Supplementary Information: Mapping individual behavior in financial markets: synchronization and anticipation

Mario Gutiérrez-Roig Javier Borge-Holthoefer Alex Arenas Josep Perelló

February 28, 2019

S1 Distribution of the symbols

The different nature of both, Position and Price time series, is manifested in the fraction of the symbols (see Figure 2 of the main paper) that appear in each type of series (see Figure S1). Actually, the two kinds of distributions exhibit an opposite behavior. In investors position series there is a predominance of symbol 0, which means no activity, with the 96% and 92% of occurrences for $m = 2$ and $m = 3$ respectively when all series are pooled together. The rest of the symbols, 1 and 2, in $m = 2$ group have almost the same 2% probability each. In the case of $m = 3$ the group of symbols with more probability apart from symbol 0 is the one made by symbols 1,2,4 and 10 which corresponds to making an operation followed by no-activity, or the other way around no-activity followed by making an operation. Finally the rest of symbols in $m = 3$ are characterized by activity in three consecutive days. Between those, symbols 5 and 12 indicate a trend and at same time are the most probable events among the full-activity subset. Regarding probability of the symbols in price series, the symbol 0 indicating no changes is the least probable either in $m = 2$ and $m = 3$. Additionally, there is no significant difference between upwards or downwards symbols. However, it is noticeable that in $m = 3$ symbols which indicate a trend, either upwards (symbol 5) or downwards (symbol 12), are twice more probable than the rest with non-identical values.

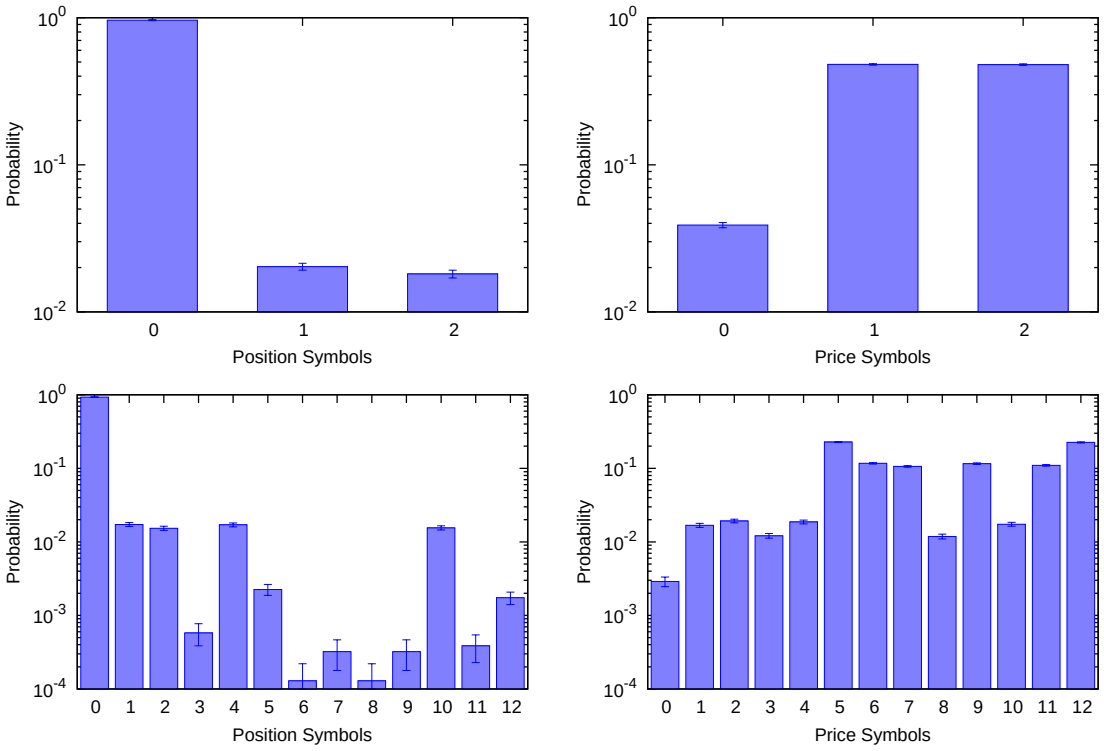

Figure S1: **Empirical Probability for Symbols.** Probability to find each symbol (see Figure 2 in the main paper) for $m = 2$ (top) and $m = 3$ (bottom) after symbolizing and aggregating the series of position (left) and price (right) of all assets.

S2 Shift of null values of Transfer of Entropy produced by different symbol proportion

In order to fairly compare different values of Transfer of Entropy, that is how much , it is important to take into account that different proportion of symbols, specially in short time series, might shift the distribution for the null values. Thus, In Figure S3 we show the results of the simulations where artificial series are generated with different symbol distribution. In the top of the Figure S3 we observe how, in spite of the distributions are different from one case to the other, each $T(\hat{X}, \hat{Y})$ value coming from a different shuffling process is distributed around zeros since the fraction of the symbols is roughly the same. Conversely, when the distribution of symbols differs from one series to the other, $T(\hat{X}, \hat{Y})$ values are peaked around a value not placed at zero (bottom part of Figure S3). This is the most general case we find in our empirical data, so we will take as a null value the median of the distribution around which all shuffled realizations oscillate.

