GUS

Indeks 371262
e-ISSN 2657-9545
ISSN 0033-2372

Przegląd Statystyczny
Statistical Review

Vol. 70 | No. 2 | 2023

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

INFORMATION FOR AUTHORS

Przegląd Statystyczny. Statistical Review publishes original research papers on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision science and data analysis. The manuscripts considered for publication should significantly contribute to the theoretical aspects of the aforementioned fields or shed new light on the practical applications of these aspects. Manuscripts reporting important results of research projects are particularly welcome. Review papers, shorter papers reporting on major conferences in the field, and reviews of seminal monographs are eligible for submission on the Editor-in-Chief's request.

Since 1 May 2019, the journal has been publishing articles in English.

Any spelling style is acceptable as long as it is consistent within the manuscript.

All work should be submitted to the journal through the Editorial System
(https://www.editorialsystem.com/pst).

For details of the submission process and editorial requirements please visit
https://ps.stat.gov.pl/ForAuthors.

Przegląd Statystyczny
Statistical Review

Vol. 70 No. 2 2023

EDITORIAL OFFICE'S ADDRESS

University of Gdańsk, ul. Armii Krajowej 101, 81-824 Sopot, Poland

Language editing: Scientific Journals Division, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Maciej Adamowicz

Zakład Wydawnictw Statystycznych

Printed and bound: Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warsaw, Poland, zws.stat.gov.pl

Website: ps.stat.gov.pl

ISSN 0033-2372
e-ISSN 2657-9545
Index 371262

Information on the sales of the journal: Statistical Publishing Establishment
phone no. +48 22 608 32 10, +48 22 608 38 10

Order no. 311/2023 – 235 printed copies

CONTENTS

Przegląd Statystyczny. Statistical Review, 2023, vol. 70, 2, 1–19 https://doi.org/10.59139/ps.2023.02.1
Received: 3.07.2023, revised: 3.10.2023, accepted: 13.10.2023

Network analysis of the intra-EU trade

Anita Modrzejewska,[a] Piotr Biskup[b]

Abstract. The research presented in the paper used tools offered by network analysis and the graph theory. The study examined the empirical properties of the intra-EU trade network in the years 1999–2019 and confirmed that the EU trade links were of a disassortative nature. The use of network indicators has proven that the European trade network was characterised by a core-periphery structure. The study shows that Germany was the undisputed leader of the EU trade network over the studied years, although its central position was weakening over the years.

Keywords: network analysis, international trade, European integration, core-periphery structure
JEL: D85, F14, F15

1. Introduction

Network analysis serves to examine the relationships between interacting units, which requires the use of a different set of methods and analytical concepts from those applied in traditional statistics. It is not the attributes of autonomous individuals, but the associations between these attributes that are studied in search of patterns or regularities in the behaviour of the tested objects (Wasserman & Faust, 1994).

Another key feature of the network perspective is its symbiotic relationship between theory (network science) and method (social network analysis – SNA). Since network science is rooted in a wide range of academic disciplines, the research on networks is inter- and multidisciplinary by nature (Jackson, 2008). The fields where SNA is applied include the Internet, the WWW, train routes, airline connections, electronic circuits, semantic structures, biological systems, social interactions, and many others. The economics literature on networks has been thriving for the past 20–25 years. As a result, such economic systems as labour markets, banking sectors or the world economy began to be considered as networked structures. Thus, a network-based approach has been employed in empirical studies of international trade (Fagiolo et al., 2010). In this context, countries represent the nodes and an export/import relation between any two countries plays the role of a trade linkage.

[a] Krakow University of Economics, Doctoral school, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: amodrzejewska2000@wp.pl, ORCID: https://orcid.org/0009-0006-4086-7452.
[b] Cracow University of Technology, Faculty of Computer Science and Telecommunications, ul. Warszawska 24, 31-155 Kraków, Poland, e-mail: piotr.biskup@pk.edu.pl, ORCID: https://orcid.org/0000-0003-3742-8388.

According to Beaudreau (2004), the EU is an example of a trade network. Generally, the main idea of European integration is based on trade between nations. In this context, determining the specific characteristics of trading partners seems to be of crucial importance. The use of network indicators allows the measurement of the number and intensity of trading relationships, of the clustering level and the centrality of each node. The analysis of the structure of the intra-EU network and the search for some potential trade patterns is based on these indicators. This study, therefore, seeks to answer the following questions: are some countries more connected than others? Do well-connected nodes trade with partners that are also well-connected or rather less-connected? Which countries occupy the most central positions in the intra-EU trade network? Does the network have a core-periphery structure? Additionally, the evolution of the network over time was observed.

Determining such topological properties provides a more detailed insight into the European integration process. It may also uncover the logic behind the propagation of macroeconomic shocks through trade channels in times of financial crises. Moreover, it shows how much stress EU countries place on the intra-EU trade in their development strategy.

The remainder of this paper is organised as follows: Section 2 introduces the main tools of trade network analysis and Section 3 briefly describes the results of the empirical studies on international trade networks. The data and methodology used are discussed in Section 4, while Section 5 presents the findings and Section 6 the conclusions.

2. A statistical analysis of undirected trade networks

A trade network can be described by means of a graph consisting of nodes (countries) connected by a set of links (export/import relations).[1] The ordinary type of graph is binary and undirected, which means that between any two nodes a tie exists or not, and the direction of the link has no meaning. If any two nodes are connected by a link, then they are called ‘partners' or ‘nearest neighbours'.

A trade network can also take the form of a matrix. To formally characterise binary undirected networks (BUNs), it is sufficient to provide symmetric $N \times N$ binary adjacency matrix $\boldsymbol{A}$, where N is the number of countries, and whose elements $a_{ij} = a_{ji} = 1$ if there is an edge (trade relation) from vertex i to vertex j, and 0 otherwise. When $i = j$, 0 is also used, i.e. diagonal elements $a_{ii} = 0$.[2]

[1] Regardless of what flow of goods we take into account in the analysis, we can read the values of exports and imports from the same matrix for a given year, because exports from country *i* to country *j* should be equal to imports to country *j* from country *i*.

[2] It is assumed that no country trades with itself.

The notation used below is provided by Fagiolo et al. (2010), unless specified otherwise. The most common nodal statistic of the BUNs is the node degree (ND). It is defined as:

$$d_i = \sum_j a_{ij} = \boldsymbol{A}_{(i)}\mathbf{1}, \tag{1}$$

where $\boldsymbol{A}_{(i)}$ is the i-th row of matrix $\boldsymbol{A}$ and $\mathbf{1}$ is the N-dimensional column vector made of all ones. ND simply stands for the number of links that a given vertex has.

It is worth emphasising that any two nodes with the same ND can assume a different importance in the network, because the connections of their partners are also significant. To measure how many partners of node i are linked with one another, the average nearest-neighbour degree (ANND) may be computed, i.e. the average of the ND of all the partners of node i:

$$annd_i = d_i^{-1} \sum_j a_{ij} d_j = d_i^{-1} \sum_j \sum_h a_{ij} a_{jh} = \frac{\boldsymbol{A}_{(i)}\boldsymbol{A}\mathbf{1}}{\boldsymbol{A}_{(i)}\mathbf{1}}. \tag{2}$$

Another significant BUN statistic is the clustering coefficient (CC). It quantifies the average probability that two neighbours of a vertex are partners themselves. The CC of node i equals:

$$C_i(\boldsymbol{A}) = \frac{\sum_j \sum_h a_{ij} a_{ih} a_{jh}}{d_i(d_i - 1)} = \frac{(\boldsymbol{A}^3)_{ii}}{d_i(d_i - 1)}, \tag{3}$$

where $(\boldsymbol{A}^3)_{ii}$ is the i-th element of the main diagonal of $\boldsymbol{A}^3 = \boldsymbol{A} \cdot \boldsymbol{A} \cdot \boldsymbol{A}$. Each product $a_{ij} a_{ih} a_{jh}$ is meant to count whether a triangle exists around i or not. Notice that the order of the subscripts is irrelevant, as all entries in $\boldsymbol{A}$ are symmetric (Fagiolo, 2007).

In the case of BUNs, the stress is put on the mere presence or absence of an interaction between any two nodes. Since trade as an economic relationship can vary in intensity, the weighted undirected network (WUN) perspective is an appropriate approach. A WUN is defined by means of symmetric $N \times N$ 'weight' matrix $\boldsymbol{W}$, whose generic $w_{ij} = w_{ji} > 0$ entry measures the intensity of a trade relation between two countries, and it is 0 if no edge exists between them (Fagiolo et al., 2010).

The equivalents of ND, ANND and CC in WUN are: the node strength (NS), the average nearest-neighbour strength (ANNS), the weighted average nearest-neighbour degree (WANND) and the weighted clustering coefficient (WCC), respectively. The first statistic, NS, designates the sum of weights associated to the links that a given node has:

$$s_i = \sum_j w_{ij} = \boldsymbol{W}_{(i)}\mathbf{1}, \tag{4}$$

where $\boldsymbol{W}_{(i)}$ is the i-th row of matrix $\boldsymbol{W}$ and $\mathbf{1}$ is the N-dimensional column vector made of all ones. The larger the NS of a country, the more intense trade relations it bears. As can be seen, nodes with the same ND do not necessarily have the same NS.

To assess the strength of node i's partners, one may compute the ANNS:

$$anns_i = d_i^{-1}\sum_j a_{ij}s_j = d_i^{-1}\sum_j\sum_h a_{ij}w_{jh} = \frac{\boldsymbol{A}_{(i)}\boldsymbol{W}\mathbf{1}}{\boldsymbol{A}_{(i)}\mathbf{1}} \tag{5}$$

or WANND:

$$wannd_i = s_i^{-1}\sum_j w_{ij}d_j = s_i^{-1}\sum_j\sum_h w_{ij}a_{jh} = \frac{\boldsymbol{W}_{(i)}\boldsymbol{A}\mathbf{1}}{\boldsymbol{W}_{(i)}\mathbf{1}}, \tag{6}$$

which determine to what extent the partners of node i themselves are represented by high degree nodes. Similarly to ND and NS, vertices with the same ANND can be characterised by a different level of ANNS or WANND.

As far as the WCC is concerned, it takes into account the weights of all the edges in a triangle and it is defined as:

$$\tilde{C}_i(\boldsymbol{W}) = \frac{\sum_j\sum_h w_{ij}^{\frac{1}{3}}w_{ih}^{\frac{1}{3}}w_{jh}^{\frac{1}{3}}}{d_i(d_i-1)} = \frac{\left(\boldsymbol{W}^{\left[\frac{1}{3}\right]}\right)_{ii}^3}{d_i(d_i-1)}, \tag{7}$$

where $\boldsymbol{W}^{\left[\frac{1}{3}\right]} = \left[w_{ij}^{\frac{1}{3}}\right]$, i.e. the matrix obtained from $\boldsymbol{W}$ by taking the 3-rd root of each entry. The WCC can be considered as the geometric average of the subgraph edge weights based on the concept of subgraph intensity (Onnela et al., 2005). The resulting values are no longer probabilities, but they can be considered as proportional to probabilities.

One of the primary uses of SNA is the identification of the nodes which occupy the most central position in the network. There are many concepts and indicators of the 'centrality of a node' (Bloch et al., 2023). In this paper, however, we concentrate only on the index called random-walk betweenness centrality (RWBC) that fits both BUN and WUN analyses (Fagiolo et al., 2010):

$$RWBC_i = \frac{\sum_h \sum_{k \neq h} I_i(h,k)}{N(N-1)}, \tag{8}$$

where the current (i.e. the intensity of the interaction) originating from node h, flowing through node i and reaching node k equals:

$$I_i(h,k) = \frac{1}{2}\sum_j \left|v_i(h,k) - v_j(h,k)\right|. \tag{9}$$

Here, $I_h(h,k) = I_k(h,k) = 1$ (Fisher & Vega-Redondo, 2006) and v_i shows the intensity (i.e. the strength) of the flow through node i from source node h to target node k, which was represented by the voltage in the original paper of Newman (2005). RWBC can be applied to networks in international trade (Fisher & Vega--Redondo, 2006). RWBC measures how often a node is traversed by a random walk between two other nodes (Newman, 2005). Vertices with a high RWBC may thus affect the spread of trade flows across the network.

Finally, in order to adopt the undirected approach, the symmetry of matrices $\boldsymbol{A}$ and $\boldsymbol{W}$ must be checked. Since trade relationships (exports/imports) are directed by nature, we must assess empirically whether the observed network is sufficiently symmetric or not to justify a BUN/WUN analysis (Fagiolo, 2006). This process is discussed in Section 4.

3. Literature review

The concept of the empirical analysis of the structure of the international trade network and its topological properties, regardless of the social or economic relationships that might underlie them, originates from econophysics. The international trade network (ITN), also called the world trade web (WTW) or the world trade network (WTN), is most often examined. Since the origins of networks go back to complexity economics (Beinhocker, 2006), many researchers focus on the complex nature of the ITN. For instance, Serrano and Boguñá (2003), who studied the WTW using the binary approach, have proven that this network displays the typical properties of complex networks (scale-free degree distribution, the small-world property and a high CC). Moreover, the WTW follows a disassortative pattern, where well-connected vertices (countries with high NDs) are joined with less-connected vertices (i.e. with low ANNDs). Garlaschelli and Loffredo (2004, 2005), Fagiolo et al. (2010) and Squartini et al. (2011a) reach similar conclusions. Furthermore, they indicate that partners of well-connected nodes are less interconnected than those of poorly connected ones, implying the occurrence of some hierarchy in the network.

A more detailed picture of ITN delivers a weighted perspective of network analysis. The reason for this is that a binary approach does not fully take advantage of the wealth of information about the intensity of trade relations represented by each edge. Thus, it tends to underestimate the role of heterogeneity in trade linkages (Fagiolo et al., 2010). According to Bhattacharya et al. (2007, 2008), the world trade is dominated by only a few top rich countries (i.e. countries with high NSs) which control half of the world's trade volume. However, the size of the rich-club shrinks with time. Fagiolo et al. (2008) show that the average ND is relatively high, while the average NS is low, which means that the majority of the existing connections are relatively weak. The average ND-NS correlation coefficient is 0.5. The correlations of WANND and ANNS with ND and NS are negative, which confirms the disassortative nature of the WTW. Moreover, the WTW displays an increasing, positive and significant correlation between NS and WCC, which means that countries with more intense relationships are more likely to form strongly connected trade triangles. In other words, trade clubs (cliques) exist in the WTW. Serrano et al. (2007) also studied the network of trade imbalances between different pairs of countries (net producers which export more than they import and net consumers for which the opposite is true). They form the backbone of the WTW characterised by a high level of heterogeneity: for each country, the profile of trade fluxes is unevenly distributed across their partners. The properties of the trade fluxes of the WTW determine a ranking of trade partnerships that highlights global interdependencies. A different approach is followed by Fagiolo (2010). The author identifies the main determinants of international trade flows using a standard gravity equation and builds a 'residual' weighted trade network by removing the whole existing structure from the data. The purpose is to check whether this ITN exhibits topological features comparable to those of the original ITN. It appears that the residual ITN has power-law distributions of link weights and node statistics (e.g. NS, WCC and RWBC) in contrast to the original ITN, which is characterised by a log-normal distribution. In other words, the weighted ITN shows signs of complexity. Moreover, the correlations between the node statistics and the correlations between the node statistics and the national GDP *per capita* are calculated. As a result, a considerably divergent picture of the architecture of the world trade is created. While the original

ITN is geographically clustered and organised around a few large-sized hubs, the residual ITN consists of many small-sized but trade-oriented countries that, regardless of their geographical position, either play the role of local hubs or attract large and rich countries in relatively complex trade-interaction patterns. The only known contribution to network analysis of the intra-EU trade is the paper of Modrzejewska and Pajor (2011). They indicate that the European trade network (ETN) also fits into a disassortative pattern of trade relationships: the correlations between ND and ANND and between NS and ANNS are strongly negative. In addition, the values of RWBC confirm the existence of a core-periphery structure at the European level, as is the case at the global level (Fagiolo et al., 2010).

Last but not least, the issue of the directed or undirected nature of the ITN should be mentioned. Researchers develop various strategies in this regard. Some examine either both the directed and undirected trade network (e.g. Squartini et al., 2011a, 2011b), or the directed trade network only (e.g. Garlaschelli & Loffredo, 2004), while others explore the conditions under which one can investigate the trade relationships as an undirected network (e.g. Fagiolo, 2006, Fagiolo et al., 2008, Fagiolo et al., 2010). The last method has been employed in this paper.

4. Data and methodology

The analysis of the intra-EU trade network is based on aggregate bilateral merchandise import data measured in current U.S. dollars, provided by the UN Comtrade Database. The database reported on 28 countries ($N = 28$) from 1999 to 2019 ($T = 21$).[3] The GDP and GDP *per capita* data were extracted from the website of the World Bank. They were also provided in current U.S. dollars.

The methodology adopted in the study is similar to that applied by Fagiolo et al. (2010) and Modrzejewska and Pajor (2011). The first step involved building a sequence of weighted adjacency matrices of the weighted directed networks defined for the years 1999–2019, in which rows represented the exporting countries and the columns referred to the importing countries. Secondly, the export and import values were divided by the GDP values of the exporter and the importer, respectively.[4] Let $\widetilde{\boldsymbol{W}} = [\widetilde{w}_{ij}]$ be a $N \times N$ weight matrix, where $\widetilde{w}_{ij} \in [0,1]$ and $\widetilde{w}_{ii} = 0$ for all i. Thus, weighted matrix $\widetilde{\boldsymbol{W}} = [\widetilde{w}_{ij}]$ was created for each year and flow type (exports and imports). Thirdly, a 'trade relationship' was defined by

[3] The study covers 28 EU countries, although the actual number of the EU members varied over time. In this analysis only import data were used because of their greater accuracy than those relating to export (Kim & Shin, 2002). However, due to the missing values of the imports, the export data were employed in the following cases: exports from Luxembourg to the Netherlands in 1999 instead of imports to the Netherlands from Luxembourg in 1999.

