Arshavskiy, Victor; Dave, Chetan; Nyarko, Yaw

**Article**

# Market integration and the law of one price in Ghana

BEROC working paper series

**Provided in Cooperation with:**
Belarusian Economic Research and Outreach Center (BEROC), Minsk

*Reference:* Arshavskiy, Victor/Dave, Chetan et. al. (2017). Market integration and the law of one price in Ghana. In: BEROC working paper series 49 S. 1 - 41.

BEROC Working Paper Series, WP no.49

September 2017

# Market Integration and the Law of One Price in Ghana[1]

Victor Arshavskiy
vya201@nyu.edu
NYUAD, BEROC

Chetan Dave
chetdav@gmail.com
NYUAD

Yaw Nyarko
yn1@nyu.edu
NYU

Abstract

Using a new unique dataset on the prices of perishable commodities traded across several markets, and associated transport costs, we investigate the extent of market integration in Ghana and test the law of one price. We find some support for standard theory, conditional on the distance between markets in a developing country context. We use our data to estimate full transaction costs and to analyze the premium demanded by traders of specific commodities and for different destination markets.

[1] Acknowledgements: We are grateful to the NYU in Abu Dhabi research Institute, CTED NYU Abu Dhabi, the IGC Ghana office, and Swedish International Development Cooperation Agency (Sida) for financial help. All errors are the responsibility of the authors

## 1. Introduction

The development economics literature forks based on whether or not one believes that smooth price adjustments or frictions are at play in an economy. For example, limited frictions imply spatial market integration (MI) such that the law of one price (LOP) holds in a region across commodities and time. We investigate both MI and the LOP for a number of perishable goods across time and geographic space in Ghana using a unique new dataset in which we observe transportation costs as well as prices.

Previous research on this topic is exemplified in an analysis by Barret and Li (2002) who employ spatial analyses to differentiate between the extent of MI and equilibrium across markets. Their methodological contribution builds on the observation that market equilibrium and MI are not necessarily the same concepts. Next, Goodwin and Piggot (2001) utilize threshold autoregression techniques to examine price linkages across markets and find evidence pointing towards MI. With respect to within-country analyses, Abdoulai (2000) examines Ghana directly and finds that maize markets in Ghana are well integrated. However, Niger, Fafchamps and Gavian (1996) find poor integration between spatially separated livestock markets. Relatedly and importantly Kaminski et al (2016) highlight the importance of seasonality in Tanzanian food markets for the ability of households to smooth consumption. Clearly food security plays an important role in the debate over African economic development (see Devereux and Maxwell (2001) for a number of essays on the topic broadly). Recent surges in food prices have also been analyzed with a view towards analyzing food security from diverse perspectives (Headey (2008)), specifically in Ghana (Alderman and Shively (1996), Quaye (2008)).

We re-evaluate existing methods of testing MI using our new dataset on food prices and transportation costs in Ghana. A main contribution we make to this literature is our ability to directly observe transport costs across commodities and locations that allow us to estimate typically unobserved transaction costs (transportation costs of course form only a component of transaction costs).

Informally, MI can be defined as two markets being connected by trade. MI in developing countries is a crucial condition under which new economic policies and technological advancements can positively affect economic growth. However for many developing countries MI is still a challenge. Poor road infrastructure is, probably, the main reason that prevents goods from freely moving between markets. This suggests that price differentials may not just be a function of transaction costs but also unobservable noise. Therefore, we start with a thorough correlation analysis as implied by MI, then conduct related co-integration and Granger causality tests. Similar to other studies on Ghana (i.e., Abdoulai (2000), Ankamah-Yeboah (2012)), we conclude that markets in Ghana are well integrated but only after conditioning on distance.

Our main contribution to the literature is our collection and use of a very detailed, unique, transportation cost dataset. Combined with our price dataset, it enables us to estimate transaction costs. We find that nominal transportation costs account for only 24% of the transaction costs. We conclude that despite co-movements, price differences exhibhit volatility that is inconsistent with the Law of One Price but consistent with the difficulties faced in transporting goods in a developing country.

The difference between transportation costs and transaction costs, among other factors, includes opportunity costs for traders. Therefore, assuming other factors are fixed across commodities and locations, we can apply our results to estimate the premium for specific commodities and city pairs. Next, conditioning on distance between city pairs, we can estimate the premium specific to a destination.

