

## Research data recycling through open sharing and reuse: A case study of sustainable digital good consumption in the sharing economy

*Ibrahim Niankara*

College of Business, Al Ain University, P.O. Box: 112612, Abu Dhabi, UAE; Tel: +97124444696, Ext: 539; Fax: +97124444304; Email: [Ibrahim.niankara@aau.ac.ae](mailto:Ibrahim.niankara@aau.ac.ae)

---

### Abstract:

In order to meet the needs of an increasingly complex research landscape, researchers engage in “collaborative presumption” through open data sharing and reuse. Although significant gains have been achieved in this regards because of growing requirements from funding agencies, governments and journals, the question of how reuse of openly available data for new research contribute to sustainability is yet to be appropriately addressed in the literature.

Therefore, relying on a three stage stratified clustered random sampling of the Journal of Applied Econometrics data archive (JAEDA), the present research provides a case study of the value of research data recycling for sustainable research and economic development. More specifically our analysis show that reformatting from wide to long format, openly shared equity price index data on eleven European countries’ extracted from JAEDA, and augmented with country level geospatial Meta data, provides a new basis for interesting descriptive analytics and spatio-temporal econometric modeling and inference.

Given the ever-increasing volume of openly available research data, our study provides a first-hand insight on open data reuse, which should benefit all stakeholders in the research community, as they seek sustainable solutions for scientific productivity and progress.

**Keywords:** *Collaborative consumption; Data sharing and reuse; Data recycling; Digital assets; United nations SDGs; Sustainability; Sustainable Development; Sustainable scholarship*

---

## Introduction

*Although there is a plethora of literature data on the value of recycling in general by scholars, there is a paradoxical paucity of literature on the value of data recycling in particular among scholars.*

With the recent developments in computational simulation and modeling, automated data acquisition and communication technologies the concept of data has evolved to be more inclusive (Curty et al, 2017; Chawinga & Zinn, 2019). In addition to digital manifestations of literature, data refers as well to forms of databases generally requiring the assistance of computational machinery and software in order to be useful (Zhang et al, 2017). Consequently, the object of data has been rendered a key infrastructural component of science (Nielsen, 2020), at the root of a new era of research called “the fourth paradigm: data-intensive scientific discovery” (Tolle et al, 2011; Pasquetto et al, 2017; Wu et al, 2019). Cornerstone of the fourth industrial revolution, and a follow up to the previous research paradigms (including theoretical and experimental), the fourth paradigm seeks to have all of science literature, and all of science data openly inter-operating with each other online (Borst & Limani, 2020). Hence the emergence of “open research data” (Mancini et al, 2020), which has been defined as structured, machine-readable, data that is actively published on the internet for public re-use, and that is freely accessible, usable, modifiable, and sharable by academic researchers (Allen & Mehler, 2019; Mayernik, 2017; Powers & Hampton, 2019; Turki et al, 2019; Zuiderwijk & Hinnant, 2019).

In this context, “Data sharing” is better understood as a problem of scarce research resources optimization. Indeed, researchers typically produce large amount of data, part of which may be useful to others for the reproduction and verification of previous research results (Hardwicke et al, 2019; Powers & Hampton, 2019), the formulation of new research questions (Boté & Térmens, 2019), and the advancement of research and innovation (Borgman et al, 2019). However, making potentially useful data available to others requires substantial investment, in documentation and data curation, beyond the actual conduct of the initial research for which the data was primarily collected (Wallis et al, 2013). Economically, this creates a misalignment in incitation, since the originating investigator bears the full cost (time and efforts) of making the data available openly, while unknown and often non-existent re-users reap the benefits. From the perspective of the researcher contemplating sharing, a valid question then is:

*“Which data are worth the investment effort, as they might get reused by others?”*

Given that specific reuse intentions by other researchers only arise when the data are discoverable, accessible and usable (Stall et al, 2019), one can ask an equally valid question:

*How can demand of data reuse be built if potential users of the data are unaware of its existence and utility?*

Hence the need for studies (including data descriptors) that showcase the re-use potential of various data sources, so as to raise awareness of their utility for prospective investigations. Moreover, pointing out the potential collaborative effort needed to address data sharing and reuse for the well-being of all stakeholders in any economic system Hanson et al, (2011) declared:

*“We must all accept that science is data and that data are science, and thus provide for, and justify the need for the support, of much-improved data curation”*; while Borgman (2012) suggested:

*“if the rewards of the data deluge are to be reaped, then researchers who produce those data must share them and do so in such a way that the data are interpretable and reusable by others”*.

The present research therefore inscribes itself in this dynamic. Although previous authors have provided insights about the motivating factors of open data sharing and re-use in other fields including archaeobotany (Lodwick, 2019), astrophysics (Zuiderwijk & Spiers, 2019), biomedical (Park et al, 2018; Heller et al, 2019), ecology (Zimmerman, 2008), psychology (Houtkoop et al, 2019), and science in general (Chawinga & Zinn, 2019; Federer et al, 2018; Tenopir et al, 2015), the economics literature is yet to properly fill this void. Indeed, although Waugh (2010) discusses the value of digital assets for sustainable economic development, to date a limited literature covers the specific value of data sharing and re-use within the field of economics (Cavallo & Rigobon, 2016; Mullainathan & Spiess, 2017). An even fewer number addresses the idea of open research data sharing and re-use for sustainable economic development (Hilbert, 2016; Jetzek et al, 2019).

*Therefore the present study aims at filling this gap in the scientific discourse, by investigating the potential of the Journal of Applied Econometrics data archive to contribute to sustainable scientific research and economic development.*

As a case study of the sharing economy, within the specific context of digital goods/assets sharing and re-using, the present study is also in line with the United Nations (UN) global agenda for Sustainable Development. Indeed, goal 12 of the UN sustainable development goals (SDGs) is broadly concerned with ensuring sustainable consumption and production patterns by 2030; while its target 12.5 specifically addresses the substantial reduction of waste generation through prevention, reduction, recycling, and reuse. In addition, SDGs target 12.8 is specifically concerned with ensuring that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature; while SDGs target 12.a looks at supporting developing nations to strengthen their scientific and technological capacity to move towards more sustainable patterns of consumption and production. Given that open data sharing in general has the potential to:

- (i) Facilitate the use of digital assets for scientific (academic) research
- (ii) Contribute to the expansion of the market for sharable “digital assets”
- (iii) Encourage multiple perspectives, discourage fraud, help identify errors
- (iv) Provide useful information for training new researchers, increase efficient use of funding and respondent population resources by avoiding duplicate data collection

The key question that will be addressed in the present study in line with our above research objective is:

*How can data sharing and re-use through the Journal of Applied Econometrics data archive contribute to the promotion of sustainable scientific research and economic development?*

To address this question, we organize the rest of the paper as follows: in section 2 we review the literature on open data sharing and re-use; in section 3 we present the adopted methodology to

bring the research to fruition; in section 4 we present and discuss the results; and finally in section 5 we conclude the analysis with policy recommendations and future directions for research.

## 2) Literature review

The topic of open data sharing and reuse has been an important and growing part of contemporary scholarly debates (Chapman et al, 2020; Pasquetto et al, 2019; Slota et al, 2020). The prospect for widely available data for all has garnered the attention of scholars from various fields of specialization (Borgman et al, 2019; Chawinga & Zinn, 2019); Wiggins et al, (2018), and featured earlier publications in *Nature* (Campbell, 2008; 2009), *Science* (Kum et al, 2011; Einav & Levin, 2014), as well as the *Journal of the American society for information science & technology* (Borgman, 2012). Testament to its ongoing significance among scholars of diverse background is the recently published comment in *nature's journal Scientific Data* (Lin et al, 2020).

Motivations for research data sharing and reusing are diverse and reflect the interests of many stakeholders including researchers, and funders (Stall et al, 2019; Wu et al, 2019; Chapman et al, 2020). Underlying the arguments for open data sharing include time efficiency gains (Pronk, 2019), providing others the ability to reproduce and verify past research (Allen & Mehler, 2019; Gray & Marwick, 2019; Hardwicke et al, 2019), allowing others to ask new questions using the data (Whitlock, 2011; Boté & Térmens, 2019; Pronk, 2019), advancing research and innovation (Borgman et al, 2019; Elsayed & Saleh, 2018), boosting faculty impact in their field of specialization through higher citations (Drachen et al, 2016; Colavizza et al, 2019; Park et al, 2018; Zeng et al, 2020), and finally making the outputs of government funded research available to the general public (Mayernik, 2017; Sholler et al, 2019; Zuiderwijk & Hinnant, 2019).