[4] Network parameters for exports were obtained by calculating them by rows, and by columns for imports.

setting 1 for all non-zero elements of matrix $\widetilde{\boldsymbol{W}}$. In this way, adjacency matrix $\widetilde{\boldsymbol{A}}$ was obtained, which enabled a binary analysis. Finally, the symmetry of the empirically-observed weighted network was verified using the method suggested by Fagiolo (2006). For this purpose, symmetry index $\tilde{S}(Q)$ was calculated (Fagiolo et al., 2010):

$$\tilde{S}(Q) = 1 - \frac{\sum_i \sum_j q_{ij} q_{ji}}{\sum_i \sum_j q_{ij}^2}, \tag{10}$$

and

$$Q = \{q_{ij}\} = \widetilde{\boldsymbol{W}} - \left(1 - \widetilde{\boldsymbol{W}}\right)\mathbf{I}_{\mathrm{N}}, \tag{11}$$

where $\mathbf{I}_{\mathrm{N}}$ was the $N \times N$ identity matrix and q_{ij}=$\widetilde{w}_{ij}$ for all $i \neq j$, while then $q_{ij} = 1$ for all i. The scaled version of $\tilde{S}(Q)$ ranged from 0 (full symmetry) to 1 (full asymmetry):

$$S(Q) = \frac{N+1}{N-1}\tilde{S}(Q). \tag{12}$$

Hence, the values of $S(Q)$ close to 0 justified a BUN/WUN analysis.[5] Therefore, the statistical properties of symmetrised ETNs would be explored. In the binary case, any entry a_{ij} of the new adjacency matrix $\boldsymbol{A}$ was set to 1 if and only if either $\tilde{a}_{ij} = 1$ or $\tilde{a}_{ji} = 1$ (and 0 otherwise). In the weighted case, the generic entry of the new weight matrix W was replaced by:

$$w_{ij}^t = \frac{1}{2}\left(\widetilde{w}_{ij}^t + \widetilde{w}_{ji}^t\right) = \frac{1}{2}\left(\frac{e_{ij}^t}{GDP_i^t} + \frac{e_{ji}^t}{GDP_j^t}\right), \tag{13}$$

where e_{ij}^t stood for the export value from country i to country j in year t.

By analogy, import values were calculated. $(imp)_{ij}^t$ was inserted instead to describe the import value to country i from country j.

All calculations presented below were done in R and Excel.

[5] The threshold of the index may be arbitrarily decided by the researcher (Fagiolo, 2006). In our case it was very small and ranged from 0.000996 to 0.001799 in exports and from 0.001284 to 0.002268 in imports, respectively.

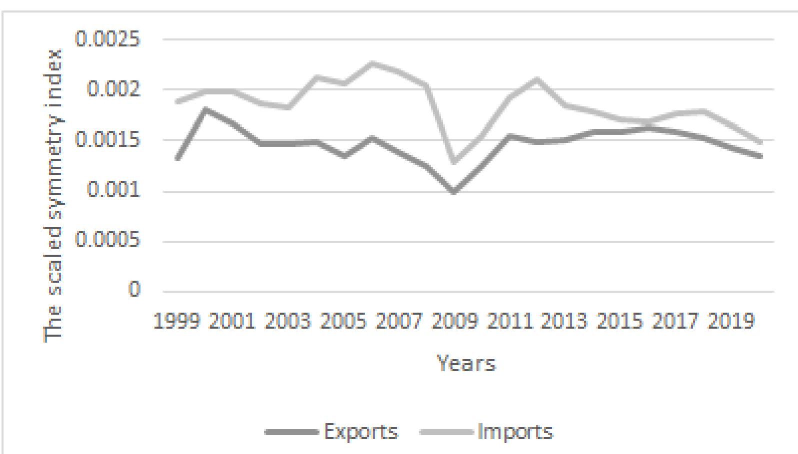
5. Findings

It should be noted that at the outset, the BUN analysis was deliberately omitted in this study, as all pairs of EU countries traded with each other during all the reported years. As a result, ND, ANND and CC took the maximum values of 27 for each country. Thus, the focus was on a WUN analysis instead.

To begin with, the symmetry of matrix $\widetilde{\boldsymbol{W}}$ was checked. As Figure 1 shows, the scaled symmetry index was close to 0 for both exports and imports. This justified a WUN approach. The obtained results indicated that the value of exports from country i to country j equalled approximately the value of exports from country j to country i, and the same applied to imports.

The distribution of NS among countries (in both exports and imports) was right-skewed with the majority of nodes characterised by weak trade relationships (see Table 1 and Table 2). The average NS oscillated between 0.2 and 0.3 (see Figure 2). The most intense trade relations were observed in Germany throughout all the studied years. Its average NS was 0.96 in exports and 1.07 in imports. At the opposite pole were Greece (the minimum value of 0.08 in exports was reported in 2006 and 0.12 in imports in 2010), Luxembourg (0.12 in imports in 2016) and Cyprus (minimum values were noted during all the remaining years in both exports and imports). These findings were consistent with the economic performance of the EU member states. Germany was the undisputed leader in the intra-EU trade. Cyprus, Greece and Luxembourg are small countries rather known for being service providers.

Figure 1. Scaled symmetry index S(Q) for WUN



Source: authors' calculation.

Figure 2. Average NS for WUN

Source: authors' calculation.

The average ANNS took similar values as the average NS (see Figure 3), whereas the WANND was much the same as the ANND and equalled 27 for all nodes in each year. There was a strong negative correlation between NS and ANNS ($r = -1$), which confirmed that the ETN was of a disassortative nature. This means that the probability of a connection between a high-strength and a low-strength node was greater than expected if the network were completely random. In other words, countries that were more closely connected (i.e. the hubs) tended to form trade relationships with more weakly connected countries. This suggests a core-periphery structure of the ETN, at least in terms of link intensity.

Table 1. Node strength of the EU countries in exports in selected years

Country	Years		
	1999	2004	2019
AT – Austria	0.199761	0.252976	0.247008
BE – Belgium	0.368554	0.453586	0.382216
BG – Bulgaria	0.102833	0.168899	0.190895
CZ – Czech Republic	0.233559	0.284407	0.392694
CY – Cyprus	0.033932	0.037207	0.048584
DE – Germany	0.900965	0.927677	0.998063
DK – Denmark	0.149741	0.162462	0.155792
EE – Estonia	0.224115	0.246399	0.194926
EL – Greece	0.062589	0.070991	0.087775
ES – Spain	0.181030	0.204553	0.223817
FI – Finland	0.178373	0.184311	0.157425
FR – France	0.369956	0.408882	0.365435
HR – Croatia	0.084321	0.088367	0.131615
HU – Hungary	0.255647	0.270027	0.382684
IE – Ireland	0.255070	0.235992	0.135715
IT – Italy	0.346196	0.408287	0.361977

Table 1. Node strength of the EU countries in exports in selected years (cont.)

Country	Years 1999	Years 2004	Years 2019
LT – Lithuania	0.136734	0.173424	0.235332
LU – Luxembourg	0.131838	0.195042	0.099752
LV – Latvia	0.161128	0.189266	0.197908
MT – Malta	0.114869	0.118795	0.072969
NL – Netherlands	0.327050	0.344058	0.412007
PL – Poland	0.137458	0.207207	0.309558
PT – Portugal	0.106953	0.111030	0.129315
RO – Romania	0.110583	0.156554	0.208952
SE – Sweden	0.220259	0.247660	0.226578
SI – Slovenia	0.177527	0.190962	0.280955
SK – Slovakia	0.182494	0.247167	0.374168
UK – United Kingdom	0.396681	0.408636	0.305586

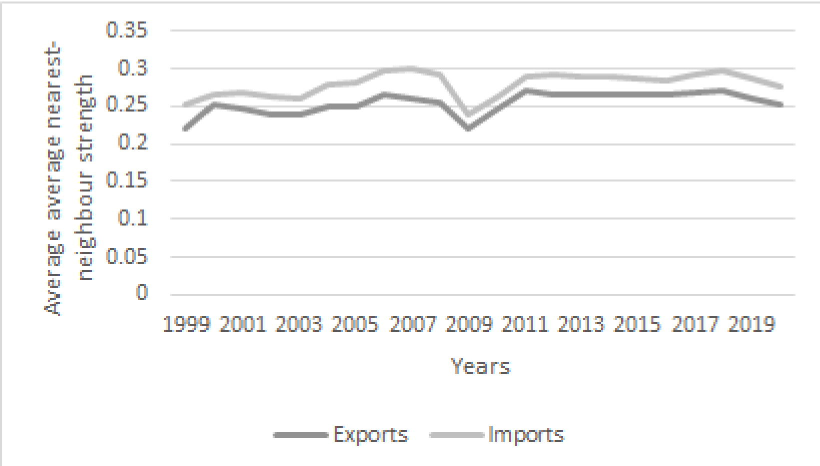
Source: authors' calculations.

Table 2. Node strength of the EU countries in imports in selected years

Country	Years 1999	Years 2004	Years 2019
AT – Austria	0.241303	0.278698	0.278546
BE – Belgium	0.429089	0.500840	0.404789
BG – Bulgaria	0.117564	0.169863	0.198306
CZ – Czech Republic	0.246599	0.303680	0.374606
CY – Cyprus	0.096285	0.113336	0.120678
DE – Germany	0.978395	1.116478	1.055845
DK – Denmark	0.163265	0.167818	0.156561
EE – Estonia	0.255140	0.223365	0.222480
EL – Greece	0.111615	0.120272	0.147078
ES – Spain	0.206490	0.237096	0.261334
FI – Finland	0.218330	0.180263	0.152572
FR – France	0.448696	0.458344	0.358984
HR – Croatia	0.130989	0.154939	0.211148
HU – Hungary	0.253992	0.284634	0.383435
IE – Ireland	0.177845	0.155241	0.129759
IT – Italy	0.439204	0.498765	0.438123
LT – Lithuania	0.172332	0.218377	0.302098
LU – Luxembourg	0.202884	0.228084	0.134828
LV – Latvia	0.168195	0.204835	0.240402
MT – Malta	0.227787	0.257335	0.174235
NL – Netherlands	0.320785	0.370702	0.437496
PL – Poland	0.170463	0.237179	0.367274
PT – Portugal	0.142948	0.140405	0.168788
RO – Romania	0.112606	0.173661	0.214797
SE – Sweden	0.226000	0.231957	0.218522
SI – Slovenia	0.206125	0.224799	0.274649
SK – Slovakia	0.168258	0.222992	0.328852
UK – United Kingdom	0.389917	0.333187	0.285058

Source: authors' calculations.

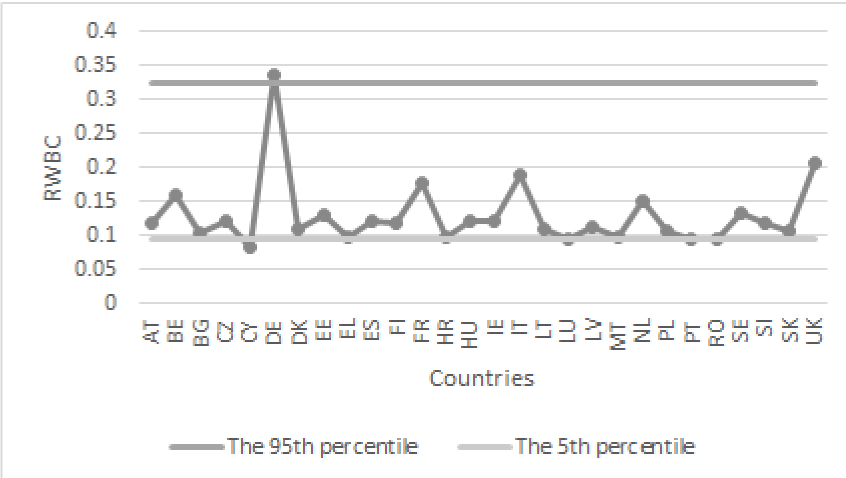
Figure 3. Average ANNS for WUN



Source: authors' calculation.

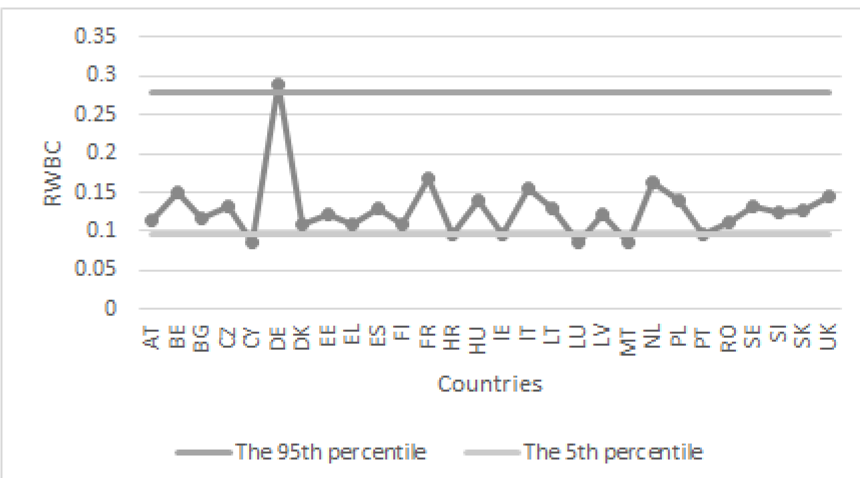
The next feature explored was the level of clustering. This raised the question of whether countries holding more intense trade relationships showed a tendency to trade with pairs of countries that themselves exchanged goods with each other. The answer was no. The average WCC ranged from 0.005 to 0.007 in exports and from 0.006 to 0.007 in imports. As regards the correlation between WCC and NS, a strong positive relationship was proven (the average of $r = 0.97$ in both exports and imports). Therefore, countries with a high NS were typically involved in highly-interconnected triples (Fagiolo et al., 2010). This pattern of nodes' behaviour could imply the occurrence of the 'rich club phenomenon'. However, the NS-GDP *per capita* and the WCC-GDP *per capita* correlations did not support this interpretation. The correlation between NS and GDP *per capita* was decreasing over time from 0.34 in 1999 to −0.02 in 2019 in exports and from 0.35 in 1999 to −0.05 in 2019 in imports. The correlation between WCC and GDP *per capita* took similar values varying from 0.39 in 1999 to −0.017 in 2019 in exports and from 0.35 in 1999 to −0.066 in 2019 in imports, respectively.

Figure 4. RWBC for WUN in exports in 1999



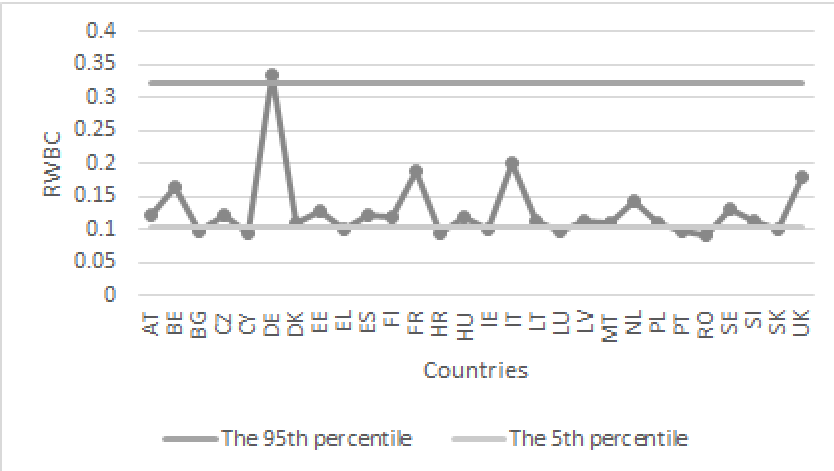
Source: authors' calculation.

Figure 5. RWBC for WUN in exports in 2019



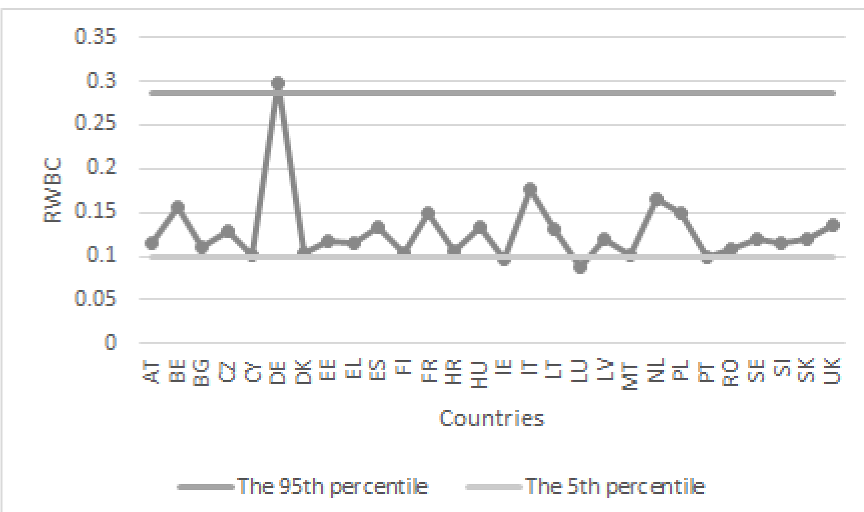
Source: authors' calculation.

