Borradores de ECONOMÍA

Extracting the sovereigns' CDS market hierarchy: a correlation-filtering approach

Por: Carlos León, Karen Leiton, Jhonatan Pérez

Núm. 766
2013

BANCO DE LA REPÚBLICA · COLOMBIA

- Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia -

Extracting the sovereigns' CDS market hierarchy:
a correlation-filtering approach[1]

Carlos León[2]
Karen Leiton[3]
Jhonatan Pérez[4]

Banco de la República (Central Bank of Colombia)

Abstract

Since correlation may be interpreted as a measure of the influence across time-series, it may be conveniently mapped into a distance and into a weighted adjacency matrix. Based on such matrix, network theory has attempted to filter out the noise in correlation matrices by extracting the dominant hierarchy (i.e. the strongest linear-dependence signals) within time-series.

The aim of this brief paper is to find the current hierarchy in the sovereigns' CDS market after the structural shift caused by the failure of Lehman Brothers. Thus, based on two different correlation-into-distance mapping techniques and a *minimal spanning tree*-based correlation-filtering methodology on 36 sovereign CDS spread time-series, the target is to identify which sovereigns are providing the strongest –less noisy- and most informative signals.

The resulting sovereigns' CDS market hierarchy agrees with prior findings of Gilmore et al. (2010) regarding sovereigns' bonds market, such as the importance of geographical clustering and the idiosyncratic nature of Japan and United States. Additionally, results (i) confirm that a small set of common factors affect the entire system; (ii) identify the relevance of credit rating clustering; (iii) identify Russia, Turkey and Brazil as regional benchmarks; (iv) suggest that *lower-medium grade* rated sovereigns are the most influential, but also the most prone to contagion; and (v) suggest the existence of a "Latin American common factor".

Keywords: correlation, minimal spanning tree, correlation-filtering, sovereign, credit default swap.

JEL classification: C32, G10, G11

---

[1] The opinions and statements are the sole responsibility of the authors and do not necessarily represent neither those of Banco de la República nor of its Board of Directors. Results are illustrative; they may not be used to infer credit quality or to make any type of assessment. Comments and suggestions from Clara Lía Machado, Silvia Juliana Mera, Alejandro Reveiz and Miguel Sarmiento are acknowledged and appreciated. As usual, any remaining errors are the authors' own.

[2] Research and Development Section Manager, Financial Infrastructure Oversight Department, Payment Systems and Banking Operation Division, Banco de la República, cleonrin@banrep.gov.co / carlosleonr@hotmail.com. [corresponding author]

[3] Specialized Professional, Department of Operations and Market Development, Monetary and International Reserves Affairs Division, Banco de la República, kleitoro@banrep.gov.co.

[4] Specialized Professional, Financial Infrastructure Oversight Department, Payment Systems and Banking Operation Division, Banco de la República, jperezvi@banrep.gov.co.

Contents

1. Introduction 2
2. Correlation as a measure of distance 4
3. From CDS spreads to the CDS hierarchical structure by means of the *minimal spanning tree* (MST) 6
   3.1. The correlation matrix 7
   3.2. The minimal spanning tree (MST) 12
4. Analysis of the resulting MST 14
5. Final remarks 19
6. References 22

1. Introduction

The correlation coefficient is a standard metric for assessing statistical dependence between assets' time-series. Correlation is commonly defined as the degree of linear association between two variables, where this association does not imply that changes in a variable cause changes in the other, but that the movements of both variables are on average related to an extent given by the correlation coefficient (Brooks, 2008).

The correlation coefficient for variables x and y ($\rho_{(x,y)}$) is estimated as in [§1], where σ_x corresponds to the standard deviation of x, and $\sigma_{(x,y)}$ to the covariance between variables x and y. By construction $\rho_{(x,y)}$ can assume values in the [-1,1] interval. The upper bound corresponds to two variables exhibiting a direct and perfect linear association, whereas the lower bound corresponds to two variables exhibiting an inverse perfect linear association. Between the upper and lower bounds diverse degrees of dependence exist, where the middle (i.e. zero correlation) corresponds to the absence of a linear association between the variables; however, the absence of correlation between two variables may not be understood as variables being independent.[5]

$$\rho = \frac{\sigma_{(x,y)}}{\sigma_x \sigma_y} \qquad [§1]$$

According to Focardi and Fabozzi (2004), correlation may also be regarded as a quantitative measure of the strength of dependence, where two variables are dependent if they move together. If the correlation coefficient is positive, they tend to move together, and they tend to be above or below their respective means in the same state; if it is negative, they still move together but their movement tends to be contrary, and one tends to be above its mean while the other tends to be below its mean. Thus, in both cases, when the correlation coefficient approximates the upper or lower bounds, the strength of the dependence increases; on the other hand, as the coefficient approximates zero, dependence weakens.

The correlation coefficient's intuitiveness and ease of estimation has been its main source of popularity for financial and economic modeling. It is beneath the most prominent valuation (e.g. Capital Asset Pricing Model – CAPM), asset allocation (e.g. Modern Portfolio Theory), and risk (e.g. variance-covariance Value at Risk) models, where it is customary to estimate the correlation matrix (P).

$$\mathrm{P} = \begin{bmatrix} 1 & \cdots & \rho_{(1,n)} & \cdots & \rho_{(1,N)} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \rho_{(n,1)} & \cdots & 1 & \cdots & \rho_{(n,N)} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ \rho_{(N,1)} & \cdots & \rho_{(N,n)} & \cdots & 1 \end{bmatrix} \qquad [§2]$$

[5] The upper bound implies that a linear equation fully (i.e. without any noise) describes the relation between two variables, with both variables increasing in tandem; in the lower bound such noiseless relation between the two variables is inverse (i.e. one increases as the other decreases). It is worth highlighting that if both variables are independent, the correlation coefficient is zero; however, the converse is not true, and variables may be non-linearly dependent despite being linearly independent.

As in [§2], if there are N-variables, (i) P contains all attainable correlations between the existing variables within a $N \times N$ matrix, where $\rho_{(x,y)}$ is the correlation between variables x and y; (ii) P is square and symmetrical ($\rho_{(x,y)} = \rho_{(y,x)}$); and (iii) since –by definition- the linear dependence between two identical time-series is 1, the main diagonal of P is constituted of 1s ($\rho_{(x,x)} = 1$).

Despite its simple estimation and interpretation[6], analyzing a non-small set of cross correlations may be challenging due to the dimensionality (i.e. $(N^2 - N)/2$) of the resulting (P) correlation matrix.[7] In this sense, the informational content of correlation declines rather rapidly when adding variables to the corresponding matrix.

However, since the correlation matrix may be interpreted as a measure of the strength or influence between variables, individual correlation coefficients may be conveniently transformed into a measure of the distance among variables; such transformation or mapping may take many mathematical forms, but should always observe an intuitive inverse relation between strength and distance (i.e. the higher the strength, the shorter the distance) and comply with some axioms regarding what a metric for distance is. After obtaining a matrix of distances, network theory techniques may be used to filter the resulting weighted adjacency matrix in search for relevant information regarding the structure of the underlying correlation matrix.

The aim of this brief paper is to find the current hierarchy in the sovereigns' CDS market after the structural shift caused by the failure of Lehman Brothers. Thus, based on a *minimal spanning tree*-based correlation-filtering methodology on 36 sovereign CDS spread time-series, the target is to identify which sovereigns are providing the strongest –less noisy- and most informative signals. [8]

However, unlike most of the existing literature (e.g. Mantegna, 1999; Mantegna and Stanley, 2000; Gilmore et al., 2010), this paper embraces a different correlation-into-distance mapping method. As explained below, instead of –customarily- mapping the lowest (-1) and the highest correlation (1) to the most distant and the closest, respectively, the authors' choice is to consider zero correlation as the most distant case, whereas the lowest and highest correlation are considered as the closest cases. This choice is rather intuitive since the presence of a strong linear dependence, positive or negative, conveys a meaningful informational content, whilst the absence of linear dependence between two variables implies the lack of informational content among them. For the sake of comparison and theoretical soundness, both transformations of correlation into distance are reported; nevertheless, as will be clear below, both are equivalent for the time-series considered.

---

[6] However, it is worth noting that the misinterpretation of the correlation coefficient is not unusual; Rebonato (2007) documents some usual misunderstandings about the interpretation of correlation coefficients.

[7] Please note that the relevant dimensionality of the correlation matrix is $(N^2 - N)/2$, where subtracting N results from the main diagonal elements being always equal to 1, whereas dividing by 2 is due to the symmetry of the matrix.

[8] The reader should be aware that using the standard (i.e. Pearson) correlation for financial time-series has been documented as displaying serious shortcomings due to the presence of non-normality and long-term dependence (e.g. León et al. 2012; León and Reveiz, 2011; Malevergne and Sornette, 2006; Holton, 1992). These issues are acknowledged but not considered in this paper.

The resulting sovereigns' CDS market hierarchy agrees with prior findings of Gilmore et al. (2010) regarding sovereign bonds market, such as the importance of geographical clustering and the idiosyncratic nature of Japan and United States. Additionally, results (i) confirm that a small set of common factors affect the entire system; (ii) identify the relevance of credit rating clustering; (iii) identify Russia, Turkey and Brazil as regional benchmarks; (iv) suggest that *lower-medium grade* rated sovereigns are the most influential, but also the most prone to contagion; and (v) suggest the existence of a "Latin American common factor".

This paper consists of four sections, where the first one is this introduction. The second section is devoted to the rationale and procedure behind transforming correlation into a distance between time-series. The third describes how to extract time-series hierarchy by means of the most documented and strict filtering technique: the *minimal spanning tree* (hereafter referred as MST). The fourth section analyzes the resulting MST. The last section discusses the main results and enumerates some caveats and challenges regarding the application of the correlation-filtering method herein presented.

2. Correlation as a measure of distance

As previously documented, correlation may be regarded as a quantitative measure of the strength of the dependence between two variables (Focardi and Fabozzi, 2004). In this sense, two variables are (i) dependent if they move together, either concurrently or contrarily, or (ii) independent, if they move without any –linear- relation between them.

If the correlation coefficient is positive they are deemed as "correlated", they tend to move concurrently, and they tend to be above or below their respective means in the same state; if it is negative they are considered as "anti-correlated", their movement tends to be contrary, and one tends to be above its mean while the other tends to be below its mean. Thus, in both cases, when the correlation coefficient approximates the upper or lower bounds, the strength of the dependence increases (i.e. the relation between the two variables is less noisy); on the other hand, as the coefficient approximates zero, linear dependence weakens (i.e. the relation becomes noisy).

