Eur. Phys. J. B **13**, 189–200 (2000)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

The correlation length of commodity markets 2. Theoretical framework

B.M. Roehner[a]

LPTHE, University Paris 7, 2 place Jussieu, 75005 Paris, France

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

Abstract. The second of this series of two papers is devoted to a theoretical analysis of spatial interaction between commodity markets. The theoretical framework that we present is referred to as the stochastic spatial arbitrage model (SSAM); it accounts for most of the empirical regularities observed in the first paper. Two basic mechanisms are found to be responsible for spatial inter-market interaction, namely (i) spatial arbitrage and hedging conducted by traders, (ii) spatial correlation between local shocks; the latter is favored by a similar economic and cultural environment. The SSAM includes both effects and offers a wide range of predictions about price volatility, trade, price correlations, price differentials. Statistical tests display a convergent array of evidence in favor of the model. However several predictions cannot be tested by lack of statistical evidence, a circumstance which shows that yet additional "experimental" work is required.

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

1 Introduction

In this series of two papers we examine spatial interactions between markets. The empirical evidence was presented in the first paper, thereafter referred to as article 1 [1]. We observed that three mechanisms seem to be responsible for the correlation between distant markets, (i) spatial arbitrage conducted by traders, (ii) the fact that local shocks are spatially correlated, (iii) exchange of information which tends to synchronize the reactions of market makers. Empirical observation has shown that for spot markets (as opposed to futures markets) the two first effects are predominant. As a result those markets are characterized by well defined spatial correlation patterns. Thus, as in fluid dynamics, a correlation length l was defined through the expression $\rho(d) = \mathrm{e}^{-d/100l}$ where $\rho(d)$ denotes the correlation between two markets which are a distance d apart. In paper I estimates for l have been obtained both for 19th century and 20th century markets. Price differentials also exhibit striking regularities.

In the present paper, starting from a reasonable assumption about the behavior of traders, we set up a theoretical framework which not only explains the above mentioned regularities but also offers several predictions about seemingly unrelated phenomena such as volume of trade or price volatility. The broad range of its predictions is a distinctive feature of the SSAM.

The paper proceeds as follows. In the next section we investigate in detail the two market-case. In particular we examine the relationship between correlation, volatility and trade with respect to transport cost; we also discuss the impact of possible correlation between local shocks. The two-market case serves as a prototype of more complicated cases. Of particular interest is the case of an infinite chain of markets. In Section 3 we confront the theoretical predictions with empirical evidence.

2 The stochastic spatial arbitrage model (SSAM)

Microeconomic spatial arbitrage models were popular in the 1960s and 1970s. A bibliographical discussion can be found in [2,3] (or in [4], Chap. 3). These models had two drawbacks however: they were deterministic and nonlinear. Nonlinearity was (as shown below) an unnecessary complication which prohibited analytical insight; even numerical solution turned out to be laborious (see in this respect [5]). Furthermore the deterministic character of these models limited their ability to make contact with statistical evidence; most often they were tested on simulated rather than on actual data. There were (at least) two reasons for this shortcoming, (i) in a deterministic framework notions such as price correlations or volatility which are so important statistically simply cannot be defined. (ii) A number of crucial variables such as transport costs, excess-supply or trade are highly fluctuating.

