Eur. Phys. J. B **13**, 175–187 (2000)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

The correlation length of commodity markets 1. Empirical evidence

B.M. Roehner[a]

LPTHE, University Paris VII, 2 place Jussieu, 75005 Paris, France

Received 17 May 1999 and Received in final form 31 May 1999

Abstract. It is a common belief nowadays that the world economy is fairly well "integrated". Yet, this belief often turns out to be in contradiction with empirical evidence. As a matter of fact the way distant markets interact is a question that has largely been ignored by economists. In this series of two papers we examine the role that space, that is to say geographical distance, plays in the economics of commodity markets. The first of these papers presents the empirical evidence while the second develops a theoretical framework. The empirical enquiry discloses several noteworthy features, *e.g.* (i) with respect to spatial interaction there is a sharp contrast between stock markets and commodity markets. While there is almost perfect spatial arbitrage in the first case, this is not true for commodity markets. (ii) In spite of their chaotic behavior in the course of time commodity prices display well defined spatial patterns, (iii) as in statistical physics and fluid dynamics interactions can be described in terms of correlation length. The correlation length of a set of markets is seen to increase along with the number of transactions; it also increases when transport costs decline as was the case during the "transportation revolution" of the mid-nineteenth century. Using the notion of correlation length one is able to give a quantitative meaning to the otherwise ill-defined concept of market integration.

PACS. 64.60.Fr Equilibrium properties near critical points, critical exponents – 87.23.Ge Dynamics of social systems – 89.40.+k Transportation

Today there is a common belief that the world economy is spatially "well integrated", an idea that the recent development of the Internet certainly helped to reinforce. Yet it is obvious that there is something wrong or at least unclear with this notion. No doubt that a major shock on the New York Stock Exchange would have world wide effects. However the 1990 collapse of the Tokyo stock market, then the largest in the world in terms of capitalization, has had only few consequences on other markets. More recently, the 1997 stock market crash in a number of Asian countries had little lasting impact on North American or European equity markets. On the basis of these examples one could think that there is at least a strong connection between western markets. This, however, is not true either, or at least not always. For instance between 1 January 1995 and 31 March 1995 the Dow Jones index gained 8.5 percent while the CAC index in Paris lost about 8 percent. If anything, these examples show that the notion of spatial interaction between markets has to be defined with great care. This is what we try in the first section of this paper; more specifically we show that different interaction concepts are usually mixed up in the loose and somewhat confusing concept of "market integration".

[a] e-mail: roehner@lpthe.jussieu.fr

1 The "taking apart" of commodity markets

Commodity prices like stock prices are known to be rather chaotic and unpredictable. But as we show in this paper, in contrast to stock prices they display spatial patterns. The basic reason is that *ultimately* commodity markets depend on a number of *physical* constraints. How long does it take to ship 10 000 tons of wheat from Omaha (Nebraska) to Portland (Oregon)? What will be the cost? Are there enough storage facilities available? Even when, as is indeed the case nowadays, transaction in futures represent 98 percent of all transactions on major commodity markets, these questions still retain their importance.

It should be emphasized from the outset that we do not claim that spatial interaction is the most important factor in the economics of commodity markets, nor do we claim that its understanding will substantially improve our ability to forecast the evolution of prices. Why then should we devote great attention to this topic? Basically because spatial effects are much "simpler" than most other phenomena which play a role in commodity markets. Let us explain that point. Commodity markets constitute networks of a high complexity level. Even if we concentrate on a single commodity leaving thus aside the connection (substitution, complementarity and so on) between different products the number of factors which play a role in the formation of prices is quite bewildering. At the production level: weather conditions, cost of factors of production (wages, oil, etc.); at the level of international trade: tariffs, quotas, bilateral trade agreements, transport and storage costs, etc.; at the level of consumption: business fluctuations, changes in consumers' taste or income, etc. Trying to model all these phenomena in a global way would lead to a phenomenological model which, in our opinion, would not provide any further understanding. As a matter of illustration that situation can be compared to trying to model the pattern of ocean streams without knowing anything about (i) the role of the sun and the moon in the formation of the tides (ii) the centrifugal or Coriolis forces due to the rotation of the earth (iii) the interaction between atmospheric movements and ocean streams. In other words the first question to address is how to decompose the complicated phenomena occurring in commodity markets into a number of simpler components. This is what in the title of this section we called the "taking apart of commodity markets".

In this series of two papers we concentrate on spatial (equilibrium) price patterns. In the first paper we survey the empirical evidence and we show that it displays a number of regularities. In the second paper we examine how these regularities can be explained in the framework of a model referred to as the stochastic spatial arbitrage model (SSAM). Restricting ourselves to that facet of the problem does not mean that we consider other aspects to be unimportant. Yet, for an econophysicist there are several good reasons to focus on that topic, (i) surprisingly enough, spatial phenomena have received little attention from economists. While a few theoretical models have been proposed very little has been done to confront them with empirical evidence. The interested reader will find a discussion of some of these attempts in chapters 1 and 3 of [1]. (ii) One fundamental question in economics is "how do economic entities interact?" It is not easy to address such a question at the micro-economic level of firms and companies because only scanty statistical information is available at that level; note however that the studies concerning the growth of firms [2,3] are promising attempts in that direction. At the macroeconomic level there are scores of data, but macroeconomics deals with aggregated variables and because of their composite nature such problems are of great complexity (see in this respect the discussion in Appendix B of [4]). Commodity market phenomena represent so to say a good compromise: they involve non-aggregated variables and nevertheless are well documented statistically. (iii) Finally, and this is certainly a point of importance for an econophysicist, physicists have developed a variety of tools to deal with spatial phenomena: propagation or diffusion equations, correlation length, scale invariance and so on. Of course none of these concepts can be applied to economic problems without a substantial amount of rethinking. It is precisely this challenge which is exciting.

In this paper we address the following questions, (i) between two markets there is a flow of information as well as a flow of goods; what are their respective roles, (ii) how quickly does a market respond to a shock originating in a distant market? (iii) Do the correlations between a large number of markets follow any specific pattern? To this last question the answer is positive and, given the complexity of the price fixing process, this comes as a good surprise; such regularities provide the starting point and building blocks for the analytical investigation undertaken in the second article.

The paper proceeds as follows. In Section 2, by examining a number of case studies, we discuss how markets interact. It will be seen that there is a sharp difference between stock markets and commodity markets. In the former arbitrage is almost instantaneous and bears a low cost. In the latter arbitrage is about two orders of magnitude less effective. In addition we explain how to distinguish between correlation due to true arbitrage and correlation due to a common economic environment. Section 3 is devoted to a statistical investigation of correlation lengths in various commodity markets and at different times. It turns out that by and large correlation lengths have been multiplied by a factor of ten since the 19th century. It will also be seen that in the event of a major speculative bubble the correlation length can become very large; the relationship with phase transitions is discussed. Finally, in the last section we examine the connection between price correlations and price differentials. From a theoretical point of view the correlations are the "natural" variables, yet for traders and arbitragers the essential indicators are the price differentials. Therefore it is important to show how the latter can be derived from the former.

2 How do markets interact?

This is certainly a very ambitious question for markets most probably interact in a number of different ways. After all in the physical world a single type of interaction is the exception rather than the rule. For instance molecules can interact attractively or repulsively depending on their distance, their form or their structure. In this section our objective is to point out some essential features and to propose a classification which constitutes a guide for further studies.