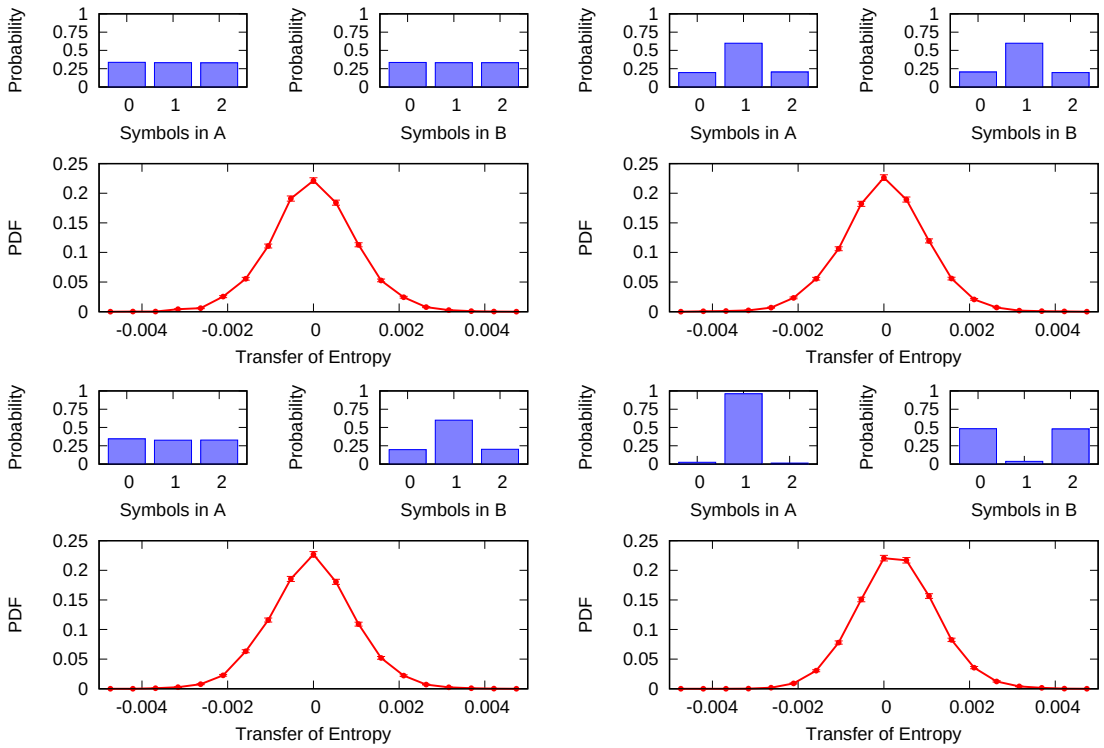

Figure S2: **Transfer Entropy Bias Depending on Underlying Symbol Distributions for $m = 2$.** Transfer Entropy distributions calculated after generating two artificial series of symbols, A and B, with flat distributions (top left), Gaussian distribution with same mean (top right) and flat-Gaussian distributions (bottom left). Length of the A and B vectors is $5,000$ and the distribution corresponds to a $n = 10,000$ calculations of TE value from A to B, shuffling the vectors each time. Finally, an example case for an individual investor is showed using the same procedure (bottom right).

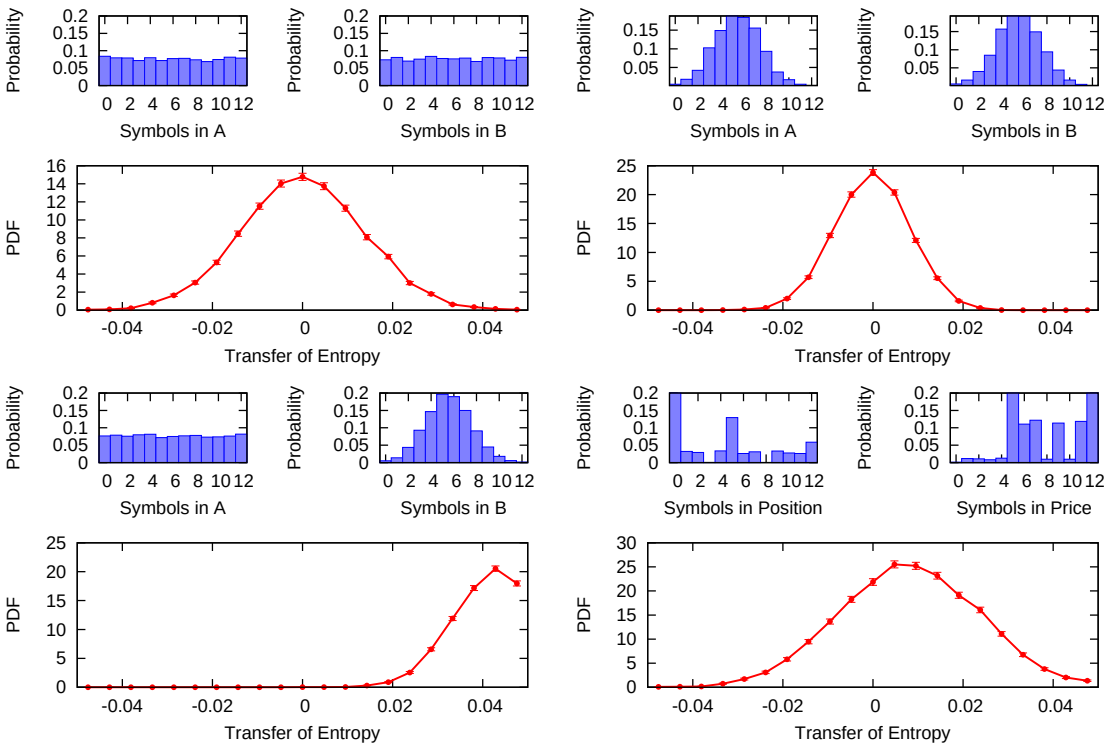

Figure S3: **Transfer Entropy Bias Depending on Underlying Symbol Distributions for $m = 3$.** Transfer Entropy distributions calculated after generating two artificial series of symbols, A and B, with flat distributions (top left), Gaussian distribution with same mean (top right) and flat-Gaussian distributions (bottom left). Length of the A and B vectors is $5,000$ and the distribution corresponds to a $n = 10,000$ calculations of TE value from A to B, shuffling the vectors each time. Finally, an example case for an individual investor is showed using the same procedure (bottom right).

We therefore conclude that we cannot assume the same Gaussian distribution for the Transfer of Entropy for each pair of investors. Instead, we need to make multiple simulations shuffling the values and then extract the values for the mean and the variance in order to define the shape of the null distribution. Then, from that null distribution we can determine whether the values are significant or not applying the techniques we have shown in the methods section.