Despite these and other arguments of the value of scientific research data beyond their original purpose (Faniel & Zimmerman, 2011; Stall et al, 2019; Wu et al, 2019), researchers are often protective of their data and may be reluctant in sharing (Piwowar & Chapman, 2010; Tenopir et al, 2011; Fecher et al, 2015; Houtkoop et al, 2018). Indeed, looking at general trends in data sharing among 1329 scientists from North America, Europe and Asia, Tenopir et al, (2011) found that despite the overwhelming support for data sharing and reuse, only 46% of the respondents made their data available openly, 36% agreed that their own data are readily accessible when needed, and less than 6% made their data available to others. Similarly, focusing on 337 researchers from three Arab universities in Egypt, Jordan, and Saudi Arabia, Elsayed & Saleh (2018) reported that 64.4% of study participants shared their data, motivated by the potential of increased research visibility and citation, along with the drive to contribute to scientific progress.

In a sub-sequent investigation 4 years later, Tenopir et al, (2015) looking at the changes in data sharing and reuse practices and perceptions among scientists worldwide found increased acceptance of and willingness to engage in data sharing behaviors. Although specific barriers to data sharing still persisted. Among the many barriers to data sharing reported by researchers recently are increased perceived risk associated with data sharing (Bektaş & Tayauova, 2019; Zipper et al, 2019), publications rights concerns (Meyer, 2018; Federer et al, 2018; Ross et al, 2018), and the level of time and effort required to make data easily findable and reusable by others

(Mayernik, 2017; Stall et al, 2019). Among the most important factors reported by researchers to hinder data re-use are trust issues in other's data content and quality (Boté & Térmens, 2019; Curty et al, 2017; Lin et al, 2020), the ability to interpret the data (Zimmerman, 2008; Boté & Térmens, 2019), and the applicability of data to the problem at hand (Hartwood et al 2012; Gregory et al, 2019a). In their follow study, Tenopir et al, (2015) also reported geographic differences in data sharing and reuse worldwide, which was attributed to collectivist vs individualist cultural differences. They further found significant cross-discipline heterogeneity in perceived constraints' and enablers of data sharing and reuse.

Despite the variety and richness of the literature on the topic, only a limited number of studies address how and when researchers reuse data they obtain from other researchers (Curty et al, 2017; Gregory et al, 2019b; Koesten et al, 2019; Borst & Limani, 2020), and even fewer studies address how such data recycling practices could potentially contribute to sustainability (Jetzek et al, 2019; Turki et al, 2019).

While the nature of the value generated through openly shared data may vary depending on the particular use case, value creation is sustainable only if the data are used and reused again and again to create long-lasting value that benefits society at large. That is, the generated value can simultaneously benefit (i) private enterprises through new funding or profits; (ii) citizens that derive utility from the provided information, products or services; and (iii) society through happier and healthier citizens, better living environments, and more efficient and sustainable economic markets. Therefore, building on Hart and Milstein (2003), definition of sustainable value as “*a contribution that simultaneously delivers both short and long-term economic, social, and environmental benefits*”, our present analysis contributes to the literature “*by looking at how data sharing and re-use through the Journal of Applied econometrics data archive could contribute to promoting sustainable scientific research and economic development*”

### 3) Methodology

#### 3.1. Data

The idea of digital data archives as knowledge infrastructures mediating data sharing and reuse among researchers has been established in the field of information science and technology (Borgman et al, 2019; Lin et al, 2020). One of the key digital data archives in the field of economics is the Journal of applied economics data archive (JAEDA), which is a secondary data source of peer reviewed publications. Its current version hosted online (see <http://qed.econ.queensu.ca/jae/>), contains data for all papers accepted for publication after January 1994. Sampled from this latter archive, the data used in the current analysis is now described in the subsections below.

##### 3.1.1 Data source and sampling frame

As of March 2020, JAEDA had 35 published volumes, with an average of 7 issues per volume. For the sake of our study, we define each publication volume as a strata (35 stratum total) of clustered (132 clusters total) publication issues. Within each clustered issue, research publications and thus research data are stratified in alphabetical order by authors' last name (about 10 publications in each cluster).

### 3.1.2 Secondary data sampling procedure

For the purpose, our study we relied on a three stage stratified clustered random sampling of JAEDA. In the first stage of the sampling, volume 34 was randomly selected between all 35 volumes (stratum), and then issue number 1 was subsequently selected at random between the 7 issues (cluster) in the volume. Within issue 1, the shared data corresponding to the second alphabetically ranked published data source “bernardi-catania” (Bernardi & Catania, 2019) was selected at random.

### 3.1.3 Brief Data content description

Our randomly sampled, openly shared data “bernardi-catania” contains a collection of time series, representing European market equity price index changes, along with those of the eleven biggest European economies. These daily fluctuations in stock market returns are specifically recorded between September 8<sup>th</sup> 1999 and October 16<sup>th</sup> 2015, across national equity markets in Austria, Belgium, Denmark, France, Germany, Hungary, Italy, the Netherlands, Spain, Sweden, and the United Kingdom respectively. In addition to the individual national market fluctuations, the data also contains daily fluctuations of the STOXX Europe 600 Index, which is used as a proxy for the overall European equity market.

In the original use of the data, Bernardi & Catania (2019) were interested in the first differences of the log returns of the various equity indexes. These were used to study the impact of the global financial crisis of 2007-2008, as well as the impact of the resulting European sovereign debt crisis that followed in 2010. Relying on the Conditional Value-at-risk (CoVaR) and the conditional Expected Shortfall (CoES) as systemic risk measures, the sharing authors analyzed the systemic spread of risks among the different European countries, through an examination of stock market co-movements overtime.

### 3.1.4 Data generating process

In the initial application, Bernardi & Catania (2019) looked at the data as a collection of time series, and therefore relied on time series methods of analysis to achieve their research objectives. In comparison, in our present analysis, we adopt a different perspective looking at the data as a collection of spatio-temporal processes, which we analyze using dynamic statistical modelling methods. In our inherently conditional view, integrated stock markets are assumed to produce spatially dependent returns that evolve overtime. As a spatio-temporal phenomenon the future evolution of stock market returns depends on their past and present values. This feature allows spatio-temporal dynamic models a better chance at establishing answers to the "why" questions (causality), which is the ultimate goal in any scientific inquiry.

Comparison between our adopted dynamic spatio-temporal modelling approach and the multivariate time series modelling approach in the previous investigation shows close features, yet two fundamental differences. The first difference is related to the fact that while not all relationships in a multivariate time-series model make economic sense, a dynamic spatio-temporal model has to represent realistically the kind of spatio-temporal interactions between the various stock market returns under investigation. The second distinguishing feature deals with dimensionality. Most often in spatio-temporal applications the dimensionality of the spatial component of the model prohibits standard inferential methods, as used in Multivariate time series



representation of the same phenomenon. In such case, special care is needed for model parametrization in order to obtain realistic yet parsimonious dynamics. The development and introduction of basis functions within hierarchical statistical modeling, has allowed for the analytical leap-frogging of the computational bottleneck caused by inverting very large covariance matrices of spatially dependent data. Therefore, in our current view of stock market returns following spatio-temporal processes, we represent these processes as mixed (linear and non-linear) models with known covariates whose coefficients are unknown and non-random, together with known basis functions whose coefficients are unknown and random. The basis functions which are usually functions of space, have coefficients defined as multivariate time series random vectors, and provide computational advantages through reduced dimensionality of the covariance matrix.

### 3.1.5 Data preparation for re-use

Because of the open source nature of the R statistical software (R Core Team, 2015), we relied on it to perform all required reformatting and data preparation for reuse. The computer codes used for all data treatments and analyses in this paper are provided in the supplementary materials of this manuscript. As shown in the supplementary materials, the originally extracted data from the Journal of Applied Econometrics Data Archive, volume 34, issue 1 (Bernardi & Catania, 2019), was provided in excel format. As a wide format collection of time series, the data represented European market equity price index changes, along with that of 11 European countries' between July 8th 1999 and October 16th 2015.

Since we depart from the time series perspective by looking at the data as a collection of spatio-temporal processes, which we analyze using dynamic statistical modelling methods, we transform the wide format time series into a long format panel data. We achieve this initial transformation by using the “melt()” function from the R library “reshape2” (Hadley, 2007).

To facilitate graphical mappings in our spatio-temporal analysis, following the long transformation of the data, we downloaded from the open source GADM database of global administrative areas (Hijmans et al, 2018), the polynomial shape files for all 11 countries in our data sample. After appropriate treatments (see supplementary R codes) these shape files were used along with the long reformatted data for spatial analytics of country level aggregates, in addition to modelling the conditional mean and variance functions of the spatio-temporal process of equity price index fluctuations among the 11 countries in the European market, between 1999 and 2015.