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

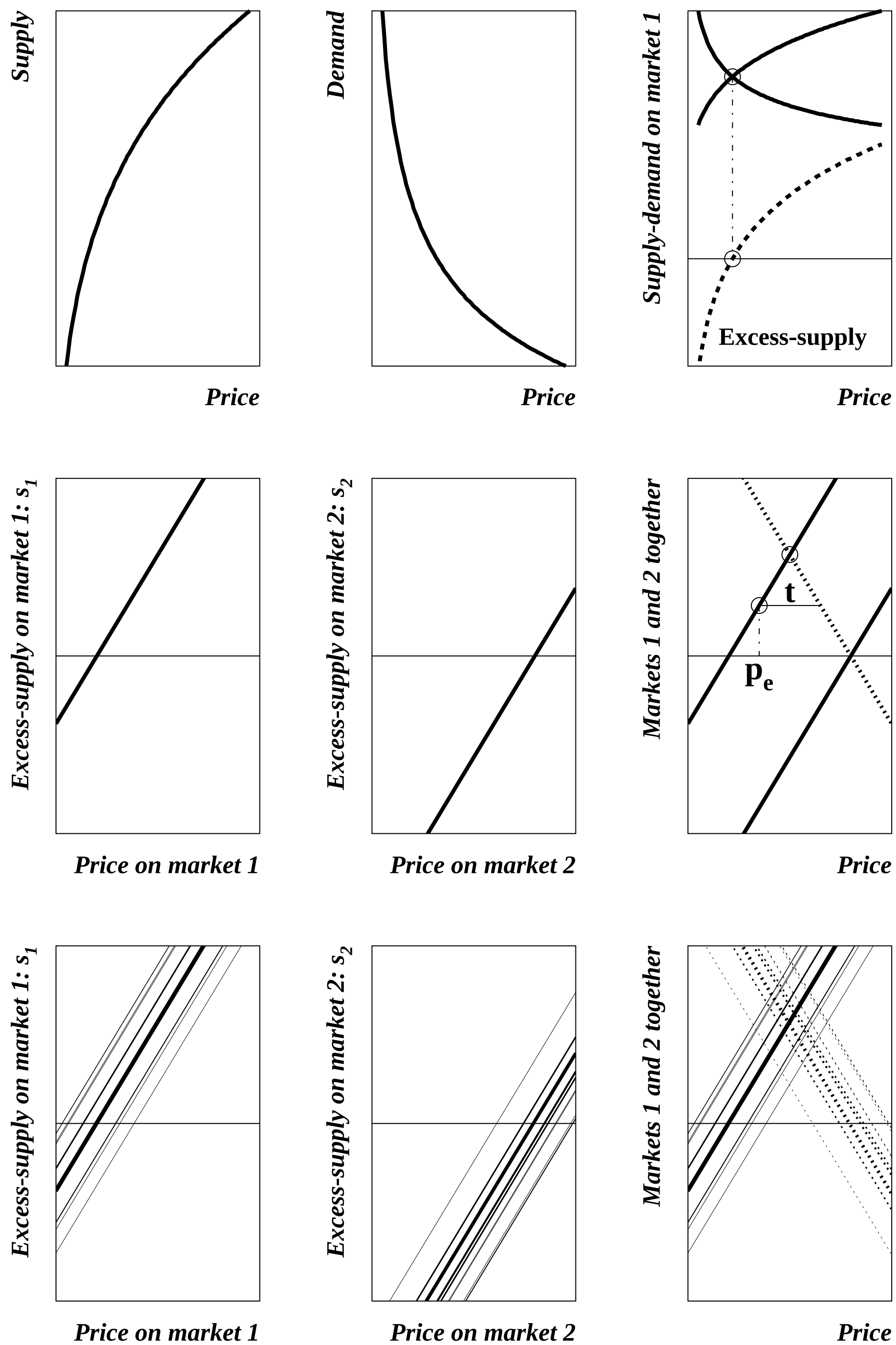


Fig. 1. Mechanism of spatial arbitrage on the example of two markets. The determination of the equilibrium price on one market is shown in the graphics of the first line. Here we have used fairly realistic power-like supply and demand functions; in subsequent graphics for the sake of simplicity we use linear functions.The determination of the equilibrium price on a set of two markets is shown in the second line. If transportation costs are assumed negligible (as is the case for stock market shares for instance) then the equilibrium condition is very much the same as on a single market, the only difference being that supply and demand functions are replaced by the excess-supply functions on each market. The equilibrium price is defined by the intersection of s_1 (solid line) and $-s_2$ (dotted line) representing the condition $s_1 + s_2 = 0$. When transport cost is equal to t the equilibrium price is p_e. The last line shows the same mechanism in a stochastic framework; the equilibrium prices on each market are then considered as random variables. As a result the equilibrium price p_e becomes a random variable too.

2.1 The two-market case

Historically the two-market case was important because it was the only case for which the nonlinear deterministic model could be solved graphically in a straightforward way. For other cases one had to rely on simulation [6] or on linear programming [7].

2.1.1 The graphical solution

Although the graphical solution is of little help in dealing with more complicated cases it gives an intuitive understanding of the mechanism of spatial arbitrage. That solution is summarized in Figure 1. The first three graphics concern a single market. If $S = S(p)$ and $D = D(p)$ denote the supply and demand functions, the equilibrium condition reads: $s(p) = S(p) - D(p) = 0$; $s(p)$ is referred to as the excess-supply function. The graphics in the second line concern the two-market case. Suppose that the equilibrium prices on market 1 and 2 (without interaction) are $\overline{p}_1$ and $\overline{p}_2$ respectively. When the markets are allowed to interact there will be a specific amount of goods transfered from one market to the other; this amount depends upon the transport cost t. If $t = 0$ the new equilibrium condition is given by $s_1 + s_2 = 0$. In spite of its simplicity this case is of interest because it describes the situation of all goods for which transportation costs are a small percentage of the price, *e.g.* shares, gold and silver, diamonds, patents, softwares, etc.

When the transportation cost is not negligible the new equilibrium is ruled by the condition:

$$p_1 - p_2 = t\psi(s) \tag{1.1}$$

where $\psi(s) = Y(s) - Y(-s)$ and $Y(s)$ denotes the Heaviside function ($\psi(s)$ represents the "sign function"). The interaction described by equation (1.1) is of a discontinuous nature: no exchange takes place until the price differential becomes equal to t. It is precisely that discontinuity which makes the problem mathematically difficult. Two simplifications are possible. The function $\psi(s)$ can be replaced by a continuous approximation; for example the function $\psi_c(s) = \tanh(cs)$ provides an approximation which can be made arbitrarily close to $\psi(s)$ when c is large enough. A second more drastic approximation is to take for $\psi_c(s)$ a linear function. However crude, that approximation leads to sensible results. As a matter of fact a comparison between the linear and the nonlinear solutions reveals that their difference is less than 20 percent ([4], p. 101).

2.1.2 Solution of the two-market case

If we introduce the supply and demand functions in the form:

$$S_i(p_i) = \gamma p_i - c_i, \quad D_i(p_i) = -\beta p_i + b_i, \quad i = 1, 2 \tag{2.1}$$

then the excess-supply function can be written:

$$s_i = S_i - D_i = a(p_i - \overline{p}_i) \tag{2.2}$$

where: $a = \gamma + \beta$, $\overline{p}_i = (c_i + b_i)/(\gamma + \beta)$, $i = 1, 2$; the $\overline{p}_i$ denote the equilibrium prices for each market when the interaction is turned off. As shown in the last three graphics of Figure 1, in this model the $\overline{p}_i$ are considered as being random variables, whether independent or dependent.

With this notations the equations corresponding to the graphical solution outlined in the previous paragraph are (see Appendix A):

$$\begin{cases} p_1(1+\theta) - p_2 = \theta\overline{p}_1 \\ p_2 - p_2(1+\theta) = \theta\overline{p}_2 \end{cases} . \tag{2.3}$$

The parameter $\theta = at$ is the product of the slope of the excess-supply function by the transportation cost per unit of weight between the two markets. If we denote by $[C]$ a given currency, and by $[M]$ the unit of mass, the dimension of a is $[M]/[C]$ while the dimension of t is $[C]/[M]$; θ is therefore a dimensionless parameter; it plays a crucial role in the SSAM.

From (2.3) it is easy to derive the following solutions:

$$p_1 = \frac{(1+\theta)\overline{p}_1 + \overline{p}_2}{2+\theta}, \qquad p_2 = \frac{\overline{p}_1 + (1+\theta)\overline{p}_2}{2+\theta} \cdot \tag{2.4a}$$

Furthermore trade between the two markets is defined by:

$$s = \frac{(\overline{p}_2 - \overline{p}_1)a}{2+\theta} \cdot \tag{2.4b}$$

In the next paragraph we examine the implications of (2.4) for price correlations, price differentials, price volatility and trade.

2.1.3 Uncorrelated local shocks

First we suppose that the local shocks $\overline{p}_1$, $\overline{p}_2$ are uncorrelated and that they have identical distributions. Mathematically the basic variables, are the covariances:

$$c_p(i,j) = E(p'_i, p'_j), \qquad i, j = 1, 2$$

where the p'_i denote the centered variables (*i.e.* $E(p'_i) = 0$). For instance the variance is given by: $\sigma_p^2 = c_p(1,1) = c_p(2,2)$ and the correlation by: $\rho_{p_1p_2} = c_p(1,2)/\sqrt{c_p(1,1)c_p(2,2)}$. Now the centered variables can be written in the form:

$$p'_1 = A\overline{p}'_1 + B\overline{p}'_2; \qquad p'_2 = B\overline{p}'_1 + A\overline{p}'_2.$$

Thus one has:

$$\sigma^2_{p'_1} = \sigma^2_{p'_2} = \sigma(A^2 + B^2), \qquad E(p'_1, p'_2) = 2AB\sigma^2$$

where σ^2 denotes the variance of the local shocks. Thus:

$$\rho_{p_1p_2} = 2\frac{1+\theta}{2+2\theta+\theta^2}, \qquad \sigma_p^2 = \sigma^2\frac{2+2\theta+\theta^2}{(2+\theta)^2} \cdot \tag{2.5}$$

In order to derive the price differential from the correlation one has to made a specific assumption regarding the distribution of prices (and therefore the distribution of the shocks). Two cases are of special interest, the Gaussian and the log-normal. The Gaussian hypothesis has the advantage of mathematical simplicity and in this case the shocks are (as linear combinations) Gaussian too. The log-normal hypothesis is more realistic for it is well known that commodity prices are fairly well described by a log-normal distribution[1]. In both cases the formulas between correlations and differentials are given in article 1 [1]. The two assumptions can in a sense be combined by supposing that the p_i are in fact logarithms of prices instead of prices; then of course the arbitrage assumption takes a multiplicative rather than an additive form. In what follows we mainly restrict ourselves to the Gaussian case.