Figure 6. RWBC for WUN in imports in 1999



Source: authors' calculation.

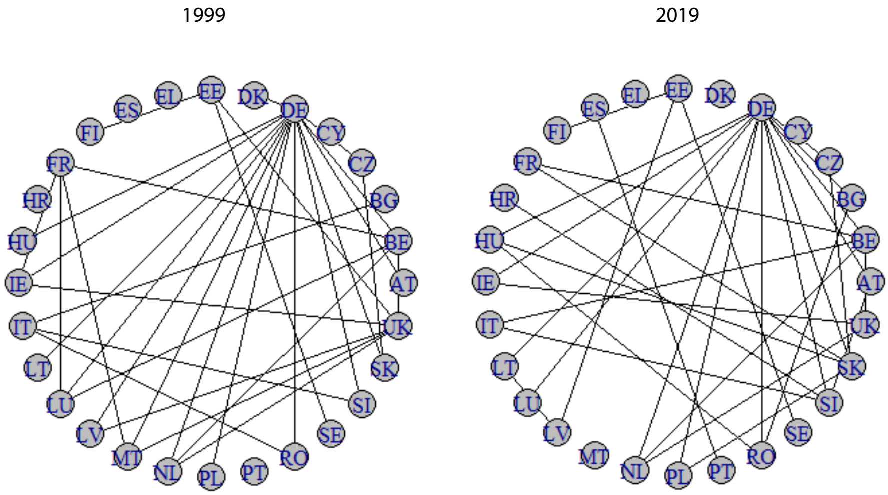
Figure 7. RWBC for WUN in imports in 2019



Source: authors' calculation.

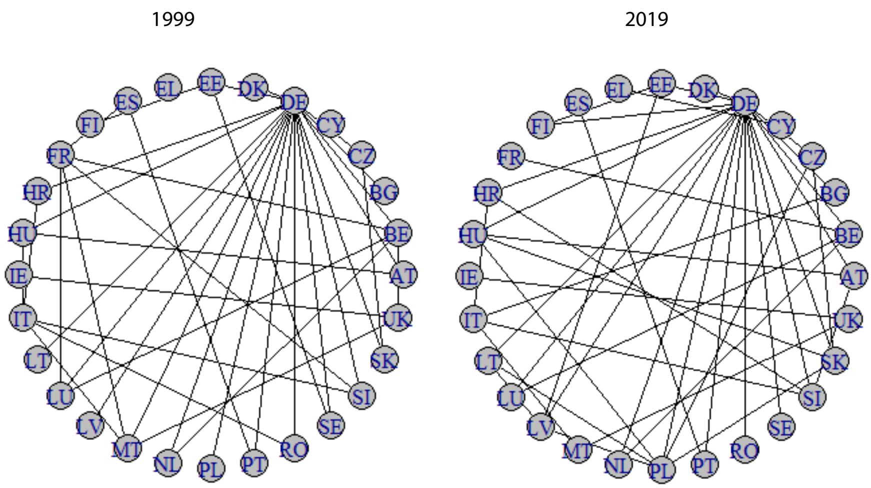
As Figures 4, 5, 6 and 7 show, the core-periphery structure of the ETN was confirmed. The RWBC index measures the likelihood of a given country to appear in a randomly selected trade chain within the network. This likelihood is determined by the number and intensity of trade relationships (Fagiolo et al., 2010). Germany occupied the most central position in the intra-EU network in both exports and imports during all the studied years. It means that this country was the most influential in the network due to a high number of direct and intense trade connections. To identify the core, a method proposed by Fagiolo et al. (2010) and Modrzejewska and Pajor (2011) was applied, i.e. a threshold at the 95th percentile of RWBC was imposed. Germany was the only country located in the core, i.e. within the top 5% of the EU countries. Nevertheless, it should be emphasised that the RWBC index for Germany was shrinking over time. It amounted to 0.34 in 1999 and 0.29 in 2019 in exports, and 0.33 in 1999 and 0.30 in 2019 in imports. In other words, dissimilarities among the EU countries decreased between 1999 and 2019. This was also proven in Modrzejewska and Pajor (2011). The nodes below the threshold of the 5th percentile of RWBC are considered the periphery. The following countries belonged to this area in various years: Croatia, Cyprus, Greece, Ireland, Luxembourg, Malta, Portugal and Romania in exports, and Bulgaria, Croatia, Cyprus, Denmark, Greece, Ireland, Luxembourg, Malta, Portugal, Romania and Slovakia in imports. The remaining countries were situated in an intermediate periphery. Some of them (e.g. the Czech Republic, Hungary, Lithuania, Latvia, the Netherlands, Poland and Slovenia in exports, and Hungary, Lithuania, the Netherlands, Poland and Spain in imports) improved their positions in the EU trade network over the years, whereas France, Italy and the United Kingdom suffered a sharp decline of RWBC in both exports and imports. Finally, the correlation between RWBC and GDP *per capita* was calculated. It revealed a similar pattern to that observed in the case of the correlation between NS and GDP *per capita*, and WCC and GDP *per capita*.

Figure 8. ETN in exports



Source: authors' calculation.

Figure 9. ETN in imports



Source: authors' calculation.

Figures 8 and 9 presenting intra-EU trade networks clearly illustrate the differences in the intensity of trade flows among EU countries. Only the flows above 4% of GDP are marked. Between 1999 and 2019, the structure of intra-EU trade did not change considerably. Some connections decreased in intensity, e.g. France maintained intense trade relations in exports with Belgium, Ireland, Luxembourg and Malta in 1999, while in 2019 with Belgium and Slovakia. In the case of imports, France had strong trade relationships with Belgium, Luxembourg, Malta and Slovenia in 1999, while in 2019, only with Belgium.

6. Conclusions

In this paper, the statistical properties of the intra-EU trade network were examined. For this purpose, data concerning import connections between all pairs of 28 EU countries from 1999 to 2019 were analysed. The ETN was conceptualised as a weighted network where countries are represented by nodes and export/import flows divided by GDP of a given country by links. Since the ETN is a symmetric network (i.e. all trade relationships appear to be reciprocated with similar intensities), the application of the WUN approach was justified. Despite a high density of the network (all pairs of countries exchanged goods with each other during all the years reported), the average NS did not exceed 0.3. The only country whose average NS oscillated around 1 was Germany. The strong negative correlation between NS and ANNS confirmed a disassortative nature of the intra-EU network, i.e. countries holding more intense trade relationships tended to exchange goods with those less-connected (having less intense trade relations). This suggests a core-periphery structure of the ETN, which was also proven by the results of the RWBC index. Germany again turned out to be the most influential country in the EU trade network. However, its central position in the ETN weakened over the years. A similar situation occurred for France, Italy and the United Kingdom. Their RWBC values decreased between 1999 and 2019. The only exception among the EU's strongest economies was the Netherlands, whose position strengthened in both exports and imports within those years. The countries situated in the periphery of the EU trade network were either small economies (e.g. Cyprus, Malta) or those specialising in trade in services (e.g. Ireland, Luxembourg). The remaining EU member states were located in the semi-periphery of the ETN, where most of the countries of Central and Eastern Europe belonged to. In general, they improved their positions in the EU trade network over the years. As Figure 8 and 9 suggest, the structure of the ETN did not change significantly. Between 1999 and 2019 the intensity of some trade connections decreased (e.g. France-Luxembourg in exports), and some increased (e.g. Estonia-Latvia in imports). Furthermore, very few countries with a high NS were involved in highly-interconnected trade triples. The average WCC was rather poor. Finally, the relationships between network properties and country income (GDP *per capita*) were studied. The 'rich club phenomenon' present in the WTW was not confirmed at the European level.

As mentioned before, this work represents a preliminary step towards the understanding of the topological properties of the ETN and its dynamics. It provides an opportunity for many possible extensions. For instance, one may consider exploring the commodity-specific trade flows in search of specialisation patterns of countries belonging to the EU trade network. Furthermore, one can examine whether the topological properties of the ETN have some policy implications for the logic of European integration.

References

Beaudreau, B. C. (2004). *World Trade. A Network Approach*. iUniverse.

Beinhocker, E. D. (2006). *The Origin of Wealth. Evolution, Complexity, and the Radical Remaking of Economics*. Harvard Business School Press.

Bhattacharya, K., Mukherjee, G., & Manna, S. S. (2007). The international trade network. In A. Chatterjee & B. K. Chakrabarti (Eds.), *Econophysics of Markets and Business Networks: Proceedings of the Econophys-Kolkata III* (pp. 139–147). Springer.

Bhattacharya, K., Mukherjee, G., Saramäki. J., Kaski, K., & Manna, S. S. (2008). The international trade network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, (2), 1–5. https://doi.org/10.1088/1742-5468/2008/02/P02002.

Bloch, F., Jackson, M. O., & Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, *61*(2), 413–453. https://doi.org/10.1007/s00355-023-01456-4.

Fagiolo, G. (2006, December 1). Directed or Undirected? A New Index to Check for Directionality of Relations in Socio-Economic Networks. *Economics Bulletin*, *3*(34), 1–12. https://doi.org/10.48550/arXiv.physics/0612017.

Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E*, *76*(2), 1–16. https://doi.org/10.1103/PhysRevE.76.026107.

Fagiolo, G. (2010). The international-trade network: gravity equations and topological properties. *Journal of Economic Interaction and Coordination*, *5*(1), 1–25. https://doi.org/10.1007/s11403-010-0061-y.

Fagiolo, G., Reyes, J., & Schiavo, S. (2008). On the topological properties of the world trade web: A weighted network analysis. *Physica A: Statistical Mechanics and its Applications*, *387*(15), 3868–3873. https://doi.org/10.1016/j.physa.2008.01.050.

Fagiolo, G., Reyes, J., & Schiavo, S. (2010). The evolution of the world trade web: a weighted network analysis. *Journal of Evolutionary Economics*, *20*(4), 479–514. https://doi.org/10.1007/s00191-009-0160-x.

Fisher, E., & Vega-Redondo, F. (2006). *The linchpins of a modern economy*. 2007 AEA Annual Meeting, Chicago. https://www.aeaweb.org/annual_mtg_papers/2007/0107_0800_1502.pdf.

Garlaschelli, D., & Loffredo, M. I. (2004). Fitness-dependent topological properties of the world trade web. *Physical Review Letters*, *93*(18), 1–4. https://doi.org/10.1103/PhysRevLett.93.188701.

Garlaschelli, D., & Loffredo, M. I. (2005). Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications*, *355*(1), 138–144. https://doi.org/10.1016/j.physa.2005.02.075.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press. https://web.stanford.edu/~jacksonm/netbook.pdf.

Kim, S., & Shin, E. H. (2002). A Longitudinal Analysis of Globalization and Regionalization in International Trade: A Social Network Approach. *Social Forces*, *81*(2), 445–468. https://doi.org/10.1353/sof.2003.0014.

Modrzejewska, A., & Pajor, A. (2011). Analiza sieciowa struktury obrotów handlowych w UE. In: S. Partycki (Ed.), *Społeczeństwo sieci. Gospodarka sieciowa w Europie Środkowo-Wschodniej* (vol. 1; pp. 643–656). Wydawnictwo KUL.

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*(1), 39–54. https://doi.org/10.1016/j.socnet.2004.11.009.

Onnela, J., Saramaki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, *71*(6), 1–5. https://doi.org/10.1103/PhysRevE.71.065103.

Serrano, M. Á., & Boguñá, M. (2003). Topology of the world trade web. *Physical Review E*, *68*(1), 1–5. https://doi.org/10.1103/PhysRevE.68.015101.

Serrano, M. Á., Boguñá, M., & Vespignani, A. (2007). Patterns of dominant flows in the world trade web. *Journal of Economic Interaction and Coordination*, *2*(2), 111–124. https://doi.org/10.1007/s11403-007-0026-y.

Squartini, T., Fagiolo, G., & Garlaschelli, D. (2011a). Randomizing world trade. I. A binary network analysis. *Physical Review E*, *84*(4), 1–17. https://doi.org/10.1103/PhysRevE.84.046117.

Squartini, T., Fagiolo, G., & Garlaschelli, D. (2011b). Randomizing world trade. II. A weighted network analysis. *Physical Review E*, *84*(4), 1–16. https://doi.org/10.1103/PhysRevE.84.046118.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478.

Przegląd Statystyczny. Statistical Review, 2023, vol. 70, 2, 20–45 https://doi.org/10.59139/ps.2023.02.2
Received: 22.07.2023, revised: 2.10.2023, accepted: 13.10.2023

Some asymptotic results of the estimators for conditional mode for functional data in the single index model missing data at random

Souad Mekkaoui,[a] Nadia Kadiri,[b] Abbes Rabhi[c]

Abstract. In this work, we consider the problem of non-parametric estimation of a regression function, namely the conditional density and the conditional mode in a single functional index model (SFIM) with randomly missing data. The main result of this work is the establishment of the asymptotic properties of the estimator, such as almost complete convergence rates. Moreover, the asymptotic normality of the constructs is obtained under certain mild conditions. We finally discuss how to apply our result to construct confidence intervals.

Keywords: functional data analysis, functional single-index process, kernel estimator, missing at random, non-parametric estimation, small ball probability

JEL: C10, C13, C14, C19, C24

1. Introduction

The Single Index Model (SIM) is a popular framework used for reducing dimensionality and modelling complex relationships between covariates and responses in a simplified way. When dealing with functional data, where each observation is a curve or a function, the SIM is extended to handle functional predictors and responses. When dealing with missing data in the SIM framework, the missingness is assumed to be at random (MAR). This means that the probability of missing values is related to the observed data but not to the missing values themselves. The key idea is that, given the observed data, the missingness mechanism is unrelated to the values that are missing. It is important to note that the choice of approach depends on the specifics of your data, the extent of the missingness, and the assumptions you are willing to make. A careful consideration of the nature of your data and consulting domain experts when handling missing data in the SIM or any other modeling framework is always recommended.

[a] Djillali LIABES University of Sidi Bel Abbes, Laboratory of Mathematics, Algeria, ORCID: https://orcid.org/0009-0005-8611-3333.

[b] Djillali LIABES University of Sidi Bel Abbes, Higher School of Economics of Oran 22000, Sidi Bel Abbes, Algeria, e-mail: nad.kad06@yahoo.com, ORCID: https://orcid.org/0000-0002-3405-7414.

[c] Djillali LIABES University of Sidi Bel Abbes, Laboratory of Mathematics, Algeria, ORCID: https://orcid.org/0000-0001-6740-0226.

The asymptotic properties of semi-parametric estimators of the conditional mode for functional data in the Single Index Model (SIM) with data missing at random (MAR) are an active area of research, and specific results may depend on the particular assumptions and estimation methods employed. However, this work provides a general overview of some relevant concepts and approaches in this context. In the SIM framework, functional data refers to observations that are functions rather than scalar values. The goal is to estimate the conditional mode of a functional-response variable given a set of functional predictors and a single-index variable.

To establish the asymptotic properties of the semi-parametric estimators of the conditional mode for functional data in the SIM with data missing at random, various theoretical conditions need to be satisfied. These conditions often involve assumptions about the functional data, the missing data mechanism, and the model specification. Some common conditions include consistency and efficiency. Specific results in this area may depend on the assumptions and estimation techniques employed in each study. Therefore, it is important to refer to the literature and research articles that focus on the specific estimation method and assumptions one is interested in to obtain more detailed and precise asymptotic properties of the estimators.

One of the most common problems in non-parametric statistics is forecasting. In some situations, regression is the best forecasting tool. Sometimes, however, e.g. in the case where the conditional density is asymmetrical or multimodal, this tool is inadequate. Therefore, the conditional quantile predicts the impact of the variable of interest Y on the explanatory variable X more efficiently. There is scarce literature investigating the statistical properties of a functional non-parametric regression model for missing data when the explanatory variable is infinite dimensional or it is of a functional nature. Recently, Ferraty et al. (2013) proposed to estimate the mean of a scalar response based on an independent and identically distributed (i.i.d) functional sample in which explanatory variables were observed for every subject, and a part of the responses were missing at random (MAR) for some of them. It generalised the results in Cheng (1994) to the case where the explanatory variables are of a functional nature.

To the best of our knowledge, the estimation of a non-parametric conditional distribution in the functional single index structure combining missing data and stationary processes of a functional nature has not yet been studied in statistical literature. Therefore, in this work we investigate a conditional quantile estimation when the data are MAR. Our aim is to develop a functional methodology for dealing with MAR samples in non-parametric problems (namely in the conditional quantile estimation). Then, the asymptotic properties of the estimator are obtained under some mild conditions. Our study considers a model in which the response variable is missing.

2. Literature review

Therefore, within this framework, the independence of the variables was assumed. As far as we know, the estimation of a conditional quantile combining censored data, an independent theory and functional data with single-index structure has not been studied in statistical literature yet. Our paper extends the work of Ling et al. (2015, 2016) and Mekki et al. (2021) to the functional single-index model case.

For the above-mentioned theoretical and application reasons, the statistical community has displayed a great interest in estimating conditional quantiles, especially the conditional median function, as an interesting, alternative predictor to the conditional mean (thanks to its robustness to the presence of outliers) (see Chaudhuri et al., 1997). The estimation of the conditional mode of a scalar response given a functional covariate has attracted the attention of many researchers. Ferraty et al. (2005) introduced a non-parametric estimator of the conditional quantile, defined as the inverse of the conditional distribution function when data are dependent. Ezzahrioui and Ould-Saïd (2008) established the asymptotic normality of the kernel conditional mode estimator. In the censored case, Ould-Saïd and Cai (2005) established a uniform strong consistency of the kernel estimator for the conditional mode function. In this context, we recommend referring to Lemdani et al. (2009) for the estimation of conditional quantiles. Other authors have been interested in the estimation of conditional models when the observations were censored or truncated, eg. Hamri et al. (2022), Liang and de Uña-Alvarez (2010), Ould-Saïd and Tatachak (2011), Ould-Saïd and Yahia (2011), Rabhi et al. (2021), etc.

