Received: 8 February 2021 | Accepted: 4 July 2022

DOI: 10.1111/rssa.12912

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY | Series A Statistics in Society

Seismonomics: Listening to the heartbeat of the economy

Luca Tiozzo Pezzoli[1] | **Elisa Tosetti**[2]

[1]Directorate A - Strategy, Work Programme and Resources, Joint Research Centre, European Commission, Ispra, Varese, Italy

[2]Department of Economics and Management, University of Padova, Padova, Italy

Correspondence
Luca Tiozzo Pezzoli, Directorate A - Strategy, Work Programme and Resources, Joint Research Centre, European Commission, Scientific Development Unit Via E. Fermi 2749., I-21027 Ispra, Varese, Italy.
Email: luca.tiozzo-pezzoli@ec.europa.eu

Abstract

Seismometers continuously record a wide range of ground vibrations that are not necessarily related to earthquake activity, but are rather caused by human activity such as industrial processes and traffic. We isolate the human-made imprints from a huge data set made of nearly 20 years of continuously recorded seismic data in Beijing, China, and construct a new daily indicator, the Vibration Index, to forecast regional industrial production. We find that our indicator closely tracks business cycle fluctuations particularly during economic crises. Our results provide policymakers with a new tool to monitor the economy at a highly granular level.

KEYWORDS

forecasting, industrial production, nowcasting, seismic noise, seismic trace

1 | INTRODUCTION

The 2008 Great Recession and the crisis generated by the novel coronavirus disease (COVID-19), although very different in nature, share similar characteristics in terms of high uncertainty and sudden, sharp collapse of markets with a significant contraction in economic activity. Economic crises exacerbate the need for policymakers to obtain accurate and timely economic data to monitor the state of the economy. However, many key economic indicators are released with significant delay and are often subject to revisions that can be substantial. The corona pandemic has exacerbated this problem, leading to the disruption of many key statistics. In a globally interconnected

world where shocks move quickly through different economies, the timeliness and accuracy of economic forecasts remain problematic when official figures are scarce. Recently, the advances in information technology and the social use of internet-related applications have allowed the collection, storage, and analysis of several new sources of information related to human behaviour and activity. Motivated by the strong interrelation between human development and economic growth (Ranis et al., 2000), many of these alternative sources have been considered for building proxies of economic conditions of a particular area, region or country. For example, mobile phone usage data (Mao et al., 2015), information about the mobility of people extracted via remote sensors, camera, satellite images (Henderson et al., 2012, Galimberti, 2020 and Bricongne et al., 2021) or Google mobility reports (Sampi & Jooste, 2020). However, the use of such alternative data sources in forecasting and nowcasting analysis is often limited by their relatively short time-series dimension. Additionally, these data often suffer from privacy issues and, in most cases, are not open source. We refer to Buono et al. (2017) for a comprehensive review of the potentials and drawbacks of unconventional datasets for economic analysis.

In this paper, we exploit the continuous recording of ground motion by seismic sensors to measure vibrations caused by human activity that can be used to build a real-time indicator of economic conditions (at the regional level). Seismic sensors continuously register a wide range of ground vibrations that are not necessarily related to earthquake activity. Rather, they are caused by natural or human-generated sources, such as glacier calving (O'Neel et al., 2007), ocean waves (McNamara et al., 2011), wind turbines (Marcillo & Carmichael, 2018), river flows (Diaz et al., 2014), industrial processes (Marcillo & MacCarthy, 2020) and traffic (Riahi & Gerstoft, 2015). We refer to Mcnamara and Boaz (2019) and Diaz (2016) for a review of existing works in this field and some interesting examples. From a technical point of view, seismologists refer to those signals as 'background seismic noise', or 'ambient seismic noise', defined as persistent vibration of the ground detected by seismic sensors in the absence of signals from earthquakes. The different sources of ambient seismic noise can be distinguished, at least in part, by the *frequency* within the full-frequency band of the seismic spectrum, with anthropogenic sources dominating at frequencies between 2 and 20 Hz (Diaz et al., 2021; Havskov & Alguacil, 2016).[1] Ambient seismic noise has been empirically observed by seismologists since long time, clearly affecting the performance of seismic instrumentation, and in particular their earthquake detection capability (see, among others, Peterson, 1993). However, its potential use for monitoring natural events (i.e. hurricanes, climate change developments) and human activities has been explored only recently. In this direction, a number of works show that the magnitude of human-induced ground vibrations measured by seismometers can be used to monitor industrial processes, urban and air traffic (Marcillo & MacCarthy, 2020; Riahi & Gerstoft, 2015). During the COVID-19 pandemic, Lecocq et al. (2020) and Xiao et al. (2020), among others, observe a remarkable reduction in seismic noise worldwide due to a slowdown in traffic and economic activities. The authors advise the use of such alternative data to monitor human behaviour in real-time. A recent, largely exploratory work, by Hong et al. (2020) suggests that the relationship existing between human activities and economic conditions could be measured using seismic data. The authors find a high correlation between annual seismic noise levels and Gross Domestic Product (GDP, particularly with its manufacturing and industrial components) for a set of countries.

We adopt a set of techniques developed within the seismic literature to disentangle sources of ground motion and propose a novel, daily indicator, that we call *Vibration Index* (VI), measuring seismic noise induced by human activity, robust to external events, such as earthquakes

[1]The frequency is the rate of repetition of complete wavelengths of seismic waves measured in cycles per second, or hertz.

or meteorological phenomena. To demonstrate the usefulness of our indicator, we collect a huge dataset made of nearly 20 years of continuously recorded seismic data in Beijing, China, and use our VI index to nowcast and forecast monthly variations in regional industrial production. To account for the mixed-frequency of the variables, we adopt a mixed-frequency data sampling (MIDAS) approach proposed by Marcellino and Schumacher (2010) and Foroni et al. (2015) and vary the forecasting horizon between 1 day up to 2.5 months before the release date of the economic variable. To compare results across forecasting horizons, we adopt the multi-horizon superior predictive ability (SPA) test recently proposed by Quaedvlieg (2021). Our results show that augmenting a simple autoregressive (AR) model with our VI index significantly enhances the in-sample and out-of-sample predictions of the economic indicator. This improvement is particularly high in the days immediately before the release date of the economic variable.

As also pointed out by recent nowcasting studies (e.g. Jardet & Meunier, 2020), high-frequency data may provide a timely signal when economic conditions suddenly deteriorate, thus allowing economic surveillance in real-time. To investigate this, we proceed in two directions. First, we study the forecasting performance of our model during three important economic contractions occurring during our sample period, namely the 2008-2009 Great Recession, the slowdown in the Chinese manufacturing industry occurring in 2015, and the economic crisis induced by the spread of COVID-19 in 2020. Second, for these episodes, we explore whether seismic data can be used as an early warning signal of these downturns by adopting a logistic model set up (e.g. Frankel & Rose 1996), within our MIDAS approach.

When focusing on the Great Recession and the 2015 economic slowdown, we observe a relatively larger gain in using our regression model augmented with the VI index across all forecasting horizons. Interestingly, during the Great Recession we also observe a considerable slowdown in business activity and seismic noise in the Summer of 2008, perhaps linked to the implementation of anti-pollution policies by local authorities in the Beijing area in view of the Olympics games. As for the logistic model, including the seismic signal helps predict the occurrence of these downturns while a simple baseline fails in both cases. Interestingly, this model also correctly detects the subsequent economic rebound in 2009 that will be otherwise missed by the benchmark model. This highlights the importance of this new high-frequency indicator also in detecting recovery periods. When looking at the COVID-19 crisis, the forecasting performance of the actual values as well as of the downturns detected by the logistic model is bad at longer forecasting horizons, while it improves considerably for relatively shorter horizons. These results highlight the difficulty in anticipating this economic crisis, perhaps because the global pandemic induced by the spread of COVID-19 has largely acted as an exogenous shock relative to our economic indicator. We refer to Bobeica and Hartwig (2021) for time series challenges related to the COVID-19 shock. However, seismic noise, by bringing fresh, up-to-date information about the state of the economy in the midst of the COVID-19 crisis, is very powerful in nowcasting and backcasting industrial production activity.

Overall, results from this paper point at seismic data as a very valuable source of information that can be exploited for monitoring regional economies, for which timely and up-to-date indicators of economic activity are often not available. Another potential application of these methods is for tracking national economies, especially in times when we observe the disruption of many key statistics.

The remaining of the paper is structured as follows. In Section 2 we introduce the data, while in Section 3 we show how to build our VI index and how we include it into the econometric model for forecasting regional industrial production. In Section 4 we summarise the main results, while in Section 5 we conclude with some concluding remarks.