Since the data are high frequency daily fluctuations of equity index values, we define three hierarchical temporal dimensions (daily, monthly, and yearly). To ease our analysis using conventional statistical methods, we proceed to various levels of data aggregations along these three hierarchical temporal dimensions. Our first data aggregation “CountryOutcDatm” averages over the days and computes the moments (mean, standard deviation, kurtosis and skewness) of monthly fluctuations in the index values for the 11 countries, over 1999 to 2015. The second level of aggregation “CountryOutcDaty” averages over the days and months and computes the moments (mean, standard deviation, kurtosis and skewness) of annual fluctuations in the index values for the 11 countries, observed between 1999 and 2015. We also provide a third level of aggregation “CountryOutcDat”, which averages across days within months, within years, and compute the

moments of the panel wide aggregated fluctuations in index values for all 11 European countries. All data aggregations and empirical moments' computations are done using functionalities from the R library “dplyr” (Hadley et al, 2019) as show in the attached supplementary R codes.

### 3.1.6 Descriptive Statistics of the reformatted Data

Table (1) summarizes the empirical distributional properties of the 12 series (daily fluctuations in equity index price), using their first four empirical moments, and the coefficient of variation. The mean results suggest that except for Italy that records an overall reduction in the average value of its national stock market index, all remaining 10 countries record an improvement between 1999 and 2015. Similarly on the variance and standard deviations of equity price changes, among all 11 national indices, that of Italy shows the greatest variations. The skewness and kurtosis results are in line with the characteristics of extreme value distributions, of which daily stock market fluctuations typically subscribe to. On the coefficient of variations, the daily fluctuations in index value appears to highlight the greatest variation above its mean value in France, while Italy records the greatest variation below its mean value.

Country	Count	Mean	Sample Standard deviation	Sample variance	Skewness	Kurtosis	Coefficient of Variation (CV)
Austria	4119	0.022152	1.102187	1.214816	-0.543877	5.805146	4975.47%
Belgium	4119	0.012306	1.050494	1.103538	-0.051448	3.607236	8536.47%
Denmark	4119	0.040774	1.286458	1.654974	-0.265674	4.667857	3155.08%
France	4119	0.002052	1.457916	2.125519	-0.179472	4.085347	71048.82%
Germany	4119	0.013624	1.318017	1.737169	-0.063657	4.384218	9674.45%
Hungary	4119	0.008229	1.030806	1.062561	-0.402209	5.120901	12525.87%
Italy	4119	-0.010520	1.531736	2.346216	-0.210106	2.777816	-14560.85%
Netherlands	4119	0.013114	1.305638	1.704690	-0.194704	3.220536	9956.23%
Spain	4119	0.015395	1.401671	1.964682	-0.165434	4.242557	9104.65%
Sweden	4119	0.028910	1.349594	1.821404	-0.078664	4.095575	4668.20%
United Kingdom	4119	0.016896	1.271958	1.617877	-0.224655	3.575333	7528.21%
Market	4119	0.014812	1.090849	1.89952	-0.306491	4.804299	7364.58 %

**Table (1):** Descriptive statistics of the raw time series (Daily equity price index fluctuations)

**Source:** Author's construction using the raw series shared in Bernardi & Catania (2019)

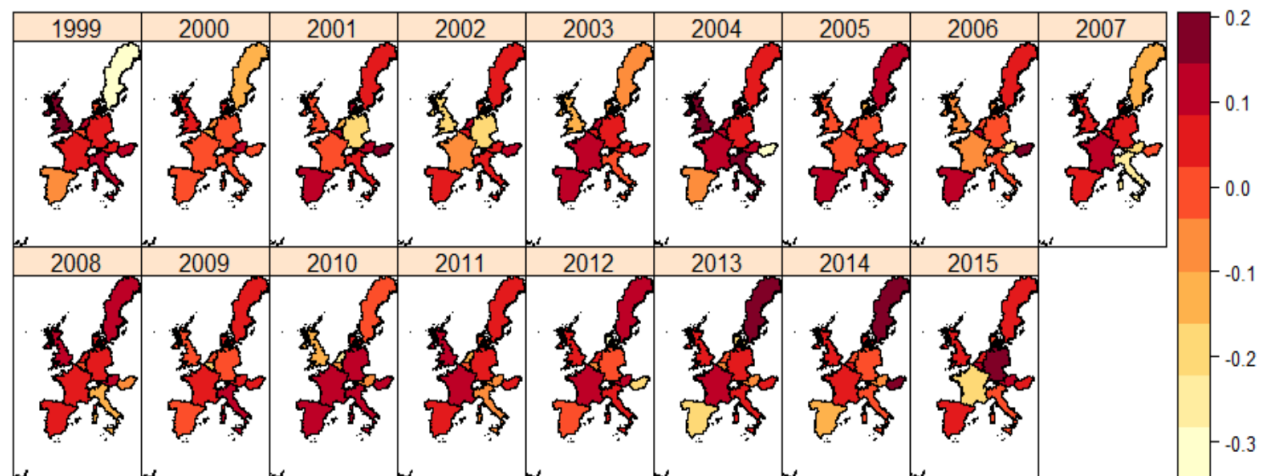
Table (2) below shows a 12 X 12 correlation matrix, summarizing the results of the test of cross-country correlation in daily equity price index fluctuations between the eleven European economies in the studied sample. The two tail test results at a critical correlation value of  $r = 1\%$  suggest the statistical significance of all the correlation coefficients in the table, further suggesting a significant positive inter-dependence in equity price index fluctuations, between the 11 European equity markets under investigation. In other words, the results suggest that the 11 European equity markets are significantly integrated. Figure (1) below provides a full graphical characterization of the process of equity index value fluctuations across the eleven European countries in our studied sample. As the figure shows the European equity index market evolves heterogeneously across both time and space, and therefore could be reasonably considered a spatio-temporal process in line with Dette et al, (2020).



	AUT	BEL	DEN	FRA	GER	HUN	ITA	NET	SPA	SWE	UK	MKT
AUT	1.000											
BEL	0.680	1.000										
DEN	0.643	0.696	1.000									
FRA	0.652	0.795	0.720	1.000								
GER	0.608	0.733	0.671	0.840	1.000							
HUN	0.441	0.448	0.466	0.474	0.451	1.000						
ITA	0.625	0.729	0.646	0.811	0.780	0.438	1.000					
NET	0.618	0.750	0.701	0.860	0.828	0.467	0.772	1.000				
SPA	0.663	0.728	0.679	0.796	0.754	0.465	0.812	0.750	1.000			
SWE	0.634	0.729	0.728	0.815	0.787	0.469	0.729	0.809	0.731	1.000		
UK	0.604	0.729	0.678	0.812	0.773	0.457	0.748	0.795	0.737	0.766	1.000	
MKT	0.764	0.859	0.820	0.929	0.891	0.585	0.880	0.904	0.879	0.886	0.875	1.000

**Table (2):** Cross-country correlation matrix for the daily equity price index fluctuations.

**Source:** Author's construction using the raw series shared in Bernardi & Catania (2019)



**figure (1):** Spatio-temporal plot of annualized average daily equity price index fluctuations.

**Source:** Author's construction using the reformatted data

### 3.2.Econometric Model

Economically, to be sustainable any society needs to take care of the needs and wants of its contemporary members, without compromising the ability of its future members to do so. Financial resource needs in modern society are usually met through financial markets where savers (investors) and borrowers (entrepreneurs) meet to trade. Each financial market is characterized by an equilibrium trade price, which evolves overtime based on both exogenous and endogenous market influences. Because of the typical integration of financial markets at various spatial locations (including regionally and internationally), equilibrium trade prices within integrated financial markets could be reasonably assumed to follow inter-related random spatio-temporal processes (RSTPs) (Dette et al, 2020), as previously discussed in the data generating process above. The mean and covariance functions of each national RSTP is assumed to depend on the regional average, but also on temporal influences (daily, monthly, and annually) as well as spatial

influences (fixed and random). The fixed spatial influences reflect national market fundamentals, while the random influences are spatial random walks, which we capture using Markov random field smoothing over the spatial random parameter space.

### 3.2.1 Model specifications

As previously discussed in the data generating process section, we represent the conditional mean and variance functions of the spatio-temporal processes of stock market returns as mixed (linear and non-linear) models with known covariates whose coefficients are unknown and non-random, together with known basis functions whose coefficients are unknown and random.

