deviations, the black dots represent the deviation of each individual month. Beside each bar, we have plotted the relative deviations in percent. Because the size of the different NACE sections differs considerably, relative deviations in small NACE sections are negligible in absolute value, whereas small relative deviations in large NACE sections could exhibit considerable absolute values; conversely, the small expected deviations for small units could appear as large relative deviations. For example, the relative deviation in NACE section C19 of  $-100\%$  corresponds to a single unit that reported zero turnover. This value was not identified as an outlier, but its value was different from zero, although moderate in size. The presentation is limited to the early reporting units for the specific month in every NACE section so that the percentages refer to all early reporting units in the respective NACE section. Figure 2 demonstrates the effectiveness of the outlier adjustment procedure in reducing these absolute and relative deviations. That said, despite concentrating on the most obvious outliers, the procedure seems to slightly over-adjust the data in several NACE sections.

### 3. Nowcasting – Multivariate Time Series

Our next step was to estimate the target variables for enterprises that had not transmitted their questionnaires in time. By 30 days after the end of the reference month, about 60% of enterprises have responded.

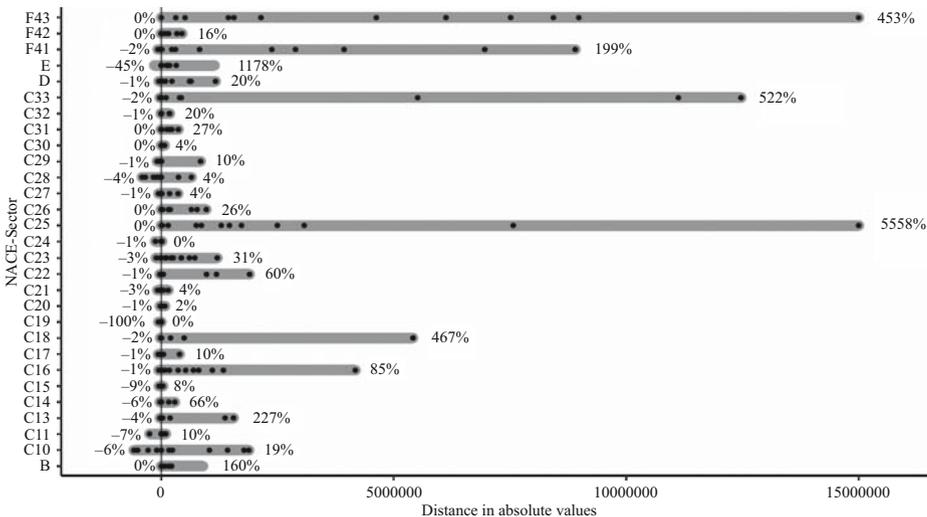


Fig. 1. Differences between the preliminary and final estimates of total turnover ( $t + 30$  versus  $t + 85$ ). The difference is plotted in absolute distances (x-axis) and relative deviations (percentages beside each bar) for each NACE section for January 2016 through December 2016. Each horizontal bar represents the range between minimum and maximum deviation in the respective time span. The x-values (absolute deviations) for NACE sections F43 and C25 have been trimmed, because the maximum values were too large to display in the graph.

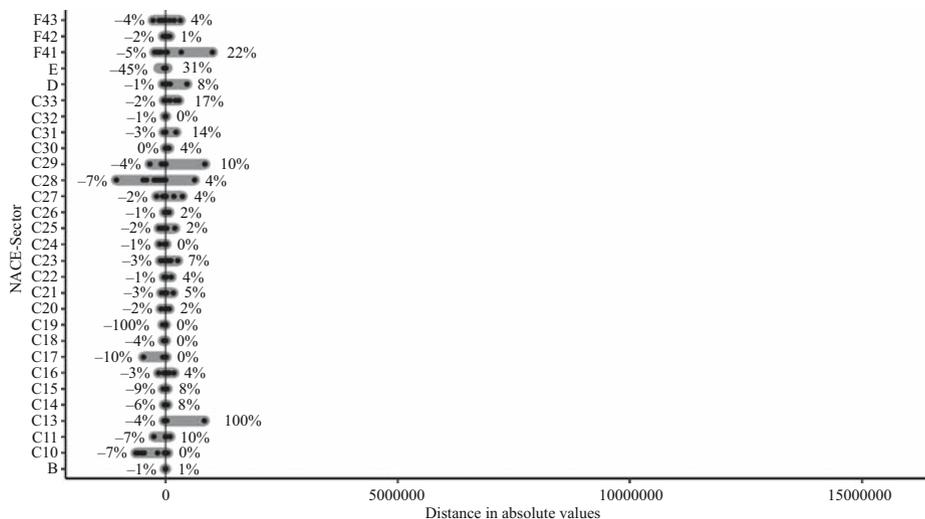


Fig. 2. Differences between the outlier-adjusted and final estimates of total turnover ( $t + 30$  versus  $t + 85$ ). The same scale was chosen as in Figure 1 to facilitate comparisons.

Recall that univariate forecasts for each time series had been generated as a byproduct of the described data-cleaning process in Section 2. As a result, the short-term indicators could technically be produced at the same level of detail as the indicators that are published 55 days after the end of the reference month. However, the quality of these estimates has not yet been analyzed systematically. Due to restrictions on computational time, this procedure was not included in the simulation study described below. In the future, we will compare the results of the estimates constructed with the one-step ahead forecasts to those obtained from the multivariate approach described below. Regardless, these naïve nowcasts will not be satisfactory when regime shifts occur; a regime shift cannot be anticipated at the end of a series because univariate models only consider historical data that are projected into the future.