The first documented attempt to map correlation coefficients into distances pertains to Mantegna (1998), who aimed to using stock prices in order to find the kind of topological arrangement in US equity markets, and to search for the existence and nature of common economic factors that drive the evolution of stock prices. After Mantegna several authors have followed a similar approach to extract the topology of foreign exchange markets (Naylor et al., 2007; Mizuno et al., 2006), interest rates (Aste and Di Matteo, 2005), government bonds (Gilmore et al., 2010) commodities (Gilmore et al., 2012) international trade (Maeng et al., 2012) and stock markets (Kullmann et al., 2002; Bonanno et al., 2004 & 2003; Eryigit and Eryigit, 2009; Onnela et al., 2003; Mantegna and Stanley, 2000; Mantegna, 1999).

All works acknowledge the potential of correlation matrices for detecting the hierarchical organization of financial markets by means of employing tools and procedures developed to model physical systems. Likewise, they all recognize the need for transforming correlation into a standard Euclidean distance metric.[9] Let $\rho_{(x,y)}$ be the correlation coefficient between assets x and y, the most common distance metric ($\dot{\delta}_{(x,y)}$) is estimated as follows:

$$\dot{\delta}_{(x,y)} = \sqrt{2\left(1 - \rho_{(x,y)}\right)} \quad [§3]$$

This non-linear transformation procedure [§3], first proposed by Mantegna (1999), results in distances assuming values in the [0,2] interval, where the lowest (highest) attainable distance corresponds to correlation's upper (lower) bound. However, against intuition, this transformation results in two strongly anti-correlated assets ($\rho_{(x,y)} \sim -1$) being considered less informative (i.e. farther located) than two non-correlated assets ($\rho_{(x,y)} \sim 0$). In other words, under this mapping procedure, two variables exhibiting linear independence are regarded as more informative than two variables displaying a strong, yet negative, dependence.[10]

If the correlation-filtering procedure is intended for asset allocation or risk management purposes –where finding strong anti-correlation is the ultimate goal- the transformation procedure in [§3] is an intuitive and convenient choice. For instance, it will allow a portfolio manager to identify those individual assets or risk factors which are farther from his current allocation according to the resulting assets' hierarchy, where the farthest will correspond to those exhibiting strong negative correlations. In other words, [§3] is particularly appropriate when correlation-filtering aims to minimizing risk via diversification.

However, away from a risk minimization objective, considering strong negative correlations as non-informative may be inconvenient. In fact, Mantegna´s first attempt (Mantegna, 1998) consisted of a different mapping method [§4], where its mathematical form (i.e. with the correlation being squared) concurs with an informational objective.

$$\ddot{\delta}_{(x,y)} = 1 - \left(\rho_{(x,y)}\right)^2 \quad [§4]$$

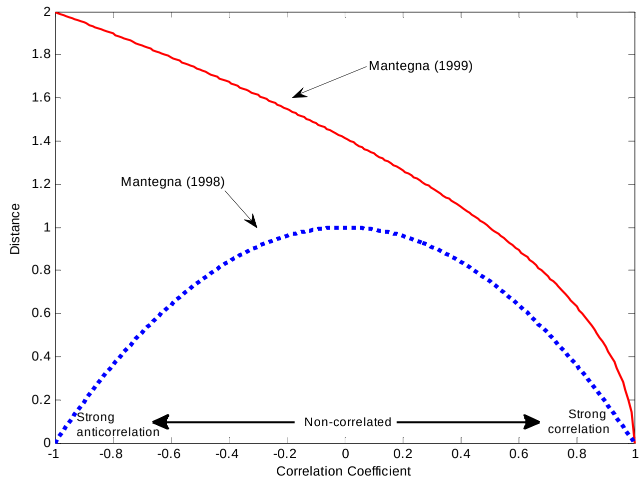
This transformation procedure, known as the squared Pearson distance (Deza and Deza, 2006), not only fulfills the three axioms of a Euclidean metric (Mantegna, 1998), but is particularly convenient when regarding distance as a metric for influence. Unlike [§3], the squared Pearson distance [§4] assumes values in the [0,1] interval, where the lowest attainable distance corresponds to perfect –positive or negative- dependence between the variables, and where linearly independent variables are considered as conveying noisy (i.e. weak) informational content. The graphical comparison of both transformations is presented in Figure 1, where the red (solid) and blue (dashed) lines correspond to [§3] and [§4], respectively.

---

[9] Let $d_{(i,j)}$ be the distance between points i and j, an Euclidean distance fulfills three axioms: (i) $d_{(i,j)} = 0$ if, and only if, $i = j$; (ii) $d_{(i,j)} = d_{(j,i)}$; and (iii) $d_{(i,j)} \leq d_{(i,k)} + d_{(k,j)}$.

[10] For instance, a correlation of -0.99, which corresponds to a particularly significant negative dependence between two series, would result in the two series being significantly apart ($\delta_{(x,y)} = 1.99$) and mutually non-informative, whereas two linearly independent series ($\rho_{(x,y)} = 0$) would be regarded as being closer ($\delta_{(x,y)} = 1.41$) and more informative.

Figure 1
Correlation-into-distance mapping methods
([§4] (Mantegna, 1998) and [§3] (Mantegna, 1999))



Source: authors' design.

Paraphrasing Mantegna (1999), the squared Pearson distance [§4] used in Mantegna (1998) complies with the axioms of a Euclidean distance, but it is only approximately correct, whereas [§3] is a theoretically rigorous definition of distance[11]; however, Mantegna (1999) also recognizes that results with both transformation procedures coincide.[12]

Consequently, in order to capture both aims (i.e. risk minimizing and informational), both mapping procedures are used in the calculations below. It is expected that if strong positive correlations dominate the dataset considered, both procedures should yield similar results, as in Mantegna (1999), and as graphically portrayed in Figure 1.[13]

3. From CDS spreads to the CDS hierarchical structure by means of the *minimal spanning tree* (MST)

The first part of this section describes the dataset and presents an initial –standard-characterization of the correlation matrix. The second part defines the MST methodology and presents the MSTs resulting from both correlation-into-distance mapping methods ([§3] and [§4]) on the selected dataset.

[11] Mantegna (1999) points out that [§4] not only complies with the axioms of Euclidean distances, but also with a more strict set of axioms that result in *ultrametric distances*. Mantegna and Stanley (2000) provide an introduction to *ultrametric distances* and their properties.
[12] Despite the source of such coincidence is not clearly stated by Mantegna (1999), based on the distances reported the authors presume that the coincidence emerges from the relative abundance of strong positive correlations (i.e. short distances), which results in both mapping procedures yielding a similar ranking for the MST.
[13] Such irrelevance for correlations above zero is important since MST algorithms are based on the ranking of distances between the time-series.

3.1. The correlation matrix

The dataset comprises 36 time-series of sovereign CDS spreads, corresponding to the following countries: Colombia, Argentina, Brazil, México, Perú, Venezuela, Chile, Portugal, Spain, Italy, Ireland, Belgium, France, United Kingdom, Austria, Germany, Netherlands, United States, Turkey, Israel, Russia, Poland, Hungary, Ukraine, Romania, Czech Republic, Bulgaria, South Africa, South Korea, Malaysia, Indonesia, Philippines, China, Australia, New Zealand and Japan. The dataset consists of daily (last) 5-year CDS spreads from November to February 2013 (i.e. 844 trading days).

The choice of the period, countries and maturity follows several considerations: (i) the aim of this paper is to find the current hierarchy in the sovereign CDS market, where the structural shift caused by the failure of Lehman Brothers and the resulting large-scale fiscal stimulus measures of the following months provide a natural breaking point[14]; (ii) all the sovereigns considered represent about 99.6% of the gross notional outstanding sovereign CDS[15]; (iii) some relevant sovereigns were discarded due to their CDS time-series being incomplete (e.g. Greece, Vietnam, Egypt)[16]; and (iv) the 5-year maturity is the most traded for CDS (Coudert and Gex, 2010).

According to the International Monetary Fund (2006), credit default swaps (CDS) are bilateral agreements to transfer the credit risk of one or more reference entities, where the structure of such agreement resembles an insurance contract in that it protects the "protection buyer" against predefined credit events –in particular the risk of default-affecting the reference entity (or entities), during the term of the contract, in return for a periodic fee paid to the "protection seller".[17] This fee results in the buyer of protection having a similar position as if he had sold short a bond issued by the reference entity, where the market price and the spread of the CDS reflect the riskiness of the underlying credit.

Working on sovereigns' CDS time-series is appealing since (i) evidence suggests a lead of CDS market –over bonds- for corporates and sovereigns, particularly for emerging markets' sovereigns (Coudert and Gex, 2010); (ii) if CDS spreads are taken as indicative of the risk of the underlying reference entity (i.e. sovereign), similar behavior of CDS spreads may suggest exposure to common risk factors (Marsh et al., 2003); (iii) CDS singles out default-risk, providing much sharper sovereigns' default-related information than the corresponding bond spreads (Neftci et al., 2003); (iv) CDS tend to be correlated (Wilmott, 2006), and this correlation has grown over the past years (Shino and Takahashi, 2010); and (v) as documented by Coudert and Gex (2010) and Shino and Takahashi (2010), despite still being outsized by corporate CDS, the sovereign CDS market has recently gained importance[18].

---

[14] There is evidence (e.g. Melo and Rincón, 2012; Fratzscher, 2011) of structural changes due to the 2007-2008 crisis.

[15] Based on the gross notional outstanding of sovereign CDS pertaining to the top-1000 CDS reference entities, as reported by DTCC as of March 22, 2013.

[16] Greece, a particularly interesting sovereign, was not included because from September 16 2011 the CDS spread is not available.

[17] As in Wagner (2008), despite the fee is often called the CDS spread, it is not quoted vis-à-vis a risk free benchmark, but it is quoted in basis points per annum of the contract's notional value; thus the term ''CDS rate'' would therefore be more appropriate.

[18] According to Coudert and Gex (2010), as of September 2009, the corporate CDS market had reached USD 9.7 trillion, whereas the corresponding long-term debt securities market had reached USD 10.0 trillion; on the other hand, the sovereign CDS market had reached USD 1.9 trillion, whereas the government debt market had reached

Log-returns were customarily calculated on the CDS spreads time-series. Their main statistical features are exhibited in Table 1.