We adopt a generalized modeling perspective that accounts for both linear as well as non-linear model representations with spatial random effects. Our adopted framework builds on generalized additive models for location, scale and shape (GAMLSS) introduced by Rigby & Stasinopoulos (2005), where model parameters can be specified as functions of additive predictors with several types of covariate effects (including linear, non-linear, random and spatial effects). Given the nature of our empirical application, GAMLSS provide indeed the ideal framework for modeling the mean and variance functions of the spatio-temporal processes of stock market returns. To this end, we define a generic predictor  $\eta_i$  as:

$$\eta_i = \beta_0 + \sum_{k=1}^K s_k(\mathbf{z}_{ki}), \forall i = 1 \dots n, \text{ and } K = 5 \quad (1)$$

Or more explicitly:

$$\eta_i = \beta_0 + s_1(\mathbf{z}_{1i}) + s_2(\mathbf{z}_{2i}) + s_3(\mathbf{z}_{3i}) + s_4(\mathbf{z}_{4i}) + s_5(\mathbf{z}_{5i}), \forall i = 1 \dots n \quad (2)$$

Where  $\beta_0$  is the overall model intercept,  $\mathbf{z}_{ki}$  denotes the  $k^{th}$  sub-vector of the complete covariate vector  $\mathbf{z}_i$ , which contains the three temporal predictors (day, month, year), one regional stock index (market) and one spatial predictor (country) as previously described. The five functions  $s_k(\mathbf{z}_{ki})$  are generic smoothed effects chosen based on the type of each of the five covariates under consideration. Each  $s_k(\mathbf{z}_{ki})$  is approximated as a linear combination of  $J_k$  basis functions  $b_{k J_k}(\mathbf{z}_{ki})$  and regression coefficients  $\beta_{k J_k} \in \mathbb{R}$ , that is:

$$\sum_{j_k=1}^{J_k} \beta_{k J_k} b_{k J_k}(\mathbf{z}_{ki}) \quad (3)$$

In this form, Equation (3) implies that the vector of evaluations  $\{s_k(\mathbf{z}_{ki}), \dots, s_k(\mathbf{z}_{kn})\}^T$  can be written as  $\mathbf{Z}_k \boldsymbol{\beta}_k$  with  $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{k J_k})^T$  and the design matrix  $Z_k[i, J_k] = b_{k J_k}(\mathbf{z}_{ki})$ . Thus allowing the predictor in equation (1) to be written generically as:

$$\boldsymbol{\eta} = \beta_0 \mathbf{1}_n + \mathbf{Z}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_3 \boldsymbol{\beta}_3 + \mathbf{Z}_4 \boldsymbol{\beta}_4 + \mathbf{Z}_5 \boldsymbol{\beta}_5 \quad (4)$$

Taking into account the actual temporal and spatial predictors, equation (4) can be specifically:

$$\boldsymbol{\eta} = \beta_0 \mathbf{1}_n + \text{Day} \boldsymbol{\beta}_1 + \text{Month} \boldsymbol{\beta}_2 + \text{Year} \boldsymbol{\beta}_3 + \text{Market} \boldsymbol{\beta}_4 + \text{Country} \boldsymbol{\beta}_5 \quad (5)$$

Where  $\mathbf{1}_n$  is an  $n$ -dimensional vector of ones. In a more compact notation, equation (4) and (5) can be rewritten as  $\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\beta}$ , with  $\boldsymbol{\beta} = (\beta_0, \beta_1^T, \dots, \beta_5^T)^T$  and  $\mathbf{Z} = (\mathbf{1}_n, \mathbf{Z}_1, \dots, \mathbf{Z}_5) = (\mathbf{1}_n, \text{Day}, \text{Month}, \text{Year}, \text{Market}, \text{Country})$ .

In this representation, the smooth functions may represent linear, non-linear; random and spatial effects. Moreover, each  $\beta_k$  has an associated quadratic penalty  $\lambda_k \beta_k^T \mathbf{D}_k \beta_k$ , which plays the role of enforcing specific properties on the  $k^{th}$  function, such as smoothness. Smoothing parameter  $\lambda_k \in [0, \infty)$  controls the trade-off between fit and smoothness, and play the role of determining the shape of  $\hat{S}_k(\mathbf{z}_{ki})$ . The overall penalty can be defined as  $\boldsymbol{\beta}^T \mathbf{D}_\lambda \boldsymbol{\beta}$ , with  $\mathbf{D}_\lambda = \text{diag}(0, \lambda_1 \mathbf{D}_1, \dots, \lambda_5 \mathbf{D}_5)$ . For identification purposes, the smooth functions are mean centered following the procedure in wood (2017).

For variables with fully linear parametric effects, there is no assigned penalty such that  $\mathbf{D}_k = 0$ , and equation (3) becomes  $\mathbf{Z}_{ki}^T \boldsymbol{\beta}_k$ , with the design matrix obtained by stacking all covariate vectors  $\mathbf{z}_{ki}$  into  $\mathbf{Z}_k$ . This is typically the case for dichotomous variables such as the temporal predictors (day, month, year), which are categorical random variables. To improve identifiability of the coefficients in those spatial predictors, a ridge penalty could be used (wood, 2017), such that  $\mathbf{D}_k = \mathbf{I}$  (the identity matrix). We achieve this latter specification in the present study by specifically entering the temporal effects as random effects. This is done in the R statistical software using the following representation for the three temporal predictors in the mean and variance functions:

$$\mu : \text{eqreturn} \sim (\text{day}, \text{bs} = \text{"re"}) + (\text{month}, \text{bs} = \text{"re"}) + (\text{year}, \text{bs} = \text{"re"}) \quad (6)$$

$$\sigma : \sim (\text{day}, \text{bs} = \text{"re"}) + (\text{month}, \text{bs} = \text{"re"}) + (\text{year}, \text{bs} = \text{"re"}) \quad (7)$$

Where “eqreturn” is the dependent variable, and more specifically the average index value or equity return; “day”, “month” and “year” are the categorical temporal predictors; “bs” is an argument specifying the type of spline basis used, which in this case is “re” for random effect.

For continuous predictors of the mean and variance functions of the spatio-temporal processes of stock market returns, such as the proxied European stock index return (market), the smooth functions are represented using the regression spline approach (Eilers & Marx, 1996). In this case,  $s_4(\mathbf{z}_{4i})$  is approximated by  $\sum_{j_k=1}^{J_k} \beta_{k J_k} b_{k J_k}(Z_{ki})$  or equivalently  $\sum_{j_k=1}^{J_k} \beta_{k J_k} b_{k J_k}(\text{Market}_i)$  where the  $b_{k J_k}(\text{Market}_i)$  are low rank thin plate regression spline basis function evaluations for each daily market return  $i$ , and are numerically stable with convenient mathematical properties (wood, 2003). This specification allow us to avoid arbitrary modeling decisions, such as selecting the appropriate degree of a polynomial or specifying cut-points, which could induce bias. Extending the mean and variance functions in equations (6) and (7), this specification is implemented in the R statistical software using the follow representations:

$$\mu : \text{eqreturn} \sim (\text{day}, \text{bs} = \text{"re"}) + (\text{month}, \text{bs} = \text{"re"}) + (\text{year}, \text{bs} = \text{"re"}) + (\text{market}, \text{bs} = \text{"tp"}, k=10) \quad (8)$$

$$\sigma : \sim (\text{day}, \text{bs} = \text{"re"}) + (\text{month}, \text{bs} = \text{"re"}) + (\text{year}, \text{bs} = \text{"re"}) + (\text{market}, \text{bs} = \text{"tp"}, k=10) \quad (9)$$

Where “market” is the continuous covariate capturing the European stock index; with the basis spline argument “bs” set to “tp” for penalized low rank thin plate spline) and 10 basis functions since “k”=10. Although not specifically used here, other available continuous basis spline functions include the cubic regression spline “cr”, and the P-spline, “ps”.

To incorporate the spatial effects into the regression model, the 11 European countries covered in the study are split into discrete contiguous geographic units, with spatial coordinates exploited through a Markov random field approach. This latter approach relies on the information contained in neighboring stock market returns from each country. In this case, equation (3) becomes  $\mathbf{Z}_{ki}^T \boldsymbol{\beta}_k$ , with  $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{kR})^T$  representing the vector of spatial effects.  $R = 11$  denotes the total number of countries and  $\mathbf{z}_{ki}$  is made up of a set of country labels. The design matrix linking each daily stock return  $i$  to the corresponding spatial effect is defined for  $r = 1, \dots, 11$  by:

$$\mathbf{z}_k[i, r] = \begin{cases} 1 & \text{if daily stock return } i \text{ belongs to country } r \\ 0 & \text{Otherwise} \end{cases} \quad (10)$$

The smoothing penalty is based on the neighborhood structure of the discrete contiguous geographic units, so that stock market returns from spatially adjacent countries share similar effects. Specifically the diagonal matrix associated with the quadratic penalty is given by:

$$\mathbf{D}_k[r, q] = \begin{cases} -1 & \text{if } r \neq q \wedge r \approx q \\ 0 & \text{if } r \neq q \wedge r \not\approx q \\ N_r & \text{if } r = q \end{cases} \quad (11)$$

Where  $r \approx q$  indicates whether any two countries  $r$  and  $q$  are adjacent neighbors, and  $N_r$  is the total number of neighbors for country  $r$ . The resulting quadratic penalty is equivalent to Rue & Held (2005) stochastic interpretation that  $\boldsymbol{\beta}_k$  follows a Gaussian Markov random field. Finally, extending the mean and variance functions in equations (8) and (9), this specification is implemented in the R statistical software using the follow representations:

$$\mu : \text{eqreturn} \sim (\text{day, bs = "re"}) + (\text{month, bs = "re"}) + (\text{year, bs = "re"}) + (\text{market, bs = "tp", k=10}) + \text{s(country, bs = "mrf", xt = xt)} \quad (12)$$

$$\sigma : \sim (\text{day, bs = "re"}) + (\text{month, bs = "re"}) + (\text{year, bs = "re"}) + (\text{market, bs = "tp", k=10}) + \text{s(country, bs = "mrf", xt = xt)} \quad (13)$$

Where “country” is a factor variable with 11 levels, indexing each country in the data. “mrf” is the Markov Random Field smoother. All described smoothers are available from the R package “GJRM”, whose documentation can be consulted for more details (see Wojtys & Marra, 2018). We estimate the parameters of the above described model specification within the R statistical software using the “gamlss()” function from the R library “SemiParBIVProbit” (see Marra & Radice, 2017).