2 | DATA

We have collected the continuous-time seismic trace of broadband seismograms for the station coded as IC.BJT located in Bejing, China, from the Incorporated Research Institutions for Seismology (IRIS) database.[2] This is an open-source data centre gathering and processing seismic data from different seismic networks around the globe. The seismic trace measures in continuous time the (velocity of) ground displacement and is expressed in m/s (metre per second). The seismic station IC.BJT is around 20 km Northwest from the centre of Bejing, the capital of China, and less than 5 km distant from the Bejing ring road and a major railway line. The station is positioned in a concrete vault located at the end of a horizontal tunnel at 60 m depth. We observe that the depth of a station does not generally prevent a seismic sensor to detect human noise (see Van Wijk et al., 2021 for more details).

We have collected the seismic traces of broadband high-gain vertical seismograms (BHZ) (see Mcnamara & Buland, 2004 for more details). The number of sampling points per unit of time (sampling rate) for the BHZ channel of the seismometer employed in this station is 20 per second, meaning that every second we obtain 20 observations of the velocity of ground displacement. We collected seismic data for the period from the 1st of January 2000 up to the 31st of August 2020. Hence, in total, we have 1,727,980 observations a day measured over 7578 days.

We expect to observe a decrease in the power of the seismic signal in periods of economic slowdown with a contraction in local production and consequent reduction in industrial processes and traffic intensity. We refer to Rothengatter (2011) for a discussion on the relationship between economic crisis and traffic.

As economic indicator, we take the 12-month growth rate of Industrial Value Added (VA) for the Beijing region available at a monthly level for the period from the 1st of January 2000 to the 31st of August 2020 and extracted from the National Bureau of Statistics of China. This region is highly populated, with over 21 million resident population at the end of 2019 and a Gross Domestic Product of over 534 million dollars.[3] One peculiarity of Chinese monthly data for industrial production as well as other key economic variables is that data are not available in levels, but rather are only released as the annual percentage growth rate (year-on-year). While one advantage of such transformation is that it attenuates the seasonality problem in the data, it does not allow complete control of the normalisation procedure of our key economic variable. We refer to Section 3.2 for further transformations and adjustments of this variable. To our knowledge, the National Bureau of Statistics of China does not give information on the release date of regional variables, although it does provide the release date of national-wide variables since 2015.[4] Hence, for this application we assume that the release date for the regional variable is the 15th of each month.

3 | METHODS

3.1 | VI construction

To obtain a daily measure of human-induced ground vibrations, we compute the daily Probability Density Function (PDF) associated with the (continuous-time) Power Spectral Density (PSD)

[2] http://ds.iris.edu/ds/

[3] https://data.stats.gov.cn/english/easyquery.htm?cn=E0103

[4] http://www.stats.gov.cn/english/pressrelease/releasecalendar/

of the seismic trace. The PSD is a measure of the seismic signal's power content (or energy) at different frequencies and is typically calculated via a finite range fast Fourier transform (Cooley & Tukey, 1965) of the seismic trace. As suggested by Mcnamara and Boaz (2006, 2019) starting from the PSD it is then possible to compute the PDF within an interval of time. Looking at the seismic noise in a probabilistic sense allows distinguishing ground vibrations due to low probability events, such as earthquakes and instrumental faults, from vibrations caused by higher probability sources, such as human activity or environmental and natural factors. Such processing is useful from a seismology perspective as it permits to assess and monitor the quality of the data produced by a seismic station (see Mcnamara & Boaz, (2019) for some practical examples) but can also provide useful information on ambient noise as a proxy for human development in a region of interest and over time.

The procedure for obtaining the PSD and corresponding PDF follows a number of steps. For each day in the sample, we first split the seismic trace into a number of 1-h overlapping segments and then apply the finite range Fourier transform to convert the displacement motion from the time series to the frequency domain. Specifically, we split the daily seismic trace into 47 1-h segments with 50% overlap, and then we further divide each 1-h segment into 13 segments with 75% overlap. The final 47 PSDs estimates are obtained by averaging the 13 spectra within each 1-h time period considered, where the power is converted into decibels (dB) with respect to the acceleration reference value of 1 $(m/s^2)^2/Hz$. The overlapping segments are considered with the intent to reduce the variance in the final PSD estimates. For a given frequency, the higher the PSD the stronger the associated seismic noise. The daily empirical PDF associated with the PSD attached to a specific frequency or an interval of frequencies can be obtained by collecting the corresponding spectra into bins and calculating the frequency distribution. Specifically, Mcnamara and Boaz (2006) suggest to first discretizing the spectrum into a smaller set of logarithmically spaced intervals of frequencies. Hence, they propose to calculate the daily PDF for the power level attached to each interval by calculating a histogram distribution using bins of 1-dB width. We refer to Mcnamara and Boaz (2006, 2019) for further details on the steps involved in the calculation of the PSD and PDF.

The top panel in Figure 1 shows the seismic trace (ground velocity) expressed in m/s against time expressed in hours (left) and the corresponding PDF (right) calculated using the Mcnamara and Boaz (2006) procedure, for a day in which no major events from a seismic point of view took place. Specifically, Figure 1 shows the seismic power (in dB, with respect to a reference value of 1 $(m/s^2)^2/Hz$) versus the frequency (in Hz), with the values of the frequencies on the x-axis ranging between 0.005 and 10 Hz. One point in the graph indicates the probability of occurrence of a given power at a particular frequency, with the colour showing the probability associated with that specific combination of power and frequency, calculated using the Mcnamara and Boaz (2006) procedure. The graph also reports the minimum, maximum and mode of the PDF. The seismic trace in this graph shows a cyclical behaviour that depends on a multitude of sources both natural and human-generated. In particular, as observed by Diaz (2016), this effect, clearly visible on quiet days but also evident during major earthquakes, corresponds to earth tide variations. A number of studies have shown that vibrations at frequencies below 2 Hz, displayed on the right portion of this graph, are usually related to natural phenomena, such as oceanic waves and large-scale meteorological events (Bonnefoy-Claudet et al., 2008; Diaz, 2016), while vibrations at frequencies above 2 Hz have been mainly linked to human activities (Diaz et al., 2021; Havskov & Alguacil, 2016), the so-called cultural or anthropogenic noise. It is interesting to note from this graph that the seismic noise has a unimodal distribution whose shape varies considerably across frequencies. We observe a more dispersed distribution for frequencies below 1 Hz, while for

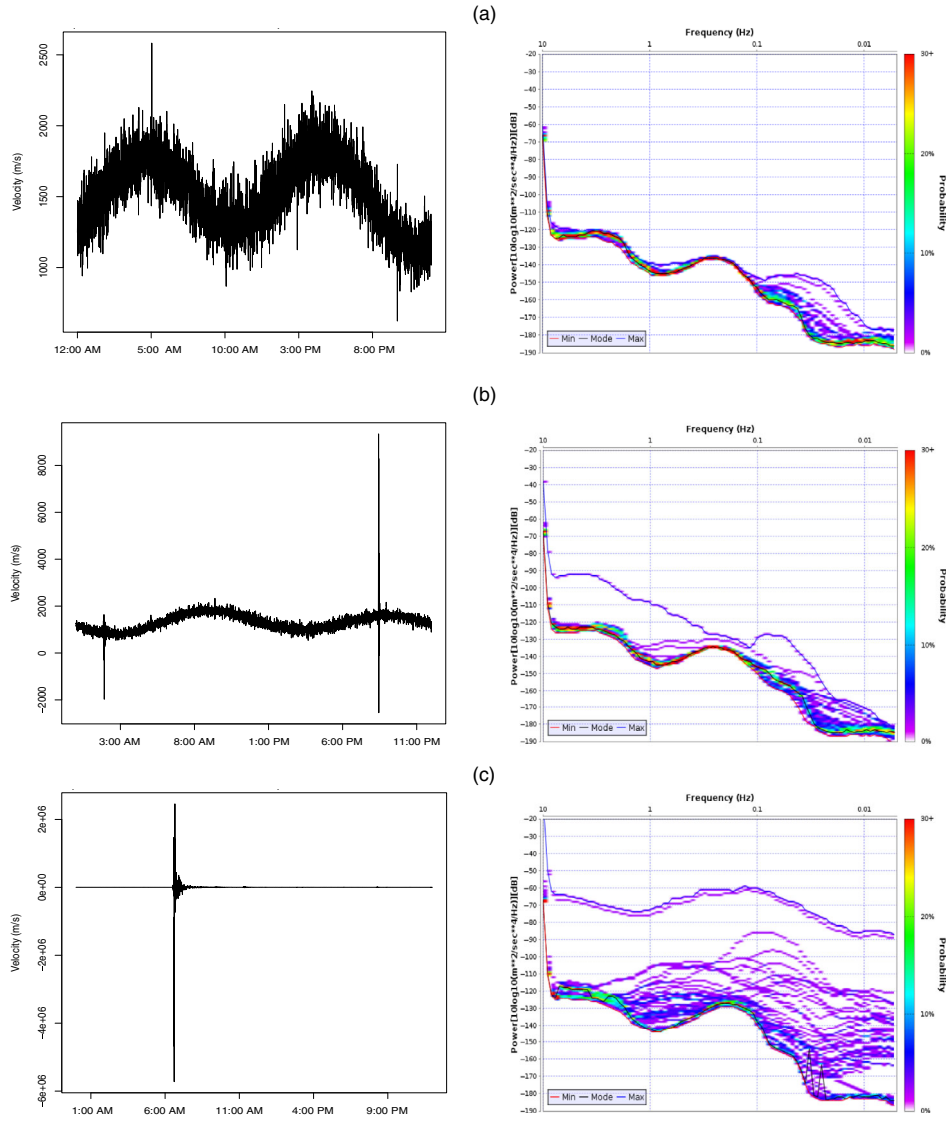

FIGURE 1 Graph of the daily seismic traces and corresponding seismic power versus frequency versus probability level in IC.BJT for three different days. For descriptive purposes, when calculating the probability density functions (PDFs) reported in these graphs, the spectrum is discretised into 88 distinct frequencies and then calculated the (daily) PDF associated to each power level. PSD-PDF graphs can be obtained online from http://service.iris.edu/mustang/noise-pdf/1/. (a) 1 May 2008; (b) 10 March 2008: Jiangsu earthquake; (c) 12 May 2008: Sichuan earthquake [Colour figure can be viewed at wileyonlinelibrary.com]