Table 1
Main statistical features of time-series' log-returns
(daily prices; November 2009 – February 2013)

Country [S&P Credit rating]	ISO Code	Share of outstanding[c]	Mean	Std. Dev.	Skewness	Kurtosis
Argentina[a] [B-]	ARG	1.23%	0.001	0.039	-1.143	33.968
Australia[b] [AAA]	AUS	1.35%	0.000	0.036	-0.331	10.315
Austria[b] [AA+]	AUT	2.16%	-0.000	0.039	0.646	10.611
Belgium[b] [AA]	BEL	2.33%	0.001	0.039	-0.111	5.311
Brazil[a] [BBB]	BRA	5.05%	0.000	0.031	0.165	9.788
Bulgaria[a] [BBB]	BGR	0.63%	-0.001	0.033	-0.270	14.058
Chile[a] [AA-]	CHI	0.21%	0.000	0.031	0.285	7.229
China[a] [AA-]	CHN	2.49%	0.000	0.034	-0.627	11.440
Colombia[a,d] [BBB-]	COL	0.97%	0.000	0.030	0.210	9.397
Czech Republic[b] [AA-]	CZH	0.52%	-0.001	0.035	-0.047	14.855
France[b] [AA+]	FRA	6.31%	0.001	0.042	-0.031	4.974
Germany[b] [AAA]	DEU	5.53%	0.001	0.038	0.010	5.714
Hungary[b] [BB]	HUN	2.42%	0.000	0.032	0.696	17.278
Indonesia[a] [BB+]	IDN	1.37%	0.000	0.034	-0.299	10.073
Ireland[b] [BBB+]	IRL	1.76%	0.000	0.038	-0.318	15.007
Israel[b] [A+]	ISR	0.48%	0.000	0.024	0.552	8.546
Italy[b] [BBB+]	ITA	14.38%	0.001	0.046	-0.726	14.448
Japan[b] [AA-]	JPN	2.89%	0.000	0.030	0.419	7.812
Malaysia[a] [A-]	MYS	0.68%	0.000	0.038	-0.732	12.180
México[a] [BBB]	MEX	4.03%	0.000	0.032	0.385	9.850
Netherlands[b] [AAA]	NDL	1.24%	0.001	0.036	-0.189	6.507
New Zealand[b] [AA]	NZL	0.13%	0.000	0.034	-0.295	11.587
Perú[a] [BBB]	PER	0.77%	0.000	0.033	0.258	7.429
Philippines[a] [BB+]	PHL	1.57%	-0.001	0.029	-1.724	20.763
Poland[b] [A-]	POL	1.56%	-0.000	0.039	-0.143	11.207
Portugal[b] [BB]	PRT	2.66%	0.002	0.045	-1.053	19.934
Romania[a] [BB+]	ROM	0.62%	-0.000	0.030	0.613	16.307
Russian Fed.[a] [BBB]	RUS	4.21%	0.000	0.036	-0.344	10.975
South Africa[a] [BBB]	ZAF	1.91%	0.000	0.033	-0.586	10.768
South Korea[b] [A+]	KOR	2.88%	0.000	0.039	-0.821	14.928
Spain[b] [BBB-]	ESP	7.45%	0.001	0.048	-0.548	9.561
Turkey[a] [BB+]	TUR	5.03%	0.000	0.030	-0.510	9.019
Ukraine[a] [B]	UKR	1.07%	-0.001	0.025	-0.661	8.450
United Kingdom[b] [AAA]	GBR	2.52%	0.000	0.032	-0.111	5.892
United States[b] [AA+]	USA	0.84%	0.000	0.030	0.518	12.102
Venezuela[a] [B+]	VEN	1.60%	-0.001	0.025	0.398	5.801

[a] Upper-middle-income or lower-middle-income country; [b] high-income country; [c] share of sovereigns' gross notional outstanding pertaining to the top-1000 reference entities, as reported by DTCC, as of March 22, 2013. [d] on April 24 2012 Colombia was upgraded to BBB. Credit ratings correspond to long-term ratings reported by Standard & Poor's, as of April 4, 2013. Source: authors' calculations, based on Bloomberg, DTCC, The World Bank and Standard & Poor's.

USD 36 trillion. As reported by Shino and Takahashi (2010), the growth rate of sovereign's CDS outstanding has increased sharply (i.e. 30.8% during 2009) due to large-scale fiscal stimulus after the failure of Lehman Brothers and concerns over sovereign risk in some European countries after the autumn of 2009.

As usual in financial data with daily frequency, (i) the standard deviation dominates the mean; (ii) skewness is non-negligible; (iii) excess kurtosis is significant; and (iv) any standard test (e.g. Jarque-Bera) rejects the null hypothesis of log-returns pertaining to a Normal –or elliptical- distribution.

Following Mantegna (1999) all time-series were transformed to individual vectors in a unitary form, namely by subtracting to each record its corresponding mean value and normalizing it to its standard deviation. Afterwards, the correlation matrix was ordinarily estimated. As expected, the result is a $N \times N$ square and symmetrical matrix (Table 2):

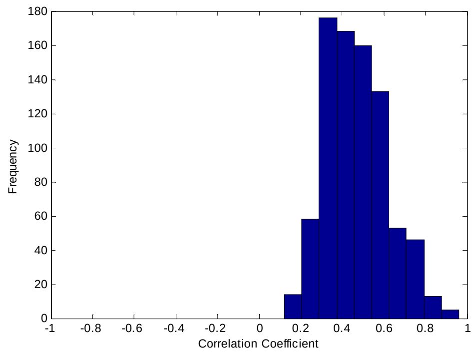
Table 2
Correlation matrix
(daily prices; November 24, 2009 – February 22, 2013)