Using the result given in article 1 [1] one obtains for the price differential:

$$E\left(|p_1 - p_2|\right) = \frac{\sigma}{\sqrt{\pi}} \frac{\theta}{1+\theta/2}.$$

As a natural definition of trade one can take: $T = (1/2)E(|s|)$, which leads to: $T = (\sigma/\sqrt{\pi})[a/(\theta+2)]$. The dependence of the correlation and differential with respect to θ is shown in Figure 2.

Two limiting cases are of particular interest, the well integrated market ($\theta \to 0$) and the segmented market ($\theta \to \infty$) :

Vanishing transport cost: $\theta \to 0$

$$r_{p_1 p_2} \sim 1 - \theta^2/2 \to 1$$
$$\sigma_p \to \sigma/\sqrt{2}$$
$$E(|p_1 - p_2|) \sim (\sigma/\sqrt{\pi})\theta \to 0$$
$$T \to (\sigma/2\sqrt{\pi})a$$

Large transport cost: $\theta \to \infty$

$$r_{p_1 p_2} \sim 2/\theta \to 0$$
$$\sigma_p \to \sigma$$
$$E(|p_1 - p_2|) \sim (2\sigma/\sqrt{\pi})(1 - 2/\theta) \to 2\sigma/\sqrt{\pi}$$
$$T \to 0.$$

2.1.4 Correlated local shocks

As already observed in article 1 [1], a substantial part of the synchronization between two markets is due to similar responses to shared shocks. This idea can be illustrated on

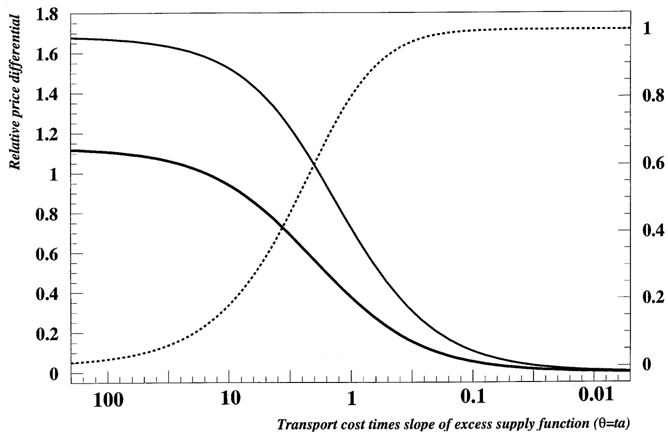


Fig. 2. Price differential and price correlation for a system of two markets. The horizontal scale represents the dimensionless parameter $\theta = at$ where a is the slope of the excess-supply function and t the transport cost; the axis is labeled from right to left in analogy with the historical evolution of transport rates; thus the left-hand side of the axis corresponds to the distant past. Thick line: price differential under the assumption of a joint Gaussian price distribution; thin line: price differential under the assumption of a joint log-normal distribution; the shape of the curve is qualitatively the same in the Gaussian or in the log-normal case. In the rest of the paper we restrict ourselves to the Gaussian case for the sake of simplicity. Broken line: price correlation (vertical right-hand scale); note that, in contrast to the price differential, the correlation does not depend upon a specific assumption about the distribution of prices; it is a "cleaner" measure of interdependence than the differential.

the two-market case. Suppose that there is a correlation r between the random variables $\overline{p}_1$ and $\overline{p}_2$.

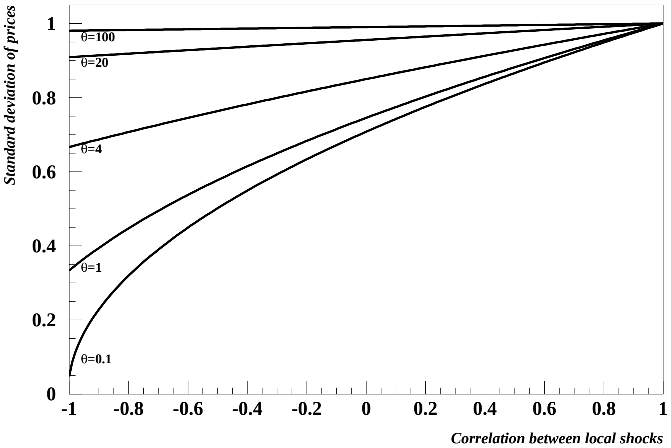
The above formulas then become:

$$r_{p_1 p_2} = \frac{r + \lambda(\theta)}{1 + r\lambda(\theta)} \qquad \text{where} \qquad \lambda(\theta) = 2\frac{1+\theta}{2 + 2\theta + \theta^2}$$
$$\sigma_p^2 = \sigma^2 \frac{2(1+r)(1+\theta) + \theta^2}{(2+\theta)^2}$$
$$E(|p_1 - p_2|) = \frac{1}{\sqrt{\pi}} \frac{\theta}{1+\theta/2} \sigma\sqrt{1-r}$$
$$T = \frac{a}{\sqrt{\pi}(\theta+2)} \sigma\sqrt{1-r}.$$

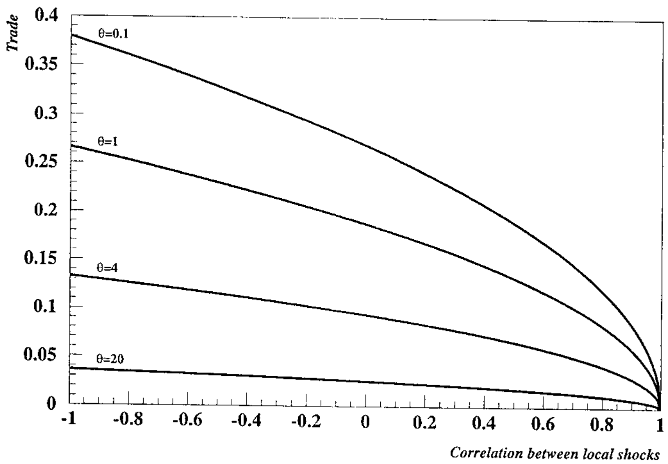
Figures 3a and 3b show σ_p and T as a function of the correlation r between local shocks.