frequencies above 1 Hz the distribution is more concentrated around the mode, with the majority of observations falling in a range between −120 and −150 dB. The graph also reports the minimum and maximum noise registered at all frequencies. These two statistics are clearly affected by major seismic episodes, such as earthquakes and other tectonic forces that produce seismic energy across all spectrum frequencies (Mcnamara & Buland, 2004). These events, however, are only limited to very short periods of time and hence are likely to belong to the tails of the

probability distribution across all frequencies. Given that in this paper we carry the analysis at a daily level, we require the PDF to be robust to seismic events. To further explore the sensitiveness of the PDF to seismic episodes, in the middle and bottom panels of Figure 1 we report the seismic trace and associated PSD-PDF calculated for 2 days in which two seismic events occurred. In particular, the middle panel shows the seismic trace and PSD-PDF for the 10th of March 2008, in which a 4.3 magnitude earthquake occurred in Jiangsu, around 200 km away from Beijing and the bottom panel reports the same graphs for the 12th of May 2008, when an 8-magnitude earthquake occurred in Sichuan, around 1500 Km away from Beijing.[5]

It is interesting to observe that a strong seismic event such as the Sichuan earthquake has a strong impact on PSD-PDF. The lines appearing in the upper portion of the spectrum can be directly attributed to the seismic event, affecting all frequencies, while the large swings observed in the bottom part, mostly at frequencies lower than 0.1 Hz, correspond to aftershocks of the large seismic event, all impacting the mode of the PDF calculated across the seismic spectrum (the black line in the graph). In contrast, a more modest seismic event relative to the Sichuan earthquake, such as that occurred in Jiangsu in March 2008, while producing visible shifts as well as aftershocks swings in the PSD, seems to impact only the tails of the PDF associated to each frequency, while leaving the central part of the distribution unaltered.

On the basis of this exploratory analysis, we have decided to drop all observations referring to days where an earthquake of magnitude larger than 5 within 1500 km ray from our station, and of magnitude larger than 8 within 2500 km ray. This has led us to discard 110 observations, spread across 20 years of data. For similar reasons, we have turned to missing the seismic data for the six bank holiday days covering the New Year Festival, with the intent to avoid the anomalous significant reductions in noise due to the suspension of industrial activities in these periods.

In our empirical application, rather than discretizing the spectrum in many intervals as suggested by Mcnamara and Boaz (2006, 2019), we first compute the PSDs attached to the band of frequencies that are linked to cultural noise, and then estimate the associated PDF. In fact, the suggested frequency discretization is arbitrary and can be avoided by integrating the PSDs across the frequencies of interest. Specifically, let $f^s_{\kappa d}$ be the PSD for the seismic power at a given frequency κ, calculated for a given day d and in a specific 1-h segment s, with $s = 1, 2, \ldots, 47$. We aggregate the PSDs that belong to the so-called cultural noise band by taking the integral:

$$f^s_d = \int_{\kappa_{\min}}^{\kappa_{\max}} f^s_{\kappa d} d\kappa, \tag{1}$$

where $[\kappa_{\min}, \kappa_{\max}]$ is the interval of frequencies belonging to the band of interest. While for this station frequencies below 10 Hz are available, in this application we set $\kappa_{\min} = 2$ and $\kappa_{\max} = 5$ Hz. In doing this we follow Xiao et al. (2020), who points out that the seismic noise from these frequencies is likely to capture ground vibrations due to traffic linked to commuting, as well as truck movement, rail freight transport and production activities of local industries. In practice, in the integral (1), we aggregate the PSDs associated to a spectrum of

[5]See https://ds.iris.edu/ds/support/faq/52/how-do-i-interpret-magnitudes-and-magnitude-types/ for a classification of earthquakes on the basis of their magnitude.

2456 different frequencies belonging to the cultural noise band, thus obtaining the empirical f_d^s. We then estimate the daily empirical PDF by collecting observations belonging to the 47 1-h segments into 1-dB length bins and calculating its histogram distribution. We apply this procedure to the daily seismic traces for the station IC.BJT over the sampling period, using the IRISSeismic package developed in *R*[6] as well as our own codes. The VI index for a given day d, VI_d, is then calculated as the mode of the empirical PDF calculated on f_d^s, for $s = 1, 2, \ldots, 47$. We take the mode of the estimated time-varying probability distribution with the intent to capture daily variations in the most probable noise power in the cultural noise band (Rastin et al., 2012).

While our procedure allows us to obtain mode time series that are not affected by extreme seismic events, we still have to clean our data to correct for a number of problems that are very typical in the context of seismic applications such as possible datalogger swaps or calibrations. First, we have corrected for two upward shifts observed in the mode time series, with the first shift recorded on the 29th of September 2007 until the 21st of July 2010, and the second shift, occurring on the 19th of April 2013. To this end, for each shift, we have subtracted the average noise observed on the last 7 days before the shift to that observed during the shift. Further, we have identified and replaced outliers by using the procedure of automatic outliers detection in time series advanced by Chen and Liu (1993) to the VI_d time series. At this point, we observe a total of 631 observations with missing values (mostly concentrated in 2005) corresponding to the 8.33% of the sample, that we have replaced with a simple moving average with 3 years rolling window. As a final step, we have carried out a number of operations to clean our seismic data from seasonal patterns. In particular, we have removed short-term seasonal effects by smoothing our daily VI with a moving average of 30 days. Then, we have got rid of long-term seasonal patterns by implementing the US Bureau of the Census X-13 ARIMA-SEATS seasonal adjustment procedure (see Sax & Eddelbuettel, (2018) for more details). Specifically, we have applied such seasonal adjustment to our index sub-sampled at monthly frequencies for each day in the sample. An alternative approach was to apply a seasonal adjustment procedure to the daily values of the series. However, we observe that the X-13ARIMA-SEATS procedure is only applicable to monthly and quarterly data.

3.2 | Calendar adjustments

When using Chinese economic data, one important challenge is deciding how to accommodate all their peculiarities. One major difficulty is related to the dates of the Chinese New Year Festival, which every year can occur either in January or in February, depending on the Lunar calendar. The significant reduction in industrial activity observed during the New Year celebrations produces a strong seasonal effect on its year-on-year growth rate. To deal with this problem, each year the National Bureau of Statistics of China releases the data relative to January and February jointly in March as a single data point. In practice, for January and February, we consider the accumulated (year-on-year) growth rate, from the beginning of the solar year. For the other months, we consider the absolute (year-on-year) growth rate. Hence, industrial production data are only released 11 times a year. We have made similar adjustments to our VI index. In particular to align our indicator with the economic variable in the context of the regression analysis, for each day in the month we have taken the average of the value of the

[6]https://cran.r-project.org/web/packages/IRISSeismic/index.html

VI index for that day in January and February. Finally, given that Industrial VA is expressed in the form of a year-on-year growth rate, we have decided to transform the VI index as 12-month growth.

3.3 | The econometric model

We wish to use real-time seismic data to predict the values of our economic variable. To this end, we adopt a MIDAS approach proposed by Marcellino and Schumacher (2010) and Foroni et al. (2015). Specifically, let y_t the value of our economic variable for the region where the seismic station is located, relative to month t and released on the rth day of month $t+1$. We are interested in forecasting the value of y_t at h days before its release, where $h = 1, 2, 3, \ldots$ is our forecasting horizon measured in days, using all information available up to that day, namely VI_{r-h} and the last released (available) economic indicator. Accordingly, our MIDAS model is as follows:

$$y_t^r = \alpha + \rho X_{r-h} + \beta_h \mathrm{VI}_{r-h} + \varepsilon_t^r, \tag{2}$$

where X_{r-h} is the lag of the dependent variable that is available to the forecaster on day $r-h$, that is, h days before the official release day r of the variable, while VI_{r-h} is our VI index measured at day $r-h$. We observe that, for $h < 15$, VI_{r-h} refers to values of the VI index measured at month $t+1$, for $15 \leq h \leq 45$, VI_{r-h} refers to values of the VI index measured at month t, while for $h \geq 45$ it is indicating values measured at time $t-1$. In our application we will consider a forecasting horizon that goes up to 2.5 months before the release, namely, $h = 1, 2, \ldots, 77$. As a benchmark, we also estimate a pure AR model, namely Equation (2) where we set $\beta_h = 0$.