## Results

Relying on the underlying philosophy of empirical process theory (Dehling & Philipp, 2002), our empirical analysis using the reformatted data, models jointly the mean function, and the variance function of the column vector of concatenated time series. In this model representation, the full vector is seen as random structure (Merlevède et al, 2019), describing the process of daily changes in equity price index across the eleven integrated European stock markets.

Making distributional assumptions about the empirical moments generating function (Cabaña & Quiroz, 2005; Collender & Chalfant, 1986) for this random vector, we are able to study the properties of its first and second order moments. These are represented by the mean and variance equations capturing respectively the conditional expected fluctuations in equity price index, and the conditional variance in equity price index fluctuations over the period of September 8<sup>th</sup> 1999 and October 16<sup>th</sup> 2015, across the 11 national equity markets in Austria, Belgium, Denmark, France, Germany, Hungary, Italy, the Netherlands, Spain, Sweden, and the United Kingdom.

**Table (2):** Estimated Spatio-temporal process of equity price index fluctuations

Mean Equation	Parametric Model		Semi-parametric Model	
	Est.	Std. err.	Edf	p-value
(intercept)	-0.7138	1.293	0.0137***	4.06e-07
day	-0.00005	0.00034	1.000	0.9903
month	0.00101	0.00092	1.000	0.3694
year	0.00035	0.00064	1.000	0.3858
Market	1.024***	0.00295	0.182***	<2e-16
Country	Edf = 2.902	p-value = 0.212	Edf = 2.832	p-value = 0.0626
Variance Equation	Est.	Std. err.	Edf	p-value
(intercept)	87.757***	2.864	-0.9599***	<2e-16
day	-0.0036***	0.00076	3.415***	<2e-16
month	0.0054**	0.00203	8.657***	<2e-16
year	-0.0441***	0.00143	8.955***	<2e-16
Market	-0.0134**	0.00422	8.052***	<2e-16
Country	Edf = 9.991	p-value <2e-16	9.992***	<2e-16
Model Performance measures	Est.	(C.I.)	Est.	(C.I.)
$\hat{\sigma}^2$	0.478	(0.456 , 0.502)	0.479	(0.439 , 0.524)
Largest Absolute gradient value	0.6973506		7.749248e-06	
Eigenvalue range	[0.1218557 , 450442249023]		[3.815519 , 1.292507e+12]	
Total edf	22.9		47.9	
N	45309		45309	
AIC	91537.82		85314.91	
BIC	91737.48		85732.68	

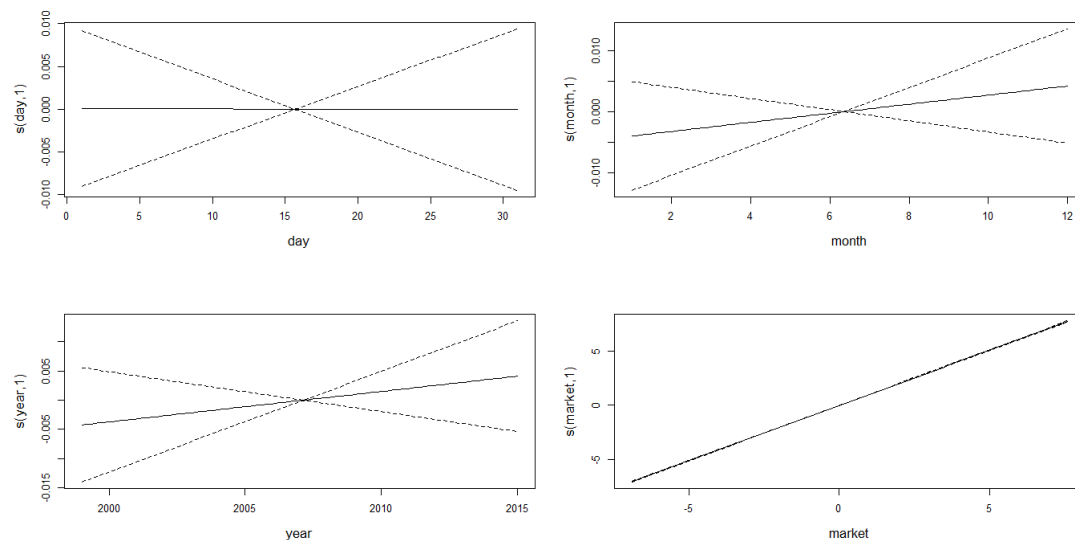
**Note:** Est. is estimated coefficient value; Std. err. Is the Standard error of the estimated coefficient; Edf is the estimated empirical density function; C.I. the confidence interval on the estimated coefficient. Significance code: \*\*\* for 0.1% level, \*\* for 1% level, and \* for 5% significance level.

For sensitivity analysis, we estimate a linear process (parametric) model, and a non-linear process (semi-parametric) model, both with random country effects. The performances of these two specifications are subsequently contrasted based on the Bayesian and the Akaike information criteria. Table (2) above summarizes the results of these estimations. Based on the model performance measures the non-linear process model is adopted as the preferred specification, with its results now discussed in the sub-sections below.

### 3.3. Conditional Mean equation results for equity price index fluctuations

Recall that in semi-parametric models with smooth function estimates, when the empirical density function (edf) is close to 1, the respective estimated effect is linear, and hence the covariate can enter the model parametrically. However, the higher the edf value the more complex is the estimated curve, and thus the corresponding covariate cannot be assumed to have a linear relationship with the outcome variable.

As shown in the lower portion of table (2) under “Semi-parametric model”, the edfs for the smooth components of day, month, and year, are all equal to 1, with statistically non-significant p-values. Conversely, the p-value of the overall European regional market effect appears significant at the 5% significance level. Although the edf value of the random country effects is 2.832, thus above 1 and suggesting its non-linear influence on the individual markets fluctuations. Since its corresponding p-value is 0.0626, this effect only reaches statistical significance at the 10% level. Together, the above suggest that expectations about future changes in equity price index, by economic actors within the individual national markets, are not significantly influenced by daily, monthly nor annual perturbations in each national economy. Instead, the mean function of the daily fluctuations in equity price index value appears to depend solely but weakly, on the fluctuations in the overall European average index price. These results are further confirmed by the smooth function plots in figure (2) below, which are estimated after fitting the spatio-temporal model.



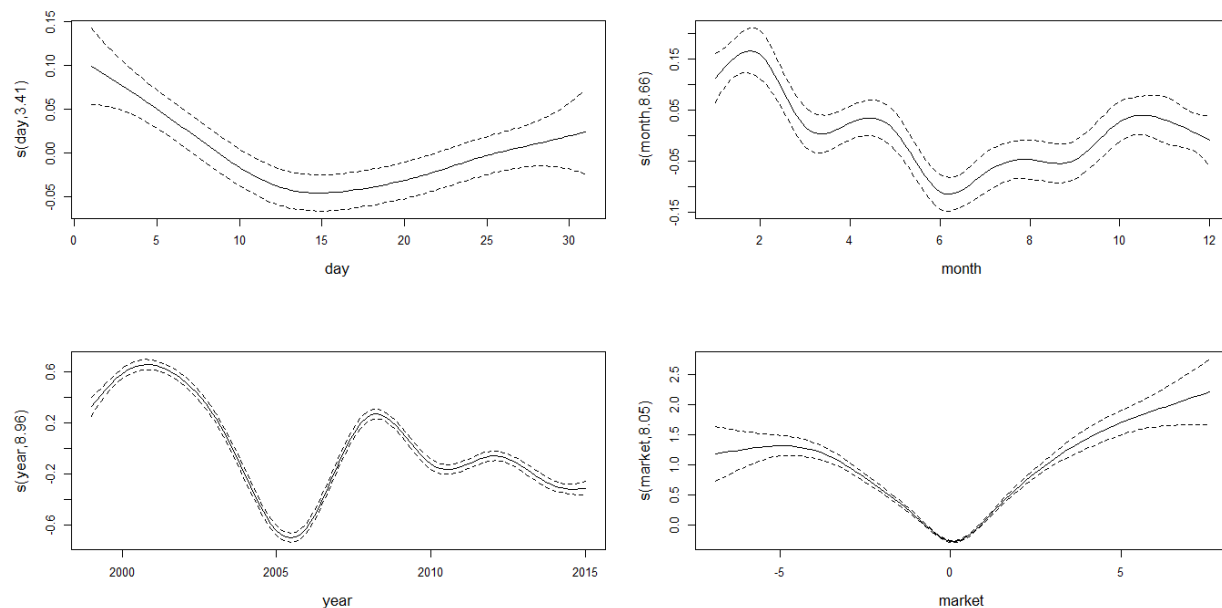
**Figure (2):** Smooth functions plots for the effects of daily, monthly, yearly, and market level perturbations on expected equity index price changes across the eleven financial markets.