For instance, Aït-Saidi et al. (2008) were interested in using SFIM to estimate the regression operator, and suggested using a cross-validation procedure allowing the estimation of the unknown link function as well as the unknown functional index. Attaoui (2014) and Attaoui and Ling (2016) studied, respectively, the estimation of the conditional density and the conditional cumulative distribution function based on a SFIM with the assumption that the data satisfy a strong mixing condition. Kadiri et al. (2018) studied the asymptotic properties of the kernel-type estimator of the conditional quantiles when the response was right-censored and the data was sampled from a strong mixing process.

The remaining part of the paper is arranged in the following way: in Section 3, we present the non-parametric estimator of the functional conditional model when the data are MAR. In Section 4, we make assumptions for the theoretical study. The point-wise almost-complete convergence and the uniform almost-complete convergence of the kernel estimator for our models (with rates) are established in Section 5.

3. Model and estimator

3.1. The functional non-parametric framework

Consider a random pair (X, Y), where Y is valued in $\mathbb{R}$ and X is valued in some infinite dimensional Hilbertian space $\mathcal{H}$ with a scalar product $<\cdot, \cdot>$. Let the $(X_i, Y_i)_{i=1,\ldots,n}$ be the statistical sample of pairs which are identically distributed like (X, Y), but not necessarily independent. Henceforward, X will be called functional random variable *f.r.v*. Let x be fixed in $\mathcal{H}$and let $F(\theta, y, x)$ be the conditional cumulative distribution function *(cond-cdf)* of T given $<\theta, X> = < \theta, x >$, specifically:

$$\forall y \in \mathbb{R},\ F\ (\theta,y,x) = \mathbb{P}(Y \leq y | <\theta, X> = <\theta, x>).$$

By the above, we are implicitly assuming the existence of a regular version of the conditional distribution Y, given $<\theta, X> = <\theta, x>$.

In our infinite dimensional purpose, we use the 'functional non-parametric' term, where the word 'functional' refers to the infinite dimensionality of the data, and the word 'non-parametric' denotes the infinite dimensionality of the model. Such 'functional non-parametric' statistics is also called 'doubly infinite dimensional' (see Ferraty & Vieu, 2003, for more details). We also use the 'operational statistics' term, since the target object to be estimated (the *cond-df* $f(\theta, ., x)$) can be viewed as a non-linear operator.

3.2. The estimators

In the case of complete data, the kernel estimator $\tilde{f}_n(\theta, ., x)$of $f(\theta, ., x)$ is presented as follows:

$$\tilde{f}(\theta, t, x) = \frac{g_n^{-1} \sum_{i=1}^{n} K\left(h_n^{-1}(|<x-X_i,\theta>|)\right) H\left(g_n^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n} K\left(h_n^{-1}(<x-X_i,\theta>)\right)},$$

where K and H are kernel functions, and h_n(resp. g_n) is a sequence of positive real numbers. Note that using similar ideas, Roussas (1969) introduced some related estimates, but in the special case where X was real, while Samanta (1989) produced an earlier asymptotic study on the subject.

Meanwhile, in an incomplete case with data missing at random for the response variable, we observe $(X_i, Y_i, \delta_i)_{1\leq i\leq n}$, where X_i is observed completely, and $\delta_i = 1$ if Y_i and $\delta_i = 0$ otherwise. We define the Bernoulli random variable δ by

$$\mathbb{P}(\delta = 1|\langle X,\theta\rangle = \langle x,\theta\rangle, Y = y) = \mathbb{P}(\delta = 1|\langle X,\theta\rangle = \langle x,\theta\rangle) = p(x,\theta),$$

where $p(x,\theta)$ is a functional operator which is conditional only on X.

Therefore, the estimator of $f(\theta, y, x)$ in the single-index model with response MAR is presented as

$$\hat{f}(\theta,t,x) = \frac{g_n^{-1}\sum_{i=1}^{n}\delta_i K\left(h_n^{-1}(|<x-X_i,\theta>|)\right)H\left(g_n^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n}\delta_i K\left(h_n^{-1}(<x-X_i,\theta>)\right)} = \frac{\hat{f}_N(\theta,y,x)}{\hat{f}_D(\theta,x)},$$

where $K_i(\theta,x) := \mathrm{K}(h_n^{-1}|< x - X_i,\theta >|)$,$H_i(y) = H\left(g_n^{-1}(y - Y_i)\right)$,

$\hat{f}_D(\theta,x) = \frac{\sum_{i=1}^{n}\delta_i K_i(\theta,x)}{n\mathbb{E}(K_1(\theta,x))}$, and $\hat{f}_N(\theta,y,x) = \frac{\sum_{i=1}^{n}\delta_i K_i(\theta,x)H_i(y)}{ng_n\mathbb{E}(K_1(\theta,x))}$.

3.3. Assumptions on the functional variable

Let N_x be a fixed neighborhood of x in $\mathcal{H}$ and let $B_\theta(x,h)$ be a ball of center x and radius h, namely $B_\theta(x,h) = \{\chi\in\mathcal{H}: 0 <| <x - \chi,\theta >| < h\}$, $d_\theta(x,X_i) = | < x - X_i,\theta>|$ denote a random variable such that its cumulative distribution function is given by $\phi_{\theta,x}(u) = \mathbb{P}(d_\theta(x,X_i) \leq u) = \mathbb{P}\left(X_i \in B_\theta(x,u)\right)$.

Now, let us consider the following basic assumptions that are necessary in deriving the main result of this paper.

(H1) $\mathbb{P}\left(X \in B_\theta(x,h_n)\right) =: \phi_{\theta,x}(h_n) > 0; \phi_{\theta,x}(h_n) \longrightarrow 0$ as$h_n \to 0$.

3.4. The non-parametric model

As is usually the case in non-parametric estimation, we suppose that the *cond-dff*$(\theta,.,x)$ verifies some smoothness constraints. Let α_1 and α_2 be two positive numbers, such that

(H2) $\forall(x_1,x_2) \in N_x \times N_x, \forall(y_1,y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}$

(i) $|f(\theta,y_1,x_1) - f(\theta,y_2,x_2)| \leq C_{\theta,x}(\|x_1 - x_2\|^{\alpha_1} + |y_1 - y_2|^{\alpha_2})$

(ii) $\int yf(\theta,y,x)dy < \infty$ for all $\theta, x \in \mathcal{H}$.

4. Asymptotic study

The objective of this paragraph is to adapt the above-mentioned ideas to the framework of a functional explanatory variable, and to construct a kernel-type estimator of the conditional distribution function $F(\theta, y, x)$ adjusted to MAR response samples. We establish an almost complete convergence[1] of our kernel estimator $\hat{F}(\theta, y, x)$ when we consider a model in which the response variable is missing. The presented results are accompanied by the data on the rate of convergence. In what follows, C and C' denote generic, strictly positive real constants, and h_n (resp. g_n) is a sequence which tends to 0 with n.

4.1. Point-wise almost-complete convergence

Besides the assumptions introduced in Section 3.4, we will need additional conditions. The assumptions we will need later, concerning the parameters of our estimator, i.e. K,H, h_n and g_n, are not very restrictive. Indeed, on the one hand they are rather inherent in the estimation problem of $f(\theta, y, x)$, and on the other they correspond to the assumptions usually made in the context of non-functional variables. More precisely, we will introduce the following conditions to ensure the performance of the estimator $\hat{f}(\theta, ., x)$:

(H3) Kernel H is a positive bounded function such that

(i) $\forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|y_1 - y_2|$, $\int |y|^{\alpha_2} H(y)dy < \infty$ and $\int yH(y)dy = 0$.

(ii) $H^{(1)}$ and $H^{(2)}$ are bounded with $\int \left(H^{(1)}(t)\right)^2 dt < \infty$.

(H4) K is a positive bounded kernel function with the support of [0,1]: $\forall u \in (0,1)$, $0 < \mathrm{K}(u)$, and the derivative K' exists on [0,1] with $K'(t) < 0$ for all $t \in [0,1]$ and $\int_0^1 \left(K^j\right)'(t)dt < \infty$ for $j = 1,2$.

(H5) $p(x, \theta)$ is continuous in the neighbourhood of x: $0 < p(x, \theta) < 1$.

THEOREM 1: Suppose that hypotheses (H1)–(H5) are satisfied if $\exists \beta > 0, n^{\beta} g_n \underset{n\to\infty}{\longrightarrow} \infty$, and if

$$\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)} \underset{n\to\infty}{\longrightarrow} 0,$$

[1] We say that a sequence $(S_n)_{n\in\mathbb{N}}$ converges almost completely to S if and only if, for any $\epsilon > 0$, we have $\sum_n \mathbb{P}(|S_n - S| > \epsilon) < \infty$.

then we have

$$\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x)-f(\theta,y,x)\right|=\mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

PROOF: The proof is based on the following decomposition, valid for any $y\in S_{\mathbb{R}}$:

$$\begin{aligned}\sup_{t\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x)-f(\theta,y,x)\right|\leq\frac{1}{\hat{f}_D(\theta,x)}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right|+\\+\frac{1}{\hat{f}_D(\theta,x)}\sup_{t\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x)\right|+\frac{f(\theta,y,x)}{\hat{f}_{\mathrm{D}}(\theta,x)}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_{\mathrm{D}}(\theta,x)-\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x)\right|.\end{aligned}\qquad(1)$$

Finally, the proof of this theorem is a direct consequence of the following intermediate results:

LEMMA 1: Suppose that hypotheses (H1)–(H3) and (H5) are satisfied, then we have

$$\sup_{y\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x)\right|=\mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right).$$

PROOF: We have

$$\begin{aligned}I=\mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x)=\mathbb{E}\left(\frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^n\delta_iK_i(\theta,x)H_i(y)\right)+\\-f(\theta,y,x)=\frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^n\mathbb{E}([\mathbb{E}(\delta_iK_i(\theta,x)H_i(y)|<\theta,X_i>)])-f(\theta,y,x)=\\=\frac{1}{g_n\mathbb{E}(K_1(\theta,x))}\mathbb{E}\left(p(x,\theta)K_1(\theta,x)\mathbb{E}(H_1(y))\right)-f(\theta,y,x).\end{aligned}$$

Moreover, by changing variables and using the fact that H is a *df* and uses a double conditioning with respect to Y_1, we can easily obtain

$$\begin{aligned}\mathbb{E}\left(H\left(g_n^{-1}(y-Y_1)\right)|<\theta,X_1>\right)=\int_{\mathbb{R}}H\left(\frac{y-u}{g_n}\right)f(\theta,u,X_1)du=\\=\int_{\mathbb{R}}H(v)f(\theta,y-vg_n,X_1)dv=\\=g_n\int_{\mathbb{R}}H(v)\left(f(\theta,y-vg_n,X_1)-f(\theta,u,x)\right)dv+g_nf(\theta,u,x)\int_{\mathbb{R}}H(v)dv.\end{aligned}$$

We can write, because of (H2) and (H3):

$$\begin{aligned}I=\frac{1}{\mathbb{E}K_1}\mathbb{E}\left(p(x,\theta)K_1(\theta,x)\int_{\mathbb{R}}H(v)\left(f(\theta,y-vg_n,X_1)-f(\theta,y,x)\right)dv\right)\leq\\\leq C_{\theta,x}\left(p(x,\theta)+o(1)\right)\int_{\mathbb{R}}H(v)\left(h_n^{\alpha_1}+|v|^{\alpha_2}g_n^{\alpha_2}\right)dv\leq\mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right).\end{aligned}$$

Finally, the proof is achieved.

LEMMA 2: Under hypotheses of Theorem 1, we have, as $n \to \infty$,

$$\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

PROOF: Using the compactness of $S_{\mathbb{R}}$, we can write that $S_{\mathbb{R}} \subset \bigcup_{j=1}^{\tau_n}\left(z_j - l_n, z_j + + l_n\right)$, with l_n and τ_n which can be chosen such that $l_n = C\tau_n^{-1} \sim Cn^{-\varsigma-1/2}$. Taking $m_y = \arg\min_{j\in\{z_1,\cdots,z_{\tau_n}\}}\left|y-m_j\right|$. Thus, we have the following decomposition:

$$\begin{aligned}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| &\le \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,t,x)-\hat{f}_N\left(\theta,m_y,x\right)\right| \\ &+ \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N\left(\theta,m_y,x\right)-\mathbb{E}\hat{f}_N\left(\theta,m_y,x\right)\right| \\ &+ \sup_{y\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N\left(\theta,m_y,x\right)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| \\ &\le B_1+B_2+B_3.\end{aligned}$$

As the first and the third term can be treated in the same manner, we deal only with the first term. By (H3)-(i), which in particular implies that H is a Hölder continuous with order one, we can write:

$$B_1 \le \frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sup_{y\in S_{\mathbb{R}}}\sum_{i=1}^{n}\delta_i\left|H_i(y)-H_i\left(m_y\right)\right|K_i(\theta,x) \le$$
$$\le \frac{C}{ng_n\mathbb{E}(K_1(\theta,x))}\sup_{y\in S_{\mathbb{R}}}\frac{\left|y-m_y\right|}{g_n}\times\sum_{i=1}^{n}\delta_iK_i(\theta,x) \le \frac{Cl_n}{ng_n^2\mathbb{E}(K_1(\theta,x))}\times\sum_{i=1}^{n}\delta_iK_i(\theta,x).$$

Using $\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) = p(x,\theta)$, (H3)-(i) and $\lim_{n\to\infty} n^{\beta}g_n = \infty$, it follows that

$$B_1 \xrightarrow[n\to\infty]{} \infty\,.$$

Thus, for n large enough, we have

$$B_1 = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

Following similar arguments, we can have

$$B_3 \le B_1.$$

Concerning B_2, let us consider $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)}}$, for all $\varepsilon_0 > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{y \in S_{\mathbb{R}}} \left|\hat{f}_N(\theta, m_y, x) - \mathbb{E}\hat{f}_N(\theta, m_y, x)\right| > \varepsilon\right) \leq \\
&\leq \mathbb{P}\left(\max_{j \in \{1, \cdots, \tau_n\}} \left|\hat{f}_N(\theta, m_y, x) - \mathbb{E}\hat{f}_N(\theta, m_y, x)\right| > \varepsilon\right) \leq \\
&\leq \tau_n \, \mathbb{P}\left(\left|\hat{f}_N(\theta, m_y, x) - \mathbb{E}\hat{f}_N(\theta, m_y, x)\right| > \varepsilon\right).
\end{aligned}$$

Applying Berstain's exponential inequality to

$$\Pi_i = \frac{1}{g_n \mathbb{E}(K_1(\theta, x))}\left[\delta_i K_i(\theta, x) H_i(m_y) - \mathbb{E}\left(\delta_i K_i(\theta, x) H_i(m_y)\right)\right].$$

Firstly, it follows from the fact that kernels K and H are bounded that we get

$$\mathbb{P}\left(\left|\hat{f}_N(\theta, m_y, x) - \mathbb{E}\hat{f}_N(\theta, m_y, x)\right| > \varepsilon\right) \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n \Pi_i\right| > \varepsilon\right) \leq 2n^{-C\varepsilon_0^2}.$$

Finally, by choosing ε_0 large enough, the proof can be concluded by the use of the Borel-Cantelli lemma, and the result can be easily deduced.

LEMMA 3: Under hypotheses (H1) and (H4)–(H5), we have, as $n \to \infty$,

(i) $\sup_{y \in S_{\mathbb{R}}} \left|\hat{f}_D(\theta, x) - \mathbb{E}\hat{f}_D(\theta, x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n \phi_{\theta,x}(h_n)}}\right)$.

(ii) $\sum_{n \geq 1} \mathbb{P}\left(\hat{f}_D(\theta, x) < 1/2\right) < \infty$.

PROOF: For the demonstration of the first part of this lemma, we use the same arguments as the previous lemma, the only change is in $\Delta_i(\theta, x)$, where

$$\hat{f}_D(\theta, x) - \mathbb{E}\hat{f}_D(\theta, x) = \frac{1}{n \mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \Delta_i(\theta, x),$$

with $\Delta_i(\theta, x) = \delta_i K_i(\theta, x) - \mathbb{E}\delta_i K_i(\theta, x)$.

All the calculus previously made with the variables $\Pi_i(\theta, x)$ remains valid for the variables $\Delta_i(\theta, x)$, and we obtain

$$\mathbb{P}\left(\left|\hat{f}_D(\theta, x) - \mathbb{E}\hat{f}_D(\theta, x)\right| > \varepsilon \sqrt{\frac{\log n}{n \phi_{\theta,x}(h_n)}}\right) \leq 2n^{-C'\varepsilon^2} < \infty.$$

For the proof of the second part of this lemma, we only need to establish $\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co.$

By the properties of the conditional expectation and the mechanism of MAR and (H5), it follows that

$$\begin{aligned}\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) &= \frac{1}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(\theta,x)\big)\\ &= \frac{1}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}[\mathbb{E}(\delta_i| <\theta,X_i>)K_i(\theta,x)]\\ &= \frac{\big(p(x,\theta)+o(1)\big)}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}\big(K_i(\theta,x)\big) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co\end{aligned}$$

Therefore (ii) of Lemma 3 follows from (i), and because $\hat{f}_{\mathrm{D}}(\theta,x) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co.$

Concerning the last part, we have

$$\begin{aligned}\big\{\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big\} \subseteq \big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - p(x,\theta)\big| > p(x,\theta)/2\big\} \Rightarrow\\ \Rightarrow \mathbb{P}\big\{\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big\} \le \mathbb{P}\big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - p(x,\theta)\big| > p(x,\theta)/2\big\} \le\\ \le \mathbb{P}\big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - \mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x)\big| > 1/2\big\},\end{aligned}$$

and because $\lim_{n\to\infty}\hat{f}_{\mathrm{D}}(\theta,x) = p(x,\theta)$, we show that

$$\sum_{n\ge 1}\mathbb{P}\big(\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big) < \infty.$$

We conclude the proof of the Theorem 1 by making use of Inequality (1), in conjunction with Lemma 1, Lemma 2 and Lemma 3.