We first carry out an in-sample analysis and estimate the model using the full sample. We then conduct an out-of-sample exercise by adopting a rolling window estimation approach. For each reference month $t = T_0 + 1, T_0 + 2, \ldots, T$, where T_0 is the timing of the first estimation window, we re-estimate the unknown parameters in Equation (2) using all information available up to day h, with $h = 1, 2, \ldots, 77$, and predict the value for y_t^r. Note that for the reference month t we keep the same target y_t^r, and re-estimate the model using information available up to h, thus producing 77 sets of estimates, forecasts and forecast errors for each target month. Given that our economic variable is published with a 15-day delay with respect to its reference month, depending on the chosen horizon h, we predict the past, current or future value of y_t^r thus leading to backcasting, nowcasting and forecasting exercises, respectively. More specifically, for a given economic variable, backcasting is the prediction of its values for the past months, nowcasting is the prediction of its values for the current month, and forecasting is the prediction of its values referring to future months (see Giannone et al., 2008). In practice, if we want to predict the value of our economic variable for August 2008 and our forecasting exercise is carried out during that month, (i.e. $15 < h \leq 45$) we do a nowcasting exercise. If instead we perform our prediction exercise before August 2008 (hence, $h > 45$) we do forecasting. Finally, if we predict our economic variable in June before the release date, namely for horizons $h < 15$, we do backcasting.

To evaluate the forecasting performance of our VI-augmented model relative to the benchmark we calculate the Root Mean Square Forecasting Error (RMSFE) of both models. The RMSFE is a common measure of forecasting, or predictive, performance of a statistical model. We refer to Bargoli et al. (2015) as an example of the use of RMFSE in a

nowcasting context. For a given model and a given forecasting horizon h, it is calculated as follows:

$$\text{RMFSE}_h = \sqrt{\frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(y_t^r - \hat{y}_t^{r-h} \right)^2}, \tag{3}$$

where $\hat{y}_t^{r-h}$ is the value for y_t^r predicted by the model using information available up to h days before the release of our economic variable. Note that we calculate one value of RMFSE_h for each forecasting horizon, h. The lower the value for the RMFSE_h the higher is the predictive performance of the model at that particular forecasting horizon. In addition to the RMFSE, we also calculate the Diebold and Mariano (1995) (DM) test statistic, which allows testing whether the predictive performance of a statistical model is significantly superior to that of a benchmark, or baseline, model. In this test, we compare the predictive performance of our VI-augmented model with that of a benchmark model where, in Equation (2), β_h is set to zero. In the DM test statistic the null hypothesis is that the two models have equal forecasting accuracy, while the alternative hypothesis is that the VI-augmented model has higher forecasting accuracy than the baseline. We observe that it is possible to calculate one DM test statistic for each forecasting horizon, h, thus leading to the calculation of 77 test statistics. As also pointed out by Quaedvlieg (2021), comparing the forecast ability across individual horizons may lead to incoherent conclusions, particularly when considering a large number of horizons, like in our case. Indeed, in finite samples we are likely to find that a misspecified model outperforms even the population model at one of the many horizons one could consider. To avoid this problem, rather than comparing the forecast ability for each individual horizon, we assess the joint ability of forecasts at the weekly level by adopting the multi-horizon SPA test recently advanced by Quaedvlieg (2021). In our application this basically consists in averaging the DM statistics across forecasting horizons of 1-week length. We calculate this test assuming a one-sided alternative hypothesis that the VI-based model outperforms the baseline. Critical values are calculated via bootstrap procedure with a moving block of length 3 months. We refer to Quaedvlieg (2021) for further details on this test and to Buono et al. (2018) for additional explanations on nowcasting using big data.

We set as first estimation window the 1st of August 2007, while our last estimation window is the 1st of July 2020, for which the economic indicator is released in mid-August 2020. We use estimation windows of length 69 months, which corresponds to roughly one-third of the sample for estimation. This choice allows to include a reasonable number of observations in the estimation windows (e.g. Choi & Varian, 2012). We observe that our rolling windows cover a period that includes three important economic contractions, namely the 2008–2009 Great Recession, the 2015 economic slowdown, and the COVID-19-induced crisis that will be investigated in depth. The timing of the 2008–2009 Great Recession (15th December 2007 to 28th February 2009) and that of the 2015 economic slowdown for China (15th January 2014 to 31st August 2016) have been extracted from the OECD Composite Leading Indicators series for China. Finally, we have manually set the timing to the COVID-19 crisis between the 1st of February 2020 up to the end of the sample period, namely the 31st of August 2020.

4 | RESULTS

We now move to our empirical results. In this section, we first analyse the relationship between industrial VA and the VI index both qualitatively and within a linear regression framework.

We then compare the point forecast performance of the VI-augmented model with that of the baseline model and further discuss their different ability to predict the direction (positive/negative) and intensity changes (the depth/amplitude of a crisis/recovery) of the economic variable of interest. Finally, we explore whether the VI index can be seen also as an early warning indicator of a downturn/upturn.

4.1 | Industrial VA and the VI index

Figure 2 shows the temporal evolution of the growth rates in Industrial VA and in our VI index. We remind that our economic variable is at monthly frequency having missing values for all Januarys, while VI is a daily indicator. Overall, the VI index seems to closely follow the economic indicator. This is particularly evident during the three sub-periods of the Great Recession, the national Chinese slowdown and COVID-19 induced crisis, reproduced at the bottom of Figure 2.

During the Great Recession, in the period between May 2008 to September 2008 the Industrial VA contracted considerably, falling to −9.1% in August. Similarly, the VI index drastically reduced by around 6.5 dB. It is interesting to note that this period partly overlaps with the timing of the Olympics Games, held in Beijing in August 2008. Starting from July 2008, the Chinese authorities put in place several measures to reduce air pollution levels such as vehicle traffic restrictions, and stopped the activities of several factories in Beijing and adjacent regions. This was a negative shock to the Beijing regional economy that was reflected in a sharp reduction in noise levels due to the imposed stop on traffic and industrial activities.

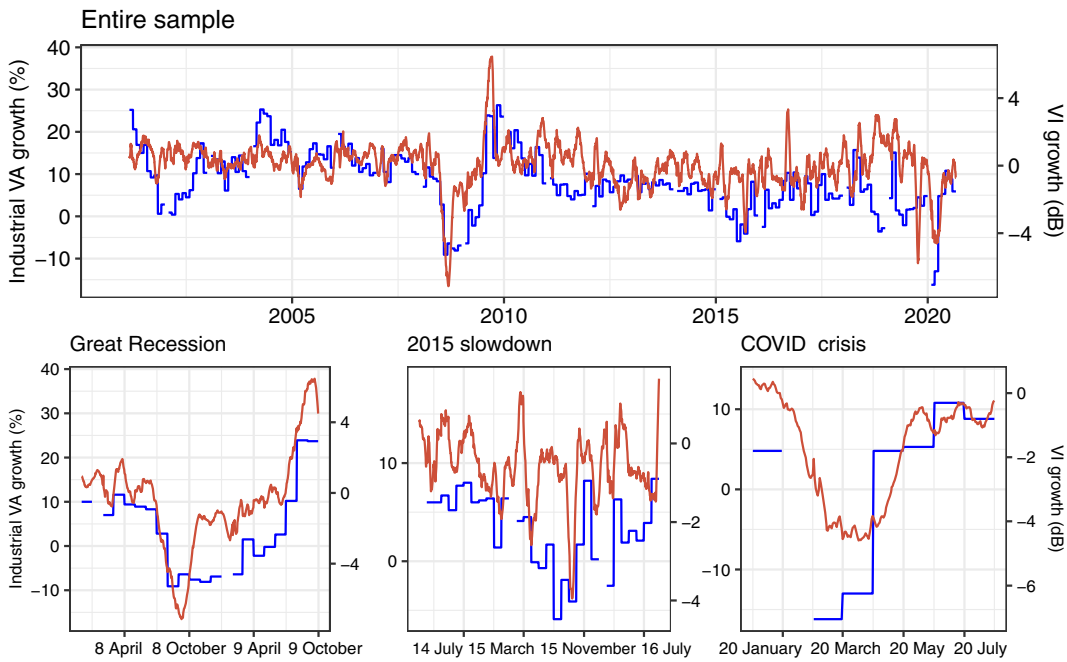

FIGURE 2 The temporal evolution of monthly Industrial Value Added growth (blue) and daily vibration index (expressed in growth) (red). [Colour figure can be viewed at wileyonlinelibrary.com]

By combining satellite images with simulations from a regional transport model, Mijling et al. (2009) found that the effect of these measures considerably reduced the level of air pollution in the months before and during the Olympics Games. The Industrial VA index remained negative until June 2009 (with the only exception of March 2009). As for the VI index, once restrictions were lifted off in September, it started increasing but did not reach the pre-Olympic Games level and remained very low until July 2009. The slow recovery of both the economy and noise levels was probably due to the effects of the global financial crisis that started off with the collapse of Lehman Brothers in the United States in September 2008. The financial collapse produced a freeze in global demand and a contraction in the Chinese export-driven economy, especially in the Eastern regions of China (Mi et al., 2017; Yu, 2009). To recover from this slump, the Chinese government implemented an expansionary fiscal policy with the aim to compensate for the drop in exports through a stimulus to the internal demand. As a result, Industrial VA for the Beijing region jumped to 10.2% in July 2009, compensating for the negative values ranging between −8.1 and −6.4 registered in the period from September to December 2008. Interestingly, our VI index closely followed the behaviour of the industrial growth rate, with a sharp rise in July 2009.