In most cases, we could construct aggregate time series estimates for every desired NACE level. Unfortunately, several changes in the economic activity classification NACE system have occurred since 1996, with a major change in 2008. For this reason, there are several divisions of NACE that are not comparable over time. In some cases, this problem can be solved by reclassifying data from 1996 to 2007 to NACE 2008. For the 1:1 relationships, the reclassification process is trivial. Unfortunately, there are many  $n:1$  and  $1:n$  reclassifications and consequently exact reclassification is not always possible.

Instead, we applied an alternative approach. At  $t + 30$  days the population for index construction for reference period  $t$  was split into groups of *early respondents* and *late respondents*, that is, enterprises/establishments that had transmitted completed questionnaires for period  $t$  and units which are obliged to report for period  $t$  but that did not yet transmit their response (see Figure 3). We aggregate the values of both groups separately for the entire time span (1996 to 2016) to produce two distinct sets of time series for each variable, with the second time series comprising  $(t - 1)$  observations with observation  $t$  to be “predicted”.

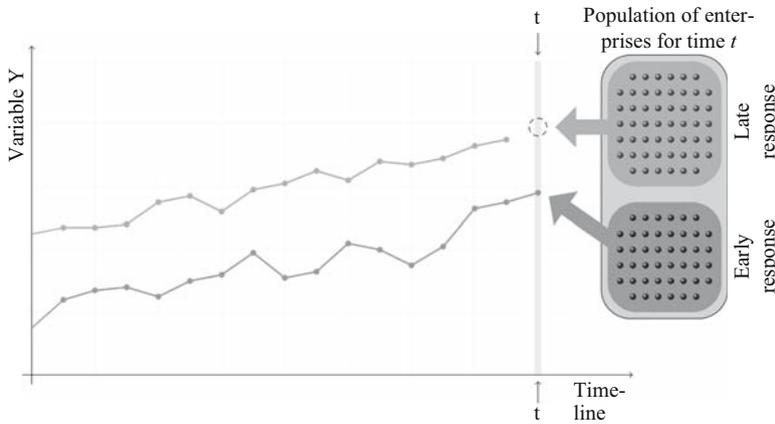


Fig. 3. Sketch of the construction of time series for early and late respondents respectively.

For this exercise, the eligible population was limited to units obliged to report for time period  $t$  only. Units that had provided data in past periods, but had no reporting obligation for reference period  $t$ , were excluded. The population, therefore, grows from the origin of the series to the end; some businesses obliged to report for time  $t$  may have been established after 1996 or may not have been included in the STS sample until later in the series. The growth of the study population is another cause of the increasing trends in series besides the dynamic development of the economy.

The variables that had to be estimated were total turnover, technical production, hours worked, number of employees, and new orders. At first, it appeared that new orders was potentially a good predictor for other variables, such as turnover or production. Unfortunately, the data quality for this variable was very poor: many units reported the same value for both turnover and new orders.

Our intention was to construct a joint model for total turnover, production, and hours worked, and to develop a separate model for number of employees. Due to its poor reliability, the *new orders* variable was omitted. Each of these variables entered the model twice, one time in the *early respondent* time series and again in the *late respondent* time series. To obtain six time series of equal length, the *early respondent* series were lagged for one period into the past, skipping the very first observation for January 1996. Thus, the *early respondent* series were designed as predictors for the *late respondent* series. Each set of time series had to be recalculated for every target period  $t$ , to accurately reflect changes in the *early* and *late respondent* populations.

A key assumption of the proposed model approach is that early responders could be used as predictors for late responders. This assumption is not consistent with findings of other authors dealing with the behavior of respondents in business surveys (see Bankowska et al. 2015 or Lineback and Thompson 2010). Figure 4 plots the response behavior in the Austrian STS by reference period (January through December 2016). We have divided the population into four size classes according to the total turnover of the enterprises. Size classes were determined using the average turnover values for the entire year, so that units are assigned to the same size class in each month. Note that small and medium enterprises tend to be overrepresented in the sample of early responders. For example, about 25% of

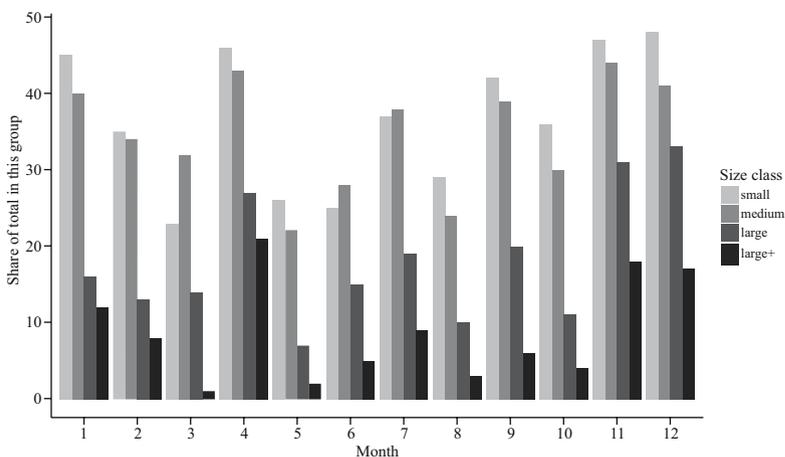


Fig. 4. Response rates for 2016 by month – share of early respondents ( $t + 30$ ) to late respondents by size class. The size classes are: Small – units with turnover smaller than the first quartile of all units; Medium – units with turnover between first and third quartile of all units; Large – units with turnover between third quartile and the 90th percentile; and Large + – units with turnover higher than the 90th percentile of all units.

the smallest enterprises, 28% of medium enterprises, 15% of large enterprises and about 5% of the largest enterprises reported their data before  $t + 30$  in June 2016. (Note that percentages are based on the number of units, not on the total turnover reported.)