### 3.4. Conditional Variance equation results for equity price index fluctuations

Also shown in the lower portion of table (2) under “Semi-parametric model”, the edfs for the smooth components of daily (3.415), monthly (8.657), yearly (8.955), European market level (8.052), and individual country level (9.992) fluctuations are all well above 1, with statistically significant p-values at the 0.1% level. These indicate the non-linear and significant effects of all the above mentioned covariates, on the variances of the individual European equity price index fluctuations. Indeed, daily variations in price index fluctuations are seen to not only depend on overall European average index fluctuations, but also on daily, monthly and annual perturbations taking place in each individual market. These algebraic results are further confirmed by the smooth functions plots in figure (3) below, which are estimated after fitting the spatio-temporal model.

The smooth function plot on the top-left panel of figure (3) shows that on the daily scale, equity price index fluctuations across the individual European stock markets, varies convexly in any given month. Indeed, the variance in daily price index fluctuations decreases on average during the first 15 days of the month, and increases back over the course of the last 15 days of the month.

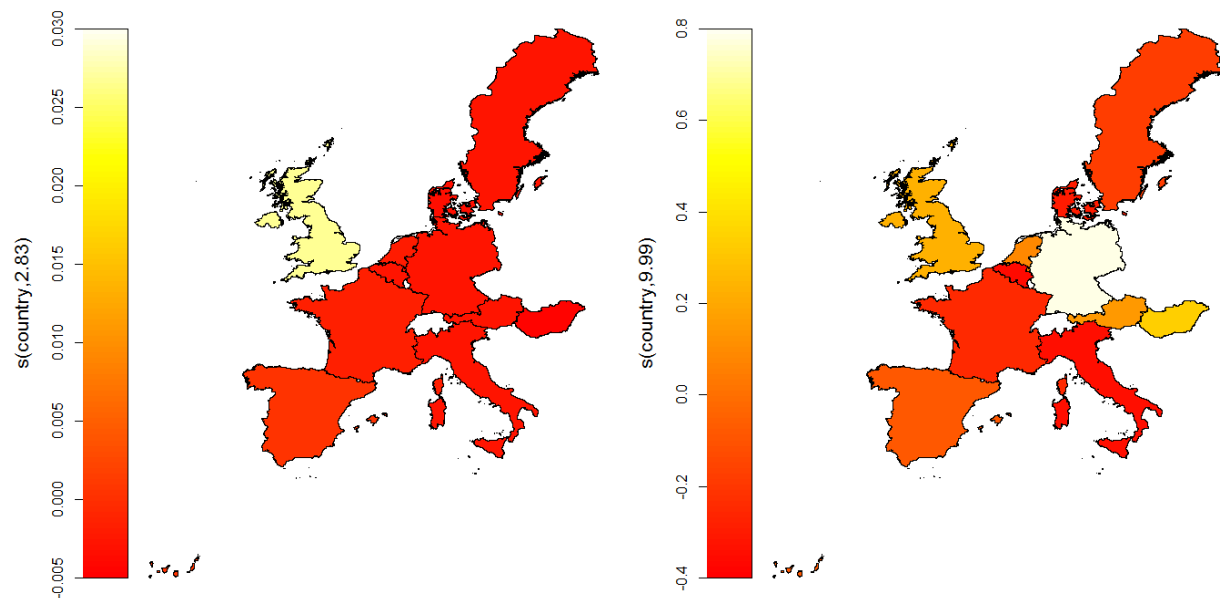


**Figure (3):** Smooth functions plots for the effects of daily, monthly, yearly, and market level perturbations on the variance of equity index price changes across the eleven financial markets.

On the monthly scale, variations in equity price index changes seem to exhibit a cascading, but overall convex shape over the course of the year. The smooth function plots on the top-right panel of figure (3) shows that the variance in daily price index fluctuations, starts the year by increasing initially between January and February. Then from February, it decreases reaching its minimum in mid-year (June); from which point it starts to increase again reaching its maximum in November. It then decreases as the year end in December, to spike back up at the beginning of the next year.

On the annual scale, variations in equity price index changes seem to exhibit a cyclical pattern over the 15 years covered by the data sample as seen in the lower-left panel of figure (3). It can be noted that variations in equity index price changes is concave between 1999 and 2005 reaching a maximum in the early 2000's. Although a local maxima appears to characterize the period of the 2008 financial crises, the level of variations in daily equity price fluctuation during this period remained less than that observed in the early 2000. A relatively steeper reduction in the variance of equity price index fluctuations is observed between 2008 and 2010. This reduction appears to have persisted between 2010 and 2015 although at a slower pace.

Finally, the lower-right panel of figure (3) suggests that changes in the overall European market average index value exert a convex effect on index price fluctuations in the individual national markets. Indeed, variations in the individual national stock market fluctuations are minimized at overall European market index price changes of zero; while strictly increasing for overall European stock market index price changes above or below zero. Further validating the above described algebraic properties in table (2), are the country level spatial random expected changes in equity price index (left panel) and expected variations in the equity price index fluctuations (right panel), which are graphically summarized in figure (4) below.



**Figure (4):** Spatial density plots of the random Mean (left panel) and Variance (right panel) of equity index returns

### 3.5. Empirical results assessment

Assessing our results through the lenses of time series modelling, it can be noted that they corroborate the theoretical predictions of stochastic mixing processes, which Meir (2000; p.13) shows converge uniformly under some regularity condition. Indeed, starting with the work of Yu (1994) on dependent sequences, scholars have shown that among time series exists a special class

of processes for which the “future” depends only weakly on the “past”, these have been referred to as “mixing processes” (Dehling & Philipp, 2002; Rio, 2017). In our analysis, the fact that the conditional expectation of equity index price changes weakly on the overall changes in the European regional average index value, while being free from individual market’s daily, monthly, and annual perturbations, suggest that European stock market index prices evolve stochastically as mixing processes. This is further reaffirmed by the results of the conditional variance equation, which shows the variations in daily equity index price in the European market to significantly depend on daily, monthly, and annual perturbations.

These results have significant implications for the stability of the European stock market and financial sector, and should provide financial regulators in the region with additional evidence for their policy efforts. In line with the vision of the European Institute of Innovation and technology (EIT)<sup>1</sup>, which was created in 2008 by the European Commission as an open innovation program in charge of translating the knowledge of open science into platforms, architectures and systems that bring significant value to society, our empirical investigation points to the potential value of creating a unique index, in the like of the existing “ethical stock market indices”, which will track the equity values of European companies involved in open innovation. Doing so will allow open innovation investors to have key market updates for their portfolio investment decisions, and may contribute to driving further the needed funds to support European open innovation initiatives, such as ATTRACT<sup>2</sup>. This latter initiative by a group of big European research labs including CERN, the European molecular biology laboratory, and the European Synchrotron Radiation Facility aims to translate the outputs of open science into open innovations, in partnership with Universities, SMEs, multinationals and private investors. According to the quadruple helix framework (Miller et al, 2018; Yun, 2019), such open science based innovation practices are more sustainable than the previous closed innovation paradigm.

#### 4. Discussion

The sharing economy as an umbrella concept that encompasses several ICT developments and technologies (Acquier et al, 2017), including collaborative consumption of goods and services through online platforms, has emerged as an economic-technological phenomenon (Laurenti et al, 2019). Its recent growth has been fueled by developments in ICT, growing consumer awareness, proliferation of collaborative web communities (Ranjbari, 2019), and increasing concerns over ecological, societal and developmental impact (Hamari et al, 2016). Though debates still persist as to whether it could be a successful pathway to sustainability in times of rapid technological developments, but constrained financial resources (Martin, 2016; Geissinger, 2019; Murillo, 2020). Although others have focused on how the data deluge along with artificial intelligence is reshaping the field of economics altogether (Athey, 2018), in this study we explored instead how sharing and reusing scientific research data, as digital goods, could contribute to sustainable research output production and economic growth. We achieved this by relying on a three stage

---

<sup>1</sup> [www.eit.europa.eu](http://www.eit.europa.eu)

<sup>2</sup> [www.attract-eu.org](http://www.attract-eu.org)

stratified clustered random sample from the Journal of Applied Econometrics Data Archive, along with descriptive analytics and spatio-temporal econometric modelling and inference.