2.1 Commodity markets versus stock markets

2.1.1 Stock markets

As a case in point we consider the price of an IBM stock on the NYSE and on the Paris stock market on the other hand. Prices in Paris which were originally expressed in French francs have been converted into dollars using daily exchange rates. The price series of daily closing prices are shown in Figure 1a. In order to estimate how "close" the two series are we use two main measures. (i) The coefficient of linear correlation, (ii) the relative price differential defined as:

$$E\left(|p_1 - p_2|\right) / \left(\left(E(p_1) + E(p_2)\right)/2\right)$$

where $E(X)$ denotes the expectation of the random variable X. The results are summarized in Table 1a.

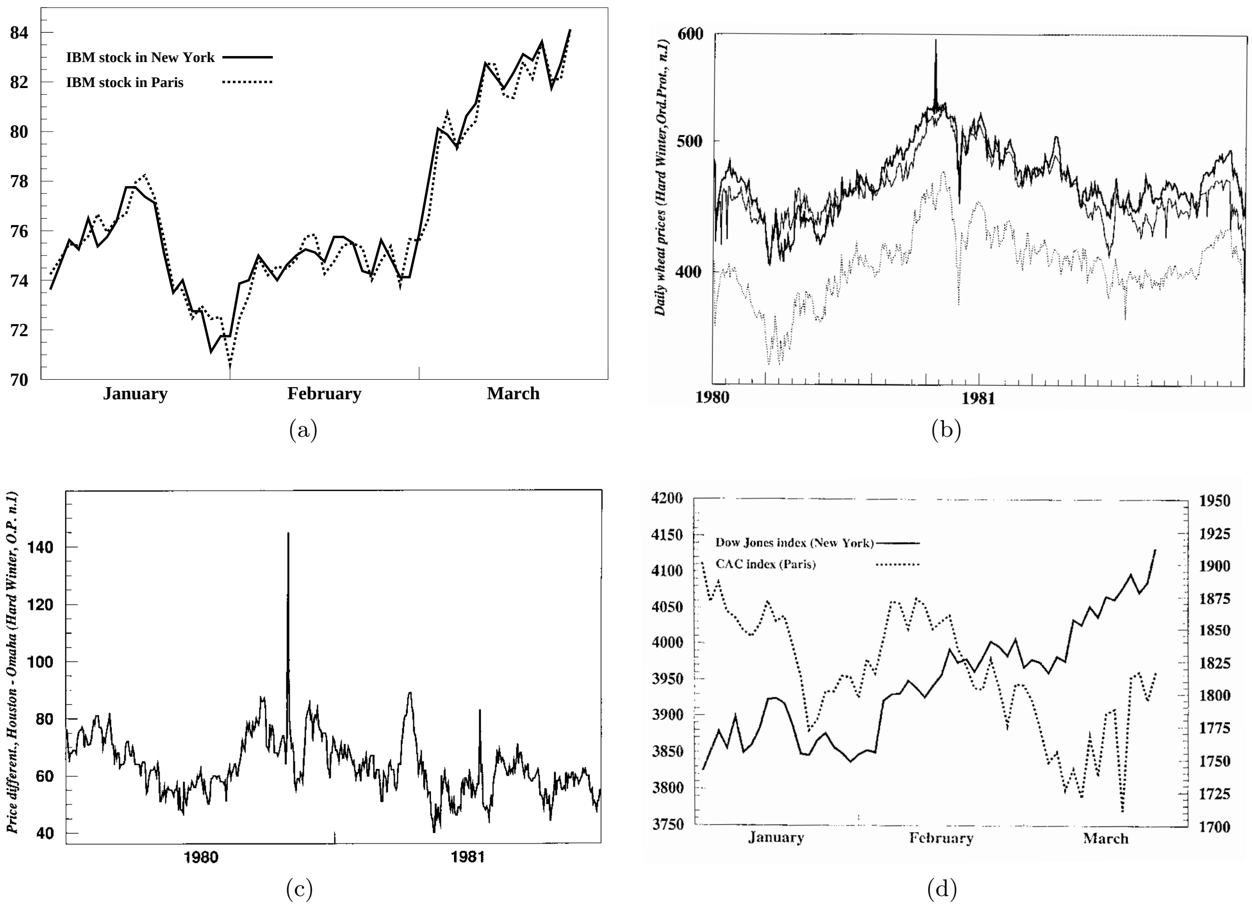


Fig. 1. (a) Comparison of the price of IBM shares in New York and in Paris (Jan.-Mar. 1995). Vertical scale: price per share in dollars. French francs have been expressed in dollars using daily (closing) exchange rates. *Source: Libération (French newspaper) Jan.-Mar. 1995.* (b) Daily wheat prices in Houston (Texas), Omaha (Nebraska) and Portland (Oregon). Thick line: Houston; thin line: Portland; broken line: Omaha. The prices are expressed in cents/bushel. Omaha is located in the center of the country almost at equal distance from the Atlantic and from the Pacific. *Source: US Dept. of Agriculture (cash grain prices).* (c) Daily price differential between Houston (Texas) and Omaha (Nebraska). Vertical scale: cents/bushel. The average price differential is equal to 63 cents, which represents a relative price differential of 14%. Note the spiky fluctuations which mean that at some moments the arbitrage mechanism lags behind price variations; the coefficient of variation of the differential is equal to 16%. *Source: same as for (b).* (d) Evolution of the Dow Jones index (New York) and of the CAC index (Paris) in 1995. The vertical scale on the left-hand side and right-hand side correspond to the Dow and to the CAC respectively. *Source: Libération (French newspaper) Jan.-Mar. 1995.*

The relative price differential is less than one percent and the correlation is as high as 0.98. Differences in share prices can be ascribed to the following three factors, (i) transaction costs: selling N IBM shares in New York and buying N shares in Paris has a cost whether in the form of brokerage charges, taxes or exchange-rate charges. Because several of these costs are fixed charges, the transaction cost per share is a decreasing function of N. For IBM shares that cost \$ 73/ share the differential of 53 cents given in the third column of Table 1a is a reasonable order of magnitude. (ii) The closing prices that we used here are only a rough approximation for the price changes that occur during the opening hours of the stock exchange. The same observation applies to exchange rates: the closing rates that we used reflect only imperfectly their evolution in the course of the trading day.

2.1.2 Cash grain prices

Let us now consider cash wheat prices in Omaha (Nebraska) and in Portland (Oregon). Omaha is an important wheat market located in the wheat belt; it is almost equally distant from the East coast (1600 km) and from the West coast (2000 km). Portland on the other hand is a market located on the Pacific coast. Figure 1b and Table 1a show that the price correlation is substantially lower than before and, more importantly, that the relative differential is about 18 times larger. It should be emphasized that this differential cannot be attributed to differences in grain quality. American grain specifications (class, grade) are very precise and the prices referred to in Table 1a refer to wheat of the same class (Hard Winter, Ordinary protein) and the same grade (No. 1);

Table 1. (a) Interaction between spatially separated markets: IBM shares *versus* wheat (daily prices), (b) interaction between spatially separated markets: stock markets *versus* commodity markets (daily prices), (c) interaction between spatially separated markets: nineteenth century wheat prices (fortnight prices).