In 2015 the Chinese economy went through another period of economic turbulence. Tensions in the domestic financial market over the summer along with a contraction in the global demand for Chinese goods against an overcapacity of Chinese factories produced an important slowdown in industrial production of the country and in the export-oriented regions. Beijing Industrial VA started to be sluggish in April and May of that year, with economic weakness becoming clearly visible over the summer, when the regional economic indicator dropped to −5.9% and −4.1% in July and September, respectively. As for the VI index, we note the first contraction in March when our indicator reached a value of −2.4 and during the summer, when it reached its minimum at around −3.9 in September. At the end of the year, an increase in industrial production is coupled with a similar increase in the VI indicator.

Finally, in February 2020, at the beginning of the COVID-19 induced economic crisis, we can observe a 16.2% fall in the Industrial VA. This is mainly due to the reduction in business activities imposed by Chinese authorities to contain the spread of the virus. As soon as the government health policies softened in April, the economic indicator registered a sharp increase to then remained stable until the end of the sample period. This fast economic recovery, clearly visible from a V-shaped dynamic in industrial production, is reflected in the temporal evolution of the VI index, which reduced by around 4.5 dB in the period February–March 2020 and then increased in April until attaining a steady level through the end of the sample.

We now turn to the estimation of Equation (2). Table 1 summarises results from Ordinary Least Squares estimation using the full sample. For descriptive purposes, in this table we report only the case where $h = 15$, so that in the estimation we use information up to the end of each reference month. Results show a significant relationship between our VI index and the Industrial VA, with 1 dB increase in the former resulting in an expected 1.25% rise in the latter. Overall, the inclusion of the VI index increases the adjusted R^2 by nearly 5 percentage points, thus indicating the good explanatory power of human-induced seismic noise in this regression. We observe that, as explained in Section 3.1, when calculating the VI index we have significantly pre-processed seismic data, for example, by dropping outliers as well as eliminating data near key holidays. With the intent to determine whether using raw seismic data deteriorates the predictive power of our VI index, thus pointing at the pre-processing of seismic data as an essential part of the analysis, we have also estimated the relationship between seismic data and our economic variable without the

TABLE 1 The relationship between Industrial value added (VA) growth and the vibration index (VI) growth with monthly data

	Dep. var.: Industrial VA growth$_t$	
	(1)	**(2)**
Industrial VA growth$_{t-1}$	0.706***	0.791***
	(0.041)	(0.041)
VI growth$_t$	1.250***	
	(0.223)	
Constant	2.363***	1.689***
	(0.442)	(0.454)
n.obs	215	215
R^2	0.680	0.633
Adj. R^2	0.677	0.631
RMSE	4.045	4.323
F -Stat	225.227***	366.8814***

Notes: Column (1) reports results for estimation of the model with $h = 15$. Column (2) shows results from estimation of our baseline model where we set $\beta_h = 0$.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

previously described preparation stage (except for the seasonal adjustment). Regressions results are very similar to those reported in Table 1, with a positive and significant coefficient attached to the VI index growth. Such finding is perhaps explained by the fact that the effect of outliers, when taking the 30-days moving-average of the daily values of VI growth in Equation (2), is considerably weakened. For the easiness of exposition results are not reported, although they are available upon request.

4.2 | Forecast analysis

Figure 3 shows the RMSFE and associated SPA statistic when forecasting y_t^r for the estimated VI-augmented and baseline models, at horizons that range from 1 day before the release of the target month ($h = 1$) up to 2.5 months before the target month ($h = 77$), for the full sample. For the easiness of exposition results for RMSFE are shown in weekly increments.

As also explained in Section 3.3, horizons in the grey shaded area ($h > 45$ days) represent forecasting, horizons in the white area ($15 < h \leq 45$) indicate nowcasting, while horizons in the orange shaded area ($h < 15$) represent backcasting. One important finding from these two graphs is that the RMSFE of the noise-augmented model is always smaller than that of the benchmark model across all forecasting horizons. The SPA statistic is statistically significant at the 5% significance level, thus the null hypothesis of superior performance of the benchmark model is rejected for several forecasting horizons. Specifically, at $t - 1$ the SPA is significant in the week just before the release of y_{t-2}^r, while at t it turns significant during the first week of the month and stays highly significant across all remaining forecasting horizons. These results indicate that augmenting the pure AR model with our VI index does significantly improve its forecasting performance, particularly in the days immediately before the release of the economic variable, and mainly for small

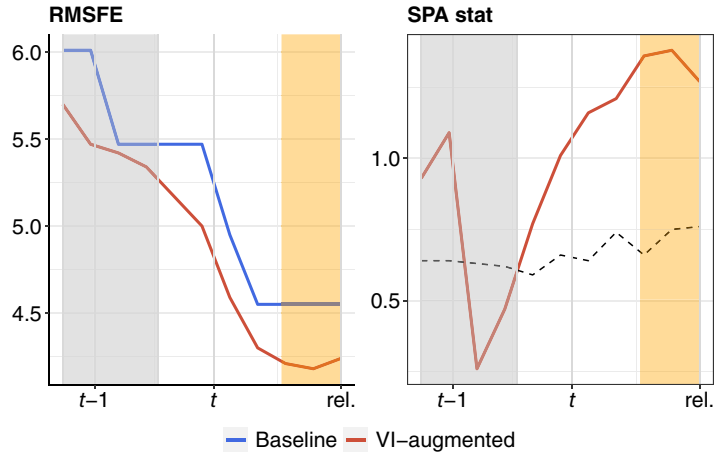

FIGURE 3 Root mean square forecasting error (RMSFE) and superior predictive ability (SPA) statistic for varying forecasting horizon. The *y*-axis of the left graph reports the RMSFE (in percentage points) over the period going from August 2007 till July 2020 while the right graph reports the value of the SPA statistic. The blue line represents the benchmark autoregressive model, while the red line represents model (2). The red line in the right graph represents the SPA statistics at different weekly horizons while the black dashed line indicates the critical value at 5% significance level for a one-sided SPA test (calculated following the bootstrap procedure of Quaedvlieg (2021)). The null hypothesis of this test is that the benchmark model is preferred over the VI-augmented one. Results are reported in weekly increments for the RMSFE. The white area represents the reference period being forecasted. Horizons in the grey area represent forecasts, those in the white area are nowcasts, and those in the orange area are backcasts. [Colour figure can be viewed at wileyonlinelibrary.com]

values of h, in particular when the forecasting horizon is very short ($h < 15$, thus backcasting) to short ($15 < h \leq 45$, thus nowcasting).

Figure 4 reports the RMSFE for the three periods of economic contractions. We observe that for these events we only report the RMSFE as we do not have enough observations to carry out reliable SPA tests. Though exploratory, these three figures indicate that the RMSFEs of our models are on average much larger than when calculated on the full sample, pointing at the difficulty in predicting economic variables at the time of crisis. However, we also observe large reductions in RMSFE for the VI-augmented model relative to the benchmark model during the three crises.

In particular, when focusing on the Great Recession, we observe a large gain across all forecasting horizons, with an average 1.93 percentage points difference in the RMSFE of the naive model and the VI-augmented specification. The largest difference, equal to 3.33 percentage points, is observed 6 weeks before the release of the economic indicator, thus in the nowcasting area.