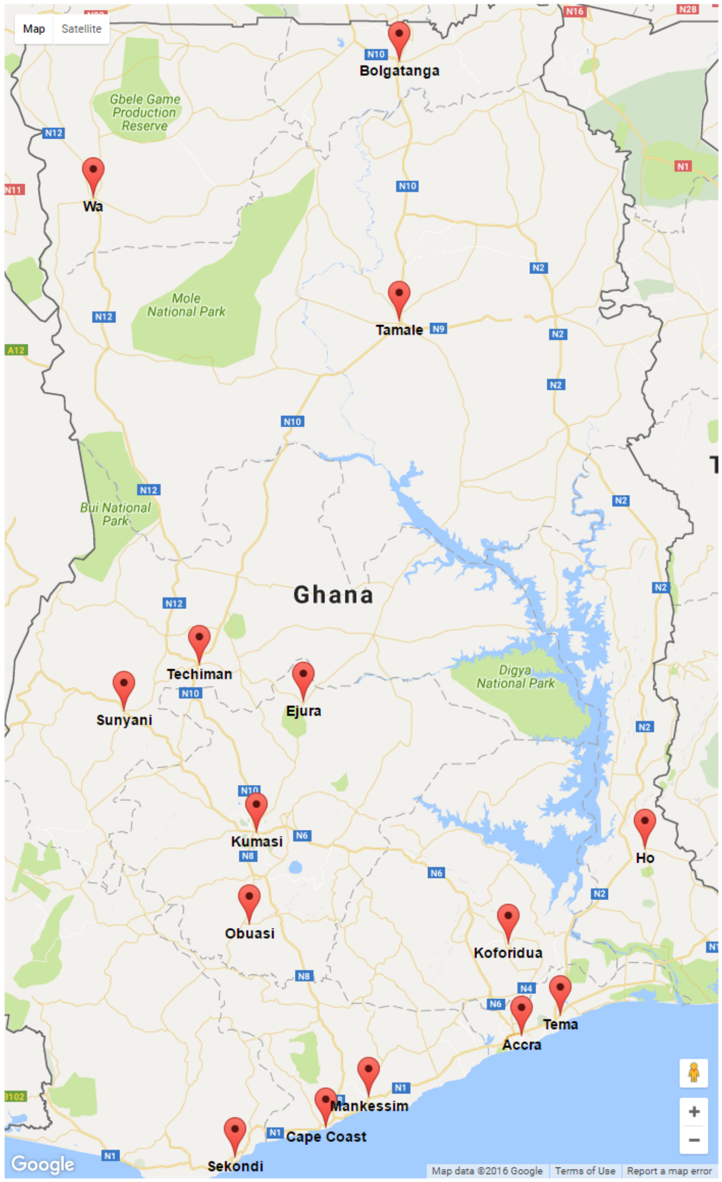
The paper is organized as follows. Section 2 describes our data. Section 3 studies correlations, co-integration of prices, and Granger causality. Section 4 then discusses the geographical pattern of MI. Section 5 describes our transportation cost dataset, estimates of the transaction costs, and provides an analysis of the premium demanded by traders for different commodities and destination markets. Section 6 concludes.

**2. The Data**

We construct a dataset for wholesale commodity prices in Ghana: a 3-dimensional panel with weekly data on prices for 19 commodities and 15 markets for almost seven years (from January 2009 to November 2015). This data has been provided by the Ministry of Food and Agriculture in Ghana (MOFA).

Traders in Ghana go to local markets to sell the commodities they buy directly from the producers (or farmers). Some farmers also go to the same markets; however, most find it more profitable to sell commodities at the farm gate to traders. Markets in Ghana usually meet once every week. On those days MOFA sends their representatives to markets who inquire about commodity prices and record them. The 15 markets for which MOFA has the most accessible data are in various regions in Ghana, shown in Figure 1.

Figure 1. Markets in Ghana




We supplement this with data on transport costs collected by the NYU Center for Technology and Economic Development (CTED) in collaboration with ESOKO, a Ghanaian technology and agricultural services company. ESOKO surveyed drivers to collect information on the costs of transportation at different locations across the country and were encouraged to do so whenever markets met on a weekly basis. The data was collected between April 2013 and

December 2014. The main variables collected were: commodity, origin, destination, date of collection, the cost of transportation, and the measure (e.g. kilograms or tubers). Finally, we obtained the distance between origin and destination markets from Google.[2]

## 3. Market Integration (MI)

If an identical commodity is sold on two spatially separated markets, MI would imply that the commodity will be transported from the market where the price is lower to the market where the price is higher until the difference in prices will be no more than the transaction costs required for a trader to transport the good from one market to another. In particular, if we define $j_1$ and $j_2$ to be indexes for two separate locations, $c$ as the commodity index, and $t$ as the current time period, we would expect the Law of One Price to hold:

$$P_{cj_2t} - P_{cj_1t} = \tau_{cj_1j_2t} \qquad (1)$$

where $P_{cj_1t}$ and $P_{cj_2t}$ are prices for commodity $c$, and $\tau_{cj_1j_2t}$ is the associated transaction cost.