S3 Synchronization and Anticipation adjacency matrices

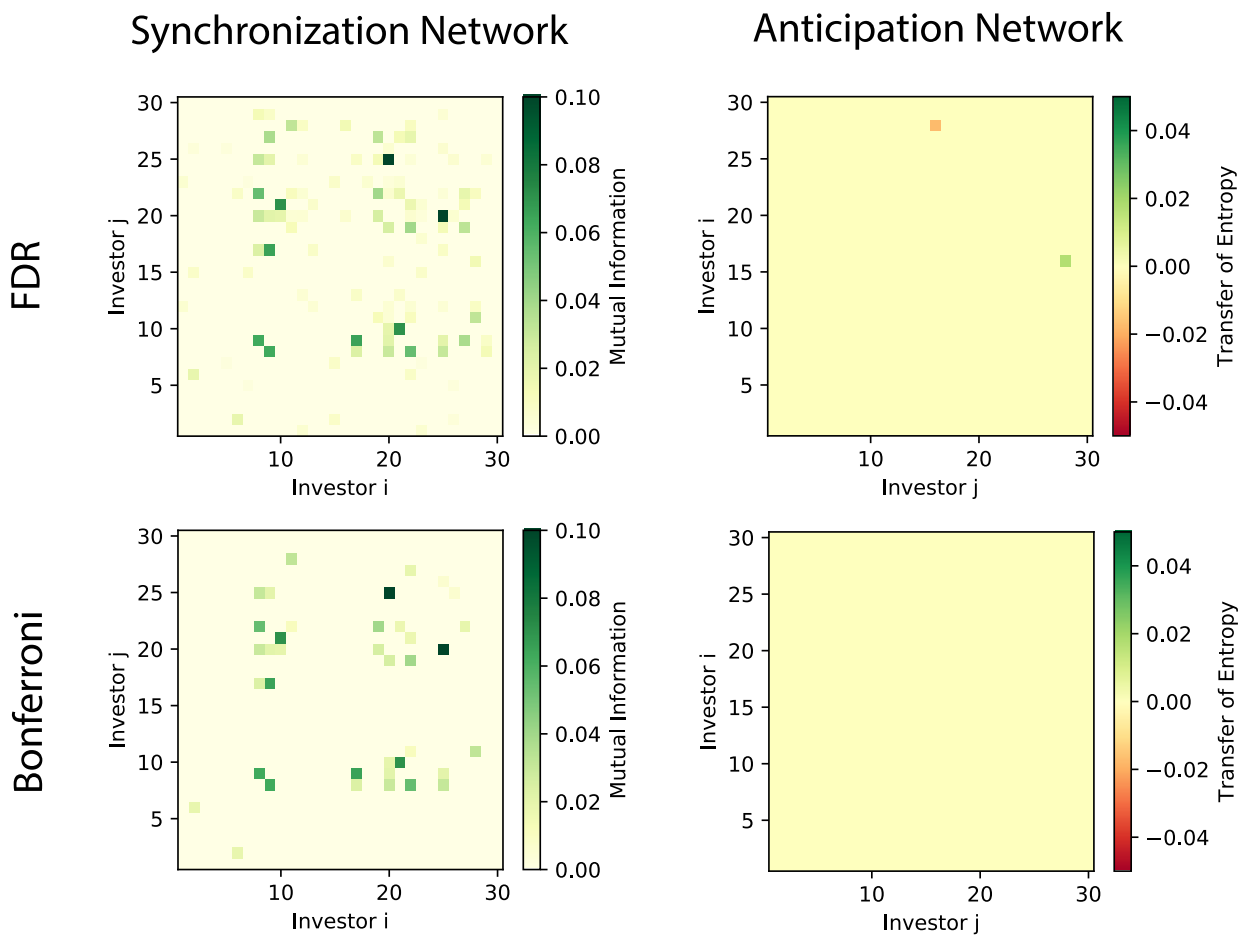

Figure S4: **Synchronization and Anticipation network heatmaps for GAS.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