Suppose that in August 2008 a policymaker is interested in knowing the state of the economy for that month. The value of the industrial VA growth for August, which will be released in mid-September, is equal to −9.1%. However, to predict this value the policymaker only knows the industrial growth rate of June (which is published on the 15th of July) if she/he is before the 15th of August, and, after this date, the July value of the economic variable. With such information, when using the pure AR model, the predicted value of the economic variable for August ranges between 10.28% and 5.53%, depending on whether the policymaker is evaluating the state

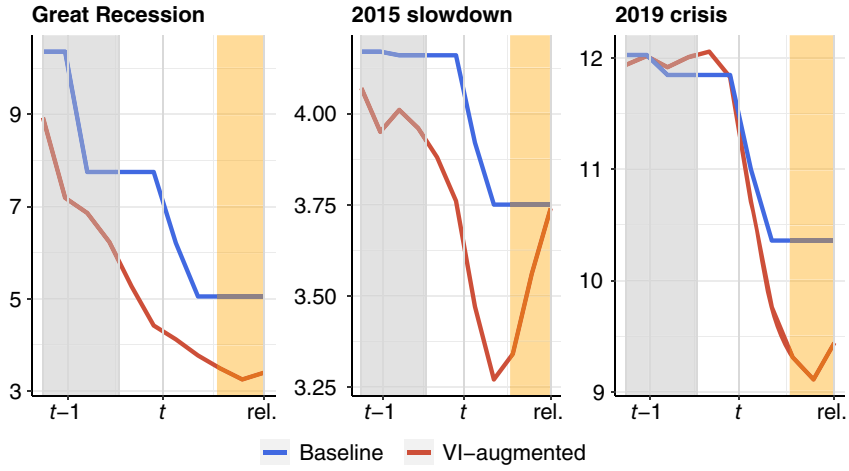

FIGURE 4 Root mean square forecasting error (RMSFE) statistic for the three economic slowdowns. Notes: The *y*-axis reports the RMSFE (in percentage points) over the period going from December 2007 till February 2009 (left graph), from February 2014 till August 2016 (centre graph) and from February 2020 till July 2020 (right graph). The blue line represents the benchmark autoregressive model, while the red line represents model (2). Results are reported in weekly increments. The white area represents the reference period being forecasted. Horizons in the grey area represent forecasts, those in the white area are nowcasts, and those in the orange area are backcasts. [Colour figure can be viewed at wileyonlinelibrary.com]

of the economy at the beginning or at the end of August. On the other hand, when adopting the VI-augmented model, the predicted industrial production will be updated daily using the observed VI index, and would range between 2.58%, at the beginning of August and −4.22% at the end of that month. We observe that, when approaching the end of the month, the forecasting error of the VI-augmented model reduces considerably because the model incorporates seismic information for the entire month of August. This example shades some lights on the type of support that a policymaker can get by using the VI-augmented model, at least for two reasons. First, the negative forecasting value of 4.22 will be a precious signal of an economic contraction that would not be detected by a naive AR model (which, in fact, still provides a positive figure). Second, the VI-augmented model reveals a more accurate estimate of the depth of the slump already in August, well before the official release of the economic variable in the following month, thus allowing policymakers to set up possible timely policy responses. Similarly, for July 2009, when the economy recovers thanks to the government's expansive policy and industrial VA growth suddenly jumps to 10.2%, the pure AR model predicts a 1.42%–3.50% growth, versus a rise between 4.79% and 8.71% estimated by the VI-augmented model. While both models predict an increase in the industrial VA growth, the seismic-augmented specification helps policymakers to produce a more accurate estimate of the amplitude of the recovery induced by the expansionary policies implemented to contrast the Great Recession.

When focusing on the national slowdown, the difference between the RMSFE of the benchmark and VI-augmented models, is on average 0.26 percentage points, which is smaller than that during the Great Recession, and more pronounced between 6 and 3 weeks before the release, where it ranges between 0.40 and 0.48. As an example, the predicted value that a policymaker will obtain from a simple AR model of the Industrial VA for July 2015, the month with the biggest drop in the economic variable namely −5.9% released in August, would range between

2.56% (6 weeks before the release) and 3.29% (4 weeks before the release) while that from the VI-augmented model would range between 2.74% and 2.83%. Although with both models we predict positive rather than negative growth, the predicted values obtained when using the VI indicator are slightly smaller than the benchmark values.

During the COVID-19 induced crisis, the difference between the RMSFE of the benchmark and VI-augmented models is small, being on average 0.34 percentage points. The forecasting performance of the VI-augmented model is very bad at longer forecasting horizons, although it improves considerably for shorter forecasting horizons, with the RMSFE dropping sharply after the release of the economic variable of the period $t-1$. Specifically, the average RMSFE difference in the last 3 weeks before the release is about 1.07. This result highlights the difficulty in anticipating such an economic crisis, perhaps because the global pandemic induced by the spread of COVID-19 has largely acted as an exogenous shock relative to our economic indicator. However, including our VI index considerably reduces the RMSFE at short and very short forecasting horizons. For example, for the 3 weeks before the release, the predicted value for Industrial VA growth in February (that considers January and February jointly) using our VI-augmented ranges between 2.17% and 2.83% while the baseline model produces a prediction of 4.57%. This is obviously very different from the real value of −16.2% that will be released in March, but still shows how our indicator better anticipates a contraction relative to a simple naive model. In March, both models provide negative predictions for the industrial production that will be released in April, that is −13%. In particular, the forecasts (for the 3 weeks before the release) range between −8.51 and 9.61 percentage points for our VI-augmented model against a predicted value of −4.87 for the benchmark specification. Again, our model better informs policymakers on the depth of the crisis than a pure AR specification.

4.3 | Predicting economic downturns and upturns

In this subsection we explore the use of our VI index to detect economic crises and recoveries in real-time, with the aim to understand whether seismic data can be used as an early warning indicator of downturns and upturns. To this end, we first define as downturn a fall in the value of y_t^r below the 10th quantile of the distribution of the time series $y_1^r, y_2^r, \ldots, y_t^r$ observed for a period longer than 2 consecutive months and summarise such information in a downturn index (Down_t^r) that is a binary variable assuming the vale of one when y_t^r falls below the 10th quantile threshold and zero otherwise. In doing so, we follow recent literature (see Barbaglia et al., 2020; Ferrara et al., 2022; Manzan, 2015 among others) which stresses the fact that high-frequency indicators could be helpful in forecasting specific portions of an economic variable distribution. Such definition is also consistent with existing definitions of troughs of a business cycle, such as the NBER recession indicator, where it is required to observe a significant decline in economic activity spread across the economy and lasting more than a few months. Similarly, we identify as upturn a rise in y_t^r that exceeds the 90th quantile of the distribution of $y_1^r, y_2^r, \ldots, y_t^r$ for a period longer than 2 consecutive months. We synthesise such information in an upturn index (Up_t^r) that is a binary variable assuming the value of one when y_t^r goes above the 90th quantile threshold and zero otherwise.

In Figure 5 we indicate in grey the downturns ($\text{Down}_t^r = 1$) and in orange the upturns ($\text{Up}_t^r = 1$) occurring during the out-of-sample period. The identified downturns occur in the months August 2008 to April 2009, April 2015 to February 2016 and February to March 2020, thus taking place during the periods of economic crises and slowdown that we have identified in the

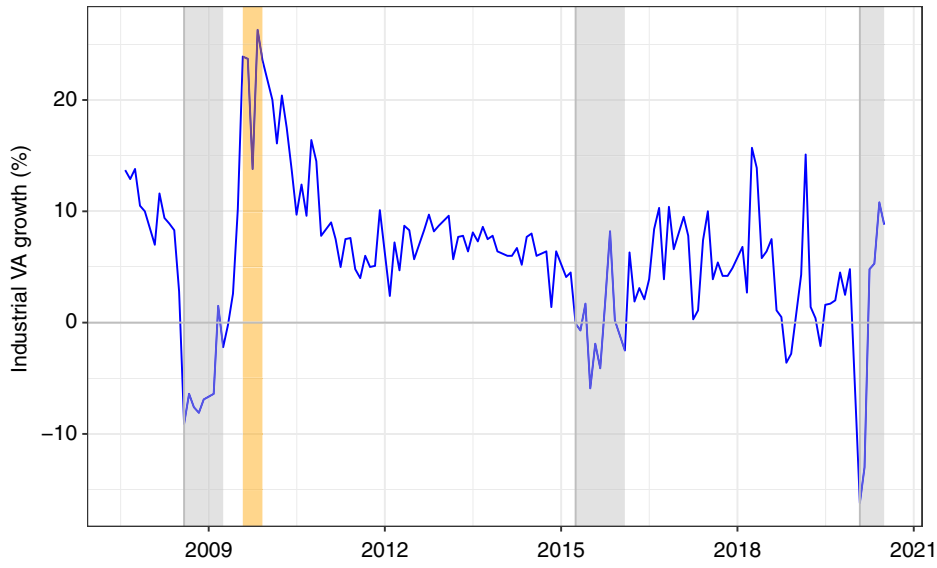

FIGURE 5 Industrial value added growth in the forecasting period with grey shadowed areas for downturns and orange shadowed areas for upturns. [Colour figure can be viewed at wileyonlinelibrary.com]

first part of the paper. Our procedure only finds one upturn, occurring in the months between August and December 2009.