Cournot (1938) states that by transaction costs we must understand "*not only the price of necessaries and the wages of the agents by whom the transportation is mechanically carried out, but also insurance premiums, and the profits of the merchant, who ought to obtain in his business the interest on the capital employed and a proper return for his industry*." We therefore view $\tau_{cj_1j_2t}$ as being the theoretical, unobserved transaction cost. Below, in section 5, we describe how we estimate these transaction costs given our data on transportation costs and other variables. For now, we assume that transaction costs are random.

[2] Using a Google Maps API, https://developers.google.com/maps/documentation/distance-matrix/start.

The Law of One Price implies that prices in different markets should co-move. The extent of co-movement of prices can be investigated in several ways. We use three popular methods in the existing literature. The first method is correlation analysis: prices for the same commodity at different markets should be significantly correlated. However, this does not recognize the time series structure of our data. Therefore, our second set of analyses are co-integration tests: prices for the same commodity at different markets should be co-integrated. Our third set of analyses are Granger Causality tests: prices for the same commodity at different markets should Granger cause each other.

Each of the above tests need also to be conditioned on the fact that in Ghana the transport network is still at a developing country phase with poor road quality and so the distance between markets may be more important than in developed countries. This feature introduces a geo-spatial dimension to our data. That is, prices between markets far away from each other might be correlated less, be less co-integrated and not Granger cause each other as much. We now turn to each of these tests with attention paid to the distance between markets and pairs thereof.

### 3.1. Correlation Analyses Implied by MI and the LOP

The simplest possible measure of price co-movement is the correlation of prices across markets. Therefore, we first compute correlations between prices across all pairs of markets. Since we have 15 markets, the total number of pairs are $\frac{15\times 14}{2} = 105$. Table 1 presents the average and standard deviation of correlations for each commodity across pairs of markets.

Table 1. Correlations across Commodities and Market Pairs

| Commodity | Average | Std. Dev. |
|---|---|---|
| Cassava | 0.616 | 0.210 |
| Cocoyam | 0.518 | 0.401 |
| Cowpea | 0.852 | 0.065 |
| Groundnuts | 0.886 | 0.068 |
| Maize | 0.867 | 0.060 |
| Millet | 0.832 | 0.090 |
| Onion | 0.653 | 0.219 |
| Oranges | 0.493 | 0.211 |
| Palm fruit | 0.314 | 0.327 |
| Palm oil | 0.797 | 0.084 |
| Dried Pepper | 0.416 | 0.224 |
| Fresh Pepper | 0.343 | 0.312 |
| Apem Plantain | 0.479 | 0.205 |
| Apentu Plantain | 0.565 | 0.166 |
| Imported Rice | 0.900 | 0.070 |
| Local Rice | 0.794 | 0.128 |
| Sorghum | 0.843 | 0.078 |
| Tomatoes | 0.622 | 0.159 |
| Yam | 0.646 | 0.147 |

From Table 1 we see that for the majority of commodities the average correlation is very high (>0.8 for cowpea, groundnuts, maize, millet, imported rice). To the extent that the Law of One Price (1) holds, it should also be true that transaction costs $\tau_{cj_1j_2t}$, which are assumed random, increase and become more variable (noisier) as the distance between markets $j_1$ and $j_2$ increases. This could partially be due to poor road infrastructure and unpredictable weather patterns so that increased trip duration increases the variability of the costs. Moreover, the uncertainty that traders face while transporting goods over a large distance increases, and, therefore, the premium they would require on top of nominal transportation costs will be higher.