Figures 5 and 6 provide examples of corresponding early and late respondent time series for turnover and production (Figure 5) and for hours worked and number of employees (Figure 6) for the July 2017 reference period. The top panels of each figures plot the corresponding early and late respondent time series for each variable. The cross-

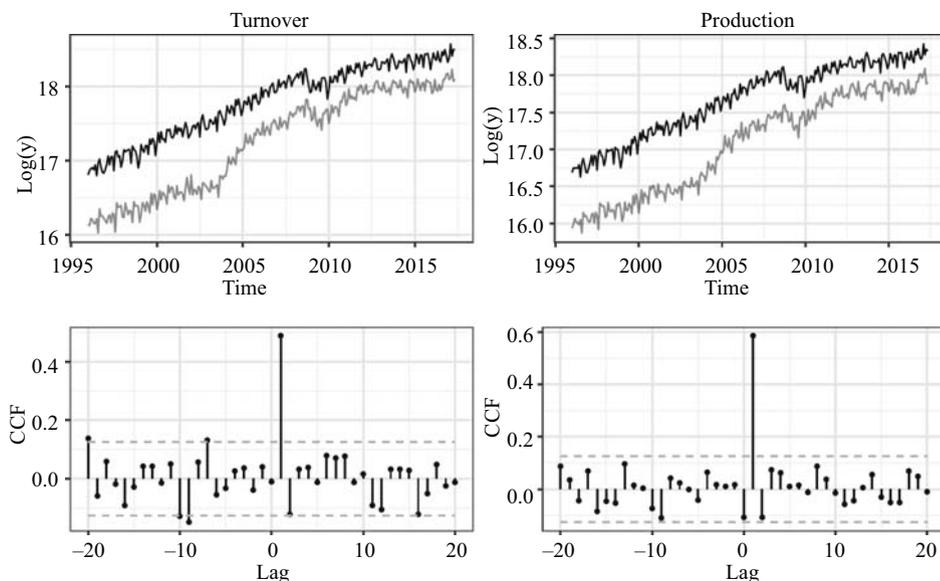


Fig. 5. Logs of original data for turnover and production from July 2017 (dark: early respondents, light: late respondents) and cross-correlations of associated pre-whitened series.

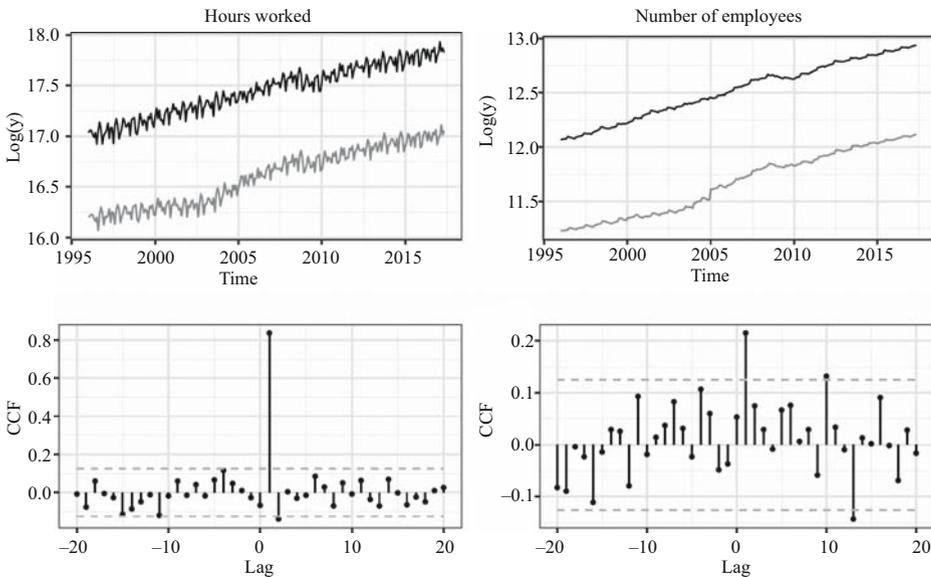


Fig. 6. Logs of original data for hours worked and number of employees from July 2017 (dark: early respondents, light: late respondents) and cross-correlations of associated pre-whitened series.

correlations are presented immediately below. Chatfield (2001) points out that the sample cross-correlations are difficult to analyze because they can be affected by auto-correlations in the component series. This problem does not arise if one of the input series is white noise, which means that it is uncorrelated with mean zero. Therefore, we transformed the first series so that the residuals were white noise, then applied the same transformation to the second series (in literature, this process is called pre-whitening). The presented cross-correlations were calculated from the two residual (“pre-whitened”) series.

A clear peak of the cross-correlograms can be observed at lag one for total turnover, production (Figure 5), and total hours worked (Figure 6). We observed similar results in other months, which we interpret as indication of the leading character of the early respondent series for these three variables. Clearly, the composition of early and late respondents changes from month to month. However, there is a substantial overlap of early respondents between different months. The simulation study presented in Subsections 3.1 and 3.2 investigate whether this presumed leading behavior in early respondents could be used to predict regime shifts like the financial crisis in November 2008.