As before one can easily derive the limiting behavior of $r, \sigma_p, E(|p_1 - p_2|)$ and T in the cases of a perfectly integrated market and of a completely segmented market. The behavior of σ_p is of particular interest; for a segmented market one has of course: $\sigma_p = \sigma$, while for a well integrated market one obtains: $\sigma_p = \sigma\sqrt{(1+r)/2}$. Thus, in the latter case the more the shocks are correlated the larger the price volatility; the smallest volatility obtains when $r = -1$, *i.e.* when the two markets are completely counter-cyclical. Intuitively this makes sense. For instance if the wheat crops in the United States and in Europe suffer from bad weather conditions in the same years

[1] For wheat prices see [8]; even for stock prices [9] shows that the log-normal distribution provides a reasonable approximation (although actual tails are "fatter" than predicted by the log-normal. By and large the following "rules of tumb" apply: for very small samples ($\lesssim 30$) the normal distribution can be used; for medium-size samples ($\lesssim 500$), the log-normal distribution provides a good approximation; for large ($\gtrsim 1000$) and very large ($\gtrsim 10\,000$) samples, the role of rare outliers (especially high price outliers due to speculative bubbles) becomes important; various specific "candidates" have been proposed for those cases in recent times.



(a)



(b)

Fig. 3. (a) Standard deviation of prices for two markets as a function of the correlation between local shocks. When $r = -1$ the standard deviation is lowest; it increases all the more quickly if θ is smaller. In other words a market which is well integrated both in terms of transportation facilities and in terms of synchronization between business fluctuations is expected to have a higher standard deviation of prices than a market well integrated in terms of transport costs but where business fluctuations are not synchronous. (b) Trade between two markets as a function of the correlation r between local shocks. The fact that trade increases when transportation cost (θ) decreases could seem obvious. It is not however; for as θ becomes smaller the price differential diminishes too and with it the incentive for traders to trade; in other words for trade to increase the differential has to fall off slower than θ, which is indeed the case.

the shocks cannot be smoothened in either of the continents by importing from or exporting to the other one.

2.2 Chain of markets

While providing an interesting insight the two-market case cannot give any information about correlation lengths. The simplest case for which a correlation length can be computed is an infinite one-dimensional chain of markets whose markets are supposed to be uniformly spaced.

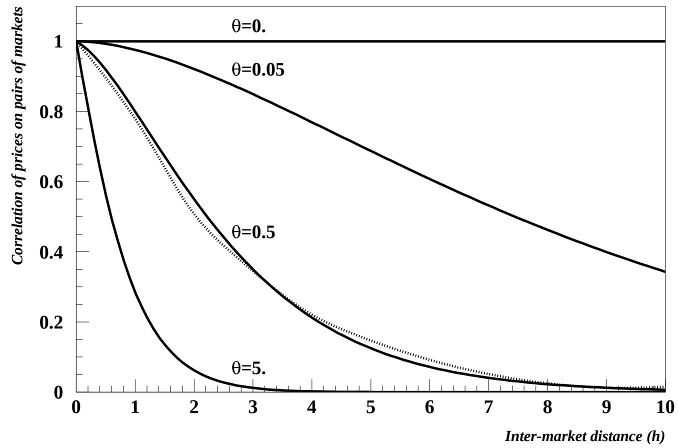


Fig. 4. Correlation of prices as a function of inter-market distance in an infinite, linear chain. Horizontal scale: number of intervals between pairs of markets (they are supposed to be uniformly spaced). The dotted curve corresponds to the case of a two-dimensional network of markets for $\theta = 0.5$; as can be seen it is very close to the curve for the one-dimensional case.

Mathematically it is the covariance function of the prices, namely:

$$c_p(h) = E(p'_{k+h}, p'_k)$$

which is the key to the problem. All other variables, and in particular the correlation length can be easily derived from $c_p(h)$. The calculation which leads to the expression of $c_p(h)$ is detailed in Appendix B. One obtains:

$$c_p(h) = \frac{\cosh\gamma \; \mathrm{e}^{-\gamma h}(1 + h\tanh\gamma)}{4\theta^2 \sinh^3\gamma} \tag{2.6a}$$

where γ is defined by the relation: $\sinh\gamma = \sqrt{\theta + \theta^2/4}$. The correlation $r_p(h) = c_p(h)/c_p(0)$ has the simpler expression:

$$r_p(h) = \mathrm{e}^{-\gamma h}(1 + h\tanh\gamma). \tag{2.6b}$$

Figure 4 gives a global view of the behavior of $r_p(h)$ with respect both to h and to θ. For a well integrated set of markets the correlation length falls off slowly with distance. Figure 4 also suggests that the properties of a two-dimensional network of markets are very similar to that of the one-dimensional chain.

From (2.6b) we can derive the correlation length. As can be remembered from article 1 one can use two alternative definition of the correlation length. If we use the definition: $r_p(h) = \mathrm{e}^{-h/L}$ one has to develop expression (2.6b) to first order with respect to the inter-market distance h. One gets in this way:

$$L = 1/(\gamma - \tanh\gamma) \underset{\gamma\to 0}{\sim} \frac{3}{\gamma^3}\cdot$$

The second option is to use the definition $\delta = \int_0^\infty r_p(h)\mathrm{d}h$ This definition relies on the behavior of $r_p(h)$ in the whole region $[0, \infty[$; it leads to:

$$\delta = \frac{\gamma + \tanh\gamma}{\gamma^2} \underset{\gamma\to 0}{\sim} \frac{2}{\gamma}\cdot$$

As we already pointed out in article 1 [1] it is the first option which is usually best suited for statistical adjustment. Thereafter we restrict ourselves to that option.

3 Comparison of the SSAM's predictions to statistical evidence

The model offers a fairly wide set of predictions. As it turns out the predictions which can be tested with acceptable accuracy are those about the spatial behavior of price correlations or price differentials. In the first paragraph we explain why the other predictions are more difficult to test. In the second paragraph we present the results of a number of tests.

3.1 Qualitative tests

Table 1 summarizes a total of 11 effects all of which could in principle be tested if adequate data were available. Unfortunately such is not the case; let us briefly explain why.

In Table 1 the dependent variable are either trade or specific functions (standard deviation, correlation or price differential) of the prices. Many long price records are available for commodities such as wheat, sugar, metals, etc.; data about international trade in such commodities have also been recorded in all industrial countries. In fact statistics about overseas trade were among the first to be collected; in Britain for instance exports and imports figures are available since 1700 ([10], p. 285). The situation is much less favorable however for the independent variables as shown by the following discussion.