4.2. Conditional mode estimation

In this section, we will study the rate of convergence of our conditional mode estimator $\widehat{M}_\theta(x)$. Obviously, obtaining these results will require more sophisticated technical developments than those presented so far. To ensure a good readability of this paragraph, we introduce conditions related to the flatness of the *cond-dff* $(\theta,.,x)$ around the conditional quantile $M_\theta(x)$.

Then a natural estimator of the conditional mode $M_\theta(x)$ is defined as:

$$\widehat{M}_\theta(x) = \arg \underset{y \in S_{\mathbb{R}}}{sup}\, \hat{f}(\theta, y, x),$$

where $M_\theta(x) = \arg \underset{y \in S_{\mathbb{R}}}{sup}\, f(\theta, y, x)$, $S_{\mathbb{R}}$ is a fixed compact subset of $\mathbb{R}$.

But a complementary way to take this local shape constraint into account is to suppose that:

(H6) The conditional density $f(\theta, ., x)$ satisfies

(i) $\exists \epsilon_0$, such that $f(\theta, ., x)$ is strictly increasing on $\left(M_\theta(x) - \epsilon_0, M_\theta(x)\right)$ and strictly decreasing on $(M_\theta(x), M_\theta(x) + \epsilon_0)$, with respect to x.

(ii) $f(\theta, y, x)$ is twice continuously differentiable around the point $M_\theta(x)$ with $f^{(1)}(\theta, M\theta(x), x) = 0$ and $f^{(2)}(\theta, ., x)$ is uniformly continuous on $S_{\mathbb{R}}$, such that $f^{(2)}(\theta, M\theta(x), x) \neq 0$, where $f^{(j)}(\theta, ., x)$ ($j = 1,2$) is the j-th order derivative of the conditional density $f(\theta, y, x)$.

(H7) $\forall \varepsilon > 0, \exists \eta > 0, \forall \varphi |M_\theta(x) - \varphi(x)| \geq \varepsilon \Rightarrow |f(\theta, \varphi(x), x) - f(\theta, M_\theta(x), x)| \geq \eta$.

The difficulty of the problem is naturally linked to the flatness of function $f(\theta, y, x)$ around mode M_θ. This flatness can be controlled by the number of vanishing derivatives at point M_θ, and this parameter will also have a significant influence on the asymptotic rates of our estimates. More precisely, we introduce the following additional smoothness condition.

(H8) There exists some integer $j > 1$, such that $\forall\, x$ and the function $f(\theta, ., x)$ is j-times continuously differentiable w.r.t y on $S_{\mathbb{R}}$ with

$$\begin{cases} f^{(j)}(\theta, M_\theta(x), x) = 0, if; 1 \leq j < l \\ f^{(j)}(\theta, ., x)\ is\ uniformly\ continuous\ on\ S_{\mathbb{R}} \\ such\ that\ f^{(j)}(\theta, M_\theta(x), x) \neq 0. \end{cases}$$

PROPOSTION 1: Suppose that the hypotheses (H1), (H3)–(H8) are satisfied if $\exists \beta > 0, n^\beta g_n \underset{n \to \infty}{\longrightarrow} \infty$, and if

$$\lim_{n \to \infty} \frac{\log n}{n g_n^{2l} \phi_{\theta,x}(h_n)} = 0,$$

then we have

$$\left|\widehat{M}_\theta(\gamma, x) - M_\theta(\gamma, x)\right| = O\left(\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right)^{\frac{1}{j}}\right) + O_{a.co.}\left(\left(\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)}\right)^{\frac{1}{2j}}\right).$$

PROOF: The proof is based on the Taylor expansion of $f(\theta,.,x)$. In the neighborhood of $M_\theta(\gamma,x)$, we get

$$\hat{f}\big(\theta,\widehat{M}_\theta(x),x\big) = f(\theta,M_\theta(x),x) + \frac{f^{(j)}(\theta,M_\theta^*(x),x)}{j!}\Big(\widehat{M}_\theta(x) - M_\theta(x)\Big)^j,$$

where $M_\theta^*(x)$ is between $M_\theta(x)$ and $\widehat{M}_\theta(x)$, combining the last equality with the fact that

$$\left|\hat{f}\big(\theta,\widehat{M}_\theta(x),x\big) - f(\theta,M_\theta(x),x)\right| \leq 2 \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|,$$

which makes it possible to write:

$$\left|M_\theta(x) - \widehat{M}_\theta(x)\right|^j \leq \frac{j!}{f^{(j)}\big(\theta,M_\theta^*(x),x\big)} \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|.$$

Using the second part of (H8), we obtain

$$\exists\delta > 0, \sum_{n\geq 1} \mathbb{P}\big(f^{(j)}(\theta,M_\theta^*(x),x) \geq \delta\big) < \infty.$$

So, we have

$$\left|\widehat{M}_\theta(\gamma,x) - M_\theta(\gamma,x)\right|^j = O_{a.co.}\left(\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|\right).$$

Finally, Proposition 1 can be deduced from Theorem 1.

COROLLARY 1: Under hypotheses of Theorem 1, we have

$$\widehat{M}_\theta(x) - M_\theta(x) \xrightarrow[n\to\infty]{} 0, a.co.$$

PROOF: The proof is based on the point-wise convergence of $\widehat{f}(\theta,.,x)$, and the Lipschitz property introduced in (H3)-(i) and hypothesis (H7), $f(\theta,t,x)$ is a continuous. We therefore have:

$\forall\, \epsilon > 0, \exists\eta(\epsilon) > 0$, such that

$$\left|f(\theta,y,x) - \hat{f}(\theta,M_\theta(x),x)\right| \leq \eta(\epsilon) \Rightarrow |y - M_\theta(x)| \leq \epsilon.$$

Therefore, for $y = \widehat{M}_\theta(x)$,

$$\mathbb{P}\left(\left|\widehat{M}_\theta(x) - M_\theta(x)\right| > \epsilon\right) \leq \mathbb{P}\left(\left| f\left(\theta, \widehat{M}_\theta(x), x\right) - f\left(\theta, M_\theta(x), x\right)\right| \geq \eta(\epsilon)\right). \quad (2)$$

Then, according to the theorem, $\widehat{M}_\theta - M_\theta$ go almost completely to 0, as n goes to infinity.

5. Asymptotic normality

The asymptotic normality of the semi-parametric estimators of the conditional mode for functional data in the Single Index Model (SIM) with missing data at random (MAR) is an important property that establishes the limiting distribution of the estimators as the sample size increases. Although specific results might vary depending on the assumptions and estimation methods used, it allows us to construct confidence intervals and hypothesis tests for the estimated mode. In this Section, the asymptotic normality of the estimator $\hat{f}(\theta,.,x)$ in the single functional index model is established.

(N1) There exists a function $\beta_{\theta,x}(\cdot)$, such that $\lim_{n\to\infty} \frac{\phi_{\theta,x}(sh_n)}{\phi_{\theta,x}(h_n)} = \beta_{\theta,x}(s)$, for $\forall s \in [0,1]$.

(N2) The bandwidth $h_n$and g_n, small ball probability $\phi_{\theta,x}(h_n)$ satisfying

(i) $ng_n^3\phi_{\theta,x}^3(h_n) \longrightarrow 0$ and $\frac{ng_n^3\phi_{\theta,x}(h_n)\log n}{\log^2 n} \longrightarrow \infty, as n \to \infty$.

(ii) $ng_n^2\phi_{\theta,x}^3(h_n) \longrightarrow 0, as n \to \infty$.

(N3) The conditional density $f(\theta,y,x)$ satisfies: $\exists \alpha > 0, \forall (y_1, y_2) \in S_\mathbb{R} \times S_\mathbb{R}$,

$$\left|f^{(j)}(\theta, y_1, x_1) - f^{(j)}(\theta, y_2, x_2)\right| \leq C(|y_1 - y_2|^\alpha), j = 1,2.$$

THEOREM 2: Under the assumptions of Theorem 1 and (N1)–(N3) for all x $\in\mathcal{H}$, and if

$$\sqrt{ng_n\phi_{\theta,x}(h_n)}\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right) \underset{n\to\infty}{\longrightarrow} 0,$$

then we have

$$\sqrt{\frac{ng_n\phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}}\left(\hat{f}(\theta,y,x) - f(\theta,y,x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\sigma^2(\theta,y,x) = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2}\frac{f(\theta,y,x)}{p(\theta,x)}\int H^2(u)du$ with $M_l(\theta,x) = K^l(1) -$ $- \int_0^1 (K^l)'(u)\,\beta_{\theta,x}(u)du$, $l = 1, 2$.

PROOF: In order to establish the asymptotic normality of $\widehat{F}(\theta,t,x)$, we need further notations and definitions. First we consider the following decomposition:

$$\hat{f}(\theta,y,x)-f(\theta,y,x)=\frac{\hat{f}_N(\theta,y,x)}{\hat{f}_{\mathrm{D}}(\theta,x)}-\frac{M_1(\theta,x)f(\theta,y,x)}{M_1(\theta,x)}=$$
$$=\frac{1}{\hat{f}_{\mathrm{D}}(\theta,x)}\Big(\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\Big)+$$
$$-\frac{1}{\hat{f}_{\mathrm{D}}(\theta,x)}\Big(M_1(\theta,x)f(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\Big)+$$
$$+\frac{f(\theta,y,x)}{\hat{f}_{\mathrm{D}}(\theta,x)}\Big(M_1(\theta,x)-\mathbb{E}\widehat{F}_D(\theta,x)\Big)-\frac{f(\theta,y,x)}{\hat{f}_{\mathrm{D}}(\theta,x)}\Big(\widehat{F}_D(\theta,x)-\mathbb{E}\widehat{F}_D(\theta,x)\Big)=$$
$$=\frac{1}{\hat{f}_{\mathrm{D}}(\theta,x)}A_n(\theta,y,x)+B_n(\theta,y,x),$$

where:

$$A_n(\theta,y,x)=\frac{1}{ng_n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\Big\{\big(H_i(t)-g_nf(\theta,y,x)\big)\delta_iK_i(\theta,x)$$
$$-\mathbb{E}\Big[\big(H_i(t)-g_nf(\theta,y,x)\big)\delta_iK_i(\theta,x)\Big]\Big\}=\frac{1}{ng_n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}N_i\,(\theta,t,x)$$

and

$$B_n(\theta,y,x)=M_1(\theta,x)f(\theta,y,x)-\ \mathbb{E}\hat{f}_N(\theta,y,x)+f(\theta,y,x)$$
$$\Big(M_1(\theta,x)-\mathbb{E}\hat{f}_D(\theta,x)\Big).$$

It follows that,

$$ng_n\phi_{\theta,x}(h_n)Var\big(A_n(\theta,y,x)\big)=\frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}Var\big(N_1(\theta,y,x)\big)=$$
$$=V_n(\theta,y,x).$$

Then, the proof of Theorem 2 can be deduced from the following Lemmas.

LEMMA 4: Under assumptions of Theorem 2, we have

$$\sqrt{ng_n\phi_{\theta,x}(h_n)}A_n(\theta,y,x)\xrightarrow{\mathcal{D}}\mathcal{N}\big(0,\sigma^2(\theta,y,x)\big).$$

PROOF:

$$V_n(\theta,y,x)=\frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}\mathbb{E}\left[\delta_1K_1^2(\theta,x)\big(H_1(y)-g_nf(\theta,y,x)\big)^2\right]$$

$$V_n(\theta,y,x) = \\ = \frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}\mathbb{E}\left[K_1^2(\theta,x)\mathbb{E}\left(\delta_1\left(H_1(y)-g_nf(\theta,y,x)\right)^2\Big|\langle\theta,X_1\rangle\right)\right] \tag{3}$$

Using the definition of conditional variance, we have

$$\mathbb{E}\left(\delta_1\left(H_1(t)-g_nf(\theta,y,x)\right)^2\Big|\langle\theta,X_1\rangle\right) = J_{1n}+J_{2n},$$

$$J_{1n} = Var(\delta_1 H_1(y)|\langle\theta,X_1\rangle), J_{2n} = [\mathbb{E}(H_1(y)|\langle\theta,X_1\rangle) - g_nf(\theta,y,x)]^2.$$

Concerning J_{1n},

$$J_{1n} = \mathbb{E}\left(H^2\left(\tfrac{y-Y_1}{g_n}\right)\Big|\langle\theta,X_1\rangle\right) - \left[\mathbb{E}\left(\delta_1 H_1\left(\tfrac{y-Y_1}{g_n}\right)\Big|\langle\theta,X_1\rangle\right)\right]^2 = \mathcal{J}_1+\mathcal{J}_2.$$

As for $\mathcal{J}_1$, by the property of double conditional expectation, we obtain

$$\mathcal{J}_1 = \mathbb{E}\left(\delta_1 H^2\left(\tfrac{y-Y_1}{g_n}\right)\Big|\langle\theta,X_1\rangle\right) = p(x,\theta)\int H^2\left(\tfrac{y-v}{g_n}\right)f(\theta,v,X_1)dv \\ = p(x,\theta)\int H^2(u)\,dF(\theta,y-ug_n,X_1).$$

On the other hand, under assumptions (H2)–(H3), we have

$$\mathcal{J}_1 = \int H^2(u)\,dF(\theta,y-ug_n,X_1) = h_n\int H^2(u)\,f(\theta,y-ug_n,X_1)du \leq \\ \leq g_n\int H^2(u)\big(f(\theta,y-ug_n,X_1)-f(\theta,y,x)\big)du + \\ +g_n\int H^2(u)\,f(\theta,y,x)du \leq \\ \leq g_n\left(C_{\theta,x}\int H^2(u)\big(h_n^{\alpha_1}+|v|^{\alpha_2}g_n^{\alpha_2}\big)du + f(\theta,y,x)\int H^2(u)\,du\right) = \\ = \mathcal{O}\big(h_n^{\alpha_1}+g_n^{\alpha_2}\big) + g_nf(\theta,y,x)\int H^2(u)\,du. \tag{4}$$

As for J_2, $\mathcal{J}_2' = \mathbb{E}(\delta_1 H_1(y)|\langle\theta,X_1\rangle) = p(x,\theta)\int H\left(\frac{y-v}{g_n}\right)f(\theta,y,X_1)dv.$

Moreover, by changing variables, we obtain:

$$\mathcal{J}'_2 = h_n \int H(u)\big(f(\theta, y - ug_n, x) - f(\theta, y, x)\big)du + g_n f(\theta, y, x) \int H(u)du.$$

The last equality is due to the fact that H is a probability density, thus we have

$$\mathcal{J}'_2 = O\big(h_n^{\alpha_1} + g_n^{\alpha_2}\big) + g_n f(\theta, y, x).$$

Finally, we get $J_2 \xrightarrow[n\to\infty]{} \infty$. As for J_{2n}, by (H1)–(H3), we obtain $J_{2n} \xrightarrow[n\to\infty]{} \infty$. Meanwile, from (H1)–(H3), it follows that

$$\frac{\phi_{\theta,x}(h_n)\mathbb{E}K_1^2(\theta,x)}{\mathbb{E}^2\big(K_1(\theta,x)\big)} \xrightarrow[n\to\infty]{} \frac{M_2(\theta,x)}{(M_1(\theta,x))^2},$$

which leads to combining equations (3) and (4):

$$V_n(\theta, t, x) \xrightarrow[n\to\infty]{} \frac{M_2(\theta,x) f(\theta,y,x)}{(M_1(\theta,x))^2 p(x,\theta)}.$$

LEMMA 5: If the assumptions (H1)–(H7) are satisfied, we have

$$\sqrt{n g_n \phi_{\theta,x}(h_n)} B_n(\theta, t, x) \to 0, \text{ in probability.}$$

PROOF: We have

$$\sqrt{n g_n \phi_{\theta,x}(h_n)} B_n(\theta, t, x) = \frac{\sqrt{n g_n \phi_{\theta,x}(h_n)}}{\hat{f}_D(\theta, x)} \{\mathbb{E}\hat{f}_N(\theta, y, x) - M_1(\theta, x) f(\theta, y, x)$$
$$+ f(\theta, y, x)\left(M_1(\theta, x) - \mathbb{E}\hat{f}_D(\theta, x)\right).$$

Firstly, it can be observed that, as $n \to \infty$,

$$\frac{1}{\phi_{\theta,x}(h_n)} \mathbb{E}\left[K^l\left(\frac{\langle \theta, x - X_i\rangle}{h_n}\right)\right] \to M_l(\theta, x), \text{for } l = 1,2, \tag{5}$$

$$\mathbb{E}\hat{f}_D(\theta, x) \to M_1(\theta, x) p(x, \theta) \ and\ \mathbb{E}\hat{f}_N(\theta, y, x) \to M_1(\theta, x) f(\theta, y, x), \tag{6}$$

can be proved in the same way as in Ezzahrioui and Ould-Saïd (2008), corresponding to their Lemmas 5.1 and 5.2, and then their proofs are omitted.