### 3.1. Simulation Study

This section describes a simulation study for nowcasting late-responding enterprises. This study used final STS data from January 1996 through June 2016. Because the production version of these data were available, we did not implement the data-cleaning step in the simulation study.

To perform the simulation study, we randomly selected a 60% sample out of the population of about 10,000 enterprises/establishments independently in each reference period. The sampled units represent the *early respondents*; the remaining 40% represent

the *late respondents*. Mimicking the procedure shown in [Figure 3](#), two sets of time series (one per variable) were constructed in each reference month, omitting the very last observation for the late respondents. We repeated this procedure independently, drawing 1,000 samples for each reference period for July 2015–June 2016 and November 2008–December 2008 respectively. The latter periods were chosen to study the performance of the nowcasting procedure in the presence of a structural change at the very end of the time series.

As one reviewer suggested, this assumption of independent nonresponse is not necessarily realistic as it is likely that response in one period might depend on prior period response. A more realistic Markov chain model could be implemented by estimating transition probabilities from historical data. We do not believe that our independent selection procedure overly affects the presented results, as many units have response probabilities of 1 or 0. However, we will consider this modification in future research.

Our objective was to implement a vector-autoregressive (VAR) model that included turnover, production, total hours worked, and number of employees expecting interrelationships between these variables. (E.g., turnover in one period could influence production in a later period. In turn, production could influence turnover and hours worked in different periods, etc.). A VAR model includes variables (time series) without deciding the direction of the relationship; there is no clear distinction between dependent and independent variables, but each variable can influence each other. A VAR model of order  $p$  has the following representation ([Lütkepohl 2005, 13](#)):

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (3)$$

where  $y_t = (y_{1t}, \dots, y_{Kt})'$  is a  $(K \times 1)$  random vector, the  $A_i$  are fixed  $(K \times K)$  coefficient matrices,  $\nu = (\nu_1, \dots, \nu_K)'$  is a fixed  $(K \times 1)$  vector of intercept terms, and  $u_t = (u_{1t}, \dots, u_{Kt})'$  is a  $K$ -dimensional white noise process, that is,  $E(u_t) = 0$ ,  $E(u_t u_t') = \Sigma_u$  and  $E(u_t u_s') = 0$  for  $s \neq t$ .

Our primary VAR model included early and late respondent estimates for turnover, production, and total hours worked (six variables). A separate VAR model was used to predict number of employees. This model also included turnover and production.

The time series included in a VAR model must be covariance stationary, meaning their first two moments must not depend on time  $t$ ; (the mean and the variance have to be constant and the autocovariances should only depend on the lag length not on the position in time). Regression of nonstationary time series on each other will generally lead to spurious results. In many cases, nonstationary time series can be made stationary by differencing prior to constructing the VAR model. However, it is likely that differencing each individual series could distort the original joint relationships.

If a nonstationary time series can be made stationary by differencing  $d$  times, the original (undifferenced) series is integrated of order  $d$ . A stationary series has integration order 0. The integration order of a time series can be determined by formal tests of stationarity such as the KPSS test proposed by [Kwiatkowski et al. \(1992\)](#). Consequently, the order of integration has to be increased by one for each significant test. Formal tests are also available to test for the presence of seasonally integrated processes  $\Delta_s x_t = x_t - x_{t-s}$  such as the HEGY test ([Hylleberg et al. 1990](#)) or the Canova-Hansen test ([Canova and](#)

Hansen 1995). We employ the latter in our applications to test the null hypothesis of deterministic seasonality (see the HEGY test, which tests for stochastic seasonality).

Cointegration provides the one exception for regressing integrated series on each other without prior differencing. Engle and Granger (1987, 253) define cointegration as follows:

“The components of the vector  $x_t$  are said to be co-integrated of order  $d$ ,  $b$ , denoted  $x_t \sim CI(d, b)$  if (i) all components of  $x_t$  are  $I(d)$ ; (ii) there exists a vector  $\alpha (\neq 0)$  so that  $z_t = \alpha' x_t \sim I(d - b)$ ,  $b > 0$ . The vector  $\alpha$  is called the co-integrating vector.”

In the case of cointegration, the series are said to be in an equilibrium relationship, which means that the series do not drift too far apart from each other, so that over time they exhibit a common stochastic trend.

The VAR model of Equation 3 can be written in the alternative form (Lütkepohl 2005),

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t \quad t = 1, 2, \dots, T \quad (4)$$

where  $y_t$  is a process of dimension  $K$ ,  $rk(\Pi) = r$ ,  $0 \leq r \leq K$  and  $u_t$  is  $K$ -dimensional white noise with mean zero and non-singular covariance matrix. This representation is called vector error correction model (VECM) for the nonstationary series  $y_t$  (Tsay 2014).

If  $y_t$  is  $I(1)$  then the term on the left-hand side of Equation 4 is stationary. In this case the terms  $\Gamma_j \Delta y_{t-j}$  and  $u_t$  (white noise) on the right hand side of the equation are stationary as well. So the equality depends on the first term on the right-hand side  $\Pi y_{t-1}$  because the  $y_t$  that enter the equation are not differenced. A cointegrated system is therefore dependent on the  $\Pi$  matrix. There are three cases to be distinguished:

- $rk(\Pi) = 0$ : this would imply that  $\Pi = 0$  and that all variables in the VAR have to be differenced – there is no cointegration.
- $0 \leq rk(\Pi) < K$ : series in the system are cointegrated, there are  $K - Rank(\Pi)$  stochastic trends in the system.
- $rk(\Pi) = K$ : All component series are already stationary – the variables are “self-cointegrated”.