(a)

	Price correlation	Price differential %	Price differential [cents]	Correlation of price changes
IBM shares New York / Paris: 6500 km (Jan.-Mar. 1995)	0.98	0.70	53	0.66
Wheat (Hard Winter, Ordinary Protein, No 1) Omaha / Portland: 2200 km (Jan.1980-Dec.1981)	0.85	13	56	0.61

Note: The relative price differential is about 20 times larger for wheat than for share prices. The correlation of price changes has been computed to control for a possible common trend. *Sources: Share prices: Libération Jan.-Mar. 1995; grain prices: United States Dept of Agriculture (cash grain prices).*

(b)

	Price correlation	Price differential %
1) Stock markets		
Eastman Kodak, New York / Paris	0.88	1.1
General Electric, New York / London	0.69	does not apply
General Electric, New York / Paris	0.88	1.0
2) Commodity markets		
Wheat, Houston / Omaha (1300 km)	0.93	14
Wheat, Houston / Portland (3000 km)	0.89	2.5

Notes: The stock prices are for the period Jan.-Mar. 1995; the wheat prices are for the period Jan. 1980-Dec. 1980. In London GE stocks are split into smaller shares than in New York and Paris; therefore the notion of price differential does not apply. The differential for grain prices is several times larger than for shares prices. It should also be emphasized that the arbitrage between New York and Paris must take into account changes in the exchange rate. This may explain the fact that the correlations are or the same order of magnitude for shares and for wheat prices; the low level of correlation for GE shares in New York *versus* London seems to be an anomaly for which we do not yet have an explanation. *Sources: Same as for (a).*

(c)

	Distance between markets [km]	Time period	Price correlation	Price differential %
Louviers / Toulouse	600	1825-1826	−0.48	35
Montauban / Toulouse	50	1825-1826	0.62	5.3
Louviers / Toulouse	600	1901-1902	0.59	9.9

Sources: Drame et al. [26].

for a more detailed discussion of wheat classes and grades see [5]. To which factors should such a large price differential be attributed? As for stock prices transaction costs certainly play a role, but the largest part of the differential is probably due to transportation costs. About fifty percent of the American wheat production is exported and for a trader having a cargo of wheat in Omaha instead of Portland makes a big difference; in the last case it can directly be shipped to Asia and to the Middle East whereas from Omaha it cannot.

To sum up there is a clear distinction between financial products for which there are no transportation costs and which are characterized by price differentials of less than one percent and commodities for which there are substantial transportation costs and which are characterized by price differentials that are several times larger.

2.1.3 Qualifications

One may wonder if the previous conclusion which was drawn from two examples only holds in other cases as well. Generally speaking the answer is "yes"; however some qualifications are in order. The Tables 1b and 1c provide additional evidence and suggest the following observations. (i) For the prices of Kodak or General Electric stocks either in New York or in Paris the price differentials are again of the order of one percent but the correlation is somewhat lower than for IBM shares. For GE stocks in New York *versus* London the correlation is even lower and we confess that at this point we have no explanation for such a low correlation level, (ii) for grain prices in Houston (Texas) *versus* Omaha the differential is again over 10%, (iii) for grain prices in Houston *versus* Portland the price differential is 2.5%; given that Houston and Portland are about 3000 km apart such a small differential could at first seem to contradict our argument about the role of transport cost, but in fact it comes as a confirmation. Indeed both Houston and Portland are ports; remember in that respect that once a cargo of wheat is on board of a freighter transportation costs per unit of weight are almost negligible (for more details about transportation costs by sea see [1], Chap. 4). (iv) Table 1c gives the correlation and price differential for French wheat markets in the 19th century. Louviers is a major wheat market in Normandy; Toulouse is also an important wheat market in the southwest of France. In the first quarter of the 19th century there was almost no correlation (at the level of fortnight prices) between those markets and the price differential was as high as 35%. Yet prices on markets separated by smaller distances were well correlated as shown by the example of Montauban and Toulouse. By the beginning of the 20th century the correlation reached the same order of magnitude between Louviers and Toulouse. A similar evolution took place in other European countries. The case of Germany has been examined in [6]. It is reasonable to attribute that evolution to the transportation revolution that took place in the second half of the 19th century.

Before we leave the comparison between stocks and commodities a last observation is in order. One may wonder whether the correlation between the prices of specific stocks in New York *versus* Paris was due to true arbitrage or rather to the fact that both markets had a similar trend. Figure 1c provides the answer by showing that in the period considered the two markets in fact moved in opposite directions. In other words the high correlation between GE, IBM and Kodak stock prices was achieved *in spite* of opposite trends. In addition this example emphasizes an important distinction. There can be no arbitrage (at least directly) between the Dow Jones index and the CAC index. Therefore it is not surprising to see these indexes moving in opposite directions. Yet, as is well known, in other time periods the Dow and the CAC happened to move in the same direction. This then is not the result of arbitrage but simply reflects the fact that both markets are in a similar phase of the business cycle. A more systematic investigation of stock index correlations has been made recently by Vandewalle and Boveroux (personal communication); it concerns the markets in Frankfurt, New York and Tokyo and covers a period of 18 years.

2.1.4 Different types of couplings

Let us try to summarize the conclusions that can be drawn from the previous observations. Spatial arbitrage is the basic mechanism which links markets together. If the price on market A is p_A and the price on market B is $p_B = p_A + t$ where t represents the transport cost from A to B, then a trader has an obvious incentive to ship goods from A to B in order to sell at price $p_A + t$ instead of selling at price p_A. This is a fairly clear mechanism; unfortunately it is often masked by other effects. The two following are of particular importance. (i) With $p_B = p_A + t$ traders on market B, knowing that the arrival of shipments from A will tend to lower p_B, may be tempted to anticipate the price reduction. If so, the price change will be triggered merely by an exchange of information (not by an actual change in supply levels). (ii) Suppose price changes on two different markets are almost synchronous; this does not necessarily mean that they are interconnected in some way. If the economic environment of both markets is the same they will be affected similarly by exogenous shocks. For instance if the traders and brokers in London, Chicago and New York have attended the same business schools, if in addition they use the same models, then they are likely to react in a similar way to any exogenous shock (*e.g.* an increase in the price of oil or a change in interest rates).

In short one can distinguish (at least) three scenarios resulting in an (apparent) interdependence between markets. Firstly interaction through a genuine arbitrage process whereby shipments of goods are transfered from one market to another; secondly interaction through the exchange of information which in fact anticipates subsequent arbitrage; and thirdly a (spurious) form of interaction based on similar responses to the same exogenous shocks.

2.2 Interaction between forward markets

On futures markets forward contracts are sold which give the buyer the right to take delivery of a given quantity

Table 2. Interaction between forward grain markets (daily prices).

	Distance between markets [km]	Time period	Price correlation	Price differential %
Chicago / New York	1200	1896-1899	0.91	14
New York / Liverpool	6000	1896-1899	0.86	11
New York / Paris	6500	1896-1899	−0.11	31
Liverpool / Paris	800	1896-1899	0.07	20

Source: Das Getreide im Weltverkehr [27].

of wheat in (for instance) 3, 6 or 9 months. Because only a small percentage of the transactions on futures markets result in an exchange of physical quantities and because contracts can be transfered from one market to another (at least if similar contracts exist on either market) one would expect interactions between futures markets to be of the same kind as interactions between stock markets. Is this confirmed by observation? An answer is provided by Table 2; it gives the correlation and price differentials for daily term prices of wheat at Chicago, New York, Liverpool and Paris. These forward contracts had to be settled either by the end of the current month or, as in Paris for the 10th, 20th or last day of the month[1]. Whereas the coupling between Chicago, New York and Liverpool seems to be fairly strong, the coupling between Paris and the other markets is almost non-existent. One could be tempted to interpret that contrast by the fact that France imported less American wheat than Britain; indeed between 1878 and 1897 American imports represented 24% of consumption in Britain but only 4.8% in France ([1], pp. 150, 250). While this may be part of the answer it cannot alone explain such a sharp difference. It seems likely that the organization of the market was different in Paris. If the contracts sold in New York were not available in similar form (quantities, qualities, duration) in Paris, that would make arbitrage difficult.