	COL [BBB-]	ARG [B-]	BRA [BBB]	MEX [BBB]	PER [BBB]	VEN [B+]	CHL [AA-]	PRT [BB]	ESP [BBB-]	DEU [AAA]	ITA [BBB+]	BEL [AA]	USA [AA+]	FRA [AA+]	GBR [AAA]	JPN [AA-]	AUS [AAA]	MYS [A-]	TUR [BB+]	RUS [BBB]	CHN [AA-]	IDN [BB+]	NZL [AA]	KOR [A+]	ZAF [BBB]	ROM [BB+]	POL [A-]	ISR [A+]	HUN [BB]	UKR [B]	PHL [BB+]	BGR [BBB]	CZH [AA-]	AUT [AA+]	NLD [AAA]	IRL [BBB+]
COL [BBB-]	1,00	0,55	0,96	0,94	0,86	0,59	0,71	0,43	0,43	0,49	0,46	0,49	0,32	0,47	0,47	0,25	0,50	0,41	0,69	0,69	0,46	0,40	0,35	0,48	0,64	0,58	0,62	0,48	0,54	0,55	0,44	0,61	0,54	0,50	0,51	0,41
ARG [B-]	0,55	1,00	0,55	0,54	0,47	0,47	0,45	0,30	0,28	0,29	0,31	0,34	0,18	0,33	0,28	0,16	0,31	0,27	0,44	0,45	0,27	0,31	0,24	0,31	0,41	0,36	0,38	0,30	0,34	0,35	0,30	0,37	0,33	0,31	0,29	0,25
BRA [BBB]	0,96	0,55	1,00	0,95	0,87	0,58	0,72	0,44	0,44	0,50	0,47	0,51	0,33	0,49	0,47	0,25	0,50	0,41	0,69	0,68	0,48	0,42	0,34	0,49	0,63	0,58	0,61	0,48	0,54	0,54	0,45	0,61	0,54	0,50	0,52	0,43
MEX [BBB]	0,94	0,54	0,95	1,00	0,85	0,58	0,71	0,43	0,44	0,48	0,47	0,50	0,31	0,46	0,47	0,24	0,50	0,42	0,68	0,68	0,47	0,42	0,34	0,49	0,63	0,58	0,61	0,47	0,55	0,53	0,45	0,60	0,54	0,50	0,51	0,42
PER [BBB]	0,86	0,47	0,87	0,85	1,00	0,52	0,62	0,41	0,40	0,42	0,41	0,42	0,26	0,39	0,40	0,21	0,43	0,36	0,60	0,59	0,42	0,37	0,29	0,43	0,56	0,51	0,54	0,44	0,48	0,45	0,40	0,54	0,48	0,43	0,46	0,38
VEN [B+]	0,59	0,47	0,58	0,58	0,52	1,00	0,51	0,26	0,28	0,30	0,32	0,32	0,16	0,29	0,31	0,24	0,34	0,33	0,48	0,49	0,34	0,32	0,32	0,40	0,44	0,39	0,40	0,31	0,35	0,41	0,30	0,43	0,36	0,34	0,32	0,27
CHL [AA-]	0,71	0,45	0,72	0,71	0,62	0,51	1,00	0,29	0,28	0,36	0,34	0,40	0,26	0,38	0,37	0,26	0,41	0,37	0,55	0,55	0,40	0,36	0,31	0,43	0,52	0,43	0,46	0,36	0,38	0,45	0,39	0,47	0,40	0,41	0,40	0,28
PRT [BB]	0,43	0,30	0,44	0,43	0,41	0,26	0,29	1,00	0,73	0,53	0,72	0,62	0,40	0,56	0,51	0,13	0,57	0,31	0,53	0,54	0,32	0,30	0,29	0,35	0,49	0,51	0,56	0,32	0,51	0,39	0,35	0,51	0,49	0,57	0,55	0,77
ESP [BBB-]	0,43	0,28	0,44	0,44	0,40	0,28	0,28	0,73	1,00	0,61	0,84	0,72	0,42	0,65	0,61	0,13	0,66	0,31	0,58	0,58	0,33	0,29	0,25	0,34	0,56	0,52	0,58	0,37	0,54	0,42	0,32	0,51	0,47	0,66	0,62	0,73
DEU [AAA]	0,49	0,29	0,50	0,48	0,42	0,30	0,36	0,53	0,61	1,00	0,65	0,74	0,51	0,74	0,68	0,22	0,75	0,35	0,55	0,55	0,34	0,30	0,29	0,38	0,52	0,51	0,53	0,37	0,53	0,42	0,33	0,51	0,46	0,75	0,75	0,60
ITA [BBB+]	0,46	0,31	0,47	0,47	0,41	0,32	0,34	0,72	0,84	0,65	1,00	0,76	0,43	0,66	0,67	0,19	0,68	0,36	0,62	0,62	0,37	0,36	0,33	0,42	0,60	0,58	0,62	0,39	0,58	0,44	0,37	0,54	0,51	0,68	0,68	0,74
BEL [AA]	0,49	0,34	0,51	0,50	0,42	0,32	0,40	0,62	0,72	0,74	0,76	1,00	0,48	0,75	0,70	0,22	0,78	0,34	0,61	0,60	0,33	0,31	0,31	0,37	0,56	0,56	0,58	0,41	0,54	0,47	0,34	0,56	0,48	0,78	0,74	0,67
USA [AA+]	0,32	0,18	0,33	0,31	0,26	0,16	0,26	0,40	0,42	0,51	0,43	0,48	1,00	0,49	0,45	0,19	0,50	0,23	0,37	0,37	0,20	0,17	0,17	0,24	0,36	0,35	0,39	0,23	0,34	0,32	0,24	0,39	0,33	0,50	0,51	0,41
FRA [AA+]	0,47	0,33	0,49	0,46	0,39	0,29	0,38	0,56	0,65	0,74	0,66	0,75	0,49	1,00	0,64	0,21	0,74	0,34	0,56	0,55	0,35	0,28	0,29	0,37	0,51	0,51	0,52	0,39	0,48	0,42	0,32	0,52	0,45	0,74	0,70	0,60
GBR [AAA]	0,47	0,28	0,47	0,47	0,40	0,31	0,37	0,51	0,61	0,68	0,67	0,70	0,45	0,64	1,00	0,22	0,67	0,33	0,56	0,56	0,33	0,28	0,29	0,37	0,55	0,48	0,52	0,35	0,47	0,45	0,31	0,47	0,43	0,67	0,65	0,57
JPN [AA-]	0,25	0,16	0,25	0,24	0,21	0,24	0,26	0,13	0,13	0,22	0,19	0,22	0,19	0,21	0,22	1,00	0,24	0,38	0,30	0,32	0,37	0,36	0,39	0,37	0,31	0,28	0,25	0,27	0,19	0,23	0,29	0,30	0,28	0,24	0,23	0,12
AUS [AAA]	0,50	0,31	0,50	0,50	0,43	0,34	0,41	0,57	0,66	0,75	0,68	0,78	0,50	0,74	0,67	0,24	1,00	0,35	0,62	0,61	0,35	0,32	0,30	0,38	0,58	0,60	0,60	0,43	0,61	0,47	0,35	0,59	0,52	1,00	0,73	0,62
MYS [A-]	0,41	0,27	0,41	0,42	0,36	0,33	0,37	0,31	0,31	0,35	0,36	0,34	0,23	0,34	0,33	0,38	0,35	1,00	0,48	0,51	0,75	0,72	0,53	0,78	0,49	0,43	0,43	0,35	0,38	0,42	0,59	0,44	0,46	0,35	0,37	0,29
TUR [BB+]	0,69	0,44	0,69	0,68	0,60	0,48	0,55	0,53	0,58	0,55	0,62	0,61	0,37	0,56	0,56	0,30	0,62	0,48	1,00	0,92	0,51	0,49	0,42	0,56	0,87	0,74	0,78	0,61	0,71	0,69	0,53	0,74	0,67	0,62	0,60	0,52
RUS [BBB]	0,69	0,45	0,68	0,68	0,59	0,49	0,55	0,54	0,58	0,55	0,62	0,60	0,37	0,55	0,56	0,32	0,61	0,51	0,92	1,00	0,53	0,51	0,47	0,59	0,88	0,76	0,80	0,59	0,73	0,72	0,54	0,75	0,69	0,61	0,61	0,53
CHN [AA-]	0,46	0,27	0,48	0,47	0,42	0,34	0,40	0,32	0,33	0,34	0,37	0,33	0,20	0,35	0,33	0,37	0,35	0,75	0,51	0,53	1,00	0,67	0,51	0,81	0,52	0,47	0,45	0,39	0,41	0,39	0,58	0,49	0,44	0,35	0,37	0,31
IDN [BB+]	0,40	0,31	0,42	0,42	0,37	0,32	0,36	0,30	0,29	0,30	0,36	0,31	0,17	0,28	0,28	0,36	0,32	0,72	0,49	0,51	0,67	1,00	0,54	0,72	0,48	0,44	0,45	0,41	0,38	0,39	0,63	0,45	0,43	0,32	0,34	0,29
NZL [AA]	0,35	0,24	0,34	0,34	0,29	0,32	0,31	0,29	0,25	0,29	0,33	0,31	0,17	0,29	0,29	0,39	0,30	0,53	0,42	0,47	0,51	0,54	1,00	0,56	0,42	0,43	0,42	0,35	0,32	0,36	0,49	0,45	0,41	0,30	0,34	0,29
KOR [A+]	0,48	0,31	0,49	0,49	0,43	0,40	0,43	0,35	0,34	0,38	0,42	0,37	0,24	0,37	0,37	0,37	0,38	0,78	0,56	0,59	0,81	0,72	0,56	1,00	0,57	0,47	0,51	0,41	0,44	0,46	0,66	0,50	0,48	0,38	0,41	0,35
ZAF [BBB]	0,64	0,41	0,63	0,63	0,56	0,44	0,52	0,49	0,56	0,52	0,60	0,56	0,36	0,51	0,55	0,31	0,58	0,49	0,87	0,88	0,52	0,48	0,42	0,57	1,00	0,73	0,76	0,56	0,68	0,67	0,51	0,69	0,67	0,58	0,57	0,49
ROM [BB+]	0,58	0,36	0,58	0,58	0,51	0,39	0,43	0,51	0,52	0,51	0,58	0,56	0,35	0,51	0,48	0,28	0,60	0,43	0,74	0,76	0,47	0,44	0,43	0,47	0,73	1,00	0,74	0,54	0,74	0,63	0,47	0,83	0,71	0,60	0,55	0,53
POL [A-]	0,62	0,38	0,61	0,61	0,54	0,40	0,46	0,56	0,58	0,53	0,62	0,58	0,39	0,52	0,52	0,25	0,60	0,43	0,78	0,80	0,45	0,45	0,42	0,51	0,76	0,74	1,00	0,54	0,75	0,62	0,47	0,74	0,67	0,60	0,55	0,57
ISR [A+]	0,48	0,30	0,48	0,47	0,44	0,31	0,36	0,32	0,37	0,37	0,39	0,41	0,23	0,39	0,35	0,27	0,43	0,35	0,61	0,59	0,39	0,41	0,35	0,41	0,56	0,54	0,54	1,00	0,45	0,48	0,40	0,52	0,50	0,43	0,40	0,34
HUN [BB]	0,54	0,34	0,54	0,55	0,48	0,35	0,38	0,51	0,54	0,53	0,58	0,54	0,34	0,48	0,47	0,19	0,61	0,38	0,71	0,73	0,41	0,38	0,32	0,44	0,68	0,74	0,75	0,45	1,00	0,58	0,40	0,69	0,62	0,61	0,52	0,54
UKR [B]	0,55	0,35	0,54	0,53	0,45	0,41	0,45	0,39	0,42	0,42	0,44	0,47	0,32	0,42	0,45	0,23	0,47	0,42	0,69	0,72	0,39	0,39	0,36	0,46	0,67	0,63	0,62	0,48	0,58	1,00	0,42	0,63	0,55	0,47	0,46	0,41
PHL [BB+]	0,44	0,30	0,45	0,45	0,40	0,30	0,39	0,35	0,32	0,33	0,37	0,34	0,24	0,32	0,31	0,29	0,35	0,59	0,53	0,54	0,58	0,63	0,49	0,66	0,51	0,47	0,47	0,40	0,40	0,42	1,00	0,45	0,44	0,35	0,38	0,34
BGR [BBB]	0,61	0,37	0,61	0,60	0,54	0,43	0,47	0,51	0,51	0,51	0,54	0,56	0,39	0,52	0,47	0,30	0,59	0,44	0,74	0,75	0,49	0,45	0,45	0,50	0,69	0,83	0,74	0,52	0,69	0,63	0,45	1,00	0,66	0,59	0,55	0,53
CZH [AA-]	0,54	0,33	0,54	0,54	0,48	0,36	0,40	0,49	0,47	0,46	0,51	0,48	0,33	0,45	0,43	0,28	0,52	0,46	0,67	0,69	0,44	0,43	0,41	0,48	0,67	0,71	0,67	0,50	0,62	0,55	0,44	0,66	1,00	0,52	0,49	0,47
AUT [AA+]	0,50	0,31	0,50	0,50	0,43	0,34	0,41	0,57	0,66	0,75	0,68	0,78	0,50	0,74	0,67	0,24	1,00	0,35	0,62	0,61	0,35	0,32	0,30	0,38	0,58	0,60	0,60	0,43	0,61	0,47	0,35	0,59	0,52	1,00	0,73	0,62
NLD [AAA]	0,51	0,29	0,52	0,51	0,46	0,32	0,40	0,55	0,62	0,75	0,68	0,74	0,51	0,70	0,65	0,23	0,73	0,37	0,60	0,61	0,37	0,34	0,34	0,41	0,57	0,55	0,55	0,40	0,52	0,46	0,38	0,55	0,49	0,73	1,00	0,61
IRL [BBB+]	0,41	0,25	0,43	0,42	0,38	0,27	0,28	0,77	0,73	0,60	0,74	0,67	0,41	0,60	0,57	0,12	0,62	0,29	0,52	0,53	0,31	0,29	0,29	0,35	0,49	0,53	0,57	0,34	0,54	0,41	0,34	0,53	0,47	0,62	0,61	1,00

Source: authors' calculations.

It is rather evident that analyzing the resulting correlation matrix is far from being straightforward. Finding the hierarchy (i.e. the ranking of influence) is intricate at best; identifying the most relevant peer(s) for each sovereign is not easy either. Figure 2 presents the frequency distribution of the correlations and the corresponding basic statistics.

Figure 2
Correlations' main statistical properties



Mean	0.470
Standard deviation	0.149
Maximum	0.960
Minimum	0.122

Source: authors' calculations.