In a first step, we had to determine the integration order of the input-series in order to guarantee that all series were integrated of order 1 or 0. For this task we employed the R-package `forecast` (Hyndman and Khandakar 2008), which offers functions for the estimation of seasonal and nonseasonal integration orders. For both cases, we selected tests with a non-unit root null hypothesis, using the KPSS test to identify the nonseasonal integration order and the extended Canova-Hansen test to identify the seasonal integration order.

For the June 2016 reference period, all 6,000 KPSS-nonseasonal tests indicated a single unit root. In contrast, the Canova-Hansen tests were inconclusive in all 6,000 series. Therefore, we assume that every time series entering the VAR model was integrated of order 1 ( $I(1)$ ).

The next step was to test the VAR model for possible cointegration. The model we used for estimation was a VECM with seasonal dummy variables and a drift term, that is,

$$\Delta x_t = \mu + \alpha \beta' x_{t-1} + \sum_{i=1}^{11} a_i d_{it} + \Psi_1 \Delta x_{t-1} + \Psi_2 \Delta x_{t-2} + u_t, \quad (5)$$

where  $x_t$  is a process of dimension six (variables turnover, production, and hours worked, each for early and late respondents),  $\alpha$  and  $\beta$  are  $(6 \times r)$  matrices of rank  $r$ ,  $d_{it}$  are seasonal dummy variables with coefficients  $a_i$ ,  $\Psi_i$  are the coefficients of the AR part in the equation and  $u_t$  is white noise. The graphs presented in Figures 5 and 6 provide indications of equilibrium relationships between selected variables confirmed by formal test results. Table 1 summarizes the results of the cointegration tests for 1,000 models estimated for each month from July 2015 through June 2016. Every single test provided evidence of cointegration, with cointegration ranks varying between one and four. In most cases, the cointegration rank was estimated to be one or two, meaning that there were five or four stochastic trends in the system. In every sample, the formal test provided confirmatory evidence of an equilibrium relationship.

In addition to the VECM shown in Equation 5, we estimated a VECM using seasonally adjusted data. For this, we seasonally adjusted every input series using `x12`; we used the automatic mode for transformation (log or none), model selection (ARIMA Model) and filter choice (seasonal and trend filters). The automatic outlier identification was used to identify all kinds of outliers (level shifts, transitory changes and additive outliers). Identified outliers were suppressed in the seasonally adjusted series. Trading day and leap year adjustment have been required. After that, we fitted a VECM, omitting the seasonal dummy variables and allowing for a higher AR order of the differenced input series, that is,

$$\Delta x_t = \mu + \alpha\beta'x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + u_t, \quad (6)$$

Table 2 presents the results of cointegration tests for this specific model. In a few cases, no cointegration relationship could be identified, that is, the rank was zero. On the other hand, the majority of series were found to have a cointegration rank of two or three. The test results for this model were generally slightly more heterogeneous than in the other model.

Engle and Yoo (1987) state that cointegrated systems should outperform VAR models on differenced data in the long run. For short prediction horizons, however, a VAR in

Table 1. Cointegration rank identified for a VECM with seasonal dummy variables. Reference periods July 2015 to June 2016 – 1,000 samples for each month.

	Month	Year	Rank 0	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
1	Jul	2015	0	586	375	38	1	0	0
2	Aug	2015	0	442	468	88	2	0	0
3	Sep	2015	0	443	475	81	1	0	0
4	Oct	2015	0	443	485	72	0	0	0
5	Nov	2015	0	440	482	76	2	0	0
6	Dec	2015	0	443	477	79	1	0	0
7	Jan	2016	0	575	379	45	1	0	0
8	Feb	2016	0	590	373	35	2	0	0
9	Mar	2016	0	605	352	43	0	0	0
10	Apr	2016	0	571	398	30	1	0	0
11	May	2016	0	605	363	32	0	0	0
12	Jun	2016	0	586	375	38	1	0	0

Table 2. Cointegration rank identified for a VECM with seasonally adjusted data. Reference periods July 2015 through June 2016 – 1,000 samples for each month.

	Month	Year	Rank 0	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
1	Jul	2015	12	60	362	403	130	33	0
2	Aug	2015	2	99	343	407	119	30	0
3	Sep	2015	3	103	377	381	113	23	0
4	Oct	2015	4	89	400	399	93	15	0
5	Nov	2015	7	92	395	384	100	22	0
6	Dec	2015	3	89	361	425	99	23	0
7	Jan	2016	4	78	370	417	115	16	0
8	Feb	2016	3	86	397	413	92	9	0
9	Mar	2016	3	97	394	406	90	10	0
10	Apr	2016	3	71	413	429	72	12	0
11	May	2016	1	90	391	428	81	9	0
12	Jun	2016	1	88	407	424	76	4	0

levels could give satisfactory results. Under this justification, we estimated a VAR in levels (Equation 3) with the same variables used in the previous VECM models, ignoring the cointegration relationship. The justification for the estimation of this model is that if the variables in an integrated VAR model are cointegrated, then the regression of the integrated variables on each other does not produce spurious results, which would be the case for an integrated VAR model without cointegration.