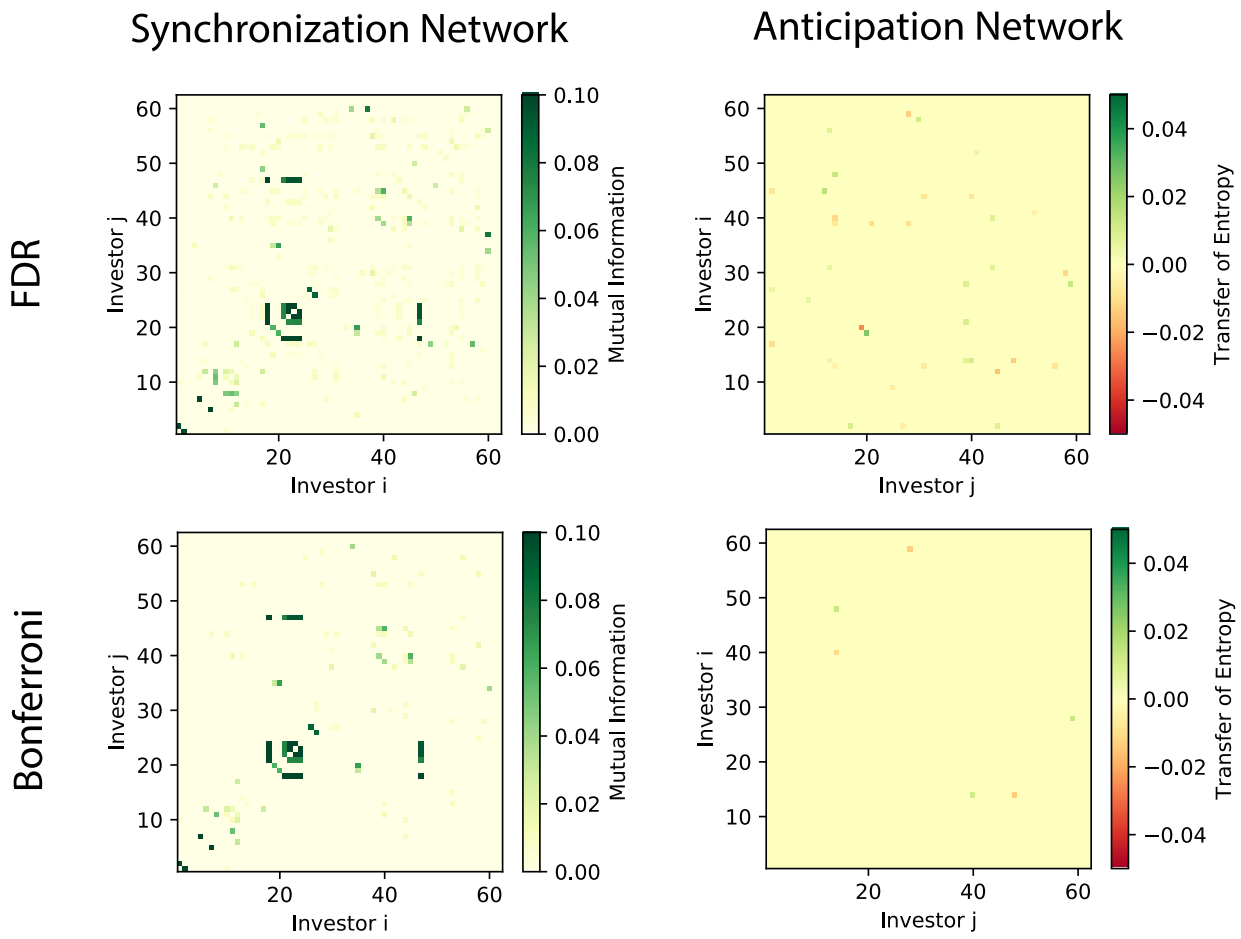

Figure S5: **Synchronization and Anticipation network heatmaps for REP.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

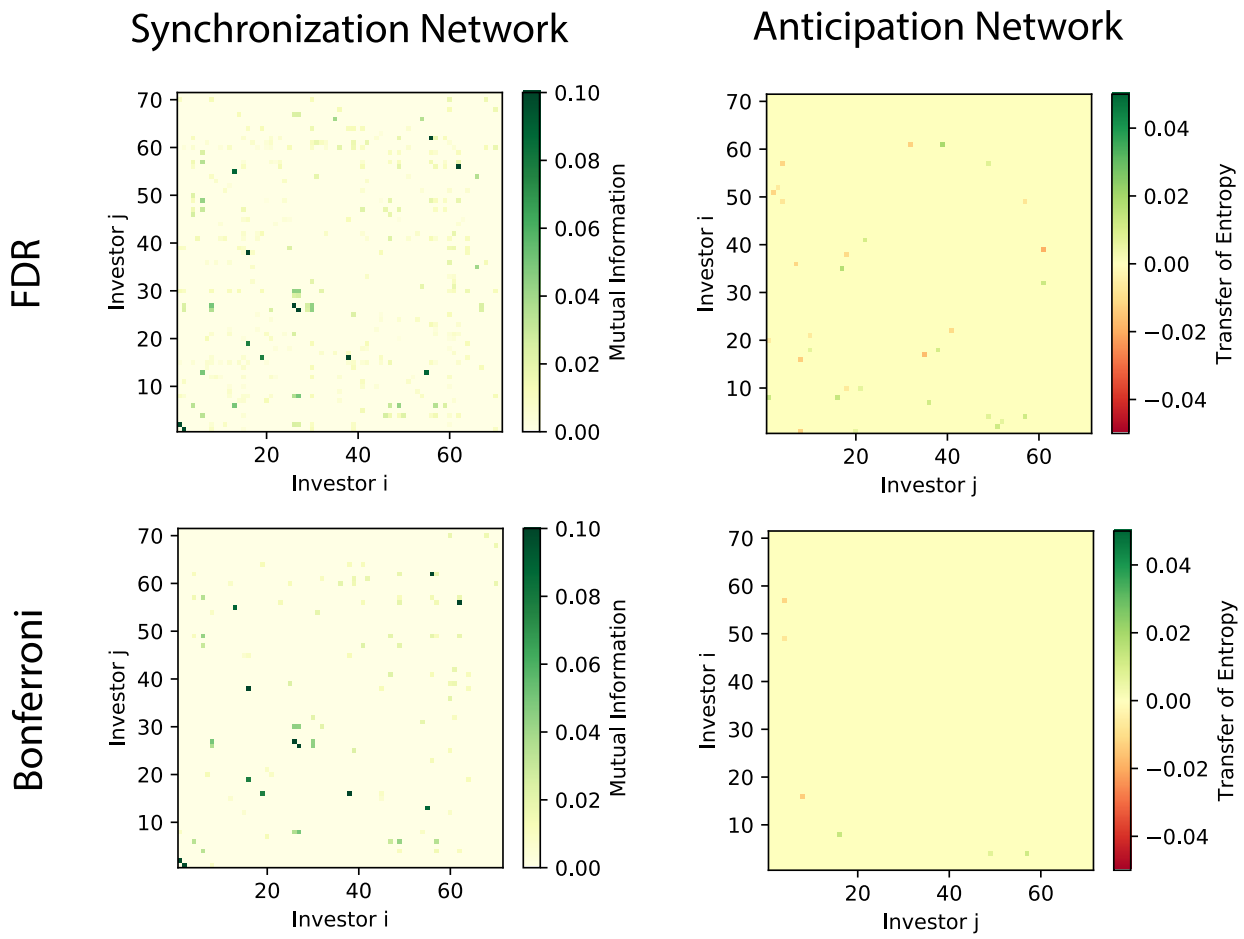

Figure S6: **Synchronization and Anticipation network heatmaps for ZEL.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