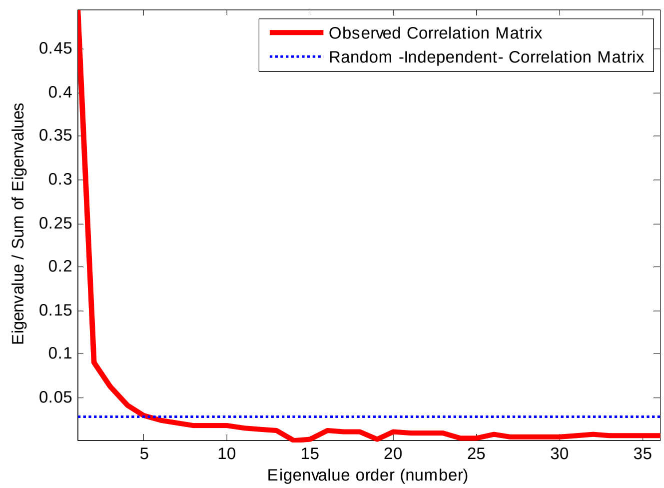
Concurrent with Mantegna and Stanley's (2000) analysis of Dow Jones Industrial Average and Standard & Poors 500 time-series, the typical minimum value of observed correlations is close to zero. However, in the CDS spread case the minimum correlation is not only close to zero, but it is also strictly above zero, which pinpoints the absence of anti-correlation in the series here considered; as previously acknowledged, the dominance of correlation over anti-correlation suggests that differences between the two correlation-into-distance mapping methods should be a matter of scale only, and thus immaterial.

Following Mantegna and Stanley (2000), the eigenvalues of correlation matrices are also informative, where the existence of dominant eigenvalues has been interpreted as evidence of a small number of economic factors driving the stochastic dynamics of asset returns in a financial market. In this sense, Figure 3 compares the eigenvalues of the observed correlation matrix (Table 2) with those resulting from a random matrix generated with a randomized version of the original time-series.[19]

Results in Figure 3 match Mantegna and Stanley's findings: *the empirical analysis […] detect[s] a prominent eigenvalue far larger than –and several other eigenvalues slightly larger than- what is expected from random matrix theory*. Likewise, the presence of a prominent eigenvalue also concurs with Bonnano et al. (2003), who points out that such finding corresponds to the collective motion of the time-series. In this case, as exhibited in Figure 3, the largest (main) eigenvalue is about 49.63% of the sum of the eigenvalues, whereas the second and third are about 8.93% and 6.29%, respectively.

[19] In this case the random correlation matrix resulted from the estimation of the mean correlation matrix based on a set of 1,000 randomly reorganized (i.e. reshuffled) original time-series; in this way the first four distributional moments of each time-series are preserved, but they are (pseudo) cross-independent.

Figure 3
Eigenvalues of the correlation matrix

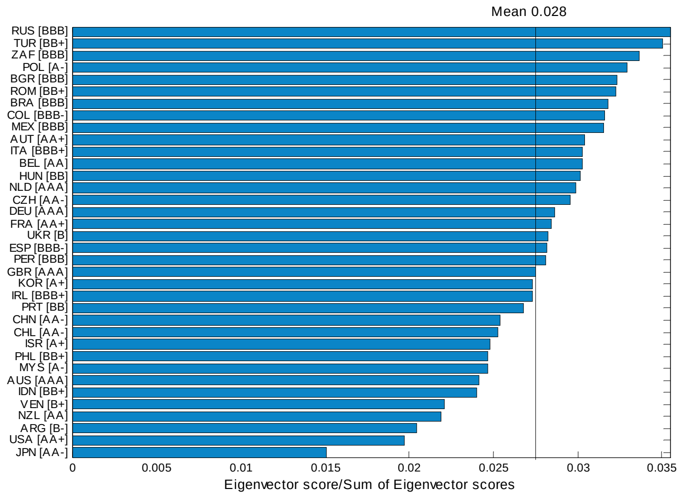


Source: authors' calculations.

Accordingly, finding a prominent eigenvalue (i) confirms that a small number of globally common factors drive a large number of CDS spreads time-series (i.e. sovereigns), as suggested by Shino and Takahashi (2010); (ii) verifies that the system under analysis is not a random –Poisson- graph; and therefore (iii) a hierarchy does exist in the corresponding correlation matrix.

Regarding the largest eigenvalue, the correspondent eigenvector provides relevant information about the contribution of each time-series to the whole system. As proposed by Bonacich (1972), the popularity (i.e. power) scores of a system are described by the eigenvector of the largest eigenvalue, where the magnitude of the latter is a measure of how good the former is in summarizing the relationships in the system. Hence, in order to visualize the influence of each time-series on the whole system, Figure 4 presents the eigenvector of the largest eigenvalue's scores, also known as eigenvector centrality scores; in this case, since the largest eigenvalue is about 49.63% of the sum of the eigenvalues, the scores in the corresponding eigenvector may be regarded as fairly capable of summarizing the system's relationships.

Figure 4
Eigenvector centrality (eigenvector of the largest eigenvalue)



Source: authors' calculations

As it is evident from Figure 4, some sovereigns display rather high eigenvector centrality scores (i.e. Russia, Turkey, South Africa, Poland, Bulgaria, Romania, Brazil, Colombia and México), whereas some others display low scores (i.e. Japan, United States, New Zealand, Argentina and Venezuela). Accordingly, it is expected that time-series with high (low) eigenvector centrality scores are more (less) influential in the correlation matrix, corresponding to central (peripheral) nodes in the MST graph.

3.2. The minimal spanning tree (MST)

The correlation matrix in Table 2 is mapped into two different weighted adjacency matrices, $\dot{\Delta}$ and $\ddot{\Delta}$, resulting from the two estimated distances $\dot{\delta}_{(x,y)}$ and $\ddot{\delta}_{(x,y)}$, as in [§3] and [§4], respectively. As depicted by Eryigit and Eryigit (2009), let N be the number of time-series (i.e. nodes), each adjacency matrix has $N(N-1)/2$ different elements (i.e. non-directed links or edges), and describes a fully connected network which is not very interesting by itself.

After obtaining the required weighted adjacency matrix, which contains all the estimated distances between each pair of nodes (i.e. time-series), it is possible to filter the information within. The most drastic but simple filtering technique is the *minimal spanning tree* (MST), which consists of choosing the minimal weights (i.e. shortest distances) of a connected system of N nodes in such a way that the resulting system is an acyclic network with $N-1$ links that minimize the system´s weight.

The MST procedure has been generally described in several ways. For instance, regarding weighted networks, Wu et al. (2006) defines the MST as a *tree including all of the nodes but only a subset of the links, which has the minimum total weight out of all possible trees that span the entire network*, where this tree may be considered as the *"skeleton" inside the network*. Similarly, Braunstein et al. (2007) states that the *MST on a weighted graph is a tree that reaches all nodes of the graph and for which the sum of the weights of all the links or nodes (total weight) is minimal*.

When applied as a correlation-filtering technique, the MST has been described in several ways too. *As a strongly reduced representative of the whole correlation matrix, MST bears the essential information about asset correlations*, where the information space is effectively reduced *from* $N(N-1)/2$ *separate correlation coefficients to* $N-1$ *tree edges, in other words, compressing the amount of information dramatically* (Onnela et al., 2003). Gilmore et al. (2010) states that *the MST reduces the information space from N(N – 1)/2 separate correlation coefficients to (N −1) tree edges, while retaining the salient features of the system*. Bonnano et al. (2003) highlights that *MST filters out the more relevant information about the correlation structure, whereas most of the correlation matrix is heavily dressed by noise*.

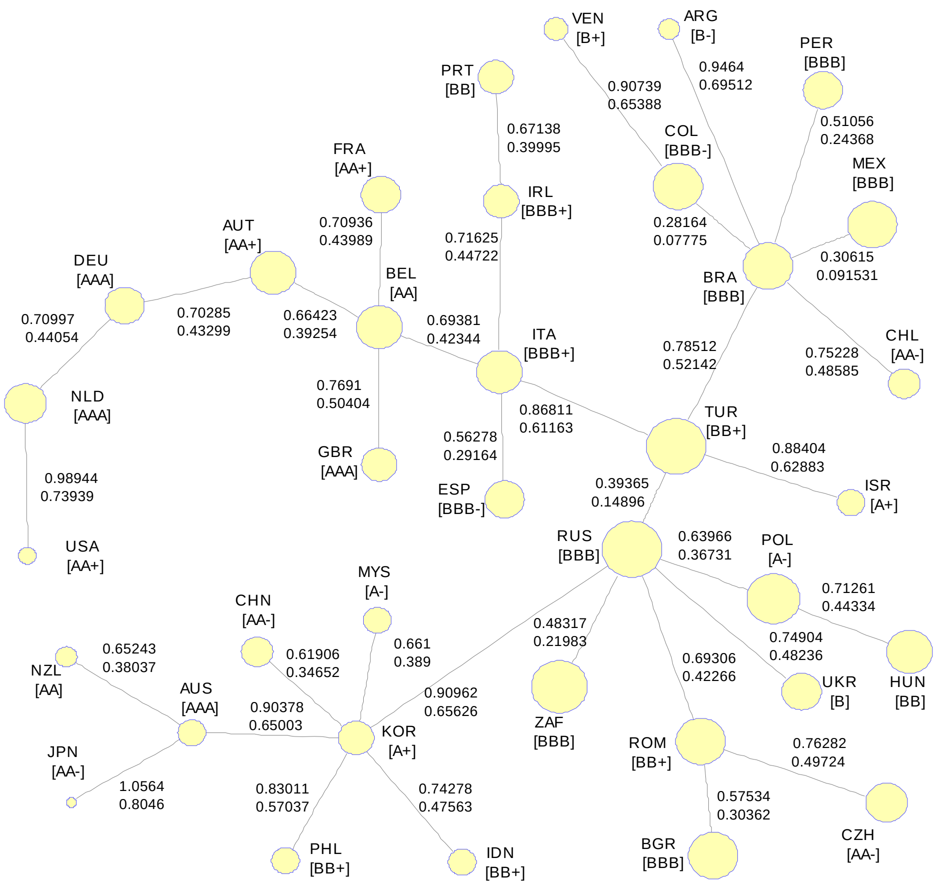
Accordingly, if each sovereign is represented by a node, constructing the MST is straightforward: (i) identifying the pair of nodes with the shortest distance (i.e. strongest – less noisy- dependence signals); (ii) identifying the pair of nodes with the second shortest distance and adding it to the previously found pair; (iii) identifying and adding the subsequent shortest distances between nodes with the condition of avoiding closed loops, until connecting all the nodes (i.e. until the tree reaches $N-1$ edges). Such incremental addition of links according to the increasing order of distances is known as the Kruskal's algorithm.[20]

Based on the two different weighted adjacency matrices, $\dot{\Delta}$ and $\ddot{\Delta}$, Figure 5 presents the resulting MSTs, $\dot{\Delta}_{MST}$ and $\ddot{\Delta}_{MST}$, corresponding to the distances estimated as in [§3] and [§4], respectively. As expected, because of the absence of anti-correlations, the two MSTs are strictly the same, but the estimated distances ($\dot{\delta}_{(x,y)}$ and $\ddot{\delta}_{(x,y)}$) are different, and simultaneously reported. The diameter of the nodes of the MST corresponds to the eigenvector centrality for each sovereign (as in Figure 4), where more central nodes have greater diameters.