Although a (slope of the excess-supply function) and t (transport cost) come into the model mainly through their product $\theta = at$ one needs separate information for each variable in order to control that a change in transport cost is not offset by a change in a. Of these two variables it is of course a which is the most difficult to estimate. If we interpret both s and p_i as being the logarithms of the corresponding items a is the elasticity of excess-supply with respect to price, that is to say the difference between the elasticity of supply and the elasticity of demand. Now elasticities are fairly fluctuating quantities which are difficult to estimate. For wheat for instance, Schultz [11] obtained estimates for the elasticity of demand ranging from 0.02 to 0.21 for the period 1880-1934. Although transport costs are somewhat easier to estimate one should bear in mind that they are extremely dependent upon business fluctuations; doubling of freight rates within a few years is not uncommon. To make things even worse actual freight rates data often are withheld by firms and are not made public (see in this respect [4], p. 80). In short it is difficult to asses the level of the variable θ even within a 100% margin.

The variance σ of local shocks or their correlation r are even more difficult to measure. True, some meteorological factors are well known. For instance it has been estimated [12] that the correlation length of rainfalls is of the order of 10 km that is to say at least five times smaller than the correlation length of wheat prices. But rainfalls are only one of the many factors that can affect the demand and supply of wheat.

Is there no way around such obstacles? One possibility is to select commodities for which one has some kind of special knowledge about the independent variables. Let us give some examples. (i) For gold, silver, platinum or diamonds it is obvious that transport costs represent only a negligible fraction of their value. (ii) The price elasticity of bananas is about 10 times larger than the price elasticity of sugar: 0.37 against 0.04 ([13], p. 394). Taking into account such orders of magnitude allows at least qualitative comparisons.

As an application we consider agricultural commodities versus mineral commodities. In the wake of the transportation revolution there has been a marked decrease in the volatility of agricultural commodities in striking contrast with what happened for minerals. How could this difference be accounted for? We argue that there are two competing effects. On one side the decrease in transport rates tended to smoothen out price fluctuations; on the other side, at least as far as non-meteorological factors are concerned, the growing globalization of the world economy made local supply/demand shocks to become more correlated. One crucial difference between agricultural and mineral materials is precisely the fact that the latter do *not* depend upon meteorological factors. This reasoning is illustrated schematically in Figure 5 by a comparison between the volatility of gold and of wheat.

Gold prices are known to be more volatile than wheat prices; everyone remembers the huge price peaks that occurred in 1968 and 1979 (for more details see Ref. [13], p. 395). Wheat in contrast has had in the 20th century a fairly low volatility. Yet, according to equation (2.5) one would expect the opposite to be true. Indeed, under the assumption that the other parameters are the same, a commodity characterized by a low transport cost should have a low volatility too. The solution of this paradox has to be found in the fact that the other parameters, namely σ and r, are *not* the same. Little can be said about σ, but there are good reasons to believe that r is larger for the gold market than for the wheat market. Remember that the production of gold is concentrated in a few countries and is in the hands of a small number of companies. As shown in Figure 5 the impact of a larger r can be more important than the transport cost effect.

3.2 Quantitative tests

There are only two cases in Table 1 where both the dependent and the independent variable can be estimated with reasonable accuracy, namely cases 2c and 3c for correlation and price differentials as a function of inter-market distance. These are the two tests that we discuss now.

3.2.1 Correlation length

As we have seen in article 1 (Figs. 4a, 4b) the price correlation is usually fairly close to one. As a result it can be

Table 1. Summary of the predictions of the stochastic spatial arbitrage model (SSAM).

	Dependent variable	Independent variable	Phenomenon
1a	Standard deviation (σ_p)	$\theta = at$	Decrease of volatility when transport costs become smaller or when the commodity becomes less elastic
1b		r	Increase of volatility with correlation of local shocks
1c		σ	Increase of volatility with standard deviation of local shocks
2a	Correlation of prices (r_p)	$\theta = at$	Increase of correlation when transport costs become smaller or when the commodity becomes less elastic
2b		r	Increase of correlation when correlation of local shocks grows
2c		h	Decrease of correlation for larger inter-market distances
3a	Price differential (D_p)	$\theta = at$	Decrease of price differentials when transport costs become smaller or when the commodity becomes less elastic
3b		r	Decrease of differentials when correlation of local shocks grows
3c		h	Increase of differentials for larger inter-market distances
4a	Trade (T)	$\theta = at$	Increase of trade when transport costs become smaller or when the commodity becomes less elastic
4b		r	Decrease of trade when correlation of local shocks grows

developed to first order:

$$r_p(h) = 1 - h\gamma^3/3 + O(h^2) = 1 - d/(100l) + O((d/100l)^2).$$

Remember that h is the number of inter-market *intervals*; thus, in order to make contact between the two expressions one has to introduce the average spacing, Δ, between markets; then:

$$\frac{d}{100l} = \frac{d/\Delta}{100\ l/\Delta} = \frac{h}{100\ l/\Delta}.$$

Taking into account that: $\theta \underset{\theta\to 0}{\sim} \gamma^2$ one obtains:

$$\theta \simeq \left(\frac{1}{33\ (l/\Delta)}\right)^{2/3}.$$

As a preliminary step one has therefore to estimate the average inter-market spacing. There are two different methods. (i) In the first method we assume that the (real) markets are more or less uniformly spaced; then knowing the total number of markets and the area of the region it is a simple matter to derive the distance between neighboring markets. Let us see how the procedure works on two examples. First we consider 19th century France. In the middle of the 19th century wheat prices were recorded in about 500 markets, which means that a wheat market was held (once or twice a week) in all towns of some importance; more precisely the figure of 500 markets corresponds to all towns with a population over 5000 (Census of 1851, Annuaire Statistique de la France 1960). Now if we schematize the territory of France by a square with a 700 km side, one gets a spacing Δ of 32 km. Next we consider the United States; in this case we used state level price data. But since there are a number of states with a fairly large area and almost no wheat production we left aside Alaska, Colorado, Nevada, New Mexico and Utah. This

Table 2. Correlation length for commodity markets: fit of parameter θ.

Commodity	Region	Year	Number of markets	Periodicity	Fit of parameter θ $[\times 10^{-2}]$
Wheat	Bavaria	1829	9	M	13 ± 7
Wheat	France (center)	1873	11	M	9.3 ± 1.6
Wheat	France	1873	12	M	7.9 ± 1.4
Wheat	France (ports)	1873	10	M	3.8 ± 0.4
Wheat	United States	1970	31	A	4.6 ± 0.2
Potato	Prussia	1837	10	A	20.8 ± 2.2
Potato	United States	1970	35	A	14.0 ± 0.5

Notes: This table parallels Table 3 of article 1. The dimensionless parameter θ is the model's most important parameter; it is equal to the product at of the slope of excess-supply by the transport cost. The fit of θ is based on the assumption of an average inter-market distance of 50 km for nineteenth century wheat markets and of 170 km for twentieth century wheat markets. The dates refer to the middle of the interval used in the calculation of the correlation.