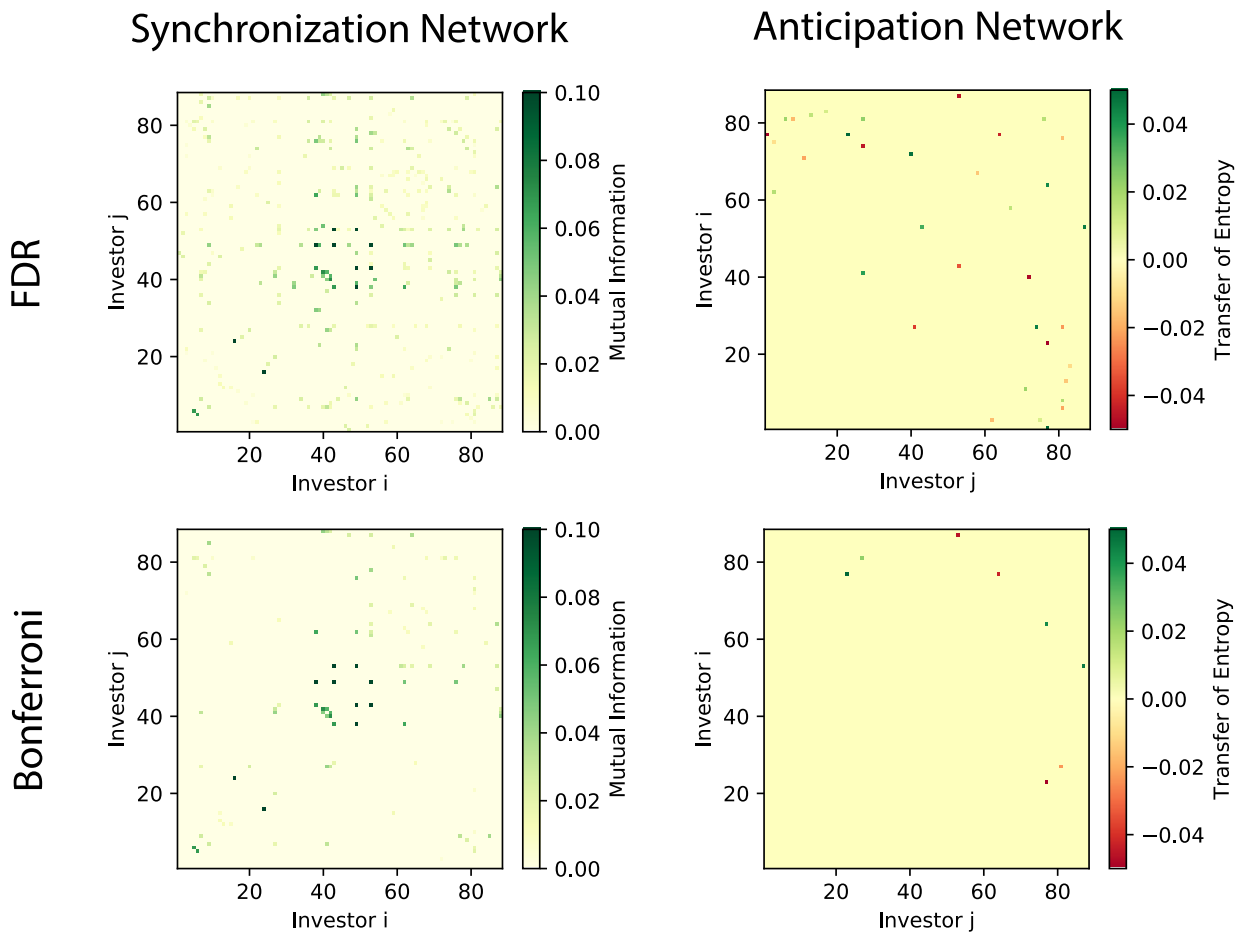

Figure S7: **Synchronization and Anticipation network heatmaps for EZE.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

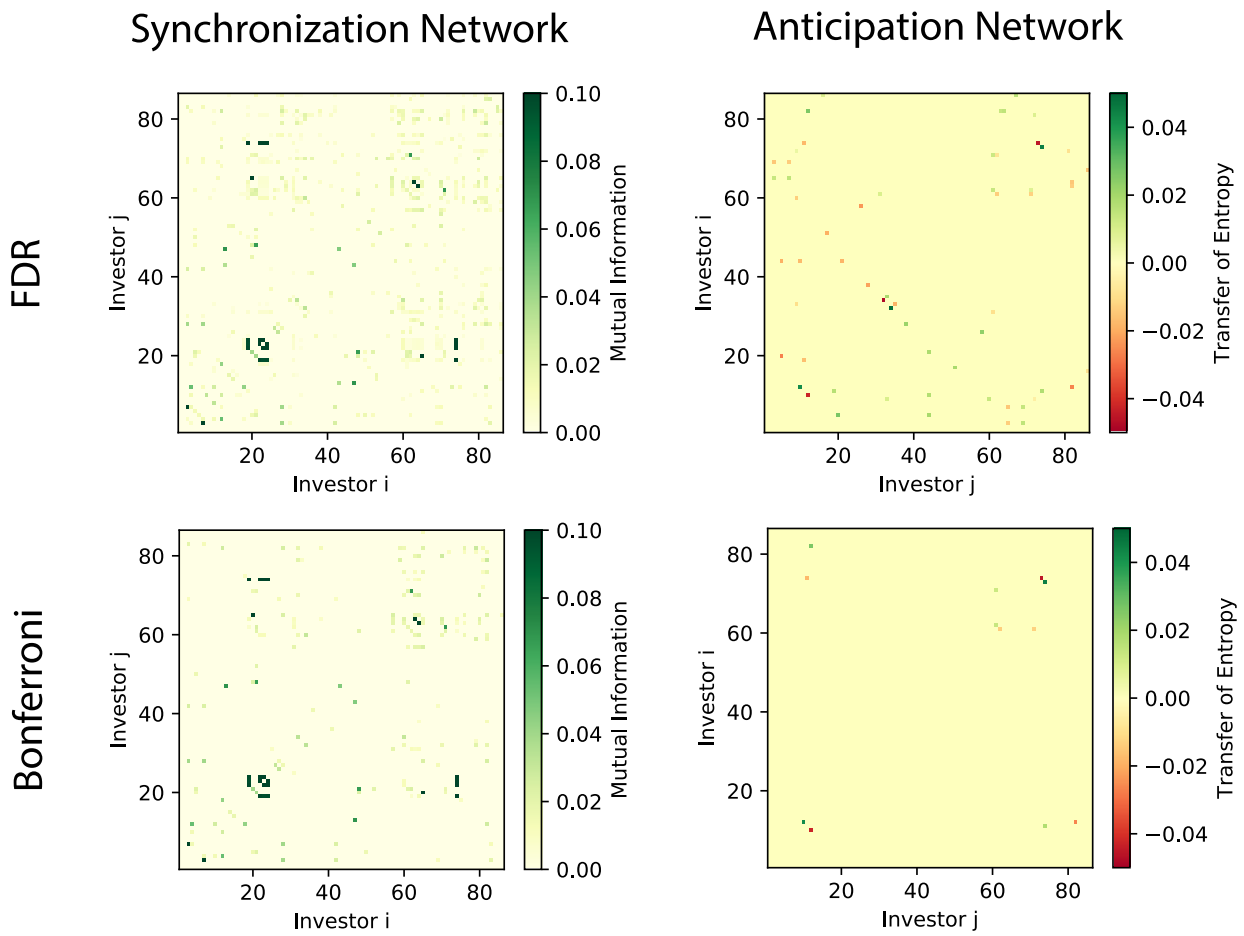

Figure S8: **Synchronization and Anticipation network heatmaps for ELE.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