We find that despite the strong deductive reasoning requirement for the successful reuse of openly shared data<sup>3</sup> in subsequent analyses, such practice does provide a viable solution for the sustainability of research output production, innovation and economic development. Sustainable value creation is not only reflected in direct economic value, but also through better social and environmental outcomes (Geissinger, 2019). This is even more apparent if we adopt a “data value-chain” perspective to link openly shared data as raw material to the digital contents, goods and services produced (De Reuver et al, 2018). Such perspective is underpinned however by the interplay of two fundamental mechanisms (Jetzek et al, 2019), the first of which is the information sharing mechanism that relates to how openly shared data are used to create informational content that creates value for society through increased transparency, reduced information asymmetry, and improved decision-making. The second channel is the market mechanism which relates to how openly shared data help make processes more efficient, and often satisfy previously unmet needs by providing the raw material for the production of digital goods and services that are subsequently sold in markets.

Sustainable value generation through openly shared data is based on creating an opportunity for anyone to reuse data beyond the organizational boundaries of the data custodian and the technical boundaries of the originating system (Murillo, 2020). Making the value creation itself sustainable requires re-users to complement open data with proprietary data sources and use the enriched data in combination with specialized algorithms and technical infrastructures for the development of digital content, products, and services (Welle Donker & van Loenen, 2017). Although such idea of “data recycling” remains at present date under-researched, its connection with “collaborative consumption” through open data sharing and reuse is increasingly regarded as a practice that engages especially environmentally and ecologically conscious consumers (Jetzek et al, 2019).

Our analysis further support the idea that viewing research data recycling as a sustainable practice, like any other recycling activity can lead to an increase in data sharing and reuse, especially when adopting this view leads to increased positive attitudes towards participation (Turki et al, 2019). Because aspirations to sustainability do not always strongly translate into action, expectations as to a wider diffusion of data sharing and reuse within the research community might be deflated. It may be that opportunistically, people seeking economic benefits end up adopting data sharing and reuse as an alternative mode of scientific research production (Houtkoop et al, 2018). Or, in a worst case scenario, some researchers might be altruistic and share openly their research data, while other researchers benefit mostly from such sharing. This situation however would undermine the sustainability of data sharing and reuse if proper compensating mechanisms are not put in place. As initiatives such as the “Research Data Alliance<sup>4</sup>”, and new

---

<sup>3</sup> Explained as “reasoning from the general to the particular, the deductive approach follows the path of logic most closely. The reasoning starts with a theory and leads to a new hypothesis. This hypothesis is put to the test by confronting it with observations that either lead to a confirmation or a rejection of the hypothesis” (Russell, 2010).

<sup>4</sup> <https://rd-alliance.org/about.html>

journals such as Elsevier’s “data in brief”, MDPI’s “data”, and Nature’s “Scientific Data” are created with inter-disciplinary focus, and allowing researchers to openly publish and get full citable credit for their research data, the practice of data recycling through open data sharing and reuse should find its way to make scientific research production more sustainable. It is also our hope that initiatives such as “Open Science Grid<sup>5</sup>”, will successfully contribute to fostering the wider, faster, and cheaper access to new knowledge, promoting more rapid understanding and use of science.

## 5. Conclusion

The sharing economy has been emerging as one of the key paradigms in support of the Fourth Industrial Revolution (Kim & Lee, 2019). As a result in March 2017, the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM) appointed an expert committee to evaluate more fully the benefits and challenges of broadening access to the results of scientific research, described as “open science”. The committee was charged with focusing on how to move toward open science as the default for scientific research results, and to indicate both the benefits of open science and the barriers to doing so. The resulting consensus report published by the NASEM pointed out significant benefits of open science moving forward (National Academies of Sciences, Engineering, and Medicine, 2018a). Since then, there have been a growing consensus in the scientific community that the transition to data-driven open science is best achieved by establishing a globally interoperable research infrastructure (National Academies of Sciences, Engineering, and Medicine, 2018b). This has led the Board on Research Data and Information (BRDI) of the NASEM to organize a follow up symposium on November 1st, 2017 to explore the issues of making research data findable, accessible, interoperable, and reusable (abbreviated as FAIR).

The current investigation was carried out within this general context, with the understanding that the value-generating mechanisms of openly shared data remains mostly unobservable, and that research on the underlying topic of data recycling for sustainability is still in a nascent state (Meijer & Grimmelikhuijsen, 2017). Relying therefore on a resourceful research approach that focused on showcasing the re-use potential of a key digital data archive, we hoped to provide potential re-users with a case study example that could assist in further prospective investigations using similar open data sources. This practice contributes to the literature on how and when researchers reuse data they obtain from openly shared sources, easing therefore the transition into the era of “citizen science” or “crowdsience” (Teo, 2020). As suggested in the recently published comment in nature’s “Scientific Data” journal (Lin et al, 2020), the issue of open data sharing and reuse, and its “trust” requirement between the different stakeholders including data repositories, needs to be at the center of ongoing scientific discourses. We hope that our current treatment, as a step in that direction, would draw further prospective interests on the topic.

---

<sup>5</sup> Opensciencegrid.org

## References

- Allen, C., & Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS biology*, 17(5), e3000246.
- Acquier, A., Daudigeos, T., & Pinkse, J. (2017). Promises and paradoxes of the sharing economy: An organizing framework. *Technological Forecasting and Social Change*, 125, 1-10.
- Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507-547). University of Chicago Press.
- Bektaş, Ç., & Tayauova, G. (2019). Science Ethics and Social Responsibilities of Scientists. *The Journal of International Scientific Researches*, 4(2), 108-120.
- Bernardi M. and Catania L. (2019) "Switching Generalised Autoregressive Score Copula Models with Application to Systemic Risk", *Journal of Applied Econometrics*, Vol. 34, No. 1, 2019, pp. 43-65.
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888-904.
- Borst, T., & Limani, F. (2020). Patterns for searching data on the web across different research communities. *LIBER Quarterly*, 30(1).
- Boté Vericad, J. J., & Térmens i Graells, M. (2019). Reusing Data: Technical and Ethical Challenges. *DESIDOC Journal of Library & Information Technology*, 2019, vol. 39, num. 6, p. 329-337.
- Cabaña, A., & Quiroz, A. J. (2005). Using the empirical moment generating function in testing for the Weibull and the type I extreme value distributions. *Test*, 14(2), 417-431.
- Campbell P. (2009) Data's shameful neglect. *Nature* 461(7261), <https://doi.org/10.1038/461145a>
- Campbell P. (2008) Community cleverness required. *Nature* 455 (7209), <https://doi.org/10.1038/455001a>
- Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2), 151-78.
- Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L. D., Kacprzak, E., & Groth, P. (2020). Dataset search: a survey. *The VLDB Journal*, 29(1), 251-272.
- Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*. 41(2), 109-122.
- Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2019). The citation advantage of linking publications to research data. *arXiv preprint arXiv:1907.02565*.
- Collender, R. N., & Chalfant, J. A. (1986). An alternative approach to decisions under uncertainty using the empirical moment-generating function. *American Journal of Agricultural Economics*, 68(3), 727-731.
- Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PloS one*, 12(12).



- De Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2), 124-135.
- Dehling, H., & Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data* (pp. 3-113). Birkhäuser, Boston, MA.
- Dette, H., Dierickx, G., & Kutta, T. (2020). Quantifying deviations from separability in space-time functional processes. *arXiv preprint arXiv:2003.12126*.
- Drachen, T., Ellegaard, O., Larsen, A., & Dorch, S. (2016). Sharing data increases citations. *Liber Quarterly*, 26(2).
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 89-102.
- Elsayed, A. M., & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA journal*, 44(4), 281-299.
- Faniel, I. M., & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, 6(1), 58-69.
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing?. *PloS one*, 10(2).
- Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *PloS one*, 13(5).
- Geissinger, A., Laurell, C., Öberg, C., & Sandström, C. (2019). How sustainable is the sharing economy? On the sustainability connotations of sharing economy platforms. *Journal of Cleaner Production*, 206, 419-429.
- Gray, C. T., & Marwick, B. (2019, July). Truth, proof, and reproducibility: there's no counter-attack for the codeless. In *Research School on Statistics and Data Science* (pp. 111-129). Springer, Singapore.
- Gregory, K. M., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2019a). Understanding data search as a socio-technical practice. *Journal of Information Science*, 0165551519837182.
- Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., & Wyatt, S. (2019b). Searching data: A review of observational data retrieval practices in selected disciplines. *Journal of the Association for Information Science and Technology*, 70(5), 419-432.
- Hadley Wickham (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(1 2), 1-20.
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. *R package version 0.8.0.1*. [Online] Retrieved March 12, 2019 from <https://CRAN.R-project.org/package=dplyr>
- Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. *Science*, Vol. 331, Issue 6018, pp. 649
- Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. (2019). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, 7(2), 190806.