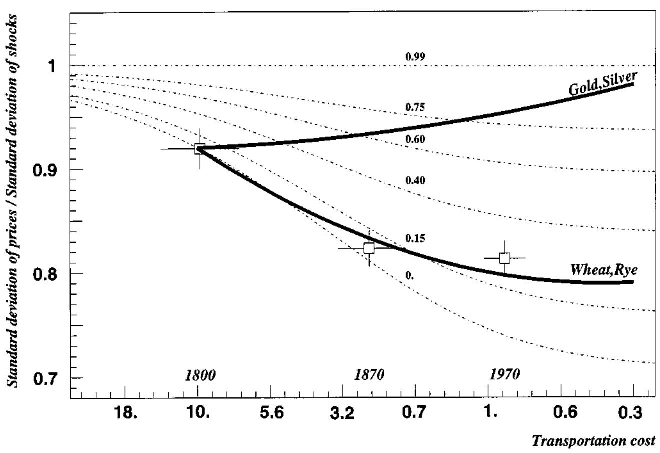


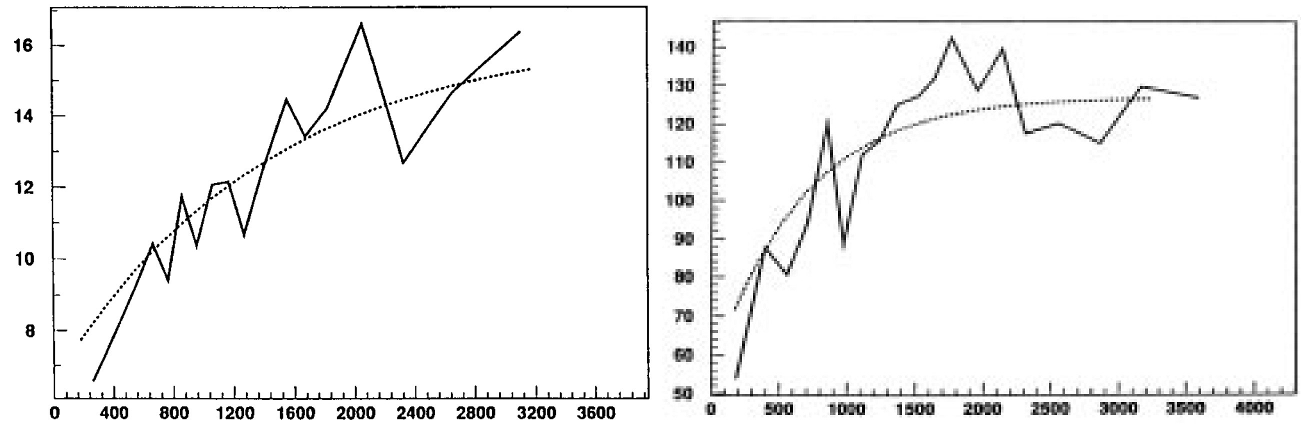
Fig. 5. Price volatility with respect to transport cost. The horizontal scale represents the dimensionless parameter $\theta = at$ where a is the slope of the excess-supply function and t the transport cost; the axis is labeled from right to left in analogy with the historical evolution of transport rates; three dates 1800, 1870 and 1970 provide approximate landmarks. The dotted curves are theoretical predictions, each for a different correlation (ranging from 0.0 to 0.99) between local shocks; the downward shape of these curves show the stabilization effect decreasing transport rates have on price fluctuations. The solid curves schematize possible evolution paths either for agricultural and for mineral materials. The squares represent average empirical estimates for grain (wheat and rye). Gold and silver provide remarkable examples of materials which, in spite of negligible transport costs (with respect to value) display huge price fluctuations; remember in this respect the twenty-fold increase of silver prices in the late 1970s.

leads to: $\Delta \sim 390$ km. Such a figure can only be a rough average of course because in this case the markets are not uniformly spaced; western states, for instance, are much larger than New England states. (ii) By fitting the *whole* price differential function one is able to estimate more than one parameter. This method which is implemented in the next paragraph gives estimates for Δ which are consistent with the orders of magnitude obtained through the first method; one gets $\Delta = 43$ km for France, and $\Delta = 170$ km for the United States.

In Table 2 we consider a number of cases which parallel those examined in Table 3 of article 1 [1]. The goodness of fit of these least-square fits can be judged on the basis of two criteria. Firstly error bars; they are on average of the order of 10%; secondly the estimates can be confronted with what is known from the general evolution of transport costs; thus one expects the following results. (i) For France (center) and France (whole country) one gets close estimates; this makes sense since both a and t should indeed be similar, (ii) for France (ports) θ is substantially smaller than in the two previous case; this again makes sense since a is the same either for ports or for other markets whereas t is likely to be smaller for ports. (iii) For Bavaria (1829) θ is larger than for France (1875); this is sensible since it is reasonable to assume that a did not change much before the end of the 19th century when wheat became a less essential factor in the people's diet; θ then is proportional to t which, as we know, decreased dramatically between 1829 and 1875. Cross comparisons between results for the 19th century and for the 20th century are more difficult because we do not know how a changed.

3.2.2 Price differentials

As far as price differentials are concerned there are two possible options, (i) if data are available only for a few markets (say less than 20) it would be unrealistic to attempt a nonlinear fit of the whole curve. In that case we restrict ourselves to a linear fit of the small-distance section of the curve. (ii) If the sample of markets is large



(a) (b)

Fig. 6. (a) Wheat price differentials as a function of inter-market distance in the United States 1951-1981. Solid line: observations; dashed line: prediction of the stochastic spatial arbitrage model (SSAM). Horizontal scale: distance in kilometers. Vertical scale: differentials in cent/bushel. The sample includes 34 states; in each distance interval differentials are averaged over 30 market pairs. The solid curve is an average over the years 1951, 1961, 1971, 1981. The theoretical curve corresponds to the parameters given in Table 3b (case 4) with σ expressed in cent/bushel. *Source: Langley et al.* [16]. (b) Potato price differential as a function of inter-market distance in the United States 1951-1981. Solid line: observations; dashed line: prediction of the stochastic spatial arbitrage model (SSAM). Horizontal scale: distance in kilometers. Vertical scale: differentials in cent/centweight. The sample includes 35 states; in each distance interval differentials are averaged over 30 market pairs. The solid curve is an average over the years 1951, 1961, 1971, 1981. The theoretical curve corresponds to the parameters given in Table 3b (case 5) with σ expressed in cent/centweight. *Source: Lucier et al.* [17].

enough (say more than 30) a nonlinear fit of the whole curve becomes possible; it provides estimates for all the parameters entering into the SSAM.