[20] Kruskal's algorithm (Kruskal, 1956) is the authors' choice. Other algorithms suitable for building the MST are Prim's algorithm and the "bombing optimization algorithm"; please refer to Braunstein et al. (2007) for a description of these algorithms. According to Kim et al. (2005) and Kim and Jeong (2004), the main features of the MST do not depend on the construction method.

Figure 5

$\dot{\Delta}_{MST}$ and $\ddot{\Delta}_{MST}$



$\dot{\delta}_{(x,y)}$ and $\ddot{\delta}_{(x,y)}$ correspond to the top and bottom reported distances, respectively.
The diameter of the nodes corresponds to the eigenvector centrality of the sovereign reported in Figure 4.
Source: authors' calculations.

Hence, due to the strict dominance of correlation over anti-correlation (Figure 2), both MSTs are graphically equivalent, and the two correlation-into-distance mapping techniques are also equivalent.[21] Consequently, without any loss of consistency, next section will analyze $\ddot{\Delta}_{MST}$.

4. Analysis of the resulting MST

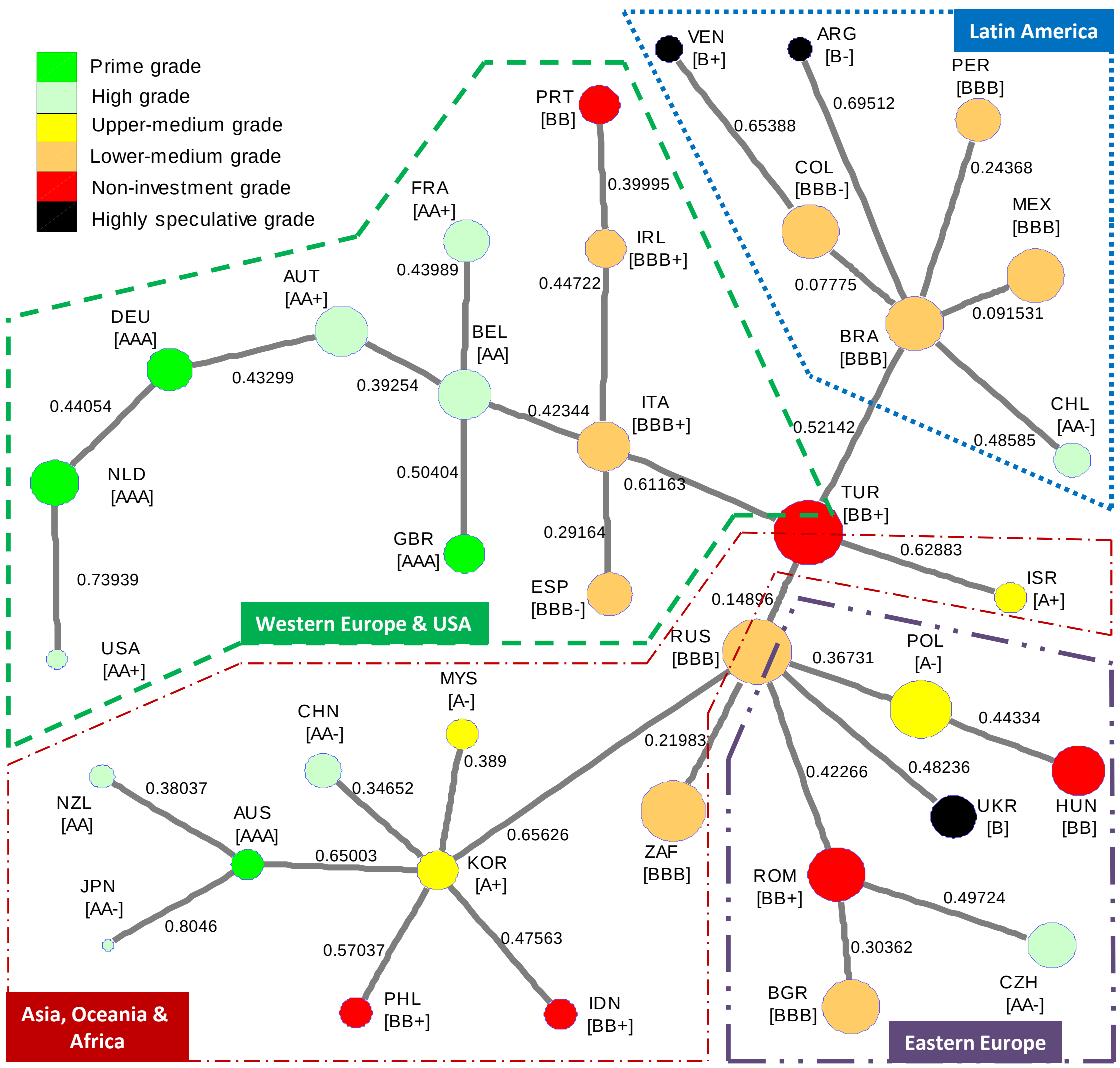
Several authors have highlighted the advantages of MST-based correlation-filtering. Most of these authors stress its usefulness for characterizing financial markets by means of identifying their underlying structure, taxonomy or hierarchy. For instance, such characterization of financial markets has allowed for (i) identifying different types of asset clusters (Naylor et al., 2007; Mantegna, 1999); (ii) finding assets' exposure to common risk

[21] It is important to stress that such equivalence may be restricted to the time-series considered. Thus, the decision on the correlation-into-distance mapping should consider the main features of the correlation matrix (i.e. the dominance of correlation over anti-correlation) and the purpose of the correlation filtering (i.e. risk minimizing or informational analysis).

factors (Marsh et al., 2003; Mantegna, 1998); detecting predominant assets (Naylor et al., 2007); (iv) identifying central and peripheral issuers (Marsh et al., 2003); and measuring financial integration (Gilmore et al., 2010). Moreover, some authors have also emphasized its usefulness to overcome the empirical problem of noise in –historical- correlation matrices (Naylor et al., 2007; Bonnano et al., 2003).

As expected, concurrent with the documented usefulness of the proposed correlation-filtering method, the estimated CDS time-series' MST ($\ddot{\Delta}_{MST}$) displays a structure worth describing and analyzing. Figure 6 exhibits $\ddot{\Delta}_{MST}$ along with some graphical aids for easing the corresponding analysis.

Figure 6
$\ddot{\Delta}_{MST}$
(with geographical and credit rating clustering)



Prime grade (AAA); High grade (AA+,AA,AA-); Upper-medium (A+, A,A-); Lower-medium (BBB+,BBB,BBB-); Non-investment (BB+,BB,BB-); Highly spreculative (B+,B,B-).
The diameter of the nodes corresponds to the eigenvector centrality of the sovereign reported in Figure 4.
Source: authors' calculations.

As is the case with most related literature on correlation-filtering, the network formed from the correlation matrix of time-series' returns has a resultant taxonomy that displays meaningful clusters. For instance, the geographical clustering is rather evident, where four main clusters exist: (i) *Western Europe & United States*, (ii) *Asia, Oceania & Africa*, (iii) *Latin America*, and (iv) *Eastern Europe*. Following Onnela et al., (2003), they all are *complete clusters* (i.e. they contain all the sovereigns belonging to that geographical zone)

Interestingly enough, the nodes connecting the four clusters are two transcontinental countries, Russia and Turkey, which have been customarily regarded as sharing features from Europe and Asia. Since Russia and Turkey serve as connectors between geographical clusters, their importance and influence is particularly high, as confirmed by their eigenvector centrality (i.e. their node's diameter), the first and second highest of the system, respectively (as in Figure 5). The link between Turkey and Russia is the third strongest (i.e. shortest) of the system. Brazil serves as the connecting node or hub for Latin American sovereigns, hence its eigenvector centrality is among the top-seven of the system; moreover, Brazil is also one of the three nodes displaying the highest degree (i.e. number of connections) of the system (6), which is an additional centrality factor within the MST graph.[22]

Turkey serves as a connector or hub for *Western Europe & United States*, connecting it with the rest of the system. It is worth noting that Mediterranean sovereigns (i.e. Italy, Spain and Portugal) and Ireland appear early in the *Western Europe & United States* cluster, whereas other Western European sovereigns and the United States appear later (farther) on the tree; this finding matches the MST analysis of government bond markets by Gilmore et al. (2010). It is also noticeable that Mediterranean sovereigns and Ireland share lower credit ratings than the rest of the cluster. Since Portugal, Ireland, Spain and Italy are closer to non-European nodes, it is conceivable that the dynamics of these sovereigns' CDS are more dependent the system's common factors, and vice versa.[23]

It is also noteworthy to find that two of the most distant sovereigns in the *Western Europe & United States* cluster are United States and Germany, presumably the most dominant economies of the cluster; this suggests that both sovereigns' CDS dynamics do not have a lot in common with their geographical peers, or with the rest of the system.[24] It is noticeable that the farthest located sovereigns in the *Europe & United States* cluster correspond to economies actually (i.e. Netherlands, Germany and United Kingdom) or until recently (i.e. United States, Austria and France[25]) considered as *prime grade* credit risk; the coincidence of their remoteness from the core of the system and their credit ratings may suggest that being considered a *safe haven* may be an important factor for the system.

[22] The degree is a standard measure of centrality for a node within a network. However, it is worth noting that the degree may be even more relevant in a MST –than in a traditional graph- since this method preserves the most significant connections for each node only. The nodes displaying the highest degree (6 links) are Brazil, Korea and Russia.

[23] In this sense, according to the resulting MST graph, the less noisy (i.e. more signaling) path between Portugal and Russia is shorter (1.607) than the less noisy path between Portugal and Germany (2.096).