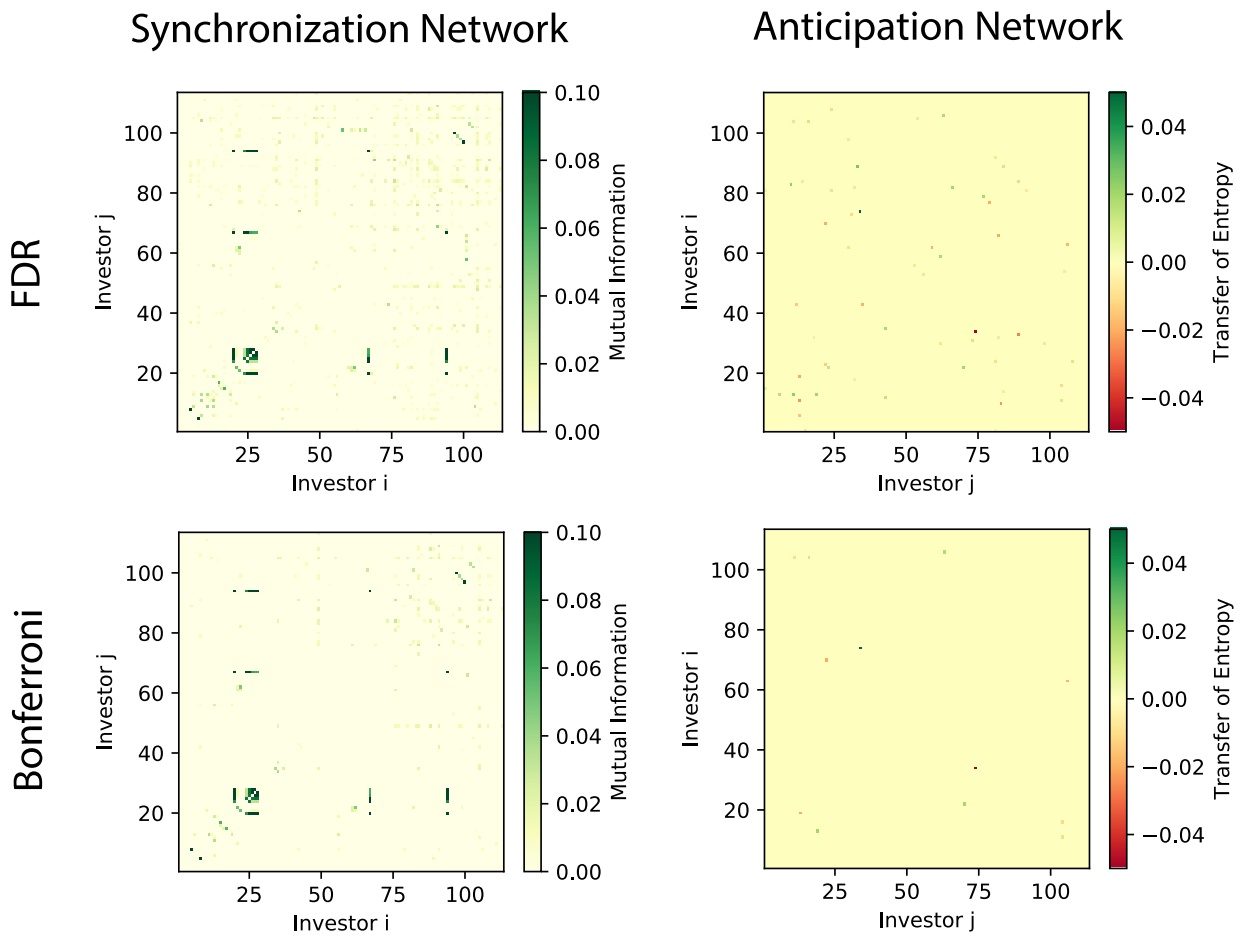

Figure S9: **Synchronization and Anticipation network heatmaps for BBVA.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