Linear fit for small distances

If v denotes the slope of the regression between $\Delta^2(|p_i - p_j|)$ against inter-market distances then, as is easy to see from the formulas for the chain of markets, the parameter θ can be estimated as:

$$\theta = \left(\frac{3\pi}{4} v\Delta\right)^{2/3} (1/\sigma_p)^{4/3}.$$

Δ again denotes the average spacing between markets. We use here the same estimates as in the previous paragraph. If a sufficiently long price series is available one can of course estimate σ_p directly. Alternatively one can use the following reasoning: it is known [8] that for nineteenth century wheat price series the coefficient of variation σ_p/m is approximately equal to 0.3. Thus σ_p can merely be derived from the average price m.

Remark

The previous procedure can be applied either to the differences of the prices or to the differences of the logarithms of the prices; theoretical arguments can be given in favor of each of the options; both options have been tried and they lead to similar results.

Are the estimates in Table 3a consistent one with another? Estimates for case 2 and especially for case 3 are substantially higher than the others. This is not surprising, however, for these estimates concern short distances for which shipping was probably carried out by small traders or even by the farmers themselves. These shippings did not benefit from the economies of scale that apply to the transportation of large quantities on longer distances.

Nonlinear fit

A comparison between observations and the fitted theoretical curve is provided in Figures 6a and 6b. The estimated values of the three parameters are given in Table 3b. We observe that the estimates for θ are in agreement with those obtained previously and with what intuition would suggest.

A special comment is in order for France 1908 (Fig. 5 of [1]). This curve has an overall shape that is drastically different from the one predicted theoretically. The concavity is downward instead of being upward; in other words the theoretical curve is of the $y = \sqrt{x}$ type while the 1908 curve is of the $y = x^2$ type. Note that this is not an isolated case but rather the result of a steady evolution that goes back to the middle of the nineteenth century: in 1825 the second derivate is negative (as predicted), in 1858 it is almost equal to zero, and in 1908 it is positive. The specific reason of that discrepancy remains an open question. It would of course be easy to list several effects which the model did not take into account; however it would be more helpful to know about other cases where the same shape is observed.

Table 3. (a) Price differentials for commodity markets: fit of parameter θ, (b) estimation of the SSAM on price differentials as a function of inter-market distance.

(a)

	Commodity	Region	Year	Number of markets	Distance range [km]	Fit of parameter θ		
						Option 1 Diff. of prices	Option 2 Diff. of logs of prices	Goodness of fit
1	Wheat	Bavaria	1815	10	< 250	0.41	0.41	0.67
			1841	10	< 250	0.49	0.42	0.66
			1855	10	< 250	0.13	0.12	0.37
2	Wheat	Côtes d'Armor	1855	7	< 100	0.44	0.34	0.43
3	Wheat	Nord	1855	7	< 100	0.80	0.65	0.83
4	Cotton	U.S.	1948	9	< 1100	0.037		
			1957		< 1100	0.057		

Notes: The dimensionless parameter θ is the model's most important parameter; it is equal to the product at of the slope of excess-supply by the transport cost. It has been computed here from the price differentials and on the basis of an average inter-market distance of 40 km for cases 1, 2, 3 and of 100 km for case 4. "Côtes d'Armor" and "Nord" are two districts in France, the first in Brittany, the second in the region of Lille. The goodness of fit is estimated from the correlation between the square of the differentials and inter-market distance.

(b)

	Commodity	Markets			Parameters			Goodness of fit		
		Region	Year	Number of markets	θ	σ	Δ [km]	η^2	k	χ^2
1	Wheat	France	1825	51	0.10	743	43	0.55	40	41
2	Wheat	France	1908	51	no adjustment possible					
3	Wheat	U.S.	1888	25	0.15	69	155	0.75	10	19.3
4	Wheat	U.S.	1966	34	0.030	50	303	0.82	18	7.6
5	Potato	U.S.	1966	35	0.13	216	170	0.74	19	10.8

Notes: θ is the product of the excess-supply slope by transportation cost; σ is the standard deviation of local shocks; Δ is the average spacing between markets. η^2 is the index of curvilinear correlation; k denotes the number of subdivisions of the distance range. The different values of σ cannot be directly compared because they are expressed in different currencies.

4 Conclusion

From a large collection of cases covering both the 19th and the 20th century there is convergent evidence in favor of the SSAM. One of the most promising features of the theory is the fact that if offers a wide range of predictions. Unfortunately, because of the lack of adequate empirical data we had to restrict ourselves to two types of predictions out of a total of about ten. Therefore it is clear that the first and most urgent task is to carry out more empirical work. Of immense help would be a sort of "Commodity handbook" (similar to the "Physical handbooks") where one would find all available "experimental" data about prices, trade, demand and supply elasticities, transport costs, storage costs, etc.; for the statistical records that are too large (as is for instance the case for high frequency price records) the data themselves could be replaced by the reference to an easily accessible (and lasting!) Internet source.

Appendix A: The two-market case

For a physicist the simplest approach is to look at the two-market case as being an electrical network. That electrical analogy is illustrated in Figure 7. The analogy goes as follows

market 1, market 2	$\longleftrightarrow$	node 1, node 2
prices	$\longleftrightarrow$	opposite of voltages (p'_i)
trade	$\longleftrightarrow$	current i
arbitrage condition	$\longleftrightarrow$	gate ψ
stochastic shocks	$\longleftrightarrow$	stochastic current generator.

The gate ψ allows current to go from node 1 to node 2 only if the absolute value of the voltage between 1 and 2 is higher than the transport cost. In the linear approximation the gate simply becomes a resistance. From the basic equations:

$$j_i = r^{-1}p'_i + a_i \quad j_1 = -i \quad j_2 = i \quad ti = p'_1 - p'_2$$
$$\text{where} \quad r^{-1} = \gamma + \beta \quad a_i = (\gamma+\beta)\overline{p}_i \quad p'_i = -p_i$$

where the p'_i denote the voltages, one gets the equations (2.3).

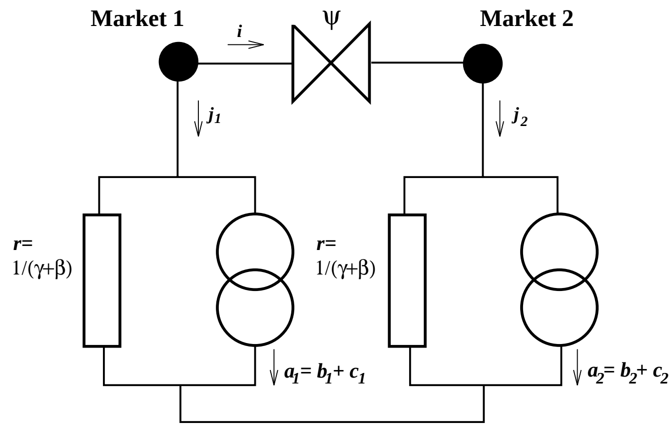


Fig. 7. Electrical analogue of a system of two markets. In the linear approximation the device ψ becomes a resistance. One of the consequence of linearity is the fact that the expectation of the prices does not depend on the statistical properties of the exogenous local shocks and in particular on their covariance matrix. Thus, in the linear approximation no price bubble can be brought about by long-range correlation between local shocks. The analysis of such a time dependent behavior is left for a subsequent paper.