A measurable implication of this is that correlations between markets close to each other should be high, while correlations between distant markets should be low. Thus, we estimate a regression of the correlation of prices between markets on the distance between markets. Previous studies (e.g. Fafchamps and Gavian (1996)) used a similar approach, however, they analyzed the relationship for each commodity separately. We employ a panel approach and estimate the following fixed effects regression:

$$Corr_{cj_1j_2} = \alpha + \beta \mathrm{DST}_{j_1j_2} + \mu_c + \varepsilon_{cj_1j_2}, \qquad (2)$$

where $c$ is the commodity index, $(j_1, j_2)$ is a pair of markets; $Corr_{cj_1j_2}$ is the price correlation between the pair for commodity $c$, $DST_{j_1,j_2}$ is the distance between the pair scaled by a factor of 100 ($DST = 1$ means markets are 100 km apart),[3] $\mu_c$ is the commodity fixed effect, and $\varepsilon_{cj_1j_2}$ is the idiosyncratic error. The results are provided in Table 2 below

Table 2. Correlation and Distances

| | Estimate | Std. Err. | \|t\|-stat. |
|---|---|---|---|
| Distance (100km) | -0.0213 | 0.00214 | 9.91 |
| Constant | 0.733 | 0.00851 | 86.15 |
| Obs. | 1,962 | | |
| Groups | 19 | | |
| F-stat. for $H_0$ of fixed effects | 106.00 | | |

We see that the effect of distance on correlation is highly significant: an extra 100 km decreases correlations by 2.1%. This confirms our hypothesis that correlations between markets close to each other are high, while correlations between distant markets are low, implied by a degree of MI from equation (1).

[3] The distance between the two furthest markets Accra and Bolgatanga is 745 km.

**3.2. Cointegration Analyses**

As argued by Harris (1979) and Blyn (1973) and emphasized by Ravallion (1986) there is a danger in using correlation analysis for market integration analyses. The following example, adapted from Ravallion (1986), explains the most important issues. Let's say we have two markets such that trade is infinitely costly between them. But at the same time, both price series heavily depend on the price of a third commodity (say, oil), or are affected by another external variable (i.e., trend or seasonality). In this instance, correlation is a poor indicator of market integration, which motivates our additional analyses below. An alternative approach to testing MI focusses on the time series structure of the data and suggests a test for co-integration of price series (Alexander and Wyeth (1994)). An "integrated" economic variable is loosely defined as a time series that has a stochastic (non-stationary) trend. If the first difference of this variable is stationary, the variable is integrated of order 1. Much of the development literature (with the notable exception of Wang and Tomek (2007)) argues that prices are commonly integrated of degree 1. The term "cointegration" is a property of two variables which have already been shown to be integrated and which, though trending, cannot drift too far from each other. Thus, Alexander and Wyeth (1994) state that when two price series are co-integrated, it follows that the markets are integrated (in the economic sense) in the long run. Our investigation proceeds in stages. We first check if price series are non-stationary, then check for stationarity of price differences. We conclude that markets are integrated if price series are non-stationary, but their difference is stationary.

Our price series are first tested for integration of degree one using an Augmented Dickey Fuller test (Dickey and Fuller, 1984). The following regression is estimated for each price series and each market:

$$\Delta y_t = \alpha + \beta\, y_{t-1} + \sum_{j=1}^{k-1} \varphi_j \Delta y_{t-j} + \varepsilon_t \qquad (3)$$

where $y_t$ is the natural logarithm of a price series. The lag $k$ is selected using the following iterative procedure. We first estimate the model with a maximum lag $k = k_{max} = 12$ weeks (i.e. three months, that is commonly used in the literature). If the last lag coefficient $\varphi_k$ is insignificant, the model is re-estimated with $k = k - 1$ and the process repeated until the last lag coefficient is significant.

Testing for integration of degree one (i.e. unit root) in these settings is equivalent to testing the hypothesis $H_0\colon \beta = 0$. We present a summary of the results organized by commodity in Table 3 below. For 49 out of 282[4] price series (17%) we reject the unit root hypothesis at a 5% confidence level. Fafchamps and Gavian (1996), found that less than 5% of price series are stationary, much smaller than 17% of stationary series in our dataset.

[4] Prices for Cocoyam in Bogatanga and Palm Fruit for Bolgatanga and Wa were not available for enough periods to perform unit root tests, therefore the total number of price series for which we could perform the test is 15×19-3=282.