- Hart, S. L., & Milstein, M. B. (2003). Creating sustainable value. *Academy of Management Executive*, 17(2), 56-67.
- Hartswood, M., Procter, R., Taylor, P., Blot, L., Anderson, S., Rouncefield, M., & Slack, R. (2012, May). Problems of data mobility and reuse in the provision of computer-based training for screening mammography. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 909-918).
- Heller, N., Rickman, J., Weight, C., & Papanikolopoulos, N. (2019). The Role of Publicly Available Data in MICCAI Papers from 2014 to 2018. In *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention* (pp. 70-77). Springer, Cham.
- Hijmans, R., Garcia, N., & Wiecek, J. "GADM database of Global Administrative Areas". Version 3.6 (released May 6, 2018). [Online] Retrieved March 12, 2019 from <https://gadm.org/data.html>
- Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1), 135-174.
- Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2019). The sustainable value of open government data. *Journal of the Association for Information Systems*, 20(6), 702-734
- Koesten, L., Gregory, K., Groth, P., & Simperl, E. (2019). Talking datasets: Understanding data sensemaking behaviours. *arXiv preprint arXiv:1911.09041*.
- Laurenti, R., Singh, J., Cotrim, J. M., Toni, M., & Sinha, R. (2019). Characterizing the sharing economy state of the research: A systematic map. *Sustainability*, 11(20), 5729.
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giarretta, D., ... & Khodiyar, V. (2020). the tRUST Principles for digital repositories. *Scientific Data*, 7(1), 1-5.
- Lodwick, L. (2019). Sowing the seeds of future research: data sharing, citation and reuse in archaeobotany. *Open Quaternary*, 5(1).
- Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the association for information science and technology*, 67(9), 2047-2059.
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in methods and practices in psychological science*, 1(1), 70-85.
- Kim, Y., & Lee, M. (2019). Typology and Unified Model of the Sharing Economy in Open Innovation Dynamics. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(4), 102.
- Kum, H. C., Ahalt, S., & Carsey, T. M. (2011). Dealing with data: Governments records. *Science*, 332(6035), 1263-1263.
- Mancini, D., Lardo, A., & De Angelis, M. (2020). Efforts Towards Openness and Transparency of Data: A Focus on Open Science Platforms. In *Exploring Digital Ecosystems* (pp. 67-84). Springer, Cham.
- Marra G. and Radice R. (2017), SemiParBIVProbit: *Semiparametric Copula Bivariate Probit Modelling*. R package version 3.8.1.

- Martin, C. J. (2016). The sharing economy: A pathway to sustainability or a nightmarish form of neoliberal capitalism?. *Ecological economics*, 121, 149-159.
- Mayernik, M. S. (2017). Open data: Accountability and transparency. *Big Data & Society*, 4(2), 2053951717718853.
- Meir, R. (2000). Nonparametric time series prediction through adaptive model selection. *Machine learning*, 39(1), 5-34.
- Merlevède, F., Peligrad, M., & Utev, S. (2019). *Functional Gaussian Approximation for Dependent Structures*. Oxford University Press.
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131-144.
- Miller, K.; McAdam, R.; McAdam, M. (2018). A systematic literature review of university technology transfer from a quadruple helix perspective: toward a research agenda. *R&D Management*, 48, 7–24.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- Murillo, D. (2020). The Politics of the Sharing Economy. In *Sharing Economy and the Impact of Collaborative Consumption* (pp. 21-36). IGI Global.
- National Academies of Sciences, Engineering, and Medicine. (2018a). *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/25116>.
- National Academies of Sciences, Engineering, and Medicine. (2018b). *International Coordination for Science Data Infrastructure: Proceedings of a Workshop—in Brief*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/25015>.
- Nielsen, M. (2020). *Reinventing discovery: the new era of networked science* (Vol. 70). Princeton University Press.
- Park, H., You, S., & Wolfram, D. (2018). Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields. *Journal of the Association for Information Science and Technology*, 69(11), 1346-1354.
- Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, 1(2).
- Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data. *Data Science Journal*, 16: 8, pp. 1–9, DOI: <https://doi.org/10.5334/dsj-2017-008>
- Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: a pilot study of associations. *Journal of informetrics*, 4(2), 148-156.
- Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological applications*, 29(1), e01822.
- Pronk, T. E. (2019). The Time Efficiency Gain in Sharing and Reuse of Research Data. *Data Science Journal*, 18(1).

- R Core Team (2015). "R: A Language and Environment for Statistical Computing", *R Foundation for Statistical Computing, Vienna, Austria*. URL: <https://www.R-project.org/>
- Ranjbari, M., Morales-Alonso, G., Esfandabadi, Z. S., & Carrasco-Gallego, R. (2019). Sustainability and the Sharing Economy: Modelling the Interconnections. *Dirección y Organización*, 33-40.
- Rio, E. (2017). *Asymptotic theory of weakly dependent random processes* (Vol. 80). Berlin: Springer.
- Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. *American Psychologist*, 73(2), 138.
- Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society, Series C*, 507-554.
- Rue, H., & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC.
- Russell, K. (2010). The Art of Being a Scientist: A Guide for Graduate Students and their Mentors by Roel Snieder, Ken Larner. *International Statistical Review*, 78(1), 159-159.
- Sholler, D., Ram, K., Boettiger, C., & Katz, D. S. (2019). Enforcing public data archiving policies in academic publishing: A study of ecology journals. *Big Data & Society*, 6(1), 2053951719836258.
- Slota, S. C., Hoffman, A. S., Ribes, D., & Bowker, G. C. (2020). Prospecting (in) the data sciences. *Big Data & Society*, 7(1), 2053951720906849.
- Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., ... & Wyborn, L. (2019). Make scientific data FAIR. *Nature* 570, 27-29 DOI: 10.1038/d41586-019-01720-7
- Teo, E. A. (2020). State-of-the-art analysis of the pedagogical underpinnings in open science, citizen science and open innovation activities. *E. Triantafyllou (Ed.). INOS Consortium*. Retrieved 30 April, 2020 from <https://inos-project.eu/>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PloS one*, 10(8).
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS one*, 6(6), e21101.
- Tolle, K. M., Tansley, D. S. W., & Hey, A. J. (2011). The fourth paradigm: Data-intensive scientific discovery [point of view]. *Proceedings of the IEEE*, 99(8), 1334-1337.
- Turki, S., Martin, S., & Renault, S. (2019). Stimulation of Open Data Ecosystems: Learnings from Theory and Practice. *Digital Innovation: Harnessing the Value Of Open Data*, 4, 41.
- Wallis JC, Rolando E, Borgman CL (2013). If we Share Data, Will Anyone Use Them ? *Data Sharing and Reuse in the Long Tail of Science and Technology*. *PLoS ONE* 8(7): e67332
- Waugh, A. (2010). Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information: Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access [Book Review]. *Archives and Manuscripts*, 38(2), 113.
- Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, 10(3), 284-306.

- Whitlock, M. C. (2011). Data archiving in ecology and evolution: best practices. *Trends in Ecology & Evolution*, 26(2), 61-65.
- Wiggins, A., Young, A., & Kenney, M. A. (2018). Exploring visual representations to support data re-use for interdisciplinary science. *Proceedings of the Association for Information Science and Technology*, 55(1), 554-563.
- Wojtyś, M., Marra, G., & Radice, R. (2018). Copula based generalized additive models for location, scale and shape with non-random sample selection. *Computational Statistics & Data Analysis*, 127, 1-14.
- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B*, 65, 95-114.
- Wu, Y., Moylan, E., Inman, H., & Graf, C. (2019). Paving the way to open data. *Data Intelligence*, 1(4), 368-380.
- Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Annals of probability*, 22, 94-116.
- Yun, J.J.; Liu, Z. (2019). Micro-and macro-dynamics of open innovation with a quadruple-helix model. *Sustainability*, 11, 3301.
- Zhang, Y., Ren, J., Liu, J., Xu, C., Guo, H., & Liu, Y. (2017). A survey on emerging computing paradigms for big data. *Chinese Journal of Electronics*, 26(1), 1-12.
- Zeng, T., Wu, L., Bratt, S., & Acuna, D. E. (2020). Assigning credit to scientific datasets using article citation networks. *Journal of Informetrics*, 14(2), 101013.
- Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33(5), 631-652.
- Zipper, S. C., Stack Whitney, K., Deines, J. M., Befus, K. M., Bhatia, U., Albers, S. J., ... & O'Donnell, F. (2019). Balancing open science and data privacy in the water sciences. *Water Resources Research*, 55(7), 5202-5211.
- Zuiderwijk, A., & Hinnant, C. C. (2019). Open data policy-making: A review of the state-of-the-art and an emerging research agenda. *Information Polity*, 117-129.
- Zuiderwijk, A., & Spiers, H. (2019). Sharing and Re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, 49, 228-241.