Ultimately, we considered the following five models in the simulation study: the VAR in levels (Equation 3) without and with seasonal dummy variables fit to the unadjusted series; the VECM (Equation 5) without and with seasonal dummy variables fit to the unadjusted series; and a VECM fit to seasonally adjusted data (Equation 6). In addition, we estimated seasonal ARIMA Models to produce univariate forecasts. We used the `vars` (Pfaff 2008) and `MTS` (Tsay 2014) R-packages for estimation, prediction, and analyzes of the VAR and VECM models. All benchmark univariate seasonal ARIMA forecasts were produced with the R-package `x12`. Each of the six individual series were seasonally adjusted (turnover, hours worked, and production, both for late and early respondents). These series were also adjusted for trading day effect and outliers (level shifts and transitory changes) to prevent the possibility of structural breaks invalidating the results of classical unit root tests.

### 3.2. Results

Figure 7 presents boxplots of the deviations of predicted values (nowcasts) to originally reported values for total turnover in reference periods July 2015 to June 2016. Prediction values were only calculated for late respondents. Ideally, the distributions should be centered around zero with low variance. Generally, we observe that the (benchmark) univariate model exhibits a larger bias than the other corresponding models. The magnitude of the bias and variance in the other models differs by month. The VECM with seasonal dummy variables exhibits less bias than the other models; however, the variability of the deviations is smaller for the VECM with seasonally adjusted data.

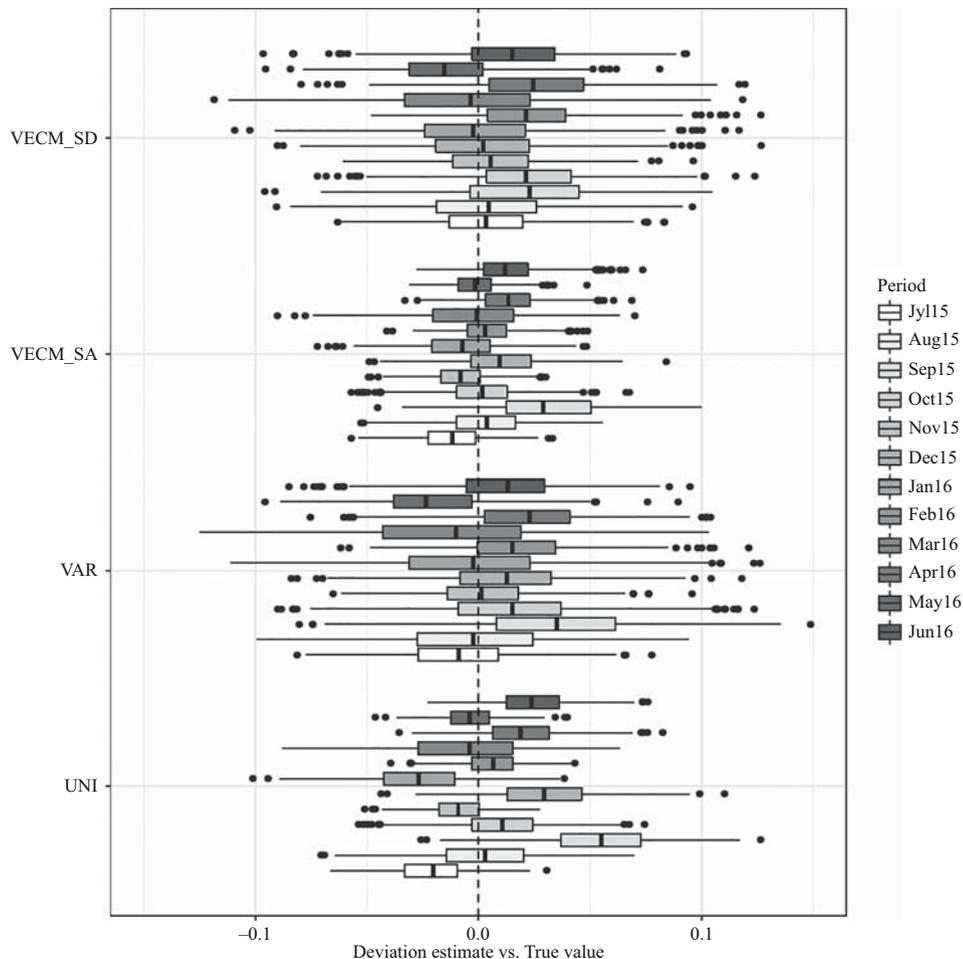


Fig. 7. Simulation results for total turnover from July 2015 to January 2016. The plot shows results of four different forecasting procedures: a VECM with seasonal dummy variables, VAR in levels, VECM with stationary AR terms applied to seasonally adjusted data, and univariate forecasts.

Forecast error patterns for total turnover, production, and hours worked were similar, although the performance of the VECM with seasonal dummy variables was not as strong for the other variables as with turnover.

To assess the performance of the candidate models when regime shifts occur, we performed a separate simulation study (using the same design as the earlier study) for November and December 2008, where the financial crises led to a distinct level shift in the studied series. [Figure 8](#) presents the results for turnover.

Obviously, the univariate model performs badly in November 2008 with a substantial positive bias (moment of structural change as a consequence of the financial crisis). Interestingly, the bias for this model remains present in December 2008. The shift in November 2008 was identified as an outlier in December 2008 by the forecast program, too late for December adjustment.