2.3 Conclusion

In this concluding paragraph we try (at least tentatively) to attribute specific weights to the three interaction effects listed above. This is not an easy task; in particular it is difficult to distinguish between true arbitrage and interaction through exchange of information. Figure 2 gives the respective weights for each factor. Let us briefly comment and justify each of these pie charts.

(i) In the case of 20th century spot markets (and a fortiori for previous centuries) spatial arbitrage seems to be the dominant effect. In a previous paper [4] it has been shown (at least for the 19th century) that price patterns are consistent with price waves travelling with a velocity of the order of 100 km/month; this strongly suggest a spatial arbitrage mechanism. As an additional proof that there is no automatic propagation of price changes in the absence of supply shifts, the following observation can be mentioned. In the fall of 1590 and spring of 1591 Paris (then ruled by the so-called Catholic League) was besieged by the king Henry IV. As a result, within one year the price of wheat was multiplied by five (weekly prices are available for that period in [7]); prices remained about three times above normal level until the end of 1592. Yet, in spite of that huge price peak in one of the largest European cities, prices in distant markets like Utrecht (400 km) or Toulouse (600 km) remained absolutely flat. Due to the siege no wheat could be brought in; therefore there was no real supply sink and the price peak did not propagate to other major wheat markets.

(ii) Regarding the impact of new information on price expectation it has been shown repeatedly that there is a striking contrast between commodity markets (and more generally industrial markets) and financial markets; for industrial markets see [8], for financial markets see for instance [9]. Financial markets respond quickly and unequivocally to new information. On the contrary the response of industrial markets is difficult to identify among a lot of other signals. Such a difference is quite understandable; industry is ruled by many constraints, delays, bottlenecks and so on. On the other hand the management of a financial portfolio is almost a pure information game. For futures markets as the settling day draws nearer one expects term prices to come in line with spot prices. As a result one would expect arbitrage mechanisms to play a greater role in forward markets than in equity markets.

In the rest of this paper we restrict ourselves to cash prices on spot markets.

3 The correlation length of commodity markets

3.1 The correlation length

The correlation length is widely used in statistical physics and in fluid dynamics as a convenient measure of the spatial extension of clusters of correlated elements. In previous publications [1,5,6] we briefly showed how it can be used for the description of correlation patterns in commodity markets. First let us recall its definition. Let X_i denote a random variable which can describe either a spin in an Ising model or the pressure at a given point in fluid

[1] I am grateful to Professor G. Persson for pointing this precision to my attention.

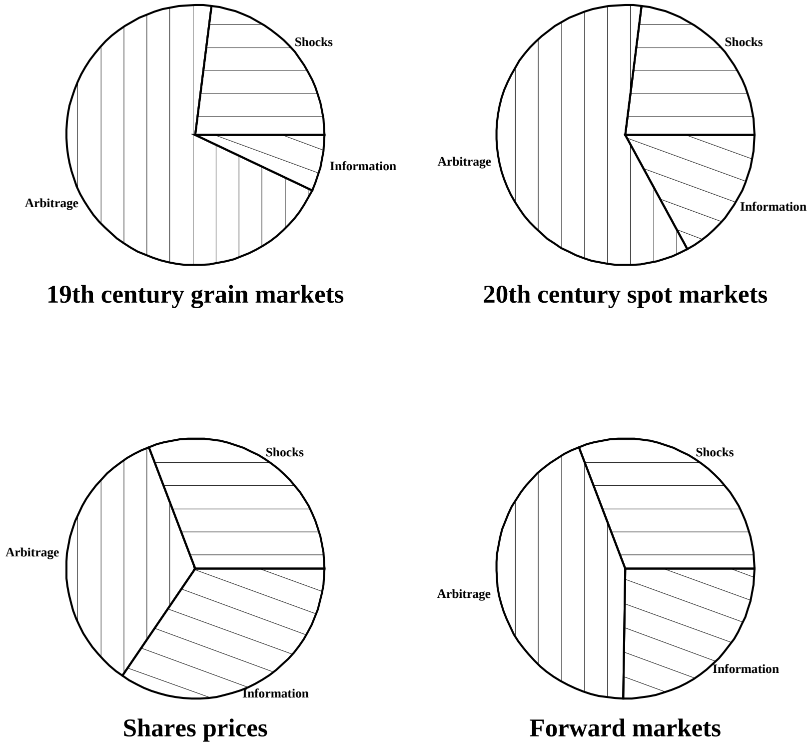


Fig. 2. Schematic representation of the weight of three causes of interdependence between markets, namely arbitrage, common exogenous shocks and transmission of information. The distinction between information and arbitrage is somewhat artificial in the case of share prices since financial portfolio management is almost a "pure information game".

or the price on a given market. Let $\rho(d)$ denote the correlation between X_i and X_j where the locations (i) and (j) are supposed a distance d apart. Observation shows that often $\rho(d)$ decreases exponentially with distance: $\rho(d) = \mathrm{e}^{-ad}$; this leads to the introduction of a correlation length L defined by:

$$\rho(d) = \mathrm{e}^{-d/L} \tag{1a}$$

This definition can be generalized to a non-exponential decrease as follows ([10], p. 249):

$$\delta = \int_0^\infty \rho(x)\mathrm{d}x \tag{1b}$$

When $\rho(x)$ is an exponential the second definition leads back to definition (1a) ($L = \delta$). In fluid dynamics ρ displays substantial time fluctuations which makes the correlation length time-dependent; in this paper however we restrict ourselves to the equilibrium case.

Figures 3a and 3b summarizes a procedure for measuring the correlation length for a set of markets. First one estimates the correlation between pairs of markets. Notice that in principle one should also consider the correlations of each market with itself; for zero-lag correlations, however, all these auto-correlations are equal to one; the situation would be different when investigating a time-dependent case. From Figure 3b the correlation length may be derived either by using a log-linear fit or by estimating the area under the curve. Statistically the second procedure is usually beset with considerable uncertainty because the low correlation tail of the curve is not well defined.

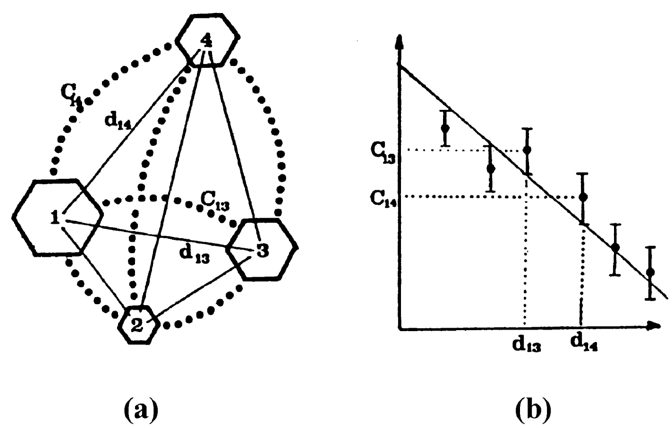
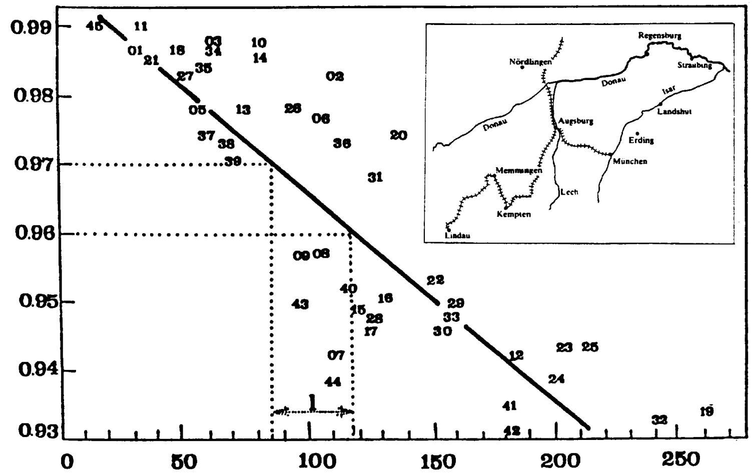