[24] For instance, United States (Germany) is located at a distance of 3.04 (1.86) from Turkey (i.e. one of the most central nodes in the system), whereas the United Kingdom (France) is at 1.53 (1.48).

[25] Standard and Poor's downgraded from AAA to AA+ the long-term sovereign ratings of the United States (August 5, 2011), Austria and France (January 13, 2012).

United States' low level of dependence and influence (i.e. its prominent distance) is confirmed by the diameter of the corresponding node in the MST graph (i.e. the eigenvector centrality of Figure 5). In this sense, the low power score –the second lowest- and the relative distance of the United States suggest that the economic factors behind its CDS dynamics are rather particular or idiosyncratic, as suggested by Shino and Takahashi (2010)[26]; on the other hand, for instance, Italy's closeness to the rest of the system signals the importance of systemic (i.e. common) factors affecting its sovereign CDS dynamics, and the influence of Italy on the rest of the system.[27]

About the *Latin American* cluster, Brazil appears as the parent node that serves as a hub for all the cluster's members, providing them the strongest connection to the rest of the system via its link to Turkey. Colombia and Mexico appear as the closest to Brazil, whereas Argentina and Venezuela are the farthest. Colombia and Brazil are not only the closest nodes in the cluster, but in the whole system.

Their *highly speculative grade* rating and other idiosyncratic factors[28] may be the origin of Argentina and Venezuela's distant relationship with their geographical peers and the whole system. Brazil, Colombia, México and Perú share *lower-medium grade* credit ratings, whereas their closeness suggests they share common economic factors. Chile, the only Latin American sovereign with a *high grade* credit rating, holds an intermediate distance to Brazil; thus, Chilean idiosyncratic factors are more important than in the case of México, Colombia or Perú, but less relevant than in the case of Argentina or Venezuela. As before, the diameter of the *Latin American* nodes (i.e. their eigenvector centrality) tend to confirm the estimated distance between the regional sovereigns, where Brazil, México and Colombia appear to be the most influential for their regional peers and for the whole system, and vice versa.

Regarding *Asia, Oceania & Africa*, Russia is the parent node of the cluster, providing its members with the strongest connection to the rest of the system; again, its transcontinental features appear to be relevant. Similar to the *Western Europe & United States* cluster, the farthest located sovereigns in the *Asia, Oceania & Africa* cluster correspond to economies awarded *prime grade* or *high grade* ratings (e.g. Australia, Japan, New Zealand, China); as before, their remoteness from the core of the system and their credit ratings may suggest that being considered a *safe haven* may be an important factor for the system.

The most degree central (i.e. with most connections) nodes within the *Asia, Oceania & Africa* cluster are Russia and South Korea. Unlike Brazil or Russia, Korea does not display a high eigenvector centrality, which may suggest that it is less dependent –and less influential- on the common factors of the system.

Japan is located at a particularly remote location from the cluster and from the system, with a rather low eigenvector centrality score (i.e. node's diameter) within the system; this

[26] Please note that this is contrary to the findings of MSTs on stock or foreign exchange markets, where the United States securities tend to appear as the most central (i.e. influential).
[27] Once again, it is important to highlight that correlation does not convey causality; therefore, a sovereign displaying a high eigenvector centrality should be interpreted as a measure of its power to influence the system and the system's power to influence the sovereign.
[28] For instance, Venezuela's political and economic stance may particularly affect market's perception of its ability and willingness to fulfill its liabilities. Likewise the recurrence of balance of payments and financial crisis may mark Argentina's CDS dynamics.

finding matches the MST analysis of government bond markets by Gilmore et al. (2010). As in the case of the United States, this suggests that Japan's CDS dynamics are dominated by idiosyncratic factors, as suggested by Shino and Takahashi (2010).[29]

Regarding the influence of each geographical cluster within the system (Table 3), *Western Europe & United States* is the one most contributing to the total sum of centrality scores (34.3%), led by Turkey, a transcontinental sovereign. However, the cluster contributing most per-sovereign is Eastern Europe, led by the other transcontinental sovereign (i.e. Russia).

Table 3
Geographical clusters' main statistics

Cluster	Latin America	Eastern Europe	Western Europe & U.S.	Asia, Oceania & Africa	All sovereigns (the system)
Members[a]	BRA, COL, MEX, PER, CHI, VEN, ARG	*RUS*, POL, HUN, UKR, ROM, BGR, CZH	ITA, ESP, POR, IRL, BEL, GBR, FRA, AUT, DEU, NLD, USA, *TUR*	*TUR*, ISR, *RUS*, ZAF, KOR, MYS, IDN, PHL, CHN, AUS, NZL, JPN	
Number of members	7	6	12	12	36
Mean distance	0.396	0.419	0.466	0.479	0.448
Minimal distance	0.078	0.303	0.291	0.149	0.078
Maximal distance	0.695	0.497	0.739	0.805	0.805
Sum of centrality scores	0.191	0.221	0.343	0.298	1.000
Centrality score per sovereign	0.027	0.031	0.029	0.026	0.028

[a] Transcontinental countries (i.e. Turkey and Russia) may appear in more than one cluster.
Source: authors' calculations.

Latin America cluster exhibits the lowest mean and minimal distances, whereas *Eastern Europe* exhibits the lowest maximal. In this sense, the MST suggests that the *Latin American* and *Eastern Europe* clusters are the tightest within the system, where the cross influences among their members are particularly strong. Consequently, there is evidence of regional common factors (i.e. a "Latin American factor" and an "Eastern Europe factor"), whilst the other two clusters display rather looser dependences (i.e. longer distances) among their members. Accordingly, *Latin America* and *Eastern Europe* clusters' members would be the most prone to contagion from within.[30]

If credit rating clusters are to be considered, Table 4 suggests that the most tightly connected (i.e. displaying lower mean and minimal distance) are those sovereigns awarded *lower-medium grade* ratings, which coincide with those contributing the most to the (total or per-sovereign) eigenvector centrality of the system. Thus, *lower-medium grade* rated sovereigns are the most influential and the most influenced, the main driving common factors and the most driven by common factors, simultaneously, and they are the most prone to contagion. On the other hand, distances increase to both ends of the credit rating spectrum, which suggests the presence of different idiosyncratic factors (e.g. political stances, willingness to pay).

[29] Japan is located at a distance about 2.11 from Russia, its cluster's parent node; as previously presented, United States' distance to its parent node (i.e. Turkey) is about 3.04. It is worth pointing out that the less noisy distance (correlation) between Japan and the United States is one of the largest (lowest) of the system, 5.30 (0.19), which further stresses the importance of idiosyncratic factors for both sovereigns.
[30] It is worth noticing that if Venezuela and Argentina are excluded from the *Latin American* cluster, the "Latin American factor" becomes the strongest of the system, with the lowest mean (0.225), minimal (0.078) and maximal (0.486) distances. In this sense, Brazil, México, Colombia, Perú and Chile would be the most prone to contagion from within.

Table 4
Credit rating clusters' main statistics[a]

Credit rating grade	Prime grade (AAA)	High (AA+, AA, AA-)	Upper-medium (A+, A, A-)	Lower-medium (BBB+, BBB, BBB-)	Non-investment (BB+, BB, BB-)	Highly speculative (B+, B, B-)	All sovereigns (the system)
Members	NLD, DEU, GBR, AUS	USA, AUT, FRA, BEL, CHL, CZH, CHN, NZL, JPN	MYS, KOR, POL, ISR	PER, MEX, COL, BRA, IRL, ITA, ESP, RUS, ZAF, BGR	PRT, TUR, HUN, ROM, IDN, PHL	VEN, ARG, UKR	
Number of members	4	9	4	10	6	3	36
Mean distance	0.507	0.501	0.510	0.296	0.412	0.610	0.448
Minimal distance	0.433	0.347	0.367	0.078	0.149	0.482	0.078
Maximal distance	0.650	0.805	0.656	0.611	0.570	0.695	0.805
Sum of centr. scores	0.110	0.226	0.110	0.310	0.173	0.071	1.000
Centr. per sovereign	0.028	0.025	0.027	0.031	0.029	0.024	0.028

[a] Distances are those corresponding to links towards the core of the system; in case of two competing distances the shortest was selected.
Source: authors' calculations.

5. Final remarks

This paper briefly presented and implemented the simplest and strictest correlation-filtering method (i.e. a *minimal spanning tree* - MST) for a non-small set of time-series consisting of 36 sovereign CDS spreads. As documented in the literature, this type of method preserves the most signaling –less noisy- correlations out of the correlation matrix, resulting in a graph displaying the dominant structure or hierarchy within the time-series.

The MST was applied to two different adjacency matrices, resulting from two distinct approaches to mapping correlations into distances, where each approach follows a different objective, namely risk minimization via traditional diversification [§3] or informational (analytical) purposes [§4]. Both approaches yielded equivalent MST graphs, which was the expected outcome because of the absence of anti-correlations in the corresponding matrix.

Analyzing the resulting MSTs highlighted several features regarding the sovereigns' CDS time-series considered. Some are worth highlighting:

i. Due to the presence of a prominent eigenvalue in the correlation matrix, the sovereigns' CDS market may be characterized as hierarchical (i.e. not random), where there is a small set of common factors affecting the entire system.
ii. Sovereigns' geographical location is key for time-series' dynamics, which resulted in evident regional –*complete*- clusters,.
iii. Sovereigns' credit rating is indicative for time-series' dynamics, which resulted in some clustering tendency of similarly rated countries, whereas sovereigns awarded the highest and lowest ratings (i.e. *prime*, *high grade*, *highly speculative*) tend to be remotely located.
iv. Transcontinental sovereigns (i.e. Russia and Turkey) have a central role within the system, either measured by their eigenvector centrality or by the mere examination of their location in the MST graph.
v. *Lower-medium grade* rated sovereigns are the most influential, but also the most prone to contagion.

vi. Japan and United States' peripheral locations within the MST graph are among the most remote within the system, and they have the lowest eigenvector centrality scores, which confirm their poor informational content for the system.
vii. Since the *Latin America* and *Eastern Europe* clusters display the lowest distances among its members (i.e. lowest mean, minimal and maximal distances), results suggest the existence of regional common factors, which would mean that these two clusters are the most prone to contagion from within.
viii. Brazil and Colombia display the shortest distance (0.078), which means that their CDS time-series are the most correlated in the sample (0.96), and that Brazil is the most informative sovereign for Colombia, and vice versa.