Table 3. Summary of Stationarity Tests

| Commodity | Av. t-stat | Av. lag | Reject 5% | Reject 1% |
|---|---|---|---|---|
| Cassava | -1.524 | 5.000 | 1 | 1 |
| Cocoyam | -1.282 | 5.071 | 1 | 0 |
| Cowpea | -0.955 | 6.600 | 0 | 0 |
| Groundnuts | -0.482 | 6.467 | 0 | 0 |
| Maize | -0.837 | 6.933 | 0 | 0 |
| Millet | -1.047 | 5.667 | 0 | 0 |
| Onion | -2.428 | 5.800 | 4 | 0 |
| Oranges | -2.099 | 4.667 | 3 | 3 |
| Palm fruit | -3.339 | 6.308 | 3 | 2 |
| Palm oil | -0.479 | 5.000 | 0 | 0 |
| Pepper (dried) | -2.003 | 6.600 | 3 | 2 |
| Pepper (fresh) | -3.573 | 4.800 | 10 | 8 |
| Plantain (apem) | -2.430 | 5.467 | 5 | 3 |
| Plantain (apentu) | -2.711 | 6.133 | 4 | 4 |
| Rice (imported) | 0.313 | 6.133 | 0 | 0 |
| Rice (local) | 0.174 | 6.333 | 0 | 0 |
| Sorghum | -0.679 | 4.867 | 0 | 0 |
| Tomatoes | -3.217 | 4.067 | 10 | 6 |
| Yam | -2.468 | 6.200 | 5 | 2 |
| Total | -1.635 | 5.69 | 49 | 31 |

Column 2 of Table 3 shows average t-statistics for the hypothesis that $\beta = 0$; column 3 shows the average lag that has been used given the iterative procedure above; columns 4 and 5 show for how many series the null hypothesis of non-stationarity has been rejected at 5% and 1% levels respectively. We see a lot heterogeneity of results between commodities. Prices for tomatoes and fresh peppers are stationary in 2/3 of all markets, whereas for yam, plantain, and onion prices in 5 and 4 markets respectively are stationary. For many goods, we have zero or only one stationary price series.

To test that our series are not integrated of degree 2 we take the first difference of prices and test them for stationarity using the same method. It turns out that 281 out of 282 differenced series are stationary. We therefore conclude that the majority of our price series are integrated of degree 1.

Next, if equation (1) holds, and we believe that transaction costs are stationary, the difference of the two price series should be stationary. This implies co-integration if the price series are non-stationary. We test for the stationarity of price differences for all combinations of commodity-market 1-market 2 for which both price series are non-stationary. It turns out that only 1392 of 1,954 combinations (71%) can be tested for co-integration. Among those 755 series (55%) are co-integrated.

The remaining 562 (29%) combinations cannot be tested for co-integration: for 464 (23%) pairs only one market is non-stationary, and co-integration tests do not make sense (the difference should always be non-stationary). For the remaining 139 pairs (6%) both price series are stationary, and, therefore, co-integration tests cannot be conducted as well. Differentiated by commodities, these results are presented in Table 4 (columns 2, 3 and 4). Columns 2-4 partition all price series for a given commodity for three sets: both series are non-stationary, only one price series is non-stationary, and both price series are stationary. Column 6 presents results for the co-integration tests, which were only conducted if price series for both markets were non-stationary.

Table 4. Unit Roots and Co-Integration

| Commodity | Both non-stationary | One stationary | Both stationary | Co-integrated |
|---|---|---|---|---|
| Cassava | 91 | 14 | 0 | 29 (32%) |
| Cocoyam | 78 | 26 | 1 | 15 (19%) |
| Cowpea | 105 | 0 | 0 | 75 (71%) |
| Groundnuts | 105 | 0 | 0 | 68 (65%) |
| Maize | 105 | 0 | 0 | 63 (60%) |
| Millet | 105 | 0 | 0 | 81 (77%) |
| Onion | 55 | 44 | 6 | 41 (74%) |
| Oranges | 66 | 36 | 3 | 44 (67%) |
| Palm fruit | 45 | 50 | 10 | 21 (47%) |
| Palm oil | 105 | 0 | 0 | 40 (38%) |
| Pepper (dried) | 66 | 36 | 3 | 27 (41%) |
| Pepper (fresh) | 10 | 50 | 45 | 4 (40%) |
| Plantain (apem) | 45 | 50 | 10 | 21 (47%) |
| Plantain (apentu) | 55 | 44 | 6 | 32 (58%) |
| Rice (imported) | 105 | 0 | 0 | 51 (49%) |
| Rice (local) | 91 | 14 | 0 | 49 (54%) |
| Sorghum | 105 | 0 | 0 | 49 (47%) |
| Tomatoes | 10 | 50 | 45 | 10 (100%) |
| Yam | 45 | 50 | 10 | 34 (76%) |
| Total | 1392 | 464 | 139 | 755 (54%) |

Again, we observe significant heterogeneity among commodities. We can see that the highest number of co-integrated series is for millet (81) followed by cowpea (75) and then groundnuts (68) and maize (63). Also, notice that 76% of yam price series and 100% of tomato price series are co-integrated, although the number of series we can test for yam and tomatoes is low.