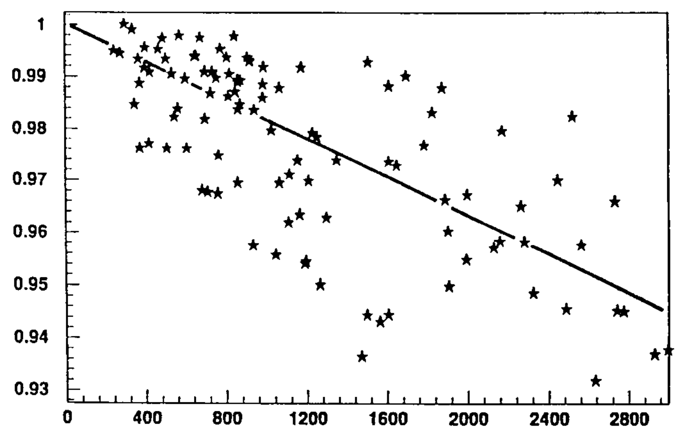
Fig. 3. Operational definition of the correlation length. This definition also suggests an estimation procedure. With a logarithmic scale on the vertical axis of figure (b) the correlation length is estimated from the distance-correlation regression as being the inverse of the slope.

3.2 Spatial patterns of price correlations

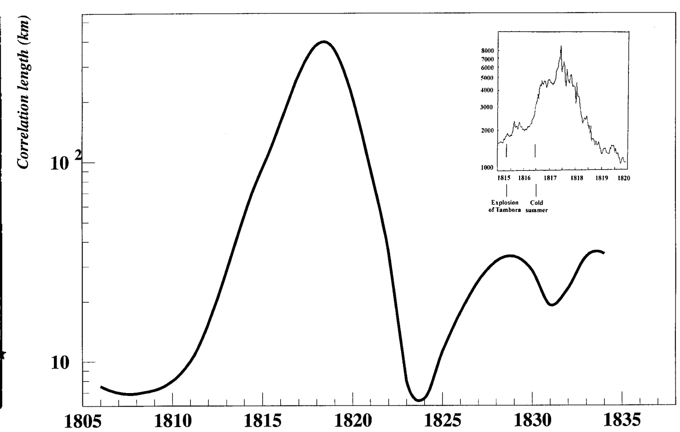
Figures 4a and 4b show two typical correlation clusters. The first concerns 19th century wheat markets in Bavaria, the second 20th century wheat markets in the United States. The scale on the vertical axis is logarithmic. The correlation of the log-linear fit is equal to -0.85 and -0.70 respectively which implies indeed a significant relationship between correlation and distance. The correlation length L would be equal to 3200 km and 54 100 km respectively; it is therefore convenient to define a correlation length



(a)



(b)



(c)

Fig. 4. (a) Decrease of wheat price correlation with distance in Bavaria (1825-1833). Horizontal scale: intermarket distance expressed in kilometers; vertical scale: inter-market correlation (logarithmic scale). The sample contains 10 markets. The plotted numbers refer to the 45 market pairs in the following way. Numbers 1-9: Munich-others; 10-17: Erding-others; 18-24: Straubing-others; 25-30: Landshut-others; 31-35: Lindau-others; 36-38: Augsburg-others; 40-42: Regensburg-others; 43-44: Nördlingen-others; 45: Memmingen-Kempten. *Source: Seuffert [23].* (b) Decrease of wheat correlation with distance in the United States (1954-1986). Horizontal scale: inter-market distance expressed in kilometers; vertical scale: inter-market correlations. The prices are annual prices received by producers averaged by state. The sample contains 15 states. *Source: Langley et al. [24].* (c) Divergence of the correlation length during a price peak. Between 1814 and 1817 wheat prices were multiplied by five as shown by the graphic in the upper right corner; the triggering factor seemed to have been the explosion of the Tambora volcano in Indonesia (1815) which was followed by the cold summer of 1816. As a result of the dearth the economy and indeed the whole society focused on the subsistance problem. The number of transactions and of traders swelled, wheat was grown even in land poorly suited for that culture. The net result of this focalisation was a multiplication of the correlation length by a factor 100 (from 5 km in 1808 to about 500 km in 1818). *Sources: Seuffert [23], Jacobs and Richter [25].*

$l = L/100$; in other words definition (1a) becomes:

$$\rho(d) = \mathrm{e}^{-d/100l}. \qquad \text{(1a')}$$

With this definition the correlation length l is equal to 32 km and 541 km respectively.

Why, could one wonder, is the dispersion much larger for the American prices. There are two main reasons, (i) in Figure 4a each correlation is computed on the basis of 108 monthly prices whereas in Figure 4b each correlation is computed from 33 annual prices. Therefore the confidence interval for each correlation is larger (approximately by a factor $\sqrt{108/33}$) in the second case than in the first. (ii) In Figure 4a all markets belong to a rather small area whereas Figure 4b covers the whole continental territory of the United States; therefore there is a larger dispersion of exogenous shocks.

The examples of Figures 4a and 4b are representative of a wide class of cases. In fact, as will be seen in the next paragraph, most commodities display similar price patterns.

Table 3. Correlation length for commodity markets.

	Year	Number of markets	Periodicity	Correlation length [km]	Goodness of fit (log-linear correlation)
1) Wheat					
Bavaria	1829	9	M	31 ± 5	(0.85)
France (center)	1873	11	M	55 ± 18	(0.87)
France	1873	12	M	71 ± 24	(0.70)
France (ports)	1873	10	M	211 ± 41	(0.58)
U.S.	1970	31	A	560 ± 41	(0.70)
2) Potato					
Prussia	1837	10	A	12 ± 4	(0.63)
U.S.	1970	31	A	96 ± 27	(0.50)

Sources: Bavaria: Seuffert [23]; France: Drame et al. [26]; Prussia: Engel ([28], p. 261); U.S. wheat: Langley et al. [24]; U.S. potato: Lucier et al. [29].

3.3 Estimating the correlation length

Estimates for the correlation length in a number of cases are shown in Table 3. Two comments are in order, (i) all log-linear regressions are significant. Not surprisingly the goodness of fit is poorer for annual series. (ii) The order of magnitude of the correlation length is in agreement with what would be expected intuitively. For instance the correlation length is larger for ports than for inland markets and, in the course of 150 years there was a ten fold increase in the correlation length.

3.4 Divergence of the correlation length and phase transition

The last major famine which occurred in Western Europe (if one leaves aside the famine in Ireland during the 1840s) was in 1817. Between 1815 and 1817 the prices of wheat in Munich were multiplied by more than 4; at the same time the correlation length was multiplied by a factor 100 (Fig. 4c). For a physicist it is tempting to interpret such a divergence of the correlation length as a phase transition. However one should be cautious about such an interpretation for at least two reasons. (i) The crisis of 1817 was triggered by a major exogenous shock. In this case it was a meteorological shock: the summer of 1816 was particularly cold to the point that 1816 was referred to as “the year without summer”. It is true that in physics phase transitions are also brought about by an exogenous perturbation, but then it is a gradual increase rather than a huge jump. (ii) During a phase transition the coupling constant of the spins does not change whereas there are good reasons to think that the coupling between markets is strengthened during major price peaks, for instance through an increase in the number of traders (see in this respect [11]).