Regarding the MST methodology and the results herein obtained, three main caveats are worth mentioning. First, as has been widely discussed in financial literature, the correlation matrix has some features not to be overlooked. Among this features it is important to highlight the following:[31] (i) since correlations are simple linear interpretations of historical movements, they are gross oversimplifications of the complex relationship between time-series, and may be inadequate descriptors of dependence; (ii) by construction correlations are confined to a Gaussian –or at least elliptical- world, where the correlation of non-Gaussian time-series (e.g. financial time-series) may not be properly defined; (iii) correlations are far from being constant, and they tend to change rather abruptly, especially during periods of financial turmoil, when the depth of the correlation matrix (i.e. its diversification potential) tends to disappear; (iv) correlations do not distinguish between dependence when moves are small and large; (v) correlations do not capture causality.

Second, as highlighted by Serrano et al. (2009), the MST is by construction acyclic, and thus this method is excessively simplifying the system and destroying features often present in real world networks (e.g. local cycles); therefore, some other –less strict and simple-correlation filtering methods should be evaluated (e.g. *planar filtered graphs*). In this sense, MST considers all links but the strongest as noise, thus discarding potentially informative connections.

Third, as stressed by Coudert and Gex (2010), despite its recent astonishing growth, the sovereigns' CDS market is still in its infancy, especially when compared to the corporates' CDS market. This may be playing a determinant factor in the dynamic of the time-series. For instance, as reported by Shino and Takahashi (2010), the outstanding level and the liquidity of sovereigns' CDS may explain the significance of fundamentals (i.e. fiscal risk) in the CDS pricing: *in countries where the amount outstanding […] is at low levels, such as Japan, the United States and the United Kingdom, CDS premiums are more directly affected by speculative flows […]*, whereas in *continental European countries and emerging economies […] CDS premiums are more likely to reflect a country's fiscal risk premiums*.[32]

However, even after these caveats are considered, the results are particularly valuable. As highlighted by Naylor et al. (2007), correlation-filtering by means of the MST technique yielded a resultant taxonomy displaying meaningful clusters, where, as in Mantegna (1998),

[31] These highlights are based on Hubbard (2009), Rebonato (2007), Bhansali (2005), Greenspan (2008), Taleb (2004 & 2007).
[32] Likewise, it is also interesting that some of the most influential (i.e. central) sovereigns in this analysis coincide with sovereigns having the highest levels of outstanding CDS (i.e. Russia, Turkey, Brazil, México, Italy), whereas some of the less influential sovereigns have low outstanding levels and a tradition of being *safe havens* (i.e. United States, Japan).

results are of great interest from an economic point of view, especially in the theoretical description of financial markets and in the search of economic common factors affecting specific groups of time-series. Therefore, due to its parsimony and meaningfulness, correlation-filtering of relevant time-series (e.g. exchange rates, asset prices) should be regarded as convenient for financial authorities and private investors, either for analytical or risk-management purposes.

Some avenues for related research may include (i) comparing the hierarchy of sovereign and corporate CDS markets before and after Lehman Brothers; (ii) comparing the hierarchy of sovereign's CDS and bond markets; (iii) comparing the hierarchy of sovereign's CDS and their corresponding currencies; and (iv) using properly lagged time-series to estimate the hierarchy behind causality relations.

6. References

Aste, T. and Di Matteo, T., "Correlation Filtering in Financial Time Series", *Noise and Fluctuations in Econophysics and Finance (Eds. Abbot, D.; Bouchaud, J.-P.; Gabaix, X.; McCauley, J.L.)*, SPO, 2005.

Bhansali, V., "Putting Economics (Back) into Quantitative Models", *Risk Magazines Annual Quant Congress*, New York, November 8 and 9, 2005.

Bonacich, P., "Factoring and weighting approaches to status scores and clique identification", *The Journal of Mathematical Sociology*, Vol.2, No.1, 1972.

Bonacich, P., "Power and centrality: A family of measures", *American Journal of Sociology*, Vol. 92, No. 5, 1987.

Bonanno, G.; Caldarelli, G.; Lillo, F.; Mantegna, R.N., "Topology of Correlation-Based minimal spanning trees in real and model markets", *Physical Review E*, Vol.68, No.4, 2003.

Bonanno, G.; Caldarelli, G.; Lillo, F.; Miccichè, S.; Vandewalle, N.; Mantegna, R.N., "Networks of equities in financial markets", *The European Physical Journal B*, Vol.38, 2004.

Braunstein, L.A.; Wu, Z.; Chen, Y.; Buldyrev, S.V.; Kalisky, T.; Sreenivasan, S.; Cohen, R.; López, E.; Havlin, S.; Stanley, E., "Optimal path and minimal spanning trees in random weighted networks", *International Journal of Bifurcation and Chaos*, Vol.17, No.7, 2007.

Brooks, C., *Introductory Econometrics for Finance*, Cambridge University Press, 2008.

Coudert, V. and Gex, M., "Credit default swap and bond markets: which lead the other", *Financial Stability Review*, No.14, Banque de France, July, 2010.

Deza, E. and Deza, M.M., *Dictionary of Distances*, Elsevier, 2006.

Eryigit, M. and Eryigit R., "Network Structure of Cross-Correlations among the World Market Indices", *Physica A*, No.388, 2009.

Focardi, S.M. and Fabozzi, F.J., *The Mathematics of Financial Modeling and Investment Management,* John Wiley and Sons, 2004.

Fratzscher, M., "Capital Flows, Push versus Pull Factors and the Global Financial Crisis", *Working Paper Series* , No.1364, European Central Bank*,* 2011.

Gilmore, C.G.; Lucey, B.M.; Boscia, M.W., "Comovements in commodity markets: A minimum spanning tree analysis", *Prepared for presentation at the Midwest Finance Conference*, 2012. < http://ssrn.com/abstract=1927811>

Gilmore, C.G.; Lucey, B.M.; Boscia, M.W., "Comovements in government bond markets: A minimum spanning tree analysis", *Physica A*, No.389, 2010.

Greenspan, A.,"We will never have the perfect model of risk", *Financial Times*, March 16, 2008.

Holton, G.A., "Time: the second dimension of risk", *Financial Analysts Journal*, No.48, Vol.6, 1992.

Hubbard, D.W., *The Failure of Risk Management*, John Wiley & Sons, 2009.

International Monetary Fund, “The influence of credit derivatives and structured credit markets on financial stability”, *Global Financial Stability Report*, International Monetary Fund, April, 2006.

Kim, D.-H. and Jeong, H., “Scale-free spanning trees of complex networks”, *Journal of the Korean Physical Society*, Vol.44, No.3, 2004.

Kim, D.-H.; Son, S.W.; Ahn, Y.-Y.; Kim, P.J.; Eom, Y.-H.; Jeong, H., “Underlying scale-free trees in complex networks”, *Progress of Theoretical Physics Supplement*, No.157, 2005.

Kruskal, J.B., “On the shortest spanning subtree of a graph and the traveling salesman problem”, *Proceedings of the American Mathematical Society*, No.1, Vol.7, 1956.

Kullmann, L.; Kertész, J.; Kaski, K., “Time dependent cross correlations getween different stock returns: a directed network of influence”, *Physical Review E*, Vol.66, No.2, 2002.

León, C.; Leiton, K., Reveiz, A., “Investment horizon dependent CAPM: adjusting beta for long-term dependence”, *Borradores de Economía*, No.730, Banco de la República, 2012.

León, C. and Reveiz, A., “Portfolio Optimization and Long-Term Dependence”, *Portfolio and risk management for central banks and sovereign wealth funds -BIS Papers*, No.58, BIS, 2011.

Maeng, S.E.; Choi, H.W.; Lee, J.W., “Complex networks and minimal spanning trees in international trade network”, *International Journal of Modern Physics - Conference Series*, Vol.16, 2012.

Malevergne, Y. and Sornette, D., *Extreme Financial Risks: From Dependence to Risk Management*, Springer-Verlag, 2006.

Mantegna, R.N. and Stanley, E., *An Introduction to Econophysics*, Cambridge University Press, 2000.

Mantegna, R.N., “Hierarchical structure in financial markets”, 1998.

Mantegna, R.N., “Hierarchical structure in financial markets”, *The European Physical Journal B*, No.11, 1999.

Marsh, I.W.; Stevens, I.; Hawkesby, C., “Large complex financial institutions: common influences on asset price behaviour?”, *Financial Stability Review*, Bank of England, December, 2003.

Melo, L. and Rincón, H., "External Shocks and Asset Prices in Latin America before and after Lehman Brothers’ Bankruptcy", *Borradores de Economía*, No. 704, Banco de la República, 2012.

Mizuno, T.; Takayasu, H.; Takayasu, M., “Correlation networks among currencies”, *Physica A*, No.364, 2006.

Naylor, M.J; Lawrence, C.R.; Moyle, B.J., “Topology of foreign exchange markets using hierarchical structure methods”, *Physica A*, No. 382, 2007.

Neftci, S.; Oliveira, A.; Yinqui, L., “Credit default swaps and financial crisis prediction”, *Working Paper Series*, No.114, National Centre of Competence in Research Financial Valuation and Risk Management, 2003.

Onnela, J.-P.; Chakraborti, A.; Kaski, K.; Kertész, J.; Kanto, A., “Dynamics of market correlations: taxonomy and portfolio analysis”, *Physical Review E*, No.68, 2003.

Rebonato, R., *Plight of the Fortune Tellers*, Princeton University Press, 2007.

Serrano, M.A.; Boguñá, M.; Vespignani, A., “Extracting the multiscale backbone of complex weighted networks”, *Proceedings of the National Academy of Sciences*, No.16, Vol.106, 2009.

Shino, J. and Takahashi, K., “Sovereign credit default swaps: market developments and factors behind price changes”, *Bank of Japan Review*, E2, April, 2010.

Taleb, N. N., *Fooled by Randomness,* Random House, 2004.

Taleb, N. N., *The Black Swan,* Random House, 2007.

Wagner, N., *Credit Risk: Models, Derivatives and Management*, CRC Press, 2008.

Wilmott, P., *Paul Wilmott on Quantitative Finance*, John Wiley & Sons, 2006.

Wu, Z.; Braunstein, L.A.; Havlin, S.; Stanley, H.E., “Transport in weighted networks: partition into superhighways and roads”, *Physical Review Letters*, No.96, 2006.