Looking further into the geo-spatial dimension of market integration, we will now analyze how co-integration is related to the distance between markets. To do so, we estimate regressions similar to (2), where we replace the dependent variable with t-statistics for unit root tests of price differences.[5] We estimate this regression conditional on the fact that co-integration tests make sense, i.e., price series for both markets are non-stationary. We therefore estimate the following panel regression with fixed effects for commodities

$$T_{cj_1j_2} = \alpha + \beta \mathrm{DST}_{j_1j_2} + \mu_c + \varepsilon_{cj_1j_2} \qquad (4)$$

where $(j_1, j_2)$ is a pair of markets; $T_{c,j_1,j_2}$ is the t-statistic of a co-integration test between prices in $j_1$ and $j_2$ for commodity $c$, $DST_{j_1,j_2}$ is the distance in kilometers between the pair scaled by a factor of 100, $\mu_c$ is the commodity fixed effect, and $\varepsilon_{c,j_1j_2}$ is the idiosyncratic error.

Table 5. Co-Integration and Distances

| | Estimate | Std. Err. | t-stat |
|---|---|---|---|
| Distance (100 km) | .1146 | .0164 | 6.96 |
| Constant | -3.52 | .0662 | -53.13 |
| Num. obs. | 1392 | | |
| Num. groups | 19 | | |
| F-stat for fixed effect | 11.70 | | |

The results, provided in Table 5, show that there is a significant relationship between price co-integration and distances between markets. Markets that are close to each other are more likely to be integrated, markets that are far from each other are less likely to be integrated.

### 3.3 Granger Causality

Another test of MI that is widely used in the literature is the Granger causality test. Granger causality is a concept that is used to determine if one statistical variable is useful in forecasting the other variable. If markets are integrated, we expect that price shocks in one market are useful in determining future prices in the other market, therefore, it is also an

[5] This is similar to the Fafchamps and Gavian (1996) regression

indirect test of equation (1), which says that prices in two markets should be connected by transportation/transaction costs.

To test for Granger causality between two markets (call them $X$ and $Y$), we use the following model.

$$\Delta y_t = \alpha + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \varepsilon_t \qquad (5)$$

where $y_t$ and $x_t$ are prices for markets $Y$ and $X$. The null hypothesis, that is, price changes in market $X$ do not cause price changes in market $Y$, is formulated as all $\gamma_j = 0$ (for all $j \in [1, k]$).

$$H_0\colon\ \gamma_j = 0\ for\ all\ j \in [1, k] \qquad (6)$$

Since we have weekly data, we take $k = 12$ weeks to include information on three past months. The specification in (5) and the hypothesis test in (6) is valid regardless of whether the prices $x_t$ and $y_t$ are stationary by themselves or not, so we conduct tests for all combinations of commodiy-market1-market2 in both directions.

The results are as follows: for 1139 pairs of markets (58%) price changes in one market cause price changes in the other market. Out of those, for 317 pairs (16%) the causality runs both ways. The results organized by commodities are presented in Table 6.

Table 6. Granger Causality Tests

| Commodity | At least one way Granger causality | Two way Granger causality |
|---|---|---|
| Cassava | 43 | 8 |
| Cocoyam | 44 | 7 |
| Cowpea | 60 | 21 |
| Groundnuts | 79 | 15 |
| Maize | 94 | 38 |
| Millet | 64 | 19 |
| Onion | 83 | 22 |
| Oranges | 62 | 18 |
| Palm fruit | 17 | 3 |
| Palm oil | 54 | 13 |
| Pepper (dried) | 52 | 16 |
| Pepper (fresh) | 58 | 16 |
| Plantain (apem) | 55 | 13 |
| Plantain (apentu) | 71 | 24 |
| Rice (imported) | 62 | 19 |
| Rice (local) | 52 | 7 |
| Sorghum | 46 | 14 |
| Tomatoes | 77 | 22 |
| Yam | 66 | 22 |
| Total | 1139 | 317 |

**Note**: The second column: total pairs where at least one market Granger-causes the other; third column: total pairs where both markets Granger-cause each other.