4 Price differentials

So far we have focused on price correlations; yet the behavior of traders is rather determined by the price differentials. For a trader it becomes worthwhile to ship a cargo of wheat from market A to market B when the price differential $p_B - p_A$ becomes larger than the transport cost from A to B. It is therefore of interest to examine the connection between price correlations and price differentials.

Being independent of the average price level, correlations provide a better statistical measure of interdependence, in particular they are free of any bias due to grade differences; but on the other hand it should be noticed that in contrast to correlations price differentials can be computed even from very short time series; a single year would in principle be sufficient although one usually wants to check that the spatial price pattern is stable in the course of time. For that reason price differential analysis has a wider range of applicability than correlation analysis.

4.1 Empirical evidence about price differentials

As will be seen in paper 2 [12] there are two different sections in the curve of the price differential as a function of distance. First it increases linearly with distance, then for large inter-market distances it reaches a plateau; therefore it is appropriate to examine separately the short and the large distance range.

4.1.1 Distances under 300 kilometers

Figure 5a is a typical example for short distances; a number of similar results are summarized in Tables 4a and 4b. Short distance shipments are likely to be carried out by small, local traders or even by the farmers themselves; therefore it is not surprising to observe a slope about 3 to 5 times larger in the (0.70 km) range than in the (50 km, 300 km) range.

4.1.2 Large distances

Figures 5b and 5c show the relationship between price differentials and inter-market distance in two cases. Figure 5b is particularly interesting for, as will turn out in

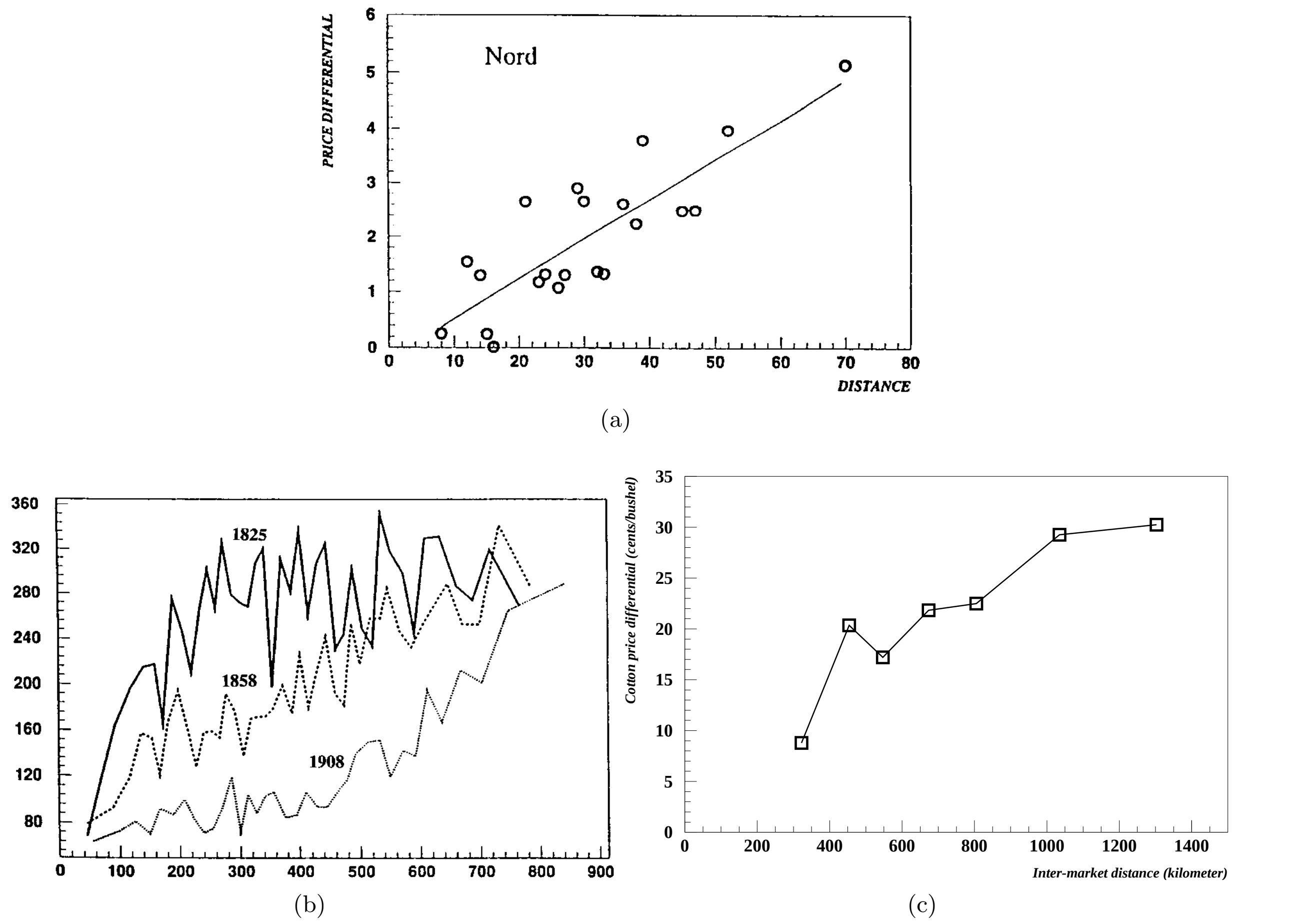


Fig. 5. (a) Price differentials within the district "Nord" (France, 1855). Horizontal scale: distance in kilometers; vertical scale: price differential expressed in French francs. The sample contains 7 markets and the prices refer to the first fortnight of January 1855. *Source: National Archives, Paris* ($F^{11*}1779-2678$). (b) Wheat price differentials as a function of distance in 19th century France. Horizontal scale: distance in kilometers; vertical scale: price differential expressed in centimes/hectoliters. The sample includes 51 markets and differentials are averaged within each distance interval over 30 market pairs. Note the change that occurred in the concavity of the curve; the 20 th century shape of the curve is not consistent with theoretical predictions. *Source: Drame et al. [26].* (c) Cotton price differentials as a function of distance in the United States (1966). The sample includes 9 markets and differentials are averaged (within each distance interval) over 5 market pairs. *Source: Cotton Statistical Bulletin 1973, 1974. US Dept. of Agriculture.*

paper 2 [12], the shape of the differential in 1908 is at variance with the theoretical prediction.

4.2 Relationship between correlation and differential

Mathematically, once a specific assumption has been made about the joint distribution of prices the correlation and the differential are related; it is important for both practical and theoretical purposes to know how this relationship reads; such results are in particular needed in [12]. To derive that relationship is basically a problem in probability calculus. Two assumptions are of particular interest: the Gaussian because of its simplicity and the log-normal because it describes fairly well the distribution of wheat prices [13]. The main steps of the calculations are summarized in Appendix A and the results are given below.