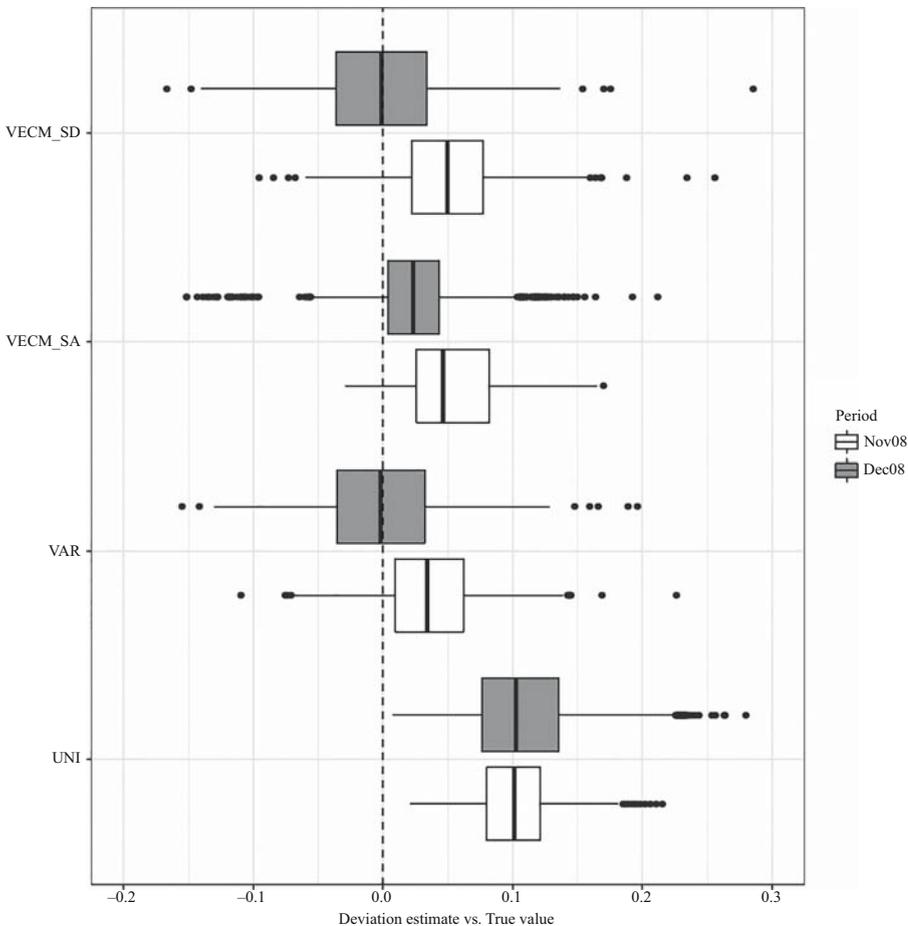


Fig. 8. Simulation results for total turnover from November 2008 and December 2008. The plot shows results of four different forecasting procedures: a VECM with seasonal dummy variables, VAR in levels, VECM with stationary AR terms applied to seasonally adjusted data and univariate forecasts.

In contrast, the multivariate models performed better for this period. It appears that the regime shift affected the early respondent series in the same direction as the late respondent series for corresponding variables, allowing for detection in the prediction models. The turnover nowcasts shown in Figure 7 still show some positive bias, although the bias is greatly reduced in the hours worked nowcasts. In part, these results could be an artifact of our simulation procedure: in particular, the cross-correlations in the simulated samples are consistently higher than those computed from the actual survey data.

#### 4. Conclusion

The methods presented in this article were motivated by the heavy demand for earlier estimates of short-term statistics. Statistics Austria publishes first estimates for STS 55 days after the end of the reference month. Earlier unsuccessful efforts to reduce the collection lag in Austrian STS have focused on improving respondent contact strategies,

for example, by motivating respondents to deliver their questionnaires earlier. In light of this, we considered using econometric models that use historic data, along with early available information for the reference period to forecast missing data. Our preliminary results indicate that this model should perform well in a stable economic environment and should be able to quickly detect and incorporate regime shift effects in the estimates.

Our proposed model could easily be adapted to other situations where actual and historic data is available to form time series and the early available information correlates with the missing data. In the case of Austrian STS, we could observe positive correlations between data from early and late respondents, further motivating the model's usage. In a simulation study, we have analyzed the performance of different models in a stable economy and during a regime shift. In the latter case, the estimation bias of the univariate model could be reduced, although not eliminated completely, whereas our multivariate models performed consistently. Finally, we note that all of our procedures (and our research) were performed with well-documented and accessible R-packages, which made the procedures portable, as well as generalizable.