Appendix B: The linear chain of markets

The equations for a linear chain of markets are obtained very much in the same way as in the two-market case. They read:

$$(2+\theta)p_k - p_{k-1} - p_{k+1} = \theta\overline{p}_k \qquad k = \ldots, -1, 0, 1, \ldots \tag{B.1}$$

In this equation the $\overline{p}_k$ are independent, identically distributed random variables. Equation (B.1) is what is called a spatial second-order autoregressive process (a general introduction to spatial autoregressive processes can be found in Chap. 7 of [4]). The boundary-stationarity conditions are such that the $p_{\pm\infty}$ are bounded. Once stationarity is ensured one knows that the solution become independent of the actual values of $p_{-\infty}$ and p_∞. If (B.1) is solved through Fourier transforms the stationary solution is selected ipso facto. Technically this goes as follows.

The first step consists in finding the Green's function G_k *i.e.* the solution of:

$$G_k - a(G_{k-1} + G_{k+1}) = \delta_{x,0} \qquad \text{where} \qquad a = \frac{1}{2+\theta}\cdot \tag{B.2}$$

To this aim we multiply both sides of (B.2) by $\mathrm{e}^{-\mathrm{i}k\omega}$ and sum over k. Next we introduce the Fourier series:

$$\widetilde{G}(\omega) = \sum_{k=-\infty}^{\infty} G_k \mathrm{e}^{-\mathrm{i}k\omega}.$$

$\widetilde{G}(\omega)$ is given by:

$$\widetilde{G}(\omega) = \frac{1}{1 - a(\mathrm{e}^{-\mathrm{i}\omega} + \mathrm{e}^{\mathrm{i}\omega})}\cdot$$

The coefficient G_k are of course expressed as:

$$G_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widetilde{G}(\omega)\mathrm{e}^{\mathrm{i}k\omega}\mathrm{d}\omega.$$

Substituting, one is lead to the following alternative expressions:

$$G_k = \frac{1}{\pi}\int_0^{\pi} \frac{\cos k\theta}{1 - 2a\cos\theta} \tag{B.3a}$$

$$G_k = \frac{1}{2\mathrm{i}\pi}\int_C \frac{z^{k-1}}{1 - a(1/z + z)}\mathrm{d}z \quad C: \text{ unit circle.} \tag{B.3b}$$

The integral (B.3a) can be obtained from a result of Gradshteyn ([14], p. 366, 3.613 (1)). Alternatively the integral (B.3b) can be computed directly using the theorem of residues. Here, however we are interested in the covariance function of the prices rather than in the Green's function. As will be seen the former can be derived from the latter. Using the fundamental property of the Green's function the covariance function $c_p(h) = E(p_k p_{k+h})$ can be written (we proceed formally, for a more rigorous derivation see [15]):

$$c_p(h) = \sum_{i,j} G_{k-i}G_{k+h-j}E(\overline{p}_i\overline{p}_j).$$

If the $\overline{p}_i$ are uncorrelated white noise: $E(\overline{p}_i\overline{p}_j) = \delta_{i,j}$. In this case $c_p(h)$ can be expressed as the convolution product of G_j and $\widehat{G}_j = G_{-j}$; thus the Fourier transform of $c_p(h)$ is given by: $\widetilde{c}(\omega) = \widetilde{G}(-\omega)\widetilde{G}(\omega) = |\widetilde{G}(\omega)|^2$, and going back to the inverse Fourier transform one obtains:

$$c_p(h) = \mathcal{F}^{-1}[\widetilde{G}(-\omega)\widetilde{G}(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\widetilde{G}(\omega)|^2 \mathrm{e}^{\mathrm{i}h\omega}\mathrm{d}\omega.$$

Therefore:

$$c_p(h) = \frac{1}{\pi}\int_0^{\pi} \frac{\cos h\theta}{(1-2a\cos\theta)^2}.$$

The integral can be obtained through the theorem of residues; alternatively it can be derived by differentiation of G_h; one obtains in this way:

$$c_p(h) = \frac{(\sqrt{1+\alpha^2}-\alpha)^h(\sqrt{1+\alpha^2}+h\alpha)}{4\alpha^3}\,\theta^2\,\sigma^2, \qquad \text{(B.4)}$$

$$\theta = at, \quad \alpha = \sqrt{\theta + \frac{\theta^2}{4}}\,.$$

From (B.4) a simple derivation leads to (2.6a).

References

1. B.M. Roehner, Eur. Phys. J. B **13**, 175 (2000).
2. R.G. Bressler, R.A. King, *Markets, prices and interregional trade* (John Wiley, New York, 1970).
3. W.C. Labys, *Dynamic commodity models: specification, estimation and simulation* (Lexington Books, Mass, 1973).
4. B.M. Roehner, *Theory of markets. Trade and space-time patterns of price fluctuations. A study in analytical economics* (Springer-Verlag, Berlin, 1995); B.M. Roehner, Am. J. Agr. Econ. **78**, 339 (1996).
5. T. Takayama, G.G. Judge, *Spatial and temporal price and allocation models* (North Holland, Amsterdam, 1971).
6. S. Enke, Econometrica **19**, 40 (1951).
7. P.A. Samuelson, Am. Econ. Rev. **42**, 283 (1952).
8. B.M. Roehner, Econom. Soc. **11**, 5 (1990).
9. E.F. Fama, J. Business **38**, 34 (1965).
10. B.R. Mitchell, P. Deane, *Abstract of British historical statistics* (Cambridge University Press, 1971).
11. H. Schultz, *The theory and measurement of demand* (University of Chicago Press, Chicago, 1938).
12. B.M. Roehner, Environm. Planning A **21**, 289 (1989).
13. B.M. Roehner, D. Sornette, Eur. Phys. J. B **4**, 387 (1998).
14. I.S. Gradshteyn, I.M. Ryzhik, *Table of integrals, series and products* (Academic Press, New York, 1980).
15. M.B. Priestley, *Spectral analysis and time series* (Academic Press, New York, 1981).
16. J. Langley, S. Langley, State-level wheat statistics 1949-1988, U.S. Dept. of Agriculture, Statistical Bulletin No. 779 (1989).
17. G. Lucier, A. Budge, C. Plummer, C. Spurgeon, *US potato statistics*, U.S. Dept. of Agriculture, Statistical Bulletin No. 829 (1991).