1) Under the assumption of a joint Gaussian distribution $(m_1, \sigma_1; m_2, \sigma_2; r)$ for the random variables X_1, X_2, the price differential is given by:

$$E\left(|X_1 - X_2|\right) = \sigma_Z \sqrt{\frac{2}{\pi}} \exp\left(\frac{-d^2}{2\sigma_Z^2}\right) + d \ \mathrm{erf}\left(\frac{d}{\sigma_Z\sqrt{2}}\right) \tag{3}$$

where:

$$\sigma_Z^2 = \sigma_1^2 - 2r\sigma_1\sigma_2 + \sigma_2^2; \qquad d = m_1 - m_2.$$

Special case: $m_1 = m_2$, $\sigma_1 = \sigma_2 = \sigma$, then:

$$E\left(|X_1 - X_2|\right) = \frac{2\sigma}{\sqrt{\pi}}\sqrt{1-r}. \tag{3a}$$

Table 4. (a) Variation of wheat price differentials with distance (distances under 80 km), (b) variation of wheat price differentials with distance (distances between 50 km and 300 km).

(a)

	Year	Number of markets	Slope of linear regression [French franc / hectol.x100 km]
France (Nord)	1855	7	7.2
France (Côtes-d'Armor)	1855	7	3.2

Notes: "Nord" is a district in the north of France; "Côtes-d'Armor" (formerly Côtes-du-Nord) is a district in Brittany. The French francs are pre-1914 francs, *i.e.* gold based francs. *Source: National Archives, Paris (call numbers: F11*1779-2678, F11*2877-2984).*

(b)

	Year	Number of markets	Slope of linear regression [French franc / hectol.x100 km]
Germany: Bavaria	1815	10	1.1
Germany: Bavaria	1841	10	0.74
France: Brittany	1855	31	0.81
France: Brittany	1875	31	0.46

Notes: The French francs are pre-1914 francs, *i.e.* gold based francs. The slopes are substantially lower than for the short distance case of the previous table; probably because short distance transports were carried out by local traders without the benefit of economies of scale. *Sources: Bavaria: Seuffert [23]; Brittany: National Archives, Paris (call numbers: F11*1779-2678, F11*2877-2984).*

2) Under the assumption of a joint log-normal distribution $(m_1, \sigma; m_2, \sigma; r)$ for the random variables X_1, X_2, the price differential is given by:

$$E\left(|X_1 - X_2|\right) = E(X_1)\ \mathrm{erf}\left(s\frac{\sigma}{2} + \frac{d}{2s\sigma}\right) + E(X_2)\ \mathrm{erf}\left(s\frac{\sigma}{2} - \frac{d}{2s\sigma}\right) \quad (4)$$

$$s = \sqrt{1-r}; \qquad d = m_1 - m_2.$$

From a practical point of view expression (4) is not completely satisfactory; indeed it gives the differential as a function of r whereas one would prefer an expression in terms of the correlation $\rho(X_1, X_2)$ which for log-normal variables is not equal to r. The relationship between $\rho(X_1, X_2)$ and r when (X_1, X_2) are jointly log-normal $(m_1, \sigma_1; m_2, \sigma_2; r)$ reads:

$$\rho_r(X_1, X_2) = \frac{\mathrm{e}^{r\sigma_1\sigma_2} - 1}{\sqrt{\mathrm{e}^{\sigma_1^2} - 1}\sqrt{\mathrm{e}^{\sigma_2^2} - 1}}. \quad (5)$$

Note that $\rho_1(X_1, X_2) = 1$ if $\sigma_1 = \sigma_2$ but that $\rho_1(X_1, X_2) < 1$ when $\sigma_1 \neq \sigma_2$.

The relationships between the differential and the correlation $\rho(X_1, X_2)$ are illustrated in Figures 6a and 6b.

5 Conclusion

In recent times the attention and interests of econophysicists have been more and more directed toward financial markets. See in that respect the stimulating contributions of [14–20]. There are certainly several good reasons for studying financial markets; one can mention for instance the fact that there are many accurate data now easily available in particular through the Internet. No doubt that the historical high levels reached by stock markets both in the United States and in Western Europe also contributed to the current fascination for stock markets. However one should also be aware that in many respects stock markets are complex and elusive open systems; indeed they are tightly connected to other speculative markets *e.g.* to bond markets, money markets or real estate markets. In addition stock markets are subject to few (if any) "physical" constraints and as a result most concepts that are familiar to physicists (*e.g.* inertia, relaxation time, damping, propagation) do not apply there.

In this paper we have shown that, as a field of investigation, commodity markets share many of the advantages of stock markets without having the same drawbacks. Scores of reliable data are available both for present time and for earlier centuries; commodity markets

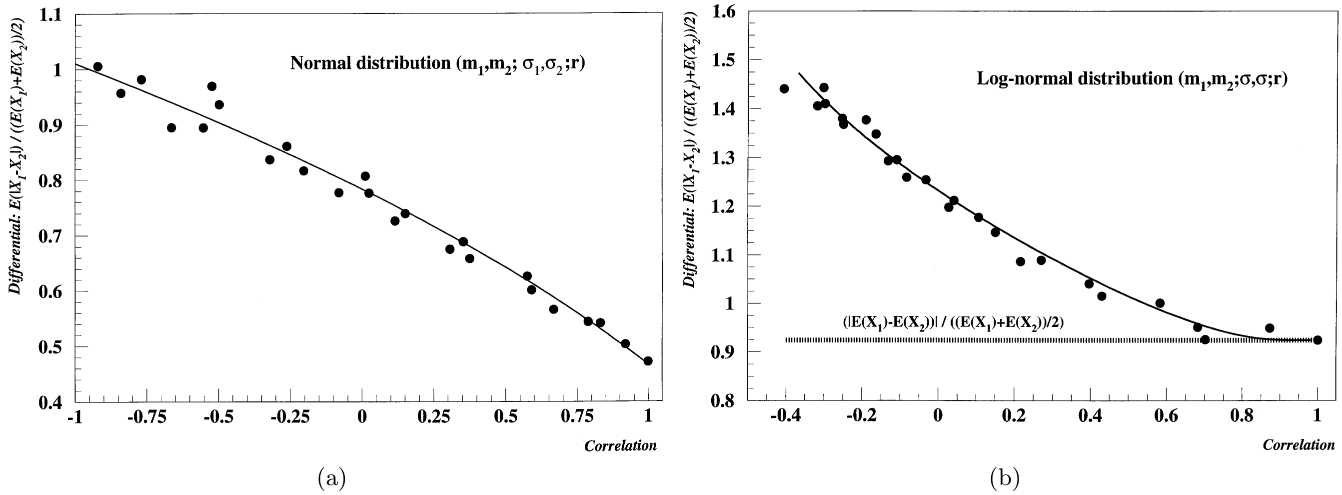


Fig. 6. (a) Relative differential for two jointly Gaussian variables as a function of their correlation r. The curve corresponds to formula (3); the dots correspond to simulated random variables. The parameters are $m_1 = 3,\ m_2 = 2,\ \sigma_1 = 2,\ \sigma_2 = 1$. (b) Relative differential for two jointly log-normal random variables $X_1,\ X_2$ as a function of their correlation $\rho(X_1, X_2)$. The curve corresponds to formula (4); the dots correspond to simulated random variables. The parameters are $m_1 = 2,\ m_2 = 3,\ \sigma = 1$. When $\rho = 1$, X_1 and X_2 are proportional and their differential reduces to the absolute value of the difference of their expectations (dotted line).

and especially spot markets have only weak connections with other speculative markets; and finally because they deal with goods that are bulky, heavy and sometimes difficult to store for a long time, commodity markets are deeply rooted in the physical space-time world.

I am indebted to Professor G. Persson for a number of stimulating discussions which provided the starting point for the writing of this paper.

Appendix A: Relationship between correlations and differentials

In this appendix we provide some details about the derivation of the formulas for differentials *versus* correlations.