The most integrated market is the market for maize: for 94 out of 105 pairs of markets we cannot reject causality at 5% confidence level.

Similar to the methodology from previous sections we next estimate a panel regression of p-values on distances with commodity fixed effects to show that the prices at markets close to each other are more likely to cause each other.

$$p_{c,j_1 j_2} = \alpha + \beta \mathrm{DST}_{j_1 j_2} + \mu_c + \varepsilon_{c,j_1 j_2} \qquad (7)$$

where $p_{c,j_1 j_2}$ is the p-value for Granger causality test for commodity $c$, between locations $j_1$ and $j_2$ and $\mu_c$ is the commodity fixed effect. The results from estimating regression (7) are presented in Table 7.

Table 7. Granger causality and distances (panel regression).
Dependent variable: p-value of Granger causality test

| | Estimate | Std. Err. | T-stat |
|---|---|---|---|
| Distance (100 km) | .0151 | .0028 | 5.34 |
| Constant | .321 | .0111 | 25.82 |
| Num. obs. | 3,906 | | |
| Num. groups | 19 | | |
| F-stat for fixed effect | 24.79 | | |

We see that overall the relationship is positive and supports the intuition that prices for markets close to each other are more likely to cause each other than markets that are far from each other. Once again, our results confirm that the degree of integration between close markets is higher than that of markets far apart, which is implied by the Law on One Price equation (1).[6]

## 4. The Geographical Pattern of Market Integration

In the context of the three measures of market connectedness, we analyze what they say about market integration across geographical space. We examine whether certain markets are more likely to be integrated with others, thereby identifying groups of markets that are jointly integrated as well as markets that are isolated.

First, we look at the average correlation for a given market.[7] The results are presented in the Table 8. Sorting the results by markets we can say which markets have the highest and

[6] We also conducted analogous tests for Wald statistics (instead of p-values) and the results are very similar.
[7] We average correlations across all market pairs that connect to this market and across all commodities.

lowest correlation on average. We see that the markets in the southern coastal areas of Ghana, many of which are large urban centers, are much more correlated with all other markets. At the same time, standard deviations indicate that this measure is imprecise, and the differences between markets are insignificant.

Table 8. Average Correlation

| Market | Average correlation | Standard deviation |
|---|---|---|
| Mankessim | 0.725 | 0.206 |
| Accra | 0.714 | 0.232 |
| Kumasi | 0.708 | 0.227 |
| Cape Coast | 0.693 | 0.285 |
| Sekondi | 0.690 | 0.268 |
| Koforidua | 0.688 | 0.233 |
| Ejura | 0.665 | 0.237 |
| Sunyani | 0.662 | 0.255 |
| Wa | 0.648 | 0.263 |
| Tema | 0.644 | 0.247 |
| Techiman | 0.641 | 0.234 |
| Ho | 0.620 | 0.246 |
| Obuasi | 0.611 | 0.268 |
| Bolgatanga | 0.595 | 0.348 |
| Tamale | 0.587 | 0.341 |

Our co-integration results show that the market that is co-integrated most with other markets is Mankessim. Mankessim is a market on the Southern coast of Ghana, directly connected to Accra, the capital of Ghana, and Kumasi, the capital of Ashanti region, sometimes called the second capital of the country. High in the list are also Koforidua, Ejura and Secondi, probably because of their direct connection to large markets. The results for co-integration are reported in Table 9.

Table 9. Percentage of co-integrated series by market

| Market | # of series that can be tested for co-integration | # of co-integrated series | % of co-integrated series |
|---|---|---|---|
| Mankessim | 196 | 120 | 61.22 |
| Koforidua | 209 | 117 | 55.98 |
| Ejura | 186 | 111 | 59.68 |
| Sekondi | 209 | 111 | 53.11 |
| Tamale | 174 | 107 | 61.49 |
| Tema | 181 | 105 | 58.01 |
| Obuasi | 156 | 103 | 66.03 |
| Kumasi | 187 | 102 | 54.55 |
| Ho | 194 | 100 | 51.55 |
| Accra | 176 | 97 | 55.11 |
| Cape Coast | 191 | 97 | 50.79 |
| Wa | 191 | 90 | 47.12 |
| Bolgatanga | 184 | 88 | 47.83 |
| Techiman | 175 | 83 | 47.43 |
| Sunyani | 175 | 79 | 45.14 |

Table 10 summarizes Granger causality tests, showing for each market the number of markets that are Granger-caused by a market (column 2) and that Granger-cause that market (column 3).