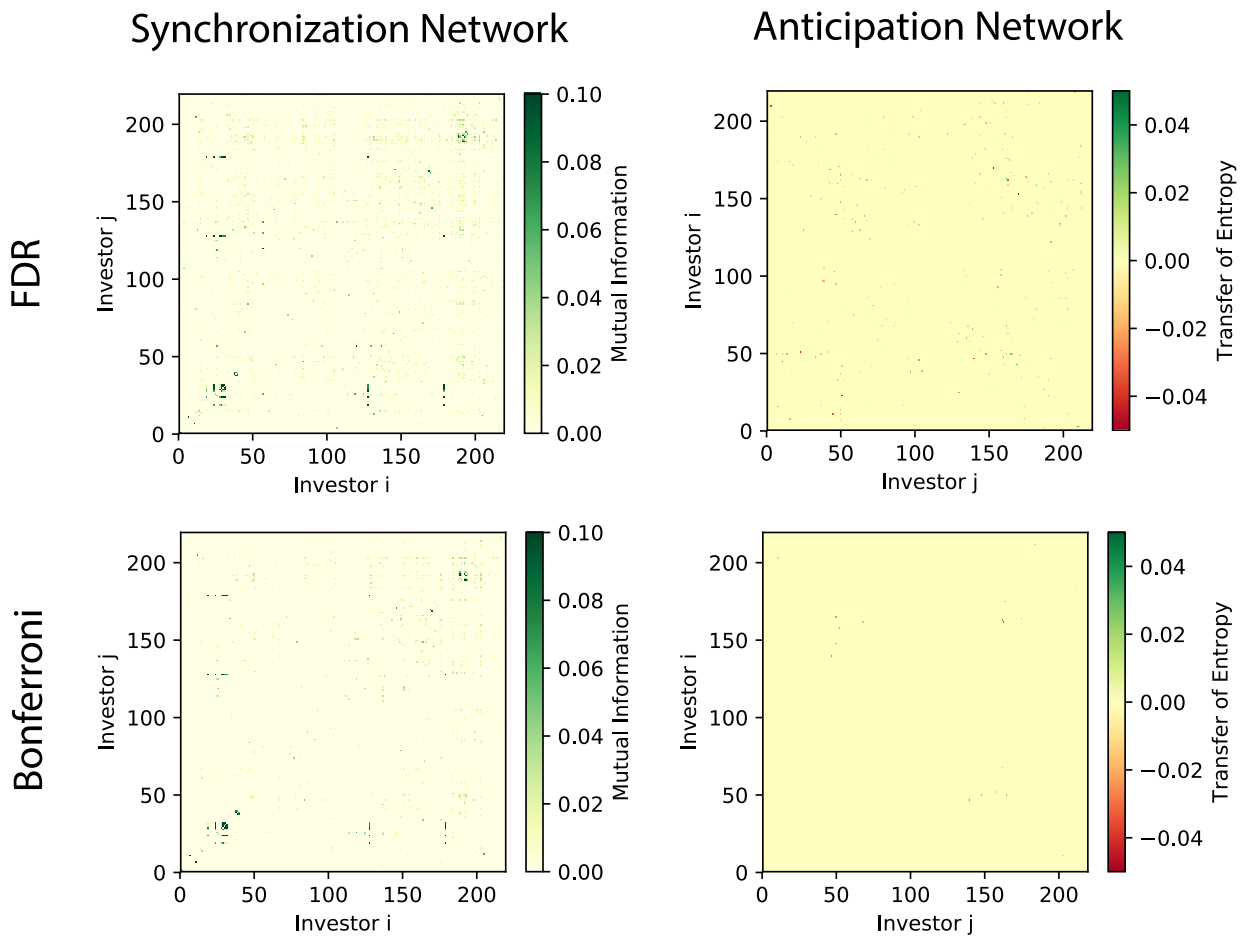

Figure S10: **Synchronization and Anticipation network heatmaps for SAN.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

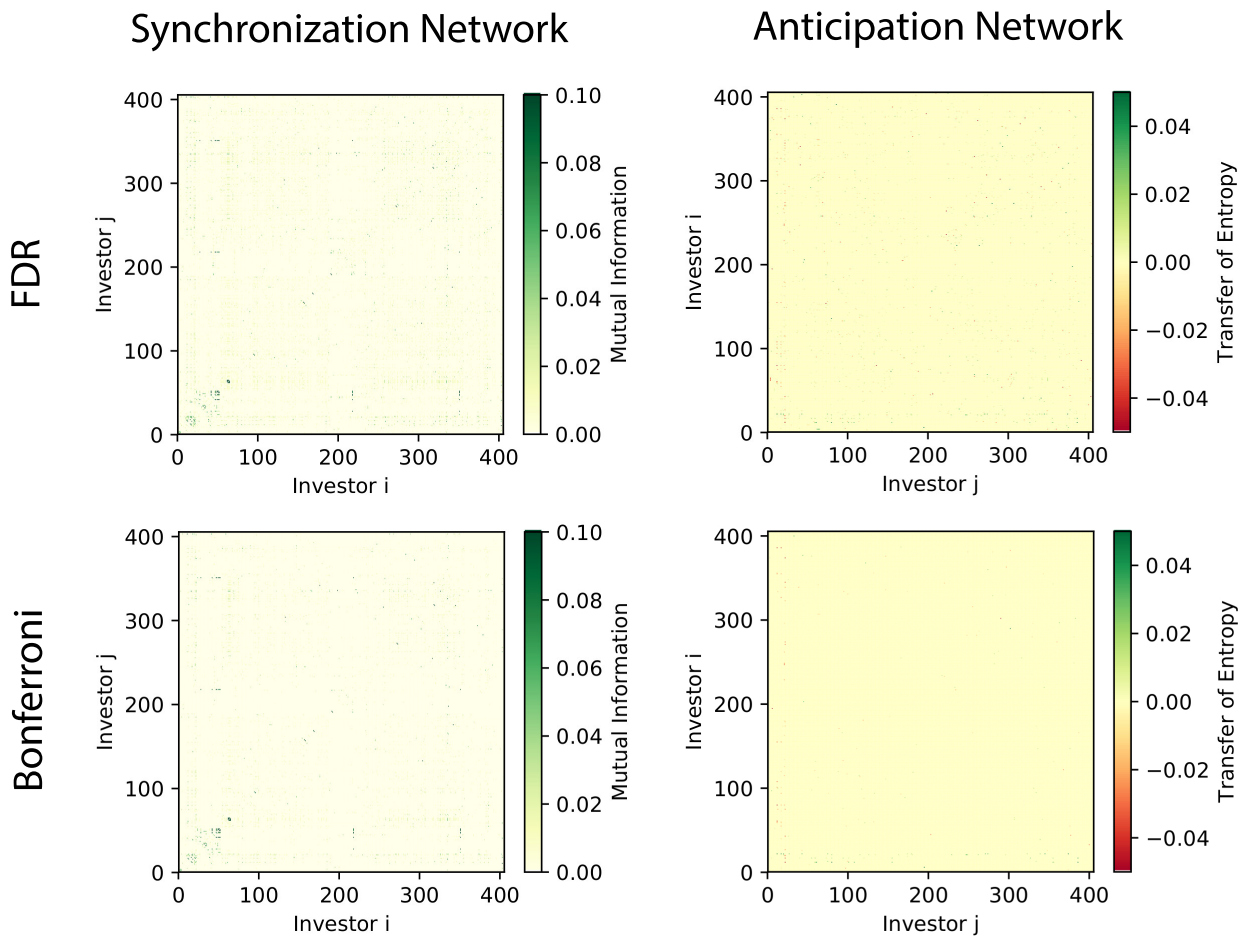

Figure S11: **Synchronization and Anticipation network heatmaps for TEF.** Heatmaps displaying I_{ij} (left column) and T_{ij} (right column) color coded in the bar on the right of each plot, for each pair of investors i and j in x-axis and y-axis respectively. In the top row we show networks generated using FDR method to control false positive, whereas in the bottom row the networks have been generated using Bonferroni.

S4 Network topological properties

Table S1: Synchronization network features using Bonferroni. Number of nodes, edges and average degree is shown in the first three columns for all networks. Fourth column refers to a Null Hypothesis testing based on Kolmogorov-Smirnov (KS) statistic for rejecting the hypothesis that underlying distribution for node out-degrees is a Poisson distribution, only p-value is shown. Fifth and sixth column show the Clustering Coefficient and Degree Assortativity of the undirected graph. The networks in this table have been built using "Bonferroni" as described in methods section.