Secondly, using (5) and (6), we have, on the one hand, as $n \to \infty$,

$$\left\{\mathbb{E}\hat{f}_N(\theta, y, x) - M_1(\theta, x) f(\theta, y, x) + f(\theta, y, x)\left(M_1(\theta, x) - \mathbb{E}\hat{f}_D(\theta, x)\right)\right\} \to 0.$$

On the other hand,

$$\frac{\sqrt{n g_n \phi_{\theta,x}(h_n)}}{\hat{f}_{\mathrm{D}}(\theta,x)} = \frac{\sqrt{n g_n \phi_{\theta,x}(h_n)}\hat{f}(\theta,y,x)}{\hat{f}_{\mathrm{D}}(\theta,x)\hat{f}(\theta,y,x)} = \frac{\sqrt{n g_n \phi_{\theta,x}(h_n)}\hat{f}(\theta,y,x)}{\hat{f}_N(\theta,y,x)}.$$

Because K and H are continuous with the support on [0,1], then from (H3) and (H4) $\exists m = \min_{[0,1]} K(t)H(t)$, it follows that

$$\hat{f}_N(\theta, y, x) \geq \frac{m}{g_n \phi_{\theta,x}(h_n)},$$

which yields

$$\frac{\sqrt{n g_n \phi_{\theta,x}(h_n)}}{\hat{f}_N(\theta,y,x)} \leq \frac{\sqrt{n g_n^3 \phi_{\theta,x}^3(h_n)}}{m}.$$

Finally, (N2)–(i) completes the proof of Lemma 5.

5.1. Application: The conditional mode in functional single-index model

The main objective of this part of our work is to establish the asymptotic normality of the conditional mode estimator of Y, given $< \theta, X >=< \theta, x >$ denoted by $M_\theta(x)$.

COROLLARY 2: Under the assumptions of Theorem 2, and if (H6) holds true, and in addition if

$$n g_n^3 \phi_{\theta,x}(h_n) \underset{n\to\infty}{\longrightarrow} 0,$$

then we have, as $n \to \infty$,

$$\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\left(\widehat{M}_\theta(x) - M_\theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \varrho^2(\theta, M_\theta(x), x)\right),$$

where

$$\varrho^2(\theta, M_\theta(x), x) = \frac{M_2(\theta,x) f(\theta,M_\theta(x),x)}{p(\theta,x)\left(M_1(\theta,x) f^{(2)}(\theta,M_\theta(x),x)\right)^2}.$$

PROOF: By the first-order Taylor expansion for $\hat{f}^{(1)}(\theta, y, x)$ at point $M_\theta(x)$, and the fact that $\widehat{f}^{(1)}\left(\theta, \widehat{M}_\theta(x), x\right) = 0$, it follows that

$$\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\left(\widehat{M}_\theta(x) - M_\theta(x)\right) = -\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\,\frac{\hat{f}^{(1)}(\theta,M_\theta(x),x)}{\hat{f}^{(2)}(\theta,M_\theta^*(x),x)},$$

where $M_\theta^*(x)$ is between $\widehat{M}_\theta(x)$ and $M_\theta(x)$. Similarly to the proof of Theorem 2, it follows that

$$-\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\,\hat{f}^{(1)}(\theta, M_\theta(x), x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \varrho_0^2(\theta, M_\theta(x), x)\right), \tag{7}$$

where

$$\varrho_0^2(\theta, M_\theta(x), x) = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2}\frac{f(\theta,M_\theta(x),x)}{p(\theta,x)}\int\left(H'(u)\right)^2 du.$$

Thus, as above, similarly to Ferraty and Vieu (2006), we can obtain $\hat{f}^{(2)}(\theta, y, x) \xrightarrow{\mathbb{P}} f^{(2)}(\theta, y, x)$, as $n \to \infty$, which implies that $\widehat{M}_\theta(x) \to M_\theta(x)$. Therefore, we get

$$\hat{f}^{(2)}(\theta, M_\theta^*(x), x) \underset{n \to \infty}{\longrightarrow} f^{(2)}(\theta, M_\theta(x), x) \neq 0. \tag{8}$$

By (H3), (H6) and (N3), similarly to the proof of lemmas, Lemma 4 and Lemma 5, respectively, (7) follows directly. Then, the proof of Corollary 2 is completed.

5.2. Confidence bands

The asymptotic variances $\sigma^2(\theta, t, x)$ and $\varrho^2(\theta, M_\theta(x), x)$ in Theorem 2 and Corollary 2 depend on some unknown quantities including M_1, M_2, $\phi(u)$, $M_\theta(x)$, $p(\theta, x)$ and $f(\theta, M_\theta(x), x)$. Therefore, $p(\theta, x)$,$M_\theta(x)$, and $f(\theta, M_\theta(x), x)$ can be estimated by $P_n(\theta, x)$, $\widehat{M}_\theta(x)$ and $\hat{f}(\theta, M_\theta(x), x)$ and $\widehat{M}_\theta(x)$, respectively. Moreover, using the decomposition given by the assumption (H1), one can estimate $\phi_{\theta,x}(.)$ by $\hat{\phi}_{\theta,x}(.)$. Because the unknown functions $M_j := M_j(\theta, x)$ and $f(\theta, y, x)$ are intervening in the expression of the variance, we need to estimate the mode $M_1(\theta, x)$, $M_2(\theta, x)$ and $f(\theta, y, x)$, respectively.

From the assumptions (H1)–(H4), we know that $M_j(\theta, x)$ can be estimated by $\widehat{M}_j(\theta, x)$, which is defined as:

$$\widehat{M}_j(\theta, x) = \frac{1}{n\widehat{\phi}_{\theta,x}(h)} \sum_{i=1}^{n} K_i^j(\theta, x), \text{where } \widehat{\phi}_{\theta,x}(h) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{|\langle x - X_i, \theta\rangle| < h\}},$$

with $\mathbf{1}_{\{.\}}$ being the indicator function. Finally, the estimator of $p(\theta, x)$ is denoted by:

$$P_n(\theta, x) = \frac{\sum_{i=1}^{n} \delta_i K\left(h_n^{-1}(<x - X_i, \theta>)\right)}{\sum_{i=1}^{n} K\left(h_n^{-1}(<x - X_i, \theta>)\right)}.$$

By applying the kernel estimator of $f(\theta, y, x)$ given above, the quantity $\sigma^2(\theta, y, x)$ can be estimated by:

$$\hat{\sigma}^2(\theta, y, x) = \frac{\widehat{M}_2(\theta,x)}{(\widehat{M}_1(\theta,x))^2} \frac{\hat{f}(\theta,y,x)}{P_n(\theta,x)} \int H^2(u) du.$$

Finally, in order to show the asymptotic $(1 - \xi)$ confidence interval of $M_\theta(x)$, we need to consider the estimator of $\varrho^2(\theta, M_\theta(x), x)$, as follows:

$$\hat{\varrho}^2(\theta, M_\theta(x), x) = \frac{\widehat{M}_2(\theta,x)}{(\widehat{M}_1(\theta,x))^2} \frac{\hat{f}(\theta,\widehat{M}_\theta(x),x)}{P_n(\theta,x)\left(\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)\right)^2} \int \left(H'(u)\right)^2 du,$$

so we can derive the corollary below.

COROLLARY 3: Under the assumptions of Theorem 2, $K'$and $(K^2)'$are integrable functions, then we get, as $n \to \infty$,

(a) $\sqrt{\frac{n g_n \widehat{\phi}_{\theta,x}(h_n)}{\hat{\sigma}^2(\theta,y,x)}} \left(\hat{f}(\theta, y, x) - f(\theta, y, x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$

(b) $\sqrt{\frac{n g_n^3 \widehat{\phi}_{\theta,x}(h_n)}{\hat{\varrho}^2(\theta,M_\theta(x),x)}} \left(\widehat{M}_\theta(x) - M_\theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$

PROOF: Observe that

$$\text{(a) } \Sigma(\theta, y, x) = \frac{\widehat{M}_1}{M_1} \sqrt{\frac{M_2}{\widehat{M}_2}} \sqrt{\frac{n g_n \widehat{\phi}_{\theta,x}(h_n) P_n(\theta,x) f(\theta,y,x)}{p(\theta,x) \hat{f}(\theta,y,x) n g_n \phi_{\theta,x}(h_n)}} \times$$
$$\times \frac{M_1}{\sqrt{M_2}} \sqrt{\frac{n g_n \phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}} \left(\hat{f}(\theta, y, x) - f(\theta, y, x)\right),$$

where $\Sigma(\theta, y, x) = \sqrt{\frac{n g_n \widehat{\phi}_{\theta,x}(h_n)}{\hat{\sigma}^2(\theta,y,x)}} \left(\hat{f}(\theta, y, x) - f(\theta, y, x)\right)$, by Theorem 2, we have, as $n \to \infty$,

$$\sqrt{\frac{ng_n\phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}}\left(\hat{f}(\theta,y,x)-f(\theta,y,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

In order to prove (a), we need to show that

$$\frac{\widehat{M}_1}{M_1}\sqrt{\frac{M_2}{\widehat{M}_2}}\sqrt{\frac{ng_n\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,y,x)}{p(\theta,x)\hat{f}(\theta,y,x)ng_n\phi_{\theta,x}(h_n)}}\left(\hat{f}(\theta,y,x)-f(\theta,y,x)\right)\xrightarrow{\mathbb{P}}\mathcal{N}(0,1),$$

using the result given by Laib and Louani (2010), we have

$$\widehat{M}_1\xrightarrow{\mathbb{P}}M_1,\widehat{M}_2\xrightarrow{\mathbb{P}}M_2 \text{ and } \frac{\widehat{\phi}_{\theta,x}(h_n)}{\sqrt{\phi_{\theta,x}(h_n)}}\xrightarrow{\mathbb{P}}1 \text{ as } n\to\infty.$$

On the other hand, from Proposition 2 in Laib and Louani (2010), it follows that

$$P_n(\theta,x)\underset{n\to\infty}{\longrightarrow}\mathbb{E}(\delta|\langle X,\theta\rangle=\langle x,\theta\rangle)=\mathbb{P}(\delta=1|\langle X,\theta\rangle=\langle x,\theta\rangle)=p(x,\theta).$$

In addition, from Theorem 1, we have $\hat{f}(\theta,y,x)\longrightarrow f(\theta,y,x)$, as $n\to\infty$. This yields the proof for the first part of Corollary 3.

(b)
$$\begin{aligned}&\frac{\widehat{M}_1\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)}{\hat{f}(\theta,\widehat{M}_\theta(x),x)}}\left(\widehat{M}_\theta(x)-M_\theta(x)\right)=\\&=\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,M_\theta(x),x)}{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}}\frac{\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{f^{(2)}(\theta,M_\theta(x),x)}\times\\&\times\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}{f(\theta,M_\theta(x),x)}}f^{(2)}(\theta,M_\theta(x),x)\left(\widehat{M}_\theta(x)-M_\theta(x)\right).\end{aligned}$$

Applying Corollary 2, we obtain

$$\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}{f(\theta,M_\theta(x),x)}}f^{(2)}(\theta,M_\theta(x),x)\left(\widehat{M}_\theta(x)-M_\theta(x)\right)\longrightarrow\mathcal{N}(0,1).$$

Further, by considering Lemma 5, (2) and (8), we obtain, as $n\to\infty$,

$$\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,M_\theta(x),x)}{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}}\frac{\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{f^{(2)}(\theta,M_\theta(x),x)}\xrightarrow{\mathbb{P}}1.$$

Hence, the proof is completed.

REMARK 1: Thus, following Corollary 3, the asymptotic $(1-\xi)$ confidence interval of the conditional density $f(\theta, y, x)$, and the conditional mode $M_\theta(x)$, respectively, are expressed as follows:

$$\hat{f}(\theta, y, x) \pm \eta_{\gamma/2}\sqrt{\frac{\hat{\sigma}^2(\theta,y,x)}{ng_n\hat{\Phi}_{\theta,x}(h_n)}} \text{and} \widehat{M}_\theta(x) \pm \eta_{\gamma/2}\sqrt{\frac{\hat{\varrho}^2(\theta,M_\theta(x),x)}{ng_n^3\hat{\Phi}_{\theta,x}(h_n)}},$$

where $\hat{\sigma}^2(\theta, y, x)$ and $\hat{\varrho}^2(\theta, M_\theta(x), x)$ are defined in Section 5.2, and $\eta_{\gamma/2}$ is the upper $\gamma/2$ quantile of the normal distribution $\mathcal{N}(0,1)$.

6. Simulation study on finite samples

6.1. A numerical study

In this Section, we will consider simulated data studied to assess the finite sample performance of the proposed estimator and compare it to the competing estimator. For studying the behavior of our estimators, and in order to illustrate our results, we evaluate the performance of our estimation approach using a single-index dimensional reduce model in order to prove the effectiveness of our model. More precisely, we will compare the finite sample behavior of estimator $\tilde{f}$ with the complete functional data and the estimator $\hat{f}$ under functional data with MAR.

Furthermore, some tuning parameters have to be specified. The kernel $K(.)$ is chosen to be the quadratic function defined as $K = \frac{3}{2}(1-u^2)\mathbf{1}_{[0;1]}$ and the cumulative distribution function (cdf)

$$H(u) = \int_{-\infty}^{u} \frac{3}{4}(1-z^2)\mathbf{1}_{[-1;1]}(z)dz.$$

The semi-metric $d(.,.)$ will be specified according to the choice of the functional space $\mathcal{H}$ discussed in the scenarios below. It is well known that some of the crucial H parameters in semi-parametric models are the smoothing parameters which are involved in defining the shape of the link function between the response and the covariate.

Now, for simplifying the implementation of our methodology, we take the bandwidths $h_K \sim h_H = h$, where h will be chosen by the cross-validation method on the k-nearest neighbors (see [12], p. 102).

Let us consider the following regression model, where the covariate is a curve and the response is a scalar:

$$Y_i = R(\langle\theta, X_i\rangle) + \epsilon_i; \quad i = 1, \dots, n = 300,$$

where ϵ_i is the error supposed to be generated by autoregressive model defined by:

$$\epsilon_i = \frac{1}{\sqrt{2}}(\epsilon_{i-1} + \eta_i), \quad i = 1, \cdots, n,$$

with $(\eta_i)_i$ is a sequence of i.i.d. random variables normally distributed with a variance equal to 0.1.

The functional covariate X is assumed to be a diffusion process defined on $[0,1]$ and generated by the following equation:

$$X(t) = \Omega(2 - \cos(\pi \Upsilon t)) + (1 - \Omega)\cos(\pi \Upsilon t), \quad t \in [0,1],$$

where Υ is generated from a standard normal distribution, and Ω is Bernoulli's law $\mathfrak{B}(0.5)$.

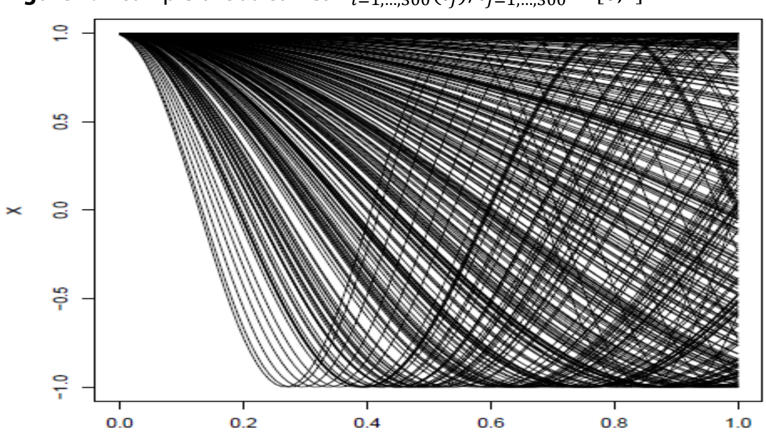
Let us take into account the smoothness of curves $X_i(t)$ (see Figure 1).We choose the distance $deriv_1$ (the semi-metric based on the first derivatives of the curves) in $\mathcal{H}$ as:

$$d(\chi_1, \chi_2) = \left(\int_0^1 (\chi_1'(t) - \chi_2'(t))^2 dt\right)^{1/2}$$

as semi-metric. Then, we consider a non-linear regression function defined as:

$$R(X) = 4exp\left(\frac{1}{2 + \int_0^{\pi/2} X^2(t)dt}\right).$$

Figure 1. A sample of 300 curves $X_{i=1,\ldots,300}(t_j)$, $t_{j=1,\ldots,300} \in [0,1]$



Source: authors' work.

The missing mechanism is similar to that in Ferraty et al. (2013):

$$p(x) = \mathbb{P}(\delta = 1|X = x) = exp\,it\left(2\alpha \int_0^1 x^2(t)dt\right),$$

where $exp\,it(v) = e^v/(1 + e^v)$ for $v \in \mathbb{R}$, and the degree of dependency between the functional covariate X and the missing variable δ is controlled by parameter α (where we compare three missing rates: strong, medium and weak cases, with α=0.05, $\alpha = 0.5$ and $\alpha = 3$ respectively). The missing proportions are quantified by the following benchmark:

$$\bar{\delta} = 1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i.$$

In practice, this parameter can be selected by a cross-validation approach (see [2]). It might be that one selected the real-valued function $\theta(t)$ from among the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}X')\langle X', . \rangle_{\mathcal{H}}]$, where $X(t)$ is a diffusion process defined on a real interval $[a, b]$, and $X'(t)$ is its first derivative (see [3]). So for a chosen training sample $\mathcal{L}$, by applying the principal component analysis (PCA) method, the computation of the eigenvectors of the covariance operator estimated by its empirical covariance operator $\frac{1}{\mathcal{L}}\sum_{i\in\mathcal{L}}(X_i' - \mathbb{E}X_i')^t\,(X_i' - \mathbb{E}X_i')$ will be the best approximation of our functional parameter θ. Now, let us denote θ^* the first eigenfunction corresponding to the first higher eigenvalue of the empirical covariance operator which will replace θ during the simulation step.