A.1 Definitions

First let us recall the following definitions. Given two interdependent random variables X_1, X_2 the correlation, differential and relative differential are defined as follows.

$$\rho(X_1, X_2) = \frac{E[(X_1 - e_1)(X_2 - e_2)]}{\sqrt{[E(X_1 - e_1)^2]\,[E(X_2 - e_1)^2]}} \qquad e_i = E(X_i)$$

$$D(X_1, X_2) = E(|X_1 - X_2|),$$

$$d(X_1, X_2) = D(X_1, X_2)/\left[(e_1 + e_2)/2\right].$$

We consider two (related) cases: the Gaussian and the log-normal.

The joint density of a pair (X_1, X_2) of Gaussian random variables $(m_1, \sigma_1; m_2, \sigma_2; r)$ reads:

$$f_{\rm G}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \exp\left[-\frac{Q(x_1, x_2)}{2(1 - r)^2}\right] \qquad \text{(A.1a)}$$

$$Q(x_1, x_2) = \left(\frac{x_1 - m_1}{\sigma_1}\right)^2 - 2r\left(\frac{x_1 - m_1}{\sigma_1}\right)\left(\frac{x_2 - m_2}{\sigma_2}\right) + \left(\frac{x_2 - m_2}{\sigma_2}\right)^2 \cdot \qquad \text{(A.1b)}$$

The joint density of a pair (X_1, X_2) of log-normal random variables is such that $(\ln X_1, \ln X_2)$ form a pair of Gaussian variables; it reads:

$$f_{\ln}(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \exp\left[-\frac{Q(\ln x_1, \ln x_2)}{2(1 - r)^2}\right] \frac{Y(x_1)Y(x_2)}{x_1 x_2} \qquad \text{(A.2)}$$

where $Y(.)$ denotes the Heaviside function.

Let us also recall that the expectation and variance of a log-normal variable (m, σ) are given by:

$$E(X) = {\rm e}^{m + \sigma^2/2}, \quad {\rm Var}(X) = E^2(X)({\rm e}^{\sigma^2} - 1).$$

The joint density of the difference $Z = X_1 - X_2$ and of its absolute value $|Z|$ read respectively [21]:

$$f_Z(z) = \int_{-\infty}^{\infty} f(y + z, y)\,{\rm d}y,$$

$$f_{|Z|}(z) = \left[f_Z(z) + f_Z(-z)\right] Y(z).$$

Therefore:

$$E(|Z|) = \int_0^\infty z\mathrm{d}z \left[\int_{-\infty}^\infty (f(y+z,y) + f(y-z,y))\,\mathrm{d}y \right]. \tag{A.3}$$

A.2 The Gaussian case

Let us first suppose that $m_1 = m_2$ and $\sigma_1 = \sigma_2 = \sigma$. Then f_G is symmetric with respect to x_1, x_2 and (A.3) takes the form:

$$E(|Z|) = 2\int_0^\infty z\mathrm{d}z \int_{-\infty}^\infty f(y+z,y)\mathrm{d}y.$$

The double integral is obtained thanks to two successive variable changes, thus leading to: $E(|Z|) = (2\sigma/\sqrt{\pi})\sqrt{1-r}$. From this special case it is easy to derive the result for the general case. One uses the fact ([21], p. 222) that the difference of two Gaussian variables is also a Gaussian variable with variance equal to: $\sigma_Z^2 = \sigma_1^2 - 2r\sigma_1\sigma_2 + \sigma_2^2$. This leads to formula (3).

A.3 The log-normal case

The calculation proceeds along the same lines as in the Gaussian case. In the case $\sigma_1 \neq \sigma_2$ one would need a closed form expression for the integral:

$$\int_0^\infty \mathrm{erfc}(x)\mathrm{e}^{-a^2x^2-bx}\mathrm{d}x$$

Unfortunately only the case $b = 0$ seems to have a simple expression ([22], p. 649). That is why we restricted ourselves to the case $\sigma_1 = \sigma_2$. Note that when X_1 and X_2 are the prices of the same commodity on two different markets that assumption is quite natural. Following the same steps as before then leads to formula (4).

References

1. B.M. Roehner, *Theory of markets. Trade and space-time patterns of price fluctuations. A study in analytical economics* (Springer-Verlag, Berlin, 1995).
2. J. Steindl, *Random processes and the growth of firms* (Charles Griffin, High Wycombe, 1965).
3. L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, E. Stanley, M.H.R. Stanley, J. Phys. I France **7**, 621 and 635 (1997).
4. B.M. Roehner, Eur. Phys. J. B **8**, 151 (1999).
5. B.M. Roehner, Am. J. Agric. Econ. **78**, 339 (1996).
6. B.M. Roehner, Environm. Plann. A **21**, 289 (1989).
7. M. Baulant, J. Meuvret, *Prix des céréales extraits de la mercuriale de Paris*, S.E.V.P.E.N., Paris, 1960.
8. M.C. Lovell, Am. Econ. Rev. **76**, 110 (1986).
9. E. Fama, L. Fisher, M. Jensen, R. Roll, Int. Econ. Rev. (1969).
10. D.J. Tritton, *Physical fluid dynamics* (Van Nostrand Reinhold Company, New York, 1977).
11. J. Meuvret, Rev. Hist. Mod. Contemp. **3**, 169 (1956).
12. B.M. Roehner, Eur. Phys. J. B **13**, 187 (2000).
13. B.M. Roehner, Economies et Sociétés **11**, 5 (1990).
14. J.-P. Bouchaud, M. Potters, *Théorie des risques financiers* (Alea-Saclay, Eyrolles, Paris, 1997).
15. J.A. Feigenbaum, P.G.O. Freund, Int. J. Mod. Phys. B **10**, 3737 (1996).
16. R.N. Mantegna, H.E. Stanley, Nature **376**, 47 (1995).
17. R.N. Mantegna, H.E. Stanley, Physica A **239**, 255 (1997).
18. S.M. de Oliveira, P.M.C. de Oliveira, D. Stauffer, *Evolution, money, wars and computers*, Teubner, Stuttgart, 1999.
19. D. Sornette, A. Johansen, Physica A **245**, 411 (1997).
20. D. Stauffer, P.M.C. de Oliveira, A.T. Bernardes, Int. J. Theor. Appl. Fin. **2**, 83 (1999).
21. A. Papoulis, *Probability, random variables and stochastic processes* (McGraw-Hill Kogakusha, Tokyo, 1965).
22. I.S. Gradshteyn, I.M. Ryzhik, *Table of integrals, series and products* (Academic Press, New York, 1980).
23. G.K.L. Seuffert, *Statistik des Getreide und Viktualien Handels im Königreiche Bayern mit Berücksichtigung des Auslandes* (Weisz, Munchen, 1857).
24. J. Langley, S. Langley, *State-level wheat statistics 1949-1988*, U.S. Dept. of Agriculture, Statistical Bulletin No. 779, (1989).
25. A. Jacobs, H. Richter, Vierteljahrshefte zur Statistik des Deutschen Reichs **44**, 273 (1935).
26. S. Drame, C. Gonfalone, J.A. Miller, B. Roehner, *Un siècle de commerce du blé en France 1825-1913*, *Les fluctuations du champ des prix* (Economica, Paris, 1991).
27. *Das Getreide im Weltverkehr*, Austrian Dept. of Agriculture, Vienna 1900.
28. E. Engel, Zeitschrift des Königl, Preussischen Statistischen Bureaus **10**, 249 (1861).
29. G. Lucier, A. Budge, C. Plummer, C. Spurgeon, *US potato statistics*, U.S. Dept. of Agriculture, Statistical Bulletin **829** (1991).