Table 10. Granger-causality by market

| Market $i$ | # of series that are Granger-caused by prices in market $i$ | # of series that Granger-cause prices in market $i$ | $Column\ 2 + Column\ 3$ | $\frac{Column\ 2}{Column\ 3}$ |
|---|---|---|---|---|
| Bolgatanga | 100 | 66 | 166 | 1.52 |
| Tamale | 110 | 77 | 187 | 1.43 |
| Cape Coast | 99 | 87 | 186 | 1.14 |
| Sekondi | 90 | 82 | 172 | 1.10 |
| Tema | 106 | 101 | 207 | 1.05 |
| Kumasi | 108 | 105 | 213 | 1.03 |
| Obuasi | 106 | 107 | 213 | 0.99 |
| Koforidua | 100 | 103 | 203 | 0.97 |
| Wa | 79 | 83 | 162 | 0.95 |
| Mankessim | 95 | 100 | 195 | 0.95 |
| Ho | 86 | 95 | 181 | 0.91 |
| Accra | 100 | 111 | 211 | 0.90 |
| Techiman | 94 | 107 | 201 | 0.88 |
| Ejura | 87 | 103 | 190 | 0.84 |
| Sunyani | 96 | 129 | 225 | 0.74 |

The markets in table 10 are sorted by the ratio of “cause to being caused” by other markets, and the interpretation of this ratio is as follows. Markets with a high ratio of cause to being caused are likely to be the source of price shocks that spread across country. The highest ratio belongs to the most northern market in Ghana, Bolgatanga. This is the origin for most crops in Ghana. Surprisingly, another northern market, Wa, does not have high ratio of cause / being caused. We attribute this to the fact that Wa is more isolated than other markets.

Overall, we conclude that markets in Ghana are generally integrated but there do exist geographical areas (Wa, Ho) where markets are poorly integrated. We conjecture that the main reason is bad road infrastructure. More critically, markets that are closer to each other are much more likely to be integrated according to all criteria.

## 5. Transaction Costs and Transportation Costs

We now describe our transportation cost dataset and estimate the full transaction cost $\tau_{cj_1j_2t}$ of equation (1). We will apply these estimates to an analysis of the premia for commodities and specific destination markets.

### 5.1 Transportation Cost Dataset

Our original transportation cost dataset consisted of 29 different commodities and 30 locations. The data are recorded on a weekly basis from April 2013 to December 2014. A few lines of the dataset are displayed in Table 11 below.

Table 11. A few observations from the transportation cost dataset

| Commodity | Origin | Destination | Date | Cost per 1 kg |
|---|---|---|---|---|
| Maize | Damango | Tamale | 29nov2014 | 0.12 |
| Maize | Damango | Tamale | 06dec2014 | 0.12 |
| Maize | Damango | Tamale | 27dec2014 | 0.12 |
| Maize | Damango | Techiman | 01jun2013 | 0.55 |
| Maize | Damango | Techiman | 08jun2013 | 0.75 |
| Maize | Damango | Techiman | 15jun2013 | 0.75 |
| Maize | Damango | Techiman | 22jun2013 | 0.75 |

After dropping commodities for which we have a very low number of observations and commodities that are not present in our price dataset, we are left with the following 7 commodities: maize, groundnuts, cowpea, yam, millet, sorghum, and local rice. Table 12 presents the list of commodities with the number of observations in our dataset, first, and last dates.

Table 12. Commodities and the number of observations

| Commodity | Number of observations | First date | Last date |
|---|---|---|---|
| Maize | 2021 | 27apr2013 | 03jan2015 |
| Groundnuts | 1653 | 01jun2013 | 03jan2015 |
| Cowpea | 1531 | 27apr2013 | 03jan2015 |
| Yam | 1368 | 27apr2013 | 03jan2015 |
| Millet | 1270 | 01jun2013 | 03jan2015 |
| Sorghum | 1186 | 27apr2013 | 03jan2015 |
| Rice (local) | 947 | 27apr2013 | 03jan2015 |

**Note**: The second column is the total number of observations for this commodity from the dataset; the third and fourth columns are first date and last date of the appearance of these commodities in our dataset.

Even after keeping only commodities with the most observations, if we consider all possible pairs of origins and destinations, we will have a dataset with only 1.9% of non-missing observations. This is significantly different from our transportation cost dataset, where we have about 90% of non-missing observations.