In our simulation, the sample size is $n = 300$. We divide it into two parts: one is a learning sample of 250 observations, and the other 50 observations are a test sample.

The first one from 1 to 250 will be used to make the simulation, and the second, from 251 to 300, will serve for the prediction. We then perform the following steps:

- **Step 1:** Compute the inner product: $\langle\theta^*, X_1\rangle, \dots, \langle\theta^*, X_{300}\rangle$, generate independently variables $\epsilon_1, \dots, \epsilon_{300}$, then simulate response variables $Y_i = r(\langle\theta^*, X_i\rangle) + \epsilon_i$, where $r(\langle\theta^*, X_i\rangle) = exp\big(10(\langle\theta^*, X_i\rangle - 0.05)\big)$, and generate independently variables $\epsilon_1, \dots, \epsilon_{300}$.
- **Step 2:** For each j in, the test sample $\mathcal{J} = 251, \dots, 300$, we compute: $\tilde{Y}_j = \widetilde{M}_{\theta^*}(X_j)$ and $\hat{Y}_j = \widehat{M}_{\theta^*}(X_j)$, where:

$$\widetilde{M}_\theta(\chi) = \arg \sup_{y\in\mathcal{S}_{\mathbb{R}}} \tilde{f}\,(\theta, y, \chi).$$

- **Step 3:** Finally, we show the results by juxtaposing the true values and the predicted values for the MSE, both for the option of having complete data and the option of having a MAR response.

$$CMSE = \frac{1}{50}\sum_{j=251}^{300}\left(Y_j - \tilde{Y}_j\right)^2 \quad and \quad MMSE = \frac{1}{50}\sum_{j=251}^{300}\left(Y_j - \hat{Y}_j\right)^2.$$

Figure 2. Complete data case

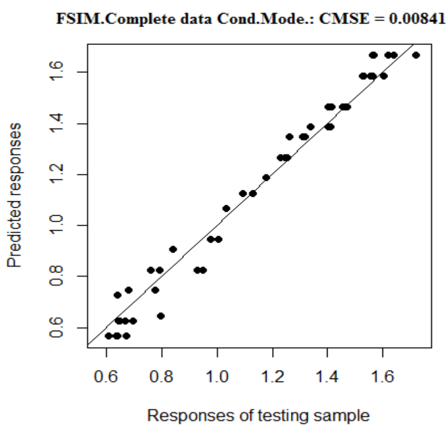
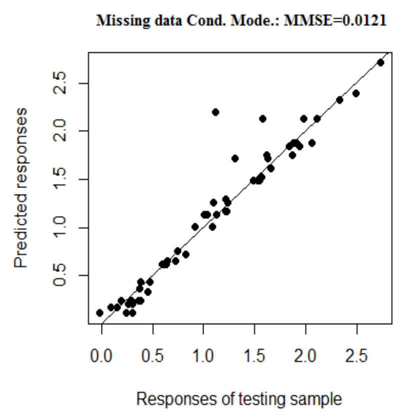


Figure 3. MAR with $\alpha = 2$



Source: authors' work.

Figure 4. MAR with $\alpha = 1.5$

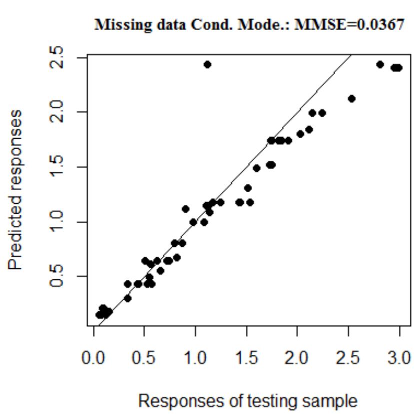
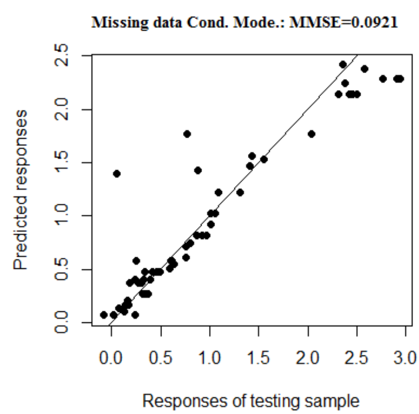


Figure 5. MAR with $\alpha = 0.5$



Source: authors' work.

In the MAR responses, the proportion $\overline{\delta}$ of missing response data is the key parameter which is shown by α. In Figures 2–5, we can see that when $\overline{\delta}$ is small (or

α is big), the $\widehat{M}_\theta(x)$ of the MAR response works almost as efficiently as if we had a complete data set and used $\widetilde{M}_\theta(x)$. In this case, when one has missing response data, the estimator $\widetilde{M}_\theta(x)$ is not useful, but the $\widehat{M}_\theta(x)$ is a benchmark analysis, and the fact that $\widehat{M}_\theta(x)$ is almost as effective as $\widetilde{M}_\theta(x)$ is what one is really expecting.

Acknowledgement

The authors are grateful to anonymous referees and the Editor-in-Chief for their valuable comments and suggestions towards the improvement of this paper.

References

Aït-Saidi, A., Ferraty, F., Kassa, R., & Vieu, P. (2008). Cross-validated estimation in the single-functional index model. *Statistics. A Journal of Theoretical and Applied Statistics*, *42*(6), 475–494. https://doi.org/10.1080/02331880801980377.

Attaoui, S.(2014). On the non-parametric conditional density and mode estimates in the single functional index model with strongly mixing data. *Sankhyã. The Indian Journal of Statistics*, *76*(2), 356–378. https://doi.org/10.1007/s13171-014-0051-6.

Attaoui, S., & Ling, N. (2016). Asymptotic results of a non-parametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications. *Metrika*. *International Journal for Theoretical and Applied Statistics*, *79*(5), 485–511. https://doi.org/10.1007/s00184-015-0564-6.

Chaudhuri, P., Doksum, K., & Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, *25*(2), 715–744. https://doi.org/10.1214/aos/1031833670.

Cheng, P. E. (1994). Non-parametric Estimation of Mean Functional with Data Missing at Random. *Journal of the American Statistical Association*, *89*(425), 81–87. https://doi.org /10.2307/2291203.

Ezzahrioui, M., & Ould-Saïd, M. (2008). Asymptotic normality of a non-parametric estimator of the conditional mode function for functional data. *Journal of Non-parametric Statistics*, *20*(1), 3–18. https://doi.org/10.1080/10485250701541454.

Ferraty, F., Rabhi, A., & Vieu, P. (2005). Conditional Quantiles for Dependent Functional Data with Application to the Climatic *El Niño* Phenomenon. *Sankhyã*. *The Indian Journal of Statistics*, *67*(2), 378–398.

Ferraty, F., Sued, F., & Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics. A Journal of Theoretical and Applied Statistics*, *47*(4), 688–706.

Ferraty, F., & Vieu, P. (2003). Functional Non-parametric Statistics. A Double Infinite Dimensional Framework. In M. G. Akritas & D. N. Politis (Eds.), *Recent Advances and Trends in Non-parametric Statistics* (pp. 61–78). Elsevier.

Ferraty, F., & Vieu, P. (2006). *Non-parametric Functional Data Analysis. Theory and Practice*. Springer. https://doi.org/10.1007/0-387-36620-2.

Hamri, M. M., Mekki, S. D., Rabhi, A., & Kadiri, N. (2022). Single functional index quantile regression for independent functional data under right-censoring. *Econometrics. Ekonometria*, *26*(1), 31–62. https://doi.org/10.15611/eada.2022.1.03.

Kadiri, N., Rabhi, A., & Bouchentouf, A. A. (2018). Strong uniform consistency rates of conditional quantile estimation in the single functional index model under random censorship. *Dependence Modeling*, *6*(1), 197–227. https://doi.org/10.1515/demo-2018-0013.

Laib, N., & Louani, D. (2010). Non-parametric Kernel Regression Estimation for Functional Stationary Ergodic Data: Asymptotic Properties. *Journal of Multivariate Analysis*, *101*(10), 2266–2281. https://doi.org/10.1016/j.jmva.2010.05.010.

Lemdani, M., Ould-Saïd, E., & Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated data. *Journalof Multivariate Analysis*, *100*(3), 546–559. https://doi.org/10.1016/j.jmva.2008.06.004.

Liang, H.-Y., & de Uña-Álvarez, J. (2010). Asymptotic normality for estimator of conditional mode under left-truncated and dependent observations. *Metrika. International Journal for Theoretical and Applied Statistics*, *72*(1), 1–19. https://doi.org/10.1007/s00184-009-0237-4.

Ling, N., Liang, L., & Vieu, P. (2015). Non-parametric regression estimation for functional stationary ergodic data with missing at random. *Journal of Statistical Planning and Inference*, *162*, 75–87. https://doi.org/10.1016/j.jspi.2015.02.001.

Ling, N., Liu, Y., & Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics. A Journal of Theoretical and Applied Statistics*, *50*(5), 991–1013. https://doi.org/10.1080/02331888.2015.1122012.

Mekki, S. D., Kadiri, N., & Rabhi, A. (2021). Asymptotic Properties of the Semi-Parametric Estimators of the Conditional Density for Functional Data in the Single Index Model with Missing Data at Random. *Statistica*, *81*(4), 399–422. https://doi.org/10.6092/issn.1973-2201/10472.

Ould-Saïd, E., & Cai, Z. (2005). Strong uniform consistency of non-parametric estimation of the censored conditional mode function. *Journal of Non-parametric Statistics*, *17*(7), 797–806. https://doi.org/10.1080/10485250500038561.

Ould-Saïd, E., & Tatachak, A. (2011). A non-parametric conditional mode estimate under RLT model and strong mixing condition. *International Journal of Statistics&Economics*, *6*(S11), 76–92.

Ould-Saïd, E., & Yahia, D. (2011). Asymptotic normality of a kernel conditional quantile estimator under strong mixing hypothesis and left-truncation. *Communications in Statistics – Theory and Methods*, *40*(14), 2605–2627. https://doi.org/10.1080/03610926.2010.489171.

Rabhi, A., Kadiri, N., & Akkal, F. (2021). On the Central Limit Theorem for Conditional Density Estimator In the Single Functional Index Model. *Applications and Applied Mathematics: An International Journal*, *16*(2), 844–866. https://digitalcommons.pvamu.edu/aam/vol16/iss2/4.

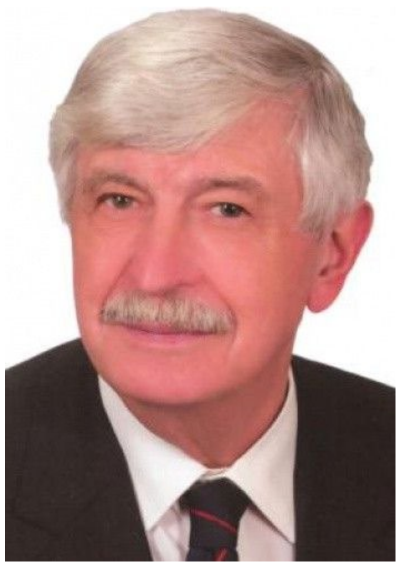
Roussas, G. G. (1969). Non-parametric estimation of the transition distribution function of a Markov Process. *The Annals of Mathematical Statistics*, *40*(4), 1386–1400. https://doi.org/10.1214/aoms/1177697510.

Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters*, *7*(5), 407–412. https://doi.org/10.1016/0167-7152(89)90095-3.

Przegląd Statystyczny. Statistical Review, 2023, vol. 70, 2, 46–56 https://doi.org/10.59139/ps.2023.02.3
Received: 16.10.2023, accepted: 23.10.2023

Professor Czesław Domański – 55 years devoted to statistics

Alina Jędrzejczak,[a] Tadeusz J. Kowaleski,[b] Aleksandra Baszczyńska[c]



Private archive

In 2023, Professor Czesław Domański celebrates his 80th birthday and the 55th anniversary of his professional career as an academic teacher. These years have been marked by numerous achievements in the field of science, education and organisation, making Professor Czesław Domański widely recognised among the academic and official statistics community. The aim of this article is to present the Professor's accomplishments in the context of his memories, shedding light on the inspirations that helped him achieve success and led to reflections on the role of statistics in the modern world.

[a] University of Lodz, Faculty of Economics and Sociology, ul. POW 3/5, 90-255 Łódź; Statistical Office in Łódź, ul. Suwalska 29, 93-176 Łódź, e-mail: alina.jedrzejczak@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-5478-9284.
[b] University of Lodz, Faculty of Economics and Sociology, ul. POW 3/5, 90-255 Łódź, e-mail: jerzy.kowaleski@wp.pl, ORCID: https://orcid.org/0000-0003-2123-0203.
[c] University of Lodz, Faculty of Economics and Sociology, ul. POW 3/5, 90-255 Łódź, e-mail: aleksandra.baszczynska@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-4477-2438.

Professor Czesław Domański was born on 11 April 1943 in the town of Dąbrówka in the Kutno district, which was annexed to Nazi Germany during the occupation. After graduating from the Pedagogical High School in Radomsko and later the Ewaryst Estkowski Teacher Training School in Łódź, he began his mathematical studies in 1963 at the Faculty of Mathematics, Physics and Chemistry of the University of Lodz. He completed his studies in 1968, earning a master's degree in mathematics for his work in the field of cybernetics and information theory entitled *Non-Comma Dictionaries* (Pol. Słowniki bezprzecinkowe), supervised by Jerzy Jaroń, PhD, DSc.

Professor Czesław Domański's life and 55-year professional career have been consistently linked to the University of Lodz. Significant stages in its development included holding the following positions:

- Vice-Director of the Institute of Econometrics and Statistics at the University of Lodz (1976–1984);
- Vice-Dean of the Faculty of Economics and Sociology at the University of Lodz (1987–1990);
- Head of the Statistical Methods Unit at the Institute of Econometrics and Statistics at the University of Lodz (1987–1991);
- Dean of the Faculty of Economics and Sociology at the University of Lodz (1990–1993);
- Head of the Department of Statistical Methods (1992–2017);
- Vice-Rector for Economic Affairs and Promotion at the University of Lodz (1993–1996);
- Director of the Institute of Econometrics and Statistics at the University of Lodz (1997–2008);
- Director of the Institute of Statistics and Demography (2009–2016).

The beginning of the Professor's academic career dates back to 1968, when he started working as an academic teacher at the University of Lodz, initially as a teaching assistant and later as an assistant professor in the Department of Demography and Statistics at the Institute of Econometrics and Statistics. This moment turned out to be pivotal in his life, as it marked the beginning of a 55-year career dedicated to the study, teaching and popularisation of statistics.

Professor Czesław Domański had encountered statistics during his university studies, but the subject didn't raise much of his interest at the time, perhaps because it was not offered as a separate subject but was only a part of Dr Tadeusz Gerstenkorn's lecture on 'Probability Calculus' (Pol. Rachunek prawdopodobieństwa), temporarily taught by Dr Ryszard Jajte. Thus, his first significant encounter with statistics was when he began working at the Department of Demography and Statistics at the University of Lodz:

> My first encounter with statistics took place on 28 September 1968 at the Department of Demography and Statistics at the University of Lodz. Professor Edward Rosset offered me the position of a teaching assistant at the Department and informed me that I would be teaching 'Mathematical Statistics' to third-year Econometrics students, starting the following week. For any current teaching matters, I was to report to Dr Zbigniew Michałkiewicz; at that time, econometrics was a prestigious field of study, not only within the department but also university-wide and across Poland. The students were then highly proficient in mathematics.[1]

Professor Domański recalls his first teaching experience as follows:

> I reported to Dr Zbigniew Michałkiewicz, who worked in room no. 6 alongside young master's graduates: Janusz Murlewski, Stanisław Bartczak, and Włodzimierz Obraniak. The future head of the Department of Demography and Statistics understood the information from Professor Rosset and informed me that, concerning the statistics classes, I should contact Jolanta Martini, MSc, a mathematician, who was also conducting the same 'Mathematical Statistics' practicals in the other group. The next day, I went to the University's HR department to navigate through a lengthy and complex formal process. Previously, obtaining a position at the university was out of the question if a candidate had participated in the student strike of March 1968 (the Students' March of 1968 was a series of protests against the ruling Polish United Workers' Party of the Polish People's Republic). After two days, I gathered the necessary opinions that did not object to my employment. In the meantime, I met Jolanta Martini, MSc, who was pleased that she would have less work and lent me the book *Probability Theory and Mathematical Statistics* (Pol. Rachunek prawdopodobieństwa i statystyka matematyczna) by Marek Fisz.

The Professor emphasised on several occasions that the cited work by Marek Fisz marked the beginning of his journey to the understanding, acquisition and expansion of his knowledge in statistics, first in the sphere of teaching and then research: 'It was the second edition of this book from 1958, somewhat thinner than the third edition consisting of 694 pages, which is on the bookshelf of my library'.

In addition to Fisz's book and other works in the field of mathematical statistics, lectures and practicals in the area of representative methods conducted by Dr Jerzy Greń played a significant role in his development as an educator. The Professor recalls:

> In the winter semester of the 1969/70 academic year, the subject of representative methods was included in the curriculum for the econometrics programme, with Dr Jerzy Greń from the Main School of Planning and Statistics as the lecturer. I was a student in these very

[1] All statements come from a private interview given to employees of the Department of Statistical Methods of the University of Lodz.

engaging lectures and discussions during my time in the Department, and I conducted the practicals for this subject.