We borrow from the extensive literature on early warning systems for financial and economic crises prediction (see Bussiere & Fratzscher 2006; Dawood et al., 2017; Frankel & Rose, 1996 and among others) and we adopt a logistic regression model with a MIDAS approach to predict the occurrence of the downturn (upturn). Specifically, the probability of downturn/upward is as follows:

$$\Pr(Y_t^r = 1) = \frac{e^{\alpha+\rho X_{r-h}+\beta_h \mathrm{VI}_{r-h}}}{1 + e^{\alpha+\rho X_{r-h}+\beta_h \mathrm{VI}_{r-h}}}, \tag{4}$$

where Y_t^r denotes either the downturn binary variable ($Y_t^r = \mathrm{Down}_t^r$) or the upturn binary index ($Y_t^r = \mathrm{Up}_t^r$), using information on our economic variable up to month t and released on the rth day of the month $t+1$. As in the linear model (2), X_{r-h} and VI_{r-h} are the economic variable and the VI index available at $r-h$, respectively. In our application we use a model with the VI index as the unique regressor ($\rho = 0$) and compare it with a benchmark logistic model that contains as explanatory variable only the lag of the industrial VA growth ($\beta_h = 0$). This is done with the intent of isolating the contribution of our seismic indicator in the downturn/upturn detection. In order words, the inherited persistence of the AR component could overwhelm the benefit of the seismic signal in turning-point detection. We guarantee a sufficient number of upward and downturn observations in the estimations (McFadden, 1984) by adopting a recursive (increasing) window approach. Finally we produce, for each reference month and binary index the predicted probabilities associated to the 77 forecasting horizons.

The graphs at the top and that at the bottom-left of Figure 6 show the estimated probability of observing a downturn in the months of August 2008, April 2015 and February 2020, respectively,

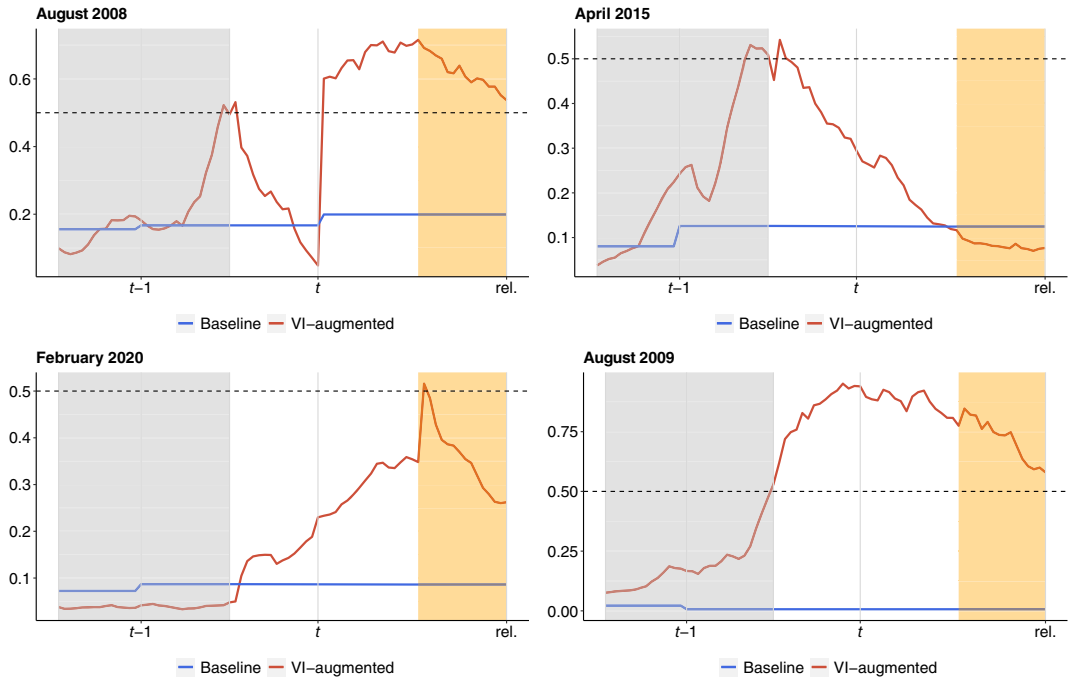

FIGURE 6 Estimated probability of observing a downturn (upturn) in four months. The *y*-axis reports the predicted probability of downturn (upper-left, upper-right and bottom-left graphs) and upturn (botton-right graph). The blue line represents the benchmark logistic model with the lagged industrial production growth as explanatory variable. The red line represents the logistic model with the vibration index as explanatory variable. The white area represents the reference period being forecasted. Horizons in the grey area represent forecasts, those in the white area are nowcasts, and those in the orange area are backcasts. The black dashed line indicates the 0.5 probability level. [Colour figure can be viewed at wileyonlinelibrary.com]

for the model employing the VI index versus the baseline one. These months are the points in time where the periods of downturn identified above begin, for which we wish to predict their occurrence in a timely manner. The graph at the bottom-right of Figure 6 reports, instead, the probability of observing an upturn in August 2009, namely the month in which the period of economic recovery begins. All graphs show the estimated probability of a downturn (upturn) calculated for each forecast horizon $h = 1, 2, \ldots, 77$.

We observe that, while the baseline model always fails to detect a downturn/upturn, for all the four time points the estimated probabilities under the model which employs the VI index rise above the standard cutoff value of 0.5 at several forecast horizons, thus correctly pointing at the occurrence of a downturn/upturn. We also observe that in August 2008 and April 2015 the estimated probability rises above the 0.5 threshold both in the forecasting and nowcasting windows. In August 2009, the upturn is identified for the nowcasting and backcasting area while for the COVID-19 crisis the model can only detect the downturn in the backcasting horizons. The difficulty in the prediction of a downturn induced by such an exogenous abrupt and deep drop in the industrial VA growth would suggest a possible extension of this work through the use of machine learning techniques such as those used by Bluwstein et al. (2020) or more formal turning point detection methods such as Markov switching techniques (Duprey & Klaus, 2017). This would allow an investigation of the non-linear effects of seismic data on the dependent variable

and will be the object of a future study. Overall, these results suggest that seismic data could be used as a timely instrument to detect the beginning of downturn and upturn periods, a task that is particularly difficult in times of structural change, where traditional, slow-moving economic indicators are not helpful.

5 | CONCLUDING REMARKS

In this paper we have explained how to use ground vibrations generated by human activities to construct a new indicator, the so-called VI, that could help nowcast and forecast an economic variable at a highly granular level, that is, regional. We apply this procedure for the case of the Beijing region and use, as economic variable, the monthly growth rate of Industrial VA for that area. The choice of Beijing provides us the opportunity to evaluate not only the performance of our indicators during periods of global economic contraction (Great Recession and the pandemic-generated crisis), but also during periods of regional slowdown in business activity such as the 2015 decrease in exports and to special events, such as the Olympics games. We compare the forecasting performances of a common benchmark (AR) model against another one augmented with our daily VI, and use a MIDAS approach to deal with the mixed frequencies of the data. We find that the inclusion of our new indicator improves the forecasting performance with respect to the benchmark model especially at very short (backcasting) and short (nowcasting) forecasting horizons. During the Great recession (including the Olympic games period) this performance extended also at longer forecasting horizons while for the COVID pandemic crisis the improvement is visible only at short horizons. We have also explored the ability of our VI index to detect economic downturns and upturns in real-time. By adopting a logistic regression method we show that the VI index is able to predict such events. Overall, our results show that the use of seismic data for nowcasting and forecasting economic variables seems to be a promising area of research, particularly useful for capturing the evolution of economic developments at high resolution (namely regional) in real-time. One interesting extension of this work is to consider a wider set of seismic stations, at the regional and national levels, and use them jointly for a more in-deep analysis of the granular variations of vibrations induced by economic activity, across different regions. However, since seismic data are strongly affected by the characteristics of the environment where the different stations are located (Hong et al., 2020), a careful selection of strategic sites should be carried out in order to exclude data recorded in areas without any interest from an economic perspective. This would allow us to pick only stations whose data are useful for economic analysis (such as airports, factories or highways) and could possibly improve the forecasting of indicators related to different economic sectors (such as tourism GDP, unemployment and industrial growth). It is also important to note that seismic data have many advantages with respect to other remote sensing data recently advanced as alternative data sets for macroeconomic analysis. Relative to satellite images, they are less sensitive to weather and low visibility (Riahi & Gerstoft, 2015). In addition, seismic data are open source and are relatively easy to collect, especially in times when key official statistics are disrupted or not promptly available. This is especially the case of regional economies, for which existing economic indicators are often released with a long delay and at lengthy time intervals. Finally, unlike many new alternative data sources, such as Google mobility reports that have been shown to be useful in nowcasting economic activity (see Sampi & Jooste, 2020), our indicator may provide a longer time series, thus well suited for macroeconomic analysis, and do not need any anonymisation processing.

ACKNOWLEDGEMENTS

The authors would like to thank the colleagues of the Centre for Advanced Studies at the Joint Research Centre of the European Commission for helpful guidance and support during the development of this research work. We are grateful to the participants of the second "Big Data & Economic Forecasting" workshop, the "Seismonomics Morning Talks" lecture held at the European Commission and the "International Symposium on Forecasting" for the numerous comments that significantly improved the paper.