Asset	Number of Nodes	Number of Edges	Average Degree	Poisson KS test p-value	Clustering Coefficient	Assortativity Coefficient
TEF	342	2401	14.04	$< 10^{-5}$	0.3	0.1
SAN	166	585	7.05	$< 10^{-5}$	0.31	-0.06
BBVA	69	138	4.00	0.00002	0.26	0.0
ELE	54	101	3.74	0.00497	0.33	0.18
EZE	45	53	2.36	0.00015	0.23	-0.16
ZEL	38	49	2.58	0.00578	0.22	0.34
REP	42	59	2.81	0.01996	0.29	0.29
GAS	15	19	2.53	0.15497	0.19	0.02

Table S2: Anticipation network features using Bonferroni. Number of nodes, edges and average degree are shown in the first three columns for all networks. Fourth column refers to a Null Hypothesis testing based on Kolmogorov-Smirnov (KS) statistic for rejecting the hypothesis that underlying distribution for node out-degrees is a Poisson distribution, only p-value is shown. Fifth column shows the Degree Assortativity of the directed graph. The networks in this table have been built using "Bonferroni" as described in methods section.

Asset	Number of Nodes	Number of Edges	Average Out-Degree	Poisson KS test p-value	Assortativity Coefficient
TEF	116	108	0.93	$< 10^{-5}$	0.04
SAN	16	9	0.56	0.00002	0.00
BBVA	11	6	0.55	0.00049	0.00
ELE	9	6	0.67	0.00822	0.00
EZE	7	4	0.57	0.01237	0.00
ZEL	5	3	0.60	0.06064	0.00
REP	5	3	0.60	0.06064	0.00
GAS	0	0	0.00	—	0.00

S5 Individual reaction to price effect on Synchronization network using Bonferroni

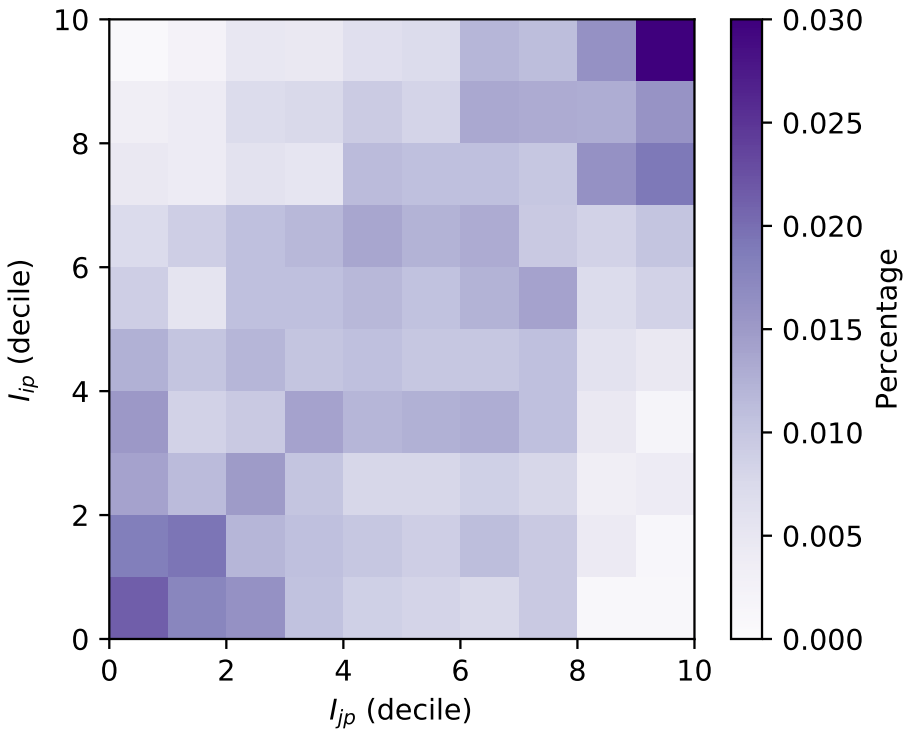

Figure S12: Synchronization with price as driver of synchronization between investors. Axis x and y represent deciles so that generate a 10x10 matrix. Each cell coordinates are given by the deciles of I_{ip} and I_{jp} respectively, and the value showing refers to the frequency of events with statistically significant Mutual Information I_{ij}, i.e. all not null edges of the Synchronization network.

S6 Individual reaction to price effect on Anticipation network using Bonferroni

Table S3: Influence of individual reaction to price over the Anticipation network. The table is divided in four groups across the 8 different markets plus the aggregation of all them. For all significant links from i to j ($T_{ij} > 0$) for each network (market), we computed the probability of i being more synchronized with price than j (first group), i being less synchronized than j (second column), i having a higher value for the STE with respect to price than j (third group) and i having a higher value for the STE with respect to price than j (fourth group). Bonferroni approach described in "Methods" section of the main paper is used here to discriminate the significant links. Numbers between brackets account for the frequency. Asterisks refer to different confidence interval levels, * for 90%, ** for 95%, and *** for 99%.

Asset	$p\,(I_{ip} > I_{jp} \vert T_{ij} > 0)$		$p\,(I_{ip} < I_{jp} \vert T_{ij} > 0)$		$p\,(T_{ip} > T_{jp} \vert T_{ij} > 0)$		$p\,(T_{ip} < T_{jp} \vert T_{ij} > 0)$	
TEF	0.57	(62)	0.43	(46)	0.49	(53)	0.51	(54)
SAN	0.67	(6)	0.33	(3)	0.33	(3)	0.67	(6)
BBVA	0.50	(3)	0.50	(3)	0.50	(3)	0.50	(3)
ELE	1.00	(6)	0.00	(0)	0.50	(3)	0.50	(3)
EZE	0.75	(3)	0.25	(1)	0.75	(3)	0.25	(1)
ZEL	1.00	(3)	0.00	(0)	0.00	(0)	1.00	(3)
REP	0.67	(2)	0.33	(1)	0.67	(2)	0.33	(1)
GAS	0.00	(0)	1.00	(1)	1.00	(1)	0.00	(0)
ALL	0.61***	(85)	0.39***	(54)	0.49	(67)	0.51	(71)