It should be mentioned here that the times in which Professor Czesław Domański began his scientific and teaching career remained challenging for statistics. In the years 1950–1956, statistics was considered a 'dangerous' science and was in different ways eliminated, restricted and adapted to meet the needs of the state authorities. As a consequence, especially in economic universities, there was a strong focus on teaching as a means of coping with these limitations. Such a strategy was employed in the Higher School of Economics in Łódź, which, for a certain period, even offered a statistics programme.

The first Polish Science Congress, held from 29 June to 2 July 1951 in Warsaw, where the Section of Economic Sciences and the Subsection of Statistics were established, ignored the existence of the Polish Statistical Association (PSA). In 1952, a Conference of Higher Schools of Economics was held, with representatives from the 15 statistics departments then existing in higher education institutions in Poland. During the conference, individual departments presented their research and work programmes.

Another conference was convened by the Department of Economic Studies of the Ministry of Higher Education, and its organisation was entrusted to the Department of Economic Statistics at the Main School of Planning and Statistics in Warsaw. The conference took place on 25–30 October 1958 in Zakopane. Representatives from 13 statistics departments and Statistics Poland participated in the event. Numerous topics related to research and organisational work were discussed, and among them the initiative to establish an Interuniversity Institute of Statistics affiliated to the Polish Academy of Sciences. The vast majority of statisticians present at this conference did not accept the idea of creating an Interuniversity Institute of Statistics, which entailed the centralisation of the planning of Polish statistics. Academic teachers skilfully overcame this impasse by introducing widespread econometrics while still teaching statistics with various 'adjectives'. The 'econometrisation' process was accelerated by the establishment of the Econometrics School by Prof. Zbigniew Pawłowski, under the auspices of the then Ministry of Higher Education and Science.

Despite the many difficulties associated with work in higher education institutions during that period, a strong emphasis was placed on teaching, and the development of the ethos of an academic teacher played a particularly important role. Professor Czesław Domański often mentions these role models:

> Superiors were very demanding and ensured that young adepts of teaching and science fully developed both intellectually and socially. They emphasised the pedagogical and educational role, as well as their influence on academic youth through a meticulous preparation for exercises, punctuality, appearance, erudition, and involvement in students' lives outside the university – as supervisors of student dormitories, as well as scientific, sports, and other clubs.

Teaching became a true passion for Professor Domański, which resulted in the preparation of a lecture and practicals programme for the subject of 'Survey sampling'. This was accompanied by the development of a script titled *A collection of exercises in survey sampling* (Pol. Zbiór zadań z metody reprezentacyjnej), of which he was the editor and co-author. However, the greatest popularity and student interest was garnered by the script first published by the University of Lodz Publishing House in 1979 under the title *A collection of statistics. Exercises* (Pol. Zbiór zadań ze statystyki). As it was modified and expanded in the subsequent editions (the latest, sixth edition in 2001 with a print run of 3,000 copies), this book, under the editorship of Czesław Domański, transformed into a statistics textbook titled *Statistical methods, theory, and exercises* (Pol. Metody statystyczne, teoria i zadania), which continues to be highly useful.

The earlier years of professor Domański's work at the Institute of Econometrics and Statistics (1970–1979), when he chaired the Institute's Committee for Didactics, resulted in the development of numerous beneficial teaching-related initiatives and actions. Thanks to his effort, collections of supplementary materials for current statistics classes were published. During his tenure as the Vice-Dean and later as the Dean of the Faculty, he initiated changes and alterations to the curriculum for all existing study programmes in line with the requirements and needs of the market economy.

Since 1997, Professor Czesław Domański has been the main organiser of the annual educational conference dedicated to 'Teaching methods for quantitative subjects'. A valuable achievement in this regard is the integration of the community of mathematicians, statisticians, economists and computer scientists – both academic teachers and representatives of the business world and various levels of local government, including the City of Łódź.

In his teaching, in addition to his professional expertise, Professor Domański emphasises the cultivation of patriotic and civic attitudes in the younger generations and the ethical aspect of acquiring and utilising knowledge gained during their studies. Therefore, he is both a teacher and a mentor.

Professor Domański's pedagogical work is closely connected with his popularisation activity. Its important aspect is the promotion of Polish statisticians and their contribution to the development of statistical research, both nationally and internationally. The Professor has authored numerous biographies of prominent figures in this field who lived in the 19th and 20th century. These biographies were included in a publication issued by Statistics Poland: *Profiles of Polish Statisticians* (Pol. Sylwetki statystyków polskich), which was published in 1984 in Polish, English and Russian, and later in 1989 and 1990, as well as in the compilation *Polish Statisticians* (Pol. Statystycy polscy), edited by Prof. Mirosław Krzyśko on the 100th anniversary of the establishment of the PSA.

In addition to his pedagogical work, Professor Czesław Domański has been developing a wide-ranging and successful scientific activity almost from the beginning of his employment at the University of Lodz.

As a result of the enduring scientific achievements over the years, a Łódź centre, managed by Professor Czesław Domański has advanced to become a significant centre of statistical research, and he himself is considered among the national academic environment as one of the founders of the Łódź School of Nonparametric Statistics.

Within the common realm of scientific and pedagogical work, Professor Czesław Domański supervises and reviews doctoral dissertations and other academic works written to attain higher academic degrees. To date, Professor Czesław Domański has supervised 29 doctoral candidates, reviewed 154 doctoral and postdoctoral dissertations and proceedings for the appointment of the title of professor. What led to these accomplishments and provided inspiration for the Professor's scientific endeavours?

Initially, when Mr Czesław Domański started his academic career, he had to delve into the fundamentals of the areas of knowledge that were new to him, but from the beginning, he took this aspect of academic work very seriously.

> In the first year of my work, I essentially continued my studies in the field of descriptive, economic and mathematical statistics, as well as in survey sampling. Simultaneously, having excellent mentors for each subject, such as Zbigniew Michałkiewicz, Władysław Welfe, Zofia Zarzycka, Tadeusz Miller and Jerzy Gren, I easily absorbed this new knowledge, and, above all, I became convinced of its immense utility in every domain of economic, social and natural life. The practical side of statistics became apparent to me most instantly in the field of medical sciences thanks to Tadeusz Miller, who introduced me to the medical environment as a statistics consultant, first for doctoral, then postdoctoral theses, as well as those needed to obtain the title of professor. Later, we conducted joint research in cooperation with the Medical Academy and the Military Medical Academy. I was associated with the latter institution for over 10 years as a lecturer in medical statistics for future medical officer students, which allowed me to gain insight into the research needs of this scientific community.

All of this led to a growing interest on the Professor's part in mathematical statistics and its applications in various fields. However, the pivotal event occurred in 1970.

> A pivotal moment in my scientific career was my encounter with Professor Zdzisław Hellwig in Krosno Wielkopolskie, near Kępno, during my days at the aforementioned Econometrics School. Professor Hellwig presented several problems that were to be the subject of doctoral research. After selecting a topic and working on its development, I was admitted to his seminars following a few hours of presenting the chosen subject at the chalkboard. The professor stated that I could participate in his seminars and emphasised that I had chosen a good topic, i.e. the 'Application of the runs theory based on the number and length of runs'. He justified this by pointing out that it provided opportunities for research considering both its analytical and applied aspects, especially in the preferred at that time field of econometrics.

> The doctoral seminars with Professor Hellwig with his scientific passion and presentation skills ignited a kind of enthusiasm in my process of learning new statistical methods and their versatile applications. Analytical problems within the scope of the runs theory were rather complex, because they dealt with the distribution of discrete variables concerning both the number and length of runs.

The new research area undertaken by Czesław Domański, MSc, covering various issues related to the construction of statistical tests based on the number or length of runs, was very well received by the scientific community. Czesław Domański obtained funding as part of Core Problem 06.0.1.06.2.04, led by the Institute of Mathematics of the Polish Academy of Sciences. The results were presented in two reports: 'On the convergence of test statistics for some tests based on the chi-square distribution' (Pol. O zbieżności sprawdzianów niektórych testów do rozkładu chi-kwadrat) and 'Tables of the unconditional distribution of the number of runs' (Pol. Tablice bezwarunkowego rozkładu liczby serii), presented before the committee chaired by Professor Jerzy Łoś. One example of these studies is the article 'A runs test based on the number of runs' (Pol. Test serii oparty na liczbie serii) (1973), published in the *Polish Statistical Review* journal.

> The chosen topic of my doctoral dissertation allowed me to present it to Professor Jacob Wolfowitz during the 40th session of the International Statistical Institute in Warsaw thanks to a recommendation from Professor Zdzisław Hellwig. At this meeting, I learned the mechanism of constructing runs distribution tables, whose first constructor was Professor Jacob Wolfowitz.

Professor Domański's thus initiated scientific career continues to this day. He obtained his doctorate in 1976 for his thesis entitled *Econometric Applications of Runs Tests* (Pol. Ekonometryczne zastosowania testów serii) written under the supervision of Professor Zdzisław Hellwig. The basis for his postdoctoral degree obtained in 1986 was the work entitled *Theoretical foundations of nonparametric tests and their application in socioeconomic sciences*, published by the University of Lodz. The book *Statistical Tests* (Pol. Testy statystyczne) from 1990, along with the earlier scientific achievements of Czesław Domański, formed the basis for obtaining the title of professor, which took place in 1991.

Among the numerous book publications, the following monographs deserve special recognition: *Nonparametric statistical tests* (Pol. Statystyczne testy nieparametryczne) (PWE, 1979), *Statistical tests* (Pol. Testy statystyczne) (PWE, 1990), *Non-classical statistical methods* (Pol. Nieklasyczne metody statystyczne) (PWE, 2000), *Statistical expert systems* (Pol. Statystyczne systemy ekspertowe) (University of Lodz Publishing, 1998; co-authored by H. Gadecki, K. Pruska, A. Rossa), *Small area statistical methods* (Pol. Metody statystyki małych obszarów) (University of Lodz Publishing, 2000; co-authored by K. Pruska), *Statistical methods of multiple inference* (Pol. Statystyczne metody wnioskowania wielokrotnego) (University of Lodz Publishing, 2007; co-authored by D. Parys), *Non-classical methods of efficiency and risk assessment. Open pension funds* (Pol. Nieklasyczne metody oceny efektywności i ryzyka. Otwarte Fundusze Emerytalne) (PWE, 2011; co-authored by J. Białek, K. Bolonek-Lasoń, A. Mikulec).

A truly exceptional aspect of Professor Czesław Domański's activity lies in his involvement in the field of official statistics. He has belonged to a variety of national and international statistical societies, and it was his efforts that led to the reactivation of the PSA. Furthermore, he was involved in the organisational work of the Association: he served as its Vice President (1985–1990) and President during the successive terms of 1994–2000, 2000–2005, and 2010–2018, as well as member of the Main Council and as Honorary President of the PSA. The Professor's achievements in these roles resulted in the intensified integration of the academic and statistical community with Statistics Poland and statistical offices in various regions of Poland. Additionally, he played an active role in the publication of works related to the history of Polish statistics and contributed to the revival of the *Statistical Quarterly* (Pol. Kwartalnik Statystyczny) journal. Professor Domański's organisational achievements include the establishment of the Bureau of Statistical Research and Analysis within the structures of the PSA. The collaboration between the PSA and Statistics Poland was based on the consolidation of Professor Czesław Domański's activities with the authorities of Statistics Poland, particularly evident in the interactions with Professor Jan Kordos. Professor Czesław Domański recalls the beginnings of this exceptional cooperation as follows:

> My first scientific article was titled *Regional income distribution* (1970), which attracted the interest of Dr Jan Kordos. This article facilitated my contact with Statistics Poland, as I was invited to participate in a seminar led by Dr Kordos, who was then deputy director and soon thereafter director of the Department of Living Conditions.

Professor Domański also took part in numerous initiatives directly serving the goals of official statistics in Poland. He was involved in the organisation of national censuses in the years 1970, 1978, 1988, 2002, 2011, and 2021; he cooperated within mathematical commissions, prepared expert opinions, worked as a consultant and engaged in activities related to the Scientific Statistical Council under the President of Statistics Poland, the Methodological Commission of Statistics Poland, and the Statistical Council under the Prime Minister. When reminiscing about the beginnings of his work in official statistics, Professor Czesław Domański stated that:

> There was excellent collaboration with the statistical office in Łódź and the authorities of Łódkie Voivodship, particularly with directors Wiktor Pietruszka, Stanisław Kwiatkowski, and Wincenty Imieniński, especially during the preparations for the 1970 National Census. All the assistants from the department underwent a one-week training programme covering the topics, organisation and methods of conducting the census. After this training, we prepared future enumerators, primarily recruited from the University of Lodz. Z. Michałkiewicz, Assoc. Prof., serving as the Rector's Commissioner for the 1970 National Census, was responsible for overseeing the enumeration districts in the city centre and Widzew district. I participated in this challenging supervision, along with Prof. Michałkiewicz at various organisational levels of the 1970 National Census in Łódź. The 1970 National Census was originally scheduled to last a week but was shortened to Wednesday due to the 'December Events' on the coast (protests sparked by a sudden increase in the prices of food and suppressed by the Polish People's Army, resulting in at least 44 people killed). Nevertheless, the census was informally extended with earlier arranged individuals. This period appeared to bring scholars closer together and expand their areas of interest, both academically and professionally. In my case, it significantly expanded my collaboration with the Director of the statistical office in Łódź, Stanisław Kwiatkowski, as part of the Łódź branch of Statistics Poland. This collaboration led to the publication of the book *Biographies of Polish Statisticians* (Sylwetki statystyków polskich) in 1984. The book received considerable attention among the international community of statisticians. The meetings which I held with my first head of department, Professor Edward Rosset, at his home over several months, featured hours-long conversations and discussions focused primarily on demography and statistics and the future development of these disciplines; they have been of significant importance to the field of statistics.

Since 2002, Professor Domański has been a regular member of the International Statistical Institute (ISI).

Professor Czesław Domański's profound knowledge and competence have been recognised in various committees responsible for the development of statistical research and broader, national-level scientific research. Particularly noteworthy is his three-term membership in the Central Commission for Academic Degrees and Titles, membership in the Methodological Committee under the President of Statistics Poland, and his appointment to the Statistics Council under the Prime Minister for the years 2001–2002. Since 1987, Professor Domański has been elected to the Committee on Statistics and Econometrics of the Polish Academy of Sciences, where he served as the Vice Chairman during the 2008–2012 term.

Professor Domański's commitment in the service of Polish statistics has been acknowledged through numerous distinctions, among which he himself considers the following as the most significant: the Medal of the National Education Commission, the Officer's Cross of the Order of Polonia Restituta, the Bolzano Medal awarded by Charles University in Prague and the Jerzy Spława-Neyman Medal.

In addition to methodological issues, such as the theoretical foundations of nonparametric tests, methods for comparing nonparametric statistical inference procedures with parametric procedures, and small area statistics, Professor Czesław Domański's scientific research focuses to a large extent on its practical application, aimed at solving real problems within various socio-economic issues, particularly in Łódź and the Łódź region. These include studies related to:

- developing the Łódź scale of child and youth development;
- the statistical assessment of newborns' health and causes of infant mortality in maternity hospitals in the city of Łódź;
- constructing the Łódź operational risk card for cardiac surgery procedures;
- creating a Polish operational risk scale for the treatment of ischemic heart disease;
- statistical models of water consumption of Łódź residents;
- predicting water consumption in Łódź based on modified statistical and econometric models;
- using machine learning for forecasting household water demand in Łódź;
- analysing energy heat forecasts for Łódź residents;
- contributing to the development of Polish statistical thought and the work of Polish statisticians;
- upholding the academic tradition of Łódź;
- ensuring the quality of economics education in response to the expectations of Łódź entrepreneurs.

An important aspect of his scientific activity is his participation in and organisation of scientific conferences. Professor Czesław Domański has been for decades the initiator and main organiser of the 'Multivariate Statistical Analysis – MSA' international conference. As many as 41 editions of this conference have been held to date.

Another significant undertaking of the Professor was organising the 'Ethics in Economic Life' conference at the Salesian University of Economics and Management, with which he was affiliated since its establishment.

The list of all of the Professor's achievements, including scientific articles is extensive and goes beyond the scope of the current presentation.[2]

Professor Czesław Domański's 55-year work in the service of statistics have indeed yielded an impressive body of work across all areas of academic life related to official statistics. Professor Czesław Domański has always emphasised the necessity of collaboration between the academic community and practitioners, which best summarises this article:

> The real confrontation of statistical methodology with specific issues related to official statistics creates a platform for discussion between empirical researchers and academic representatives. The experience gained, and above all, the successes in solving complex organisational and methodological issues have been the driving force for further research.
> It seems that the diversity of the problems to be solved, extensive discussions among practitioners and statisticians generate great satisfaction from participating in such events. My scientific research, as well as my organisational and educational path have always been associated with statistics thanks to many outstanding mathematicians, statisticians, econometricians and scientists whom I have met throughout my academic journey and whom I will always remember.

[2] A detailed description in this regard is provided in the works by Jędrzejczak, A. (2014). Profesor Czesław Domański – twórca łódzkiej szkoły statystyki nieparametrycznej. *Przegląd Statystyczny. Statistical Review*, *61*(4), 453–458 and Jędrzejczak, A., & Kowaleski, J. T. (2013). Profesor Czesław Domański – życiorys naukowy. *Acta Universitatis Lodziensis. Folia Oeconomica*, (280), 5–33.