## 5. References

- Bankowska, K., M. Osiewicz, and S. Perez-Duarte. 2015. "Measuring Nonresponse Bias in a Cross-Country Enterprise Survey." *Statistics Paper Series*, European Central Bank. Doi: <http://dx.doi.org/10.17713/ajs.v44i2.60>.
- Bodo, G., A. Cividini, and L. Signorini. 1991. "Forecasting the Italian Industrial Production Index in Real Time." *Journal of Forecasting* 10: 285–299. Doi: <https://doi.org/10.1002/for.3980100305>.
- Box, G., G. Jenkins, and G. Reinsel. 2008. *Time Series Analysis*. Wiley.
- Canova, F. and B. Hansen. 1995. "Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability." *Journal of Business and Economic Statistics* 13: 237–252. Doi: <http://dx.doi.org/10.1080/07350015.1995.10524598>.
- Chang, I., G. Tiao, and C. Chen. 1988. "Estimation of Time Series Parameters in the Presence of Outliers." *Technometrics* 30: 193–204. Doi: <http://dx.doi.org/10.2307/1270165>.
- Chatfield, C. 2001. *Time Series Forecasting*. Chapman Hall/CRC.
- Cleveland, R., W. Cleveland, J. McRae, and I. Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." *Journal of Official Statistics* 6: 3–73.
- Engle, R. and C. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55: 251–276.
- Engle, R. and B. Yoo. 1987. "Forecasting and Testing in Cointegrated Systems." *Journal of Econometrics* 35: 143–159. Doi: [https://doi.org/10.1016/0304-4076\(87\)90085-6](https://doi.org/10.1016/0304-4076(87)90085-6).
- European Commission. 1998. "Council Regulation (EC) No 1165/98 concerning Short-Term Statistics." *Official Journal of the European Communities*.
- European Commission. 2005. "Regulation (EC) No 1158/2005 of the European Parliament and of the Council, amending Council Regulation (EC) No 1165/98 concerning Short-Term Statistics." *Official Journal of the European Communities*.

- Findley, D., B. Monsell, W. Bell, M. Otto, and B. Chen. 1998. "New Capabilities and Methods of the X12-ARIMA Seasonal-Adjustment Program." *Journal of Business and Economic Statistics* 16: 127–152. Doi: <http://dx.doi.org/10.2307/1392565>.
- Friedman, J. 1984a. *A Variable Span Scatterplot Smoother*. Laboratory for Computational Statistics, Stanford University – Technical Report No. 5.
- Friedman, J. 1984b. *SMART User's Guide*. Laboratory for Computational Statistics, Stanford University – Technical Report No. 1.
- Fröhlich, M., J. Hameseder, and L. Milota. 2010. "Neue Substitutionsmethode für die Konjunkturstatistik im Produzierenden Bereich." *Statistische Nachrichten*. Statistics Austria.
- Hylleberg, S., R. Engle, C. Granger, and B. Yoo. 1990. "Seasonal Integration and Cointegration." *Journal of Econometrics* 44: 215–238. Doi: [https://doi.org/10.1016/0304-4076\(90\)90080-D](https://doi.org/10.1016/0304-4076(90)90080-D).
- Hyndman, R. and Y. Khandakar. 2008. "Automatic Time Series Forecasting: The Forecast Package for R." *Journal of Statistical Software* 27(3): 1–22. Doi: <http://dx.doi.org/10.18637/jss.v027.i03>.
- Kowarik, A., A. Meraner, M. Templ, and D. Schopfhauser. 2014. "Seasonal Adjustment with the R-Packages x12 and x12GUI." *Journal of Statistical Software* 62. Doi: <http://dx.doi.org/10.18637/jss.v062.i02>.
- Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. How sure are we that Economic Time Series have a Unit Root." *Journal of Econometrics* 54: 159–178. Doi: [http://dx.doi.org/10.1016/0304-4076\(92\)90104-Y](http://dx.doi.org/10.1016/0304-4076(92)90104-Y).
- Ladiray, D. and D. O'Brian. 2003. "Nowcasting Eurozone Industrial Production." Working papers and studies. Available at: <http://ec.europa.eu/eurostat/documents/3888793/5816153/KS-AN-03-034-EN.PDF/f4017640-a2da-44f1-859d-566450c905be?version=1.0> (accessed May 2018).
- Lineback, J. and K. Thompson. 2010. "Conducting Nonresponse Bias Analysis for Business Surveys." Joint Statistical Meetings (JSM, Vancouver – Canada). Available at: [https://ww2.amstat.org/sections/srms/Proceedings/y2010/Files/306113\\_55883.pdf](https://ww2.amstat.org/sections/srms/Proceedings/y2010/Files/306113_55883.pdf) (accessed May 2018).
- Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer.
- Mazzi, G. and R. Cannata, 2016. *Rapid Estimates: Different Products for Different Purposes*. Handbook on Rapid Estimates. Doi: <http://dx.doi.org/10.2785/488740>.
- Pfaff, B. 2008. "VAR, SVAR and SVEC Models: Implementation within R Package vars." *Journal of Statistical Software* 27/4: 1–32. Doi: <http://dx.doi.org/10.18637/jss.v027.i04>.
- R Core Team. 2010. R: A Language and Environment for Statistical Computing. R. Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/> (accessed May 2018).
- Statistics Finland. 2004a. "Country-Stratified European Sample for Retail Trade Index – Finland." Statistics Finland, Business Trends – TK-66-783-02. Available at: <https://www.oecd.org/std/36236417.pdf> (accessed April 2018).
- Statistics Finland. 2004b. "Turnover in other Services and Production in Construction – Improvement of Timeliness." Statistics Finland. Business Trends – TK-66-463-03. Available at: <https://www.oecd.org/sdd/35234210.pdf> (accessed May 2018).

Stier, W. 2001. *Methoden der Zeitreihenanalyse*. Springer.

Tsay, R.S. 2014. *Multivariate Time Series Analysis*. Wiley.

U.S. Census Bureau. 2017. *X13-ARIMA-SEATS Reference Manual*. Version 1.1 edition.

Available at: <https://www.census.gov/ts/x13as/docX13AS.pdf> (accessed May 2018).

Received November 2016

Revised November 2017

Accepted January 2018