CONFLICT OF INTEREST

The views expressed are purely those of the writers and may not in any circumstance be regarded as stating an official position of the European Commission.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in https://services.iris.edu/mustang/.

ORCID

Luca Tiozzo Pezzoli https://orcid.org/0000-0002-9737-0213
Elisa Tosetti https://orcid.org/0000-0003-0979-2828

REFERENCES

Barbaglia, L., Consoli, S. & Manzan, S. (2020) Forecasting with economic news. Available at: https://ssrn.com/abstract=3698121.

Bargoli, D., Metelli, L. & Modugno, M. (2015) The importance of updating: evidence from a Brazilian nowcasting model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2015(1), 1–22.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S. & Simsek, O. (2020) *Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach*. Bank of England Working Paper 848.

Bobeica, E. & Hartwig, B. (2021) *The COVID-19 shock and challenges for time series models*. ECB Working Paper Series 2558.

Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y. (2008) The nature of noise wavefield and its applications for site effects studies: a literature review. *Earth Science Review*, 79, 205–227.

Bricongne, J.-C., Meunier, B. & Pical, T. (2021) *Can satellite data on air pollution predict industrial production?* Working Paper Banque de France 847.

Buono, D., Kapetanios, G., Marcellino, M., Mazzi, G. & Papailias, F (2018) *Big data econometrics: nowcasting and early estimates*. Technical Report.

Buono, D., Kapetanios, G., Marcellino, M., Mazzi, G. & Papailis, F. (2017) Big data types for macroeconomic nowcasting. *EURONA - Eurostat Review on National Accounts and Macroeconomic Indicators*, 1-2017, 94–145.

Bussiere, M. & Fratzscher, M. (2006) Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25(6), 953–973.

Chen, C. & Liu, L.-M. (1993) Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421), 284–297.

Choi, N. & Varian, H. (2012) Predicting the present with google trends. *The Economic Record, Special Issue*, 88, 2–9.

Cooley, J. & Tukey, J. (1965) An algorithm for the machine calculation of complex Fourier series. *Mathematics and Computation*, 19, 297–301.

Dawood, M., Horsewood, N. & Strobel, F. (2017) Predicting sovereign debt crises: an early warning system approach. *Journal of Financial Stability*, 28, 16–28.

Diaz, J. (2016) On the origin of the signals observed across the seismic spectrum. *Earth-Science Reviews*, 106, 224–232.

Diaz, J., Ruiz, M., Crescentini, L. & Gallart, J. (2014) Seismic monitoring of an Alpine mountain river. *JGR Solid Earth*, 119(4), 3276–3289.

Diaz, J., Ruiz, M. & Jara, J.-A. (2021) Seismic monitoring of urban activity in Barcelona during COVID-19 lockdown. *Solid Earth*, 12, 725–739.

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

Duprey, T. & Klaus, B. (2017) *How to predict financial stress? An assessment of Markow switching models*. ECB Working Paper Series 2057.

Ferrara, L., Mogliani, M. & Sahuc, J.-G. (2022) High-frequency monitoring of growth at risk. *International Journal of Forecasting*, 38(2), 582–595.

Foroni, C., Marcellino, M. & Schumacher, C. (2015) Unrestricted mixed data sampling (MIDAS): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1), 57–82.

Frankel, J. & Rose, A. (1996) Currency crashes in emerging markets: an empirical treatment. *Journal of International Economics*, 41(3-4), 351–366.

Galimberti, K.J. (2020) Forecasting economic growth from outer space. *Oxford Bulletin of Economics and Statistics*, 82(4), 697–722.

Giannone, D., Reichlin, L. & Small, D. (2008) Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676.

Havskov, J. & Alguacil, G. (2016) *Seismic noise*. Cham: Springer International Publishing, pp. 101–111.

Henderson, J.V., Storeygard, A. & Weil, D.N. (2012) Measuring economic growth from outer space. *American Economic Review*, 102(2), 994–1028.

Hong, T.K., Lee, J., Lee, G., Lee, J. & Park, S. (2020) Correlation between ambient seismic noises and economic growth. *Seismological Research Letters*, 91(3), 2343–2354.

Jardet, C., & Meunier, B. (2020) *Nowcasting world gdp growth with high-frequency data*. Banque de France Working Paper 788.

Lecocq, T., Hicks, S.P., Van Noten, K., van Wijk, K., Koelemeijer, P., De Plaen, R.S.M. et al. (2020) Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures. *Science* 369(6509), 1338–1343.

Manzan, S. (2015) Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics*, 33(1), 144–164.

Mao, H., Shua, X., Ahn, Y.-Y. & Bollen, J. (2015) Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire. *EPJ Data Science*, 82(4), 4–15.

Marcellino, M. & Schumacher, C. (2010) Factor Midas for nowcasting and forecasting with ragged-edgedata: a model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.

Marcillo, O. & Carmichael, J. (2018) The detection of wind-turbine noise in seismic records. *Seismological Research Letters*, 89, 1826–1837.

Marcillo, O. & MacCarthy, J. (2020) Mapping seismic tonal noise in the contiguous united states. *Seismological Research Letters*, 91, 1707–1716.

McFadden, D. (1984) Econometric analysis of qualitative response models. In: Griliches, Z. & Intriligator, M. (Eds.) *Handbook of econometrics*, Vol. 2. Amsterdam: North-Holland.

Mcnamara, D., & Boaz, R. (2006) *Seismic noise analysis system, power spectral density probability density function: stand-alone software package*. USGS Report 2005-1438.

Mcnamara, D. & Boaz, R. (2019) Visualization of the seismic ambient noise spectrum. In: Nakata, N., Gualtieri, L. & Fichtner, A. (Eds.) *Seismic ambient noise*. Cambridge: Cambridge University Press.

Mcnamara, D. & Buland, R. (2004) Ambient noise levels in the continental united states. *Bulletin of the Seismological Society of America*, 94, 1517–1527.

McNamara, D.E., Ringler, A.T., Hutt, C.R. & Gee, L.S. (2011) Seismically observed seiching in the panama canal. *Journal of Geophysical Research*, 116, B04312.

Mi, Z., Meng, J., Guan, D., Shan, Y., Song, M., Wei, Y.-M. et al. (2017) Chinese CO2 emission flows have reversed since the global financial. *Nature Communications*, 8, 1712.

Mijling, B., van der A, R.J., Boersma, K.F., Van Roozendael, M., De Smedt, I. & Kelder, H.M. (2009) Reductions of NO2 detected from space during the 2008 Beijing Olympic Games. *Geophysical Research Letters*, 36, L13801.

O'Neel, S., Marshall, H.P., McNamara, D.E. & Pfeffer, W.T. (2007) Seismic detection and analysis of Icequakes at Columbia Glacier, Alaska. *Journal of Geophysical Research*, 112, F03S23.

Peterson, J. (1993) Observations and modeling of seismic background noise. *Geological Survey Open-File Report*, 322, 93–322.

Quaedvlieg, R. (2021) Multi-horizon forecast comparison. *Journal of Business and Economic Statistics*, 39(1), 40–53.

Ranis, G., Stewart, F. & Ramirez, A. (2000) Economic growth and human development. *World Development*, 28(2), 197–219.

Rastin, S.J., Unsworth, C.P., Gledhill, K.R. & McNamara, D.E. (2012) A detailed noise characterization and sensor evaluation of the North island of New Zealand using the PQLX data quality control system. *Bulletin of the Seismological Society of America*, 102(1), 98–113.

Riahi, N. & Gerstoft, P. (2015) The seismic traffic footprint: tracking trains, aircraft, and cars seismically. *Geophysical Research Letters*, 42(8), 2674–2681.

Rothengatter, W. (2011) Economic crisis and consequences for the transport sector. In: Rothengatter, S.W. & Hayashi, Y. (Eds.) *Transport moving to climate intelligence. Transportation research, economics and policy*. New York, NY: Springer, pp. 9–28.

Sampi, J., & Jooste, C. (2020) *Nowcasting economic activity in times of COVID-19: an approximation from the Google community mobility report*. Policy Research Working Paper 9247.

Sax, C. & Eddelbuettel, D. (2018) Seasonal adjustment by x-13arima-seats in r. *Journal of Statistical Software*, 87(11), 1–17.

Van Wijk, K., Chamberlain, C.J., Lecocq, T. & Van Noten, K. (2021) Sesmic monitoring of the Auckland volcanic field during new Zealand's COVID-19 lockdown. *Solid Earth*, 12, 363–373.

Xiao, H., Cohen Elion, Z., Ji, C. & Tanimoto, T. (2020) COVID-19 societal response captured by seismic noise in China and Italy. *Sesmological Research Letters*, 91(5), 2757–2768.

Yu, J. (2009) China's policy responses to the global financial crisis. *Richard Snape Lecture, Productivity Commission, Melbourne* (25 November).

How to cite this article: Tiozzo Pezzoli, L. & Tosetti, E. (2022) Seismonomics: Listening to the heartbeat of the economy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(Suppl. 2), S288–S309. Available from: https://doi.org/10.1111/rssa.12912