Supplementary Material to "The impact of COVID-19 on subjective well-being: evidence from Twitter data"

TIZIANA CARPI[1], AIRO HINO[2], STEFANO MARIA IACUS*[3], AND GIUSEPPE PORRO[4]

[1]DEPARTMENT OF STUDIES IN LANGUAGE MEDIATION AND INTERCULTURAL COMMUNICATION, UNIVERSITY OF MILAN, PIAZZA INDRO MONTANELLI 1, 20099 MILAN, ITALY.

[2]SCHOOL OF POLITICAL SCIENCE AND ECONOMICS, WASEDA UNIVERSITY, 1-6-1 NISHIWASEDA SHINJUKU-KU, TOKYO 169-8050, JAPAN.

[3]INSTITUTE FOR QUANTITATIVE SOCIAL SCIENCE, HARVARD UNIVERSITY, 1737 CAMBRIDGE ST, CAMBRIDGE, MA 02138, USA.

[4]DEPARTMENT OF LAW, ECONOMICS AND CULTURE, UNIVERSITY OF INSUBRIA, VIA SANT'ABBONDIO 12, 22100 COMO, COMO, ITALY.

Construction of the Twitter indicators

The interested readers can find full details on the way SWB-I and SWB-J are constructed through a supervised sentiment analysis approach, called ISA (Ceron et al., 2016), in the following already published papers: Iacus et al. (2019), Iacus et al. (2020a), Iacus et al. (2020b) for the Italian indicator and Carpi et al. (2022) for the Japanese case.

Stochastic analysis

In order to capture the trend of the two time series SWB-I and SWB-J we use a dynamical system approach inherited from finance. We model each SWB-I/J series via a stochastic differential equation (SDE) model (Iacus, 2008; Iacus and Yoshida, 2018). Without going too much into details, a SDE is a mathematical way to describe a continuous time dynamical system subject to noise. If X_t denotes the value of the time series at time t and $\mathrm{d}X_t = X_{t+\mathrm{d}t} - X_t$ the increment of the series in the short time interval $[t, t + \mathrm{d}t)$, the SDE is written as follows:

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + a(X_t)\mathrm{d}W_t, \quad X_0 = x_0,$$

where $\mathrm{d}W_t$ is a the increment of the Gaussian noise (also known as the Brownian motion or Wiener process) in the same time interval, such that $\mathrm{d}W_t \sim N(0, \mathrm{d}t)$. The function $b(\cdot)$ is called the *drift* coefficient, the function $a(\cdot)$ is called the *diffusion* coefficient and x_0 is some initial condition. As the two time series exhibit a negative trend in the first half of 2020 and then seem to oscillate around a new mean value, we assume a drift coefficient of mean-reverting type of the form: $b(X_t) = \alpha(\beta - X_t)$ where β represents the long term mean around which the time series X_t oscillates and α is called the *speed* of mean reversion: the higher α, the faster the process is pushed back to (or attracted towards) its long run mean. We will consider four SDE candidate models with the same drift coefficient but different diffusion coefficients and by model selection we will establish which model fits better which time series. The first one is the so called Ornstein-Uhlenbeck (Uhlenbeck and Ornstein, 1930) or Vasicek (VAS) model (Vasicek, 1977), which is usually interpreted as the continuous-time version of the AR(1) model (for $\beta = 0$):

$$\mathrm{d}X_t = \alpha(\beta - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 = x_0 \tag{S1}$$

where $\sigma > 0$ is the scaling factor and x_0 is some initial condition. In this case $a(X_t) = \sigma$, simply. The second one is the geometric Brownian motion (GBM) (Black and Scholes, 1973; Merton, 1973) model:

$$\mathrm{d}X_t = \alpha(\beta - X_t)\mathrm{d}t + \sigma X_t \mathrm{d}W_t, \quad X_0 = x_0 \tag{S2}$$

where the term $\sigma X_t \mathrm{d}W_t$ represents the feedback in the system, in the sense that the noise $\mathrm{d}W_t$ interacts with the state of the system X_t, i.e., the larger X_t, the higher the impact of the noise.

*Corresponding author. Email: siacus@iq.harvard.edu

In this case $a(X_t) = \sigma X_t$. The third model is the CIR (Cox-Ingersol-Ross) model (Cox et al., 1985) which differs from the previous model for the $\sqrt{X_t}$ term in the diffusion coefficient, i.e., $a(X_t) = \sigma\sqrt{X_t}$. The role of the square root is to dump a bit the feedback effect:

$$\mathrm{d}X_t = \alpha(\beta - X_t)\mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t, \quad X_0 = x_0. \tag{S3}$$

And finally, because we are uncertain about the amount of the dumping effect in the diffusion term, we also consider the CKLS (Chan, Karolyi, Longstaff and Sanders) model (Chan et al., 1992) adding an exponent $0 < \gamma < 2$ to the diffusion coefficient, i.e., $a(X_t) = \sigma X_t^\gamma$:

$$\mathrm{d}X_t = \alpha(\beta - X_t)\mathrm{d}t + \sigma X_t^\gamma \mathrm{d}W_t, \quad X_0 = x_0 \tag{S4}$$

Clearly, CKLS embeds VAS (for $\gamma = 0$), CIR (for $\gamma = 0.5$) and GBM (for $\gamma = 1$). The main characteristics of the first three models is that VAS model has Gaussian increments, GBM has log-Normal increments and CIR has non-central Chi-Squared increments.

We fitted SDE models presented in the above on the two SWB-I and SWB-J time series thorough the `yuima` R package (Brouste et al., 2014; Iacus and Yoshida, 2018) via quasi-maximum likelihood estimation. The results of the estimation are given in Table S1.

Table S1: Fitting different SDE models. Standard errors of the estimates are in parentheses and $(\cdot)$ means estimates of variance-covariance matrix did not converge.

Index	α	β	σ	γ	Model	AIC
SWB-I	3.16	38.99	14.7		VAS	787.0
	(2.41)	(6.22)	(0.56)			
	3.57	39.54	0.33		GBM	782.9
	(2.46)	(4.60)	(0.01)			
	3.34	39.28	2.20		CIR	781.9
	(2.42)	(5.35)	(0.08)			
	3.42	39.37	1.12	0.68	CKLS	783.6
	(2.44)	(5.08)	$(\cdot)$	$(\cdot)$		
SWB-J	12.92	28.44	26.25		VAS	1090.0
	(5.52)	(2.19)	(1.05)			
	11.46	28.43	0.83		GBM	1027.4
	(5.71)	(2.23)	(0.03)			
	11.98	28.42	4.62		CIR	1055.1
	(5.61)	(2.21)	(0.19)			
	11.64	28.41	0.05	1.84	CKLS	1010.7
	(5.93)	(2.22)	(0.01)	(0.08)		

It seems clear that the models agree on the α and β parameters, meaning that the drift term is correctly identified. Indeed, β correspond to about 39 for SWB-I and 28 for SWB-J across models with about 11 percentages points of difference between the two countries, and α varies in the interval [3.16,3.57] for SWB-I and in the interval [11.46,12.92] for SWB-J, meaning that the convergence to the long run mean of the SWB index is between 3 and 4 times faster in Japan than in Italy. Figure 1 in the main corpus of the manuscript shows also the estimated limit deterministic dynamical systems obtained taking the limit as $\sigma \to 0$ of the stochastic differential equation models which best fit, in terms of AIC, the set of data (respectively, CIR for SWB-I and CKLS for SWB-J). The limiting dynamical system is a way to represent the structural component of each time series in this stochastic analysis approach. What this analysis shows is that the decline of subjective well-being in Italy is much more emphasized, in 2020, compared to what happens in Japan whilst the convergence to the long run mean is faster for Japan and that their two structural components are different.

In summary, this analysis confirms the drop of the indicators during 2020, despite the high variability of the indicators themselves. We do attribute the daily fluctuations to the emotional component of the well-being as captured by the social media indicators, and the trend value to the structural component of well-being. One might be tempted to assume this limiting dynamical value as an estimate of life evaluation or life evaluation. As we previously emphasized, life satisfaction is not simply the average of the emotional, short run, values of subjective well-being. We can however propose β as a rough proxy of life evaluation, considering that β is an estimate of the long run trend of the process: to some extent, β summarizes the path of the process in the past and captures the value the process is going to assume in the long run although, in this specific COVID-19 application, we don't know if this new state is persistent or just temporary.

List of external factors

Table S2: Google search topics used in the analysis.

Search term	country	keyword type	framework	type of search
Coronavirus	IT/JP	topic	pandemic	web
CoronavirusNews	IT/JP	topic	pandemic	news
(コロナ) Corona	JP	topic	pandemic	web
(コロナ) CoronaNews	JP	topic	pandemic	news
Covid	IT/JP	topic	pandemic	web
CovidNews	IT/JP	topic	pandemic	news
Rt	IT	keyword	pandemic	web
Wuhan	IT/JP	topic	pandemic	web
Unemployment	IT/JP	topic	economics	web
UnemploymentNews	IT/JP	topic	economics	news
Economy	IT/JP	topic	economics	web
EconomyNews	IT/JP	topic	economics	news
GDP	IT/JP	topic	economics	web
GDPNews	IT/JP	topic	economics	news
Depression	IT/JP	topic	health	web
Stress	IT/JP	topic	health	web
Insomnia (disorder)	IT/JP	topic	health	web
Health	IT/JP	topic	health	web
Solitude	IT/JP	topic	health	web
AdultContent	IT/JP	topic	leisure	web

Dynamic Elastic Net

In our setup, the Elastic Net method is a penalized least squares method which adds L_1 and L_2 penalization terms to the classical objective function of the OLS (ordinary least squares), i.e., it corresponds to the optimization problem in (S5) below:

$$\underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{k} \beta_j^2 + \alpha \sum_{j=1}^{k} |\beta_j| \right) \right\}. \tag{S5}$$

where y_i is the dependent variable (either SWB-I or SWB-J), and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})'$ is the vector of covariates for unit i in the sample of n observations, β_j's are the regression coefficients, $\lambda > 0$ is a penalization factor and $\alpha \in [0, 1]$ is a tuning parameter. For $\alpha = 1$ this method corresponds to the classical LASSO (Least Absolute Shrinkage and Selection Operator) algorithm (Tibshirani, 1996), while for $\alpha = 0$ it corresponds to the Ridge estimation (Hoerl and Kennard, 1970). Loosely speaking, while the LASSO method tends to estimate as zero as much coefficients as possible and, in case of multicollinearity, selects arbitrarily one single variable in a set of correlated covariates, the Ridge regression is able to accommodate for the multicollinearity by keeping the correlated variables and "averaging" the estimated coefficients. For exactly this reason, Elastic Net include both L_1 (LASSO) and L_2 (Ridge) penalty terms. Usually LASSO is used to succinctly explain the correlation effects

Table S3: Times series used in the analysis. In the analysis Japanese variables start with j and the Italian ones with i, e.g., `iCases`, `jCases`.

Variable	area	Source
SWB-I, SWB-J	subjective well-being	Twitter
Solitude	mental health	Google Trends
Depression	mental health	Google Trends
Stress	mental health	Google Trends
Insomnia	mental health	Google Trends
Health	health	Google Trends
PM2.5	environment	WAQI
Temperature	environment	WAQI
Cases	pandemic	WHO
Deaths	pandemic	WHO
Coronavirus, CoronavirusNews	pandemic	Google Trends
(コロナ) Corona, CoronaNews	pandemic	Google Trends
Covid, CovidNews	pandemic	Google Trends
Rt	pandemic	Google Trends
Wuhan	pandemic	Google Trends
Unemployment, UnemploymentNews	economy	Google Trends
Economy, EconomyNews	economy	Google Trends
GDP, GDPNews	economy	Google Trends
FTSEMIB	economy	Yahoo! Finance
Nikkei	economy	Yahoo! Finance
AdultContent	leisure	Google Trends
FB.CLI	health	Facebook
FB.ILI	health	Facebook
FB.MC	behavioural	Facebook
FB.DC	behavioural	Facebook
FB.FH	economy	Facebook
Residential	human mobility	Google
Workspace	human mobility	Google

and Ridge is more suitable for forecasting. Overall, these methods are biased (although adaptive versions of Elastic Net also exist) but - with low variance (as they are also shrinkage methods) and hence most of the time - the Elastic Net performs quite well in terms of mean squared error compared to classical OLS. The value of $\alpha = 0.5$, usually denotes the proper Elastic Net. The penalty term λ is a tuning parameter and it is chosen through cross-validation methods. In this study we make use of the package `glmnet` developed by Friedman et al. (2010), which is computationally efficient and state of the art for this technique, to run several times the Elastic Net model. In particular, instead of a simple regression, we estimate a one step ahead forecasting model of this form:

$$\operatorname*{argmin}_{\beta} \left\{ \frac{1}{2\cdot 30} \sum_{d=t-29}^{t} (y_d - x'_{d-1}\beta)^2 + \lambda_t \left(\frac{1-\alpha}{2} \sum_{j=1}^{k} \beta_j^2 + \alpha \sum_{j=1}^{k} |\beta_j| \right) \right\}, \tag{S6}$$

with $t =$ 2020-12-02, $\ldots$, and λ_t is calculated for each varying time t through cross-validation minimizing the mean squared error of the forecast.

In the set of covariates we also include the lagged value of the index (`swbLag`) and compare the forecasting against an ARIMA(1,0,1) model. We tested the average mean squared error of the forecast against that of the ARIMA(1,0,1) and they are quite close, with the latter being slightly better: the average MSE is 0.01734827 for Elastic Net and 0.0152734 for ARIMA(1,0,1) for the Italian data and 0.08026962 and 0.06737153 respectively for the Japanese data. Clearly the ARIMA(1,0,1) model does not include any covariate but takes advantage of modeling the serial correlation, whilst Elastic Net in (S6) essentially assumes independent observations - obviously a simplification of the reality - but it allows for the inclusion of explanatory variables. Fig. S1 shows the relative performance of the Elastic Net compared to the ARIMA(1,0,1) model for both SWB-I and SWB-J. Although we are not interested in forecast *per se*, these plots show that the fitting of the Elastic

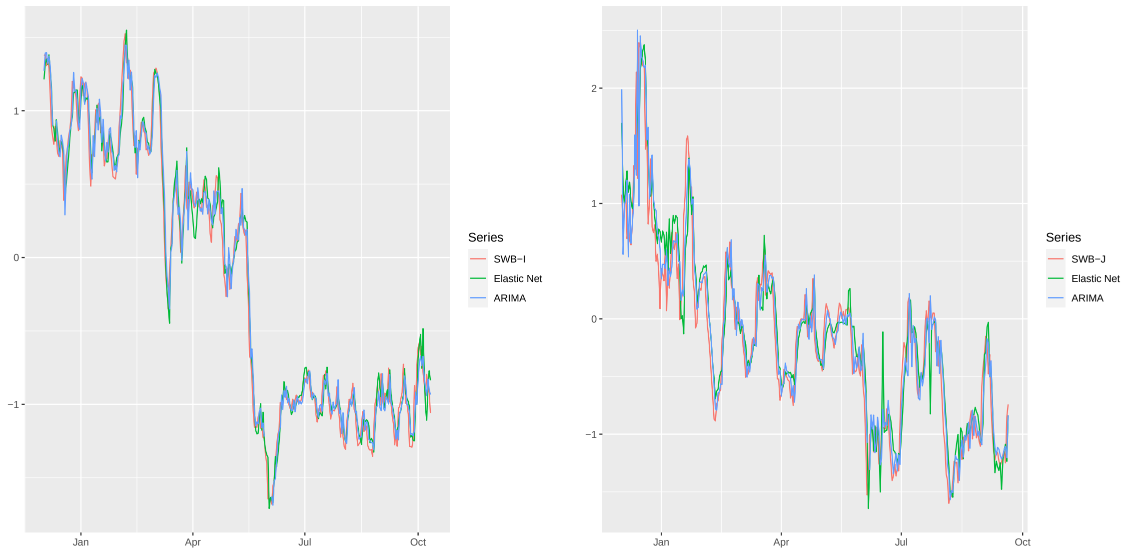

Figure S1: Forecast performance of Elastic Net and ARIMA(1,0,1). Standardized data.

Net model is not only accurate but also avoids over fitting. For completeness, we also tested the three values of $\alpha = 0$, 0.5 and 1 and obtained that, as expected, $\alpha = 0.5$ gives the best performance. A final remark: as the real scope of this study is to evaluate the impact of the covariates on the SWB indexes, the actual values of the forecasts are not important, so we standardize the data, as Elastic Net works better with standardized data. Moreover, as the standard error of the Elastic Net estimates is quite difficult to obtain and it is not a typical outcome of the procedure, via the standardized data we also have standardized coefficients which make the evaluation of the impact of each variable easier to interpret (see Figures S2 and S3).

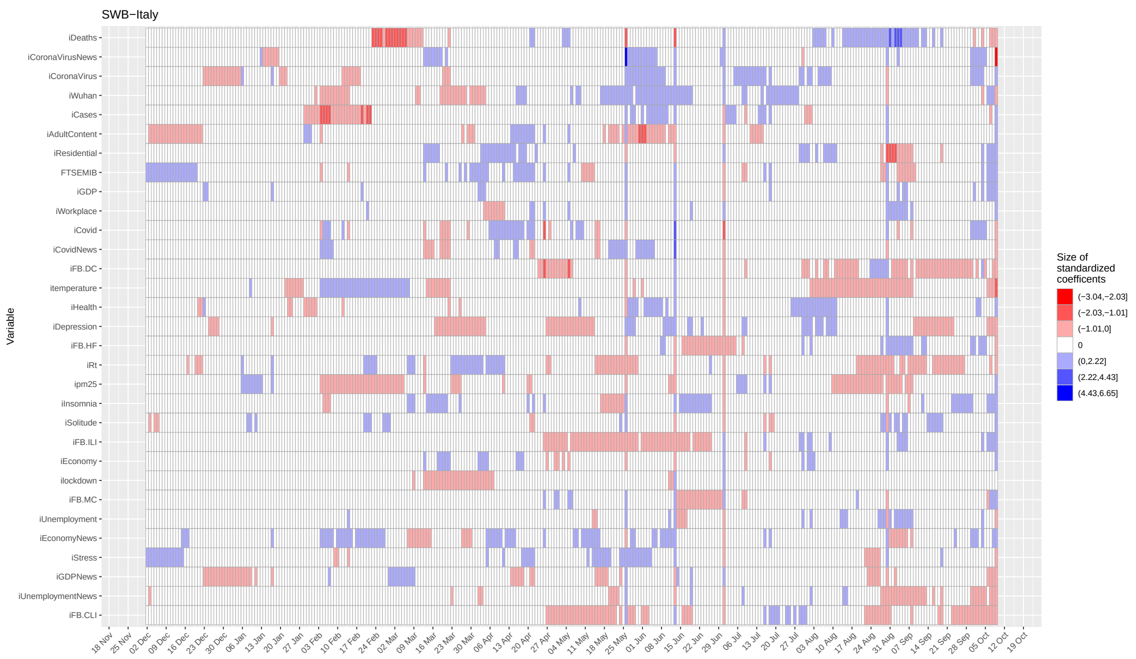

Figure S2: **Sign of coefficients of SWB-I analysis.** Standardized coefficients of the covariates selected by the Elastic Net for SWB-I through time.

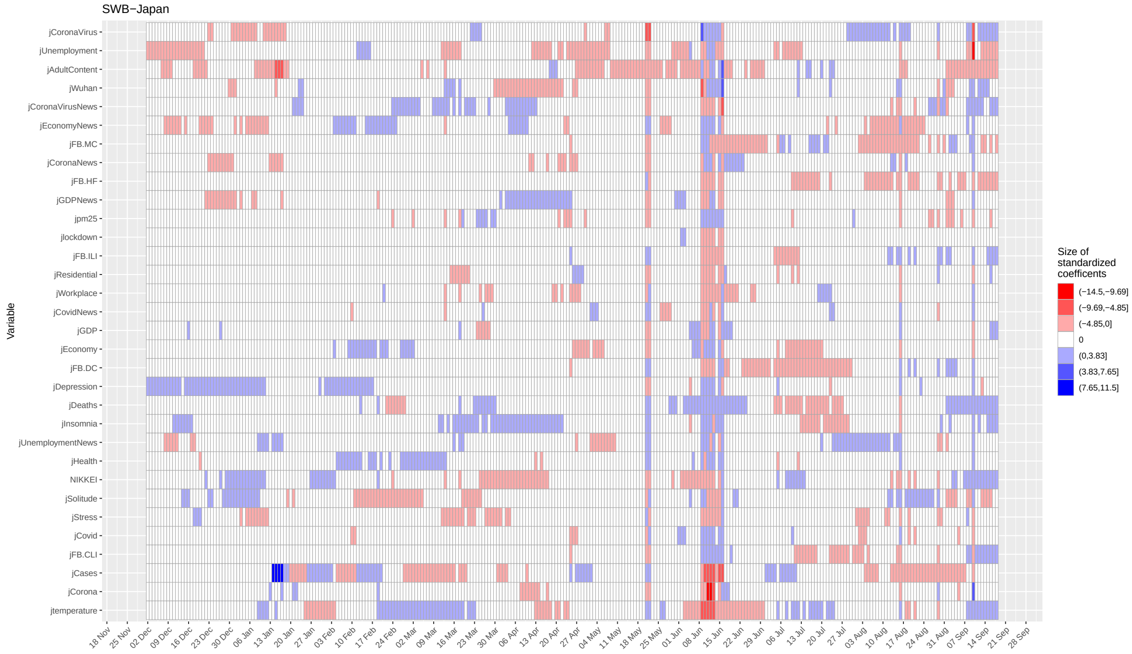

Figure S3: **Sign of coefficients of SWB-J analysis.** Standardized coefficients of the covariates selected by the Elastic Net for SWB-J through time.

Dynamic variable selection for the SWB-I indicator

The application of the dynamic elastic net method shows that (see also Figure S2), the variable `iDeaths` has a negative impact on SWB-I from mid February till the end of March (at the peak of the first wave), then it appears isolated on the 25th of May and again at the end of September and early October, when the second wave restarted. Interestingly enough, `iCases` shows up much earlier around the end of January (notice that all flights from China to Italy were stopped on January 21st) and vanishes as `iDeaths` pops up. Then, its impact is again positive from mid May to mid June, when the good news of the decreasing in number of cases arrived. A similar pattern is exhibited by the `iWuhan` variable. It is worth noting that the `iRt` variable becomes negative by the end of April and quite persistent starting from the summer. This is clearly a media effect, as the notion of R_t itself became popular within the general public only in July, when both opinion makers and virologists started discussing about the R_t indicator also on prime time tv shows.

The variable `ilockdown` and `iDepression` appear in mid March (the lockdown in Italy started officially on March 15th), and the latter has also a persistently negative impact between the end of April and mid May. Notice that, the variable `iDepression` looks way more important than `iSolitude`, `iInsomnia` and `iStress` in explaining the SWB-I indicator (see also Figure 4 in the main manuscript).

The Facebook survey variables `iFB.ILI` and `iFB.CLI` appear from their beginning (April 27th) and their impact remain almost always negative (apart around the last weeks of July), meaning that the symptoms clearly impacted negatively the reported well-being.

The variable `iUnemploymentNews` has mostly a negative impact from August, i.e. when the largest economic impact on several jobs, including seasonal jobs, was registered and frequently reported in the news. It can be noted, moreover, that the `iUnemployment` effect is much less sharp, suggesting a higher incidence of the labor market perspectives, with respect to the impact of the current employment conditions.

The `itemperature` variable has a positive impact in winter (which has been mild in Italy) and negative in summer, which is somewhat confirmed by common experience, and the air pollution variable `ipm25` has almost always a negative impact, when it is significant.

The mobility variables also show interesting patterns: `iResidential` has often a positive impact except in August, when staying at home was quite challenging, while `iWorkplace` had a negative impact around the last week of March and the first week of April, then generally positive when the mobility constraints were lifted in most regions.

The `iAdultContent` variable is always negatively correlated with the well-being, confirming that the consumption of these products are closely related to unsatisfactory levels of happiness.

The impact of economy related news on the indexes are somehow questionable as they switch sign or are scattered along the time axis. This may indicate that these effects are extremely volatile and mainly determined by the correspondent news, as we may infer from `iGDP` and `iGDPNews`: in fact, while `iGDP` has sometimes a positive impact, the expected negative role seems played by `iGDPNews`. On the other hand, one may think that the impact of the economic factors is a long-run one, and hence is not properly captured by a short-term SWB indicator.

Dynamic variable selection for the SWB-J indicator

The first wave of the pandemic started in Japan with the beginning of 2020, while - after a slowdown - a second wave arose in June. For the Japan data, the variable `jCases` is quite often present from January till May, then in June and August, and with almost always negative impact, whilst the `jDeaths` variable seems to be selected much less frequently and with alternate sign (see Figure S3). This is probably because the number of deaths in Japan is way less than the number of cases. Also, the cases appeared in Japan much earlier than in Italy, and this is reflected by the selection of the variables.

Psychological aspects captured by the Google search `jStress` and `jSolitude` are mostly present with negative correlation while, surprisingly enough, `jInsomnia` and `jDepression` seems to be positively correlated with the SWB-J index, and the latter only up to mid February. Notice further that, the variable `jStress` looks way more important than `jSolitude`, `jInsomnia` and `jDepression` in explaining the SWB-J indicator (see also Figure 4 in the main manuscript).

The economic variables - stock market index `NIKKEI`, `jEconomy`, `jFB.HF` and `jGDPNews` - are often selected and with negative impact on subjective well-being: in particular, labor market related covariates (`jUnemployment`, `jUnemploymentNews`) frequently appear, even with alternate sign, to measure concerns for current and future employment perspectives.

As expected, `jAdultContent` is negative when present, and seems to be relevant before the start of the pandemic and in between the two waves mostly. COVID-19 symptoms, as registered through the Facebook survey, seem to be less important apart for the period mid July - mid August, when `jFB.CLI` appears with negative sign.

The mobility variables `jWorkplace`, `jResidential` and `jlockdown` appear to have a negative impact but they are not very persistent in time like it happens for the Italian case. In fact, the mobility restrictions in Italy were much stronger and this may be the reason of this different impact. On the contrary, the social distancing proxies `jFB.MC` and `jFB.DC` appear and remain persistent during the second wave of the pandemic, and have both a negative impact on the SWB-J index.

The search terms related to Covid or coronavirus are also scattered and with alternate sign, showing a volatile relationship with well-being, not so different from what one may observe in the Italian case.

The variable `jtemperature` behaves similarly to the Italian case: it has often a positive impact in winter/spring and negative in June and August, and again positive in the first half of July. The air quality proxy `jpm2.5` is rarely significant, compared to the Italian case.

Neither the postponement of the Olympic Games 2020, at the end of March, nor the resignation by the prime minister Shinzo Abe (end of August) seem to be associated with the selection of specific covariates. On the other hand, the well-being reaction to the second wave outbreak in June is accompanied by a selection of all the covariates, some of which (`jCases`, `jCorona`) with a strongly negative impact.

Structural equation model

This section contains the results of the fitted SEM models for SWB-I and SWB-J. The results are summarised in Table S4 for Italy and Table S5 for Japan. Figures 7 and 8 in the main manuscript give a graphical representation of the same fitting. The models have been fitted using the `lavaan` package (Rosseel, 2012) and plots have been generated through the `semPlot` package (Epskamp, 2019).

Table S4: **The fitting results of the SEM model for the Italian data.**

Relationship			Coefficient	Std.Err.
VirusSearch	$\mapsto$	iCoronaVirus	0.226***	0.043
VirusSearch	$\mapsto$	iCovid	−0.638***	0.063
VirusSearch	$\mapsto$	iRt	−0.372***	0.057
PsySearch	$\mapsto$	iStress	0.602***	0.053
PsySearch	$\mapsto$	iInsomnia	0.410***	0.055
PsySearch	$\mapsto$	iSolitude	1.003***	0.047
PsySearch	$\mapsto$	iDepression	−0.186***	0.056
HealthStatus	$\mapsto$	iFB.CLI	0.929***	0.047
HealthStatus	$\mapsto$	iFB.ILI	1.010***	0.044
Mobility	$\mapsto$	iResidential	0.791***	0.033
Mobility	$\mapsto$	iWorkplace	−0.726***	0.039
Mobility	$\mapsto$	ilockdown	0.762***	0.036
Finance	$\mapsto$	FTSEMIB	−0.778***	0.052
Finance	$\mapsto$	iFB.HF	0.521***	0.053
Finance	$\mapsto$	iUnemployment	0.421***	0.051
SocDist	$\mapsto$	iFB.MC	0.882***	0.047
SocDist	$\mapsto$	iFB.DC	0.976***	0.044
WellBeing	$\mapsto$	SWB-I	0.248***	0.019
WellBeing	$\leftarrow\!\shortmid$	VirusSearch	0.307***	0.075
WellBeing	$\leftarrow\!\shortmid$	HealthStatus	−0.167***	0.085
WellBeing	$\leftarrow\!\shortmid$	Mobility	−0.290***	0.124
WellBeing	$\leftarrow\!\shortmid$	Finance	−0.524***	0.101
WellBeing	$\leftarrow\!\shortmid$	SocDist	−3.345***	0.336
PsySearch	$\leftarrow\!\shortmid$	WellBeing	0.065***	0.016
iAdultContent	$\leftarrow\!\shortmid$	WellBeing	−0.136***	0.016
PsySearch	cov	Mobility	0.003	0.008
Mobility	cov	SocDist	−0.412***	0.049
Mobility	cov	iCases	0.356***	0.034
VirusSearch	cov	Mobility	−0.293***	0.081
VirusSearch	cov	Finance	−1.227***	0.091
VirusSearch	cov	SocDist	−1.089***	0.072
HealthStatus	cov	Mobility	−0.002	0.057
HealthStatus	cov	Finance	0.230***	0.069
HealthStatus	cov	SocDist	0.314***	0.055
Mobility	cov	Finance	0.850***	0.033
Finance	cov	SocDist	0.305***	0.069
PsySearch	cov	iAdultContent	0.608***	0.046

Note: *p<0.1; **p<0.05; ***p<0.01

Table S5: **The fitting results of the SEM model for the Japanese data.**

Relationship			Coefficient	Std.Err.
VirusSearch	↦	jCoronaVirus	0.867***	0.050
VirusSearch	↦	jCovid	0.204***	0.061
VirusSearch	↦	jCorona	0.714***	0.054
PsySearch	↦	jStress	0.401***	0.055
PsySearch	↦	jInsomnia	0.749***	0.049
PsySearch	↦	jSolitude	0.763***	0.049
PsySearch	↦	jDepression	0.545***	0.052
HealthStatus	↦	jFB.CLI	1.016***	0.042
HealthStatus	↦	jFB.ILI	0.947***	0.045
Mobility	↦	jResidential	1.071***	0.039
Mobility	↦	jWorkplace	−0.902***	0.046
Mobility	↦	jlockdown	0.708***	0.050
Finance	↦	NIKKEI	0.836***	0.050
Finance	↦	jFB.HF	−0.441***	0.054
Finance	↦	jUnemployment	−0.694***	0.053
SocDist	↦	jFB.MC	0.959***	0.044
SocDist	↦	jFB.DC	0.903***	0.047
WellBeing	↦	SWB-J	0.674***	0.036
WellBeing	↤	VirusSearch	−0.368***	0.134
WellBeing	↤	HealthStatus	−0.617***	0.240
WellBeing	↤	Mobility	−0.077	0.121
WellBeing	↤	Finance	0.052	0.165
WellBeing	↤	SocDist	−1.795***	0.298
PsySearch	↤	WellBeing	−0.342***	0.052
jAdultContent	↤	WellBeing	−0.456***	0.039
PsySearch	cov	Mobility	0.002	0.023
Mobility	cov	SocDist	0.087	0.054
Mobility	cov	jCases	0.235***	0.030
VirusSearch	cov	HealthStatus	0.536***	0.049
VirusSearch	cov	Mobility	0.192***	0.056
VirusSearch	cov	Finance	−0.947***	0.031
VirusSearch	cov	SocDist	−0.566***	0.051
HealthStatus	cov	Mobility	−0.278***	0.047
HealthStatus	cov	Finance	−0.269***	0.062
HealthStatus	cov	SocDist	−0.928***	0.010
Mobility	cov	Finance	−0.676***	0.034
Finance	cov	SocDist	0.450***	0.059
PsySearch	cov	jAdultContent	0.247***	0.047

Note: *p<0.1; **p<0.05; ***p<0.01

Data Availability

All data and R scripts are available here https://github.com/siacus/swbCovid.

Authors' contribution

SI and GP conceived the project. TC, AH, SI and GP designed the study. TC and AH wrote the intercultural and socio-linguistic analysis. SI and GP did the statistical analyses of the data. TC, AH, SI, and GP interpreted the results and wrote and revised the article.

References

Black F, Scholes M (1973). The pricing of options and corporate liabilities. Journal of Political Economy., 81: 637–65.

Brouste A, Fukasawa M, Hino H, Iacus S, Kamatani K, Koike Y, et al. (2014). The yuima project: A computational framework for simulation and inference of stochastic differential equations. Journal of Statistical Software, Articles, 57(4): 1–51. https://www.jstatsoft.org/v057/i04.

Carpi T, Hino A, Iacus SM, Porro G (2022). A Japanese Subjective Well-Being Indicator Based on Twitter Data. Social Science Japan Journal, 25(2): 273–296.

Ceron A, Curini L, Iacus SM (2016). iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content. Information Science, 367(C): 105‑–124.

Chan KC, Karolyi GA, Longstaff FA, Sanders AB (1992). An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance, 47(3): 1209–1227.

Cox J, Ingersoll J, Ross S (1985). A theory of the term structure of interest rates. Econometrica, 53: 385–408.

Epskamp S (2019). semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. CRAN. R package version 1.1.2.

Friedman J, Hastie T, Tibshirani R (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, Articles, 33(1): 1–22.

Hoerl AE, Kennard RW (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1): 55–67.

Iacus SM (2008). Simulation and Inference for Stochastic Differential Equations. With R Examples. New York: Springer.

Iacus SM, Porro G, Salini S, Siletti E (2019). Social networks data and subjective well-being. an innovative measurement for Italian provinces. Scienze Regionali, Italian Journal of Regional Science, 2019(Speciale): 667–678.

Iacus SM, Porro G, Salini S, Siletti E (2020a). Controlling for selection bias in social media indicators through official statistics: a proposal. Journal of Official Statistics, 36(2): 315–338.

Iacus SM, Porro G, Salini S, Siletti E (2020b). An Italian composite subjective well-being index: The voice of Twitter users from 2012 to 2017. Social Indicators Research, 161: 1–31.

Iacus SM, Yoshida N (2018). Simulation and inference for stochastic processes with YUIMA: a comprehensive R framework for SDEs and other stochastic processes. Springer, New York.

Merton R (1973). Theory of rational option pricing. Bell J. Econ. Manage., 4(1): 141–183.

Rosseel Y (2012). lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48(2): 1–36.

Tibshirani R (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society (Series B), 58: 267–288.

Uhlenbeck G, Ornstein L (1930). On the theory of Brownian motion. Phys. Rev., 36: 823–841.

Vasicek O (1977). An equilibrium characterization of the term structure. Financial Econ., 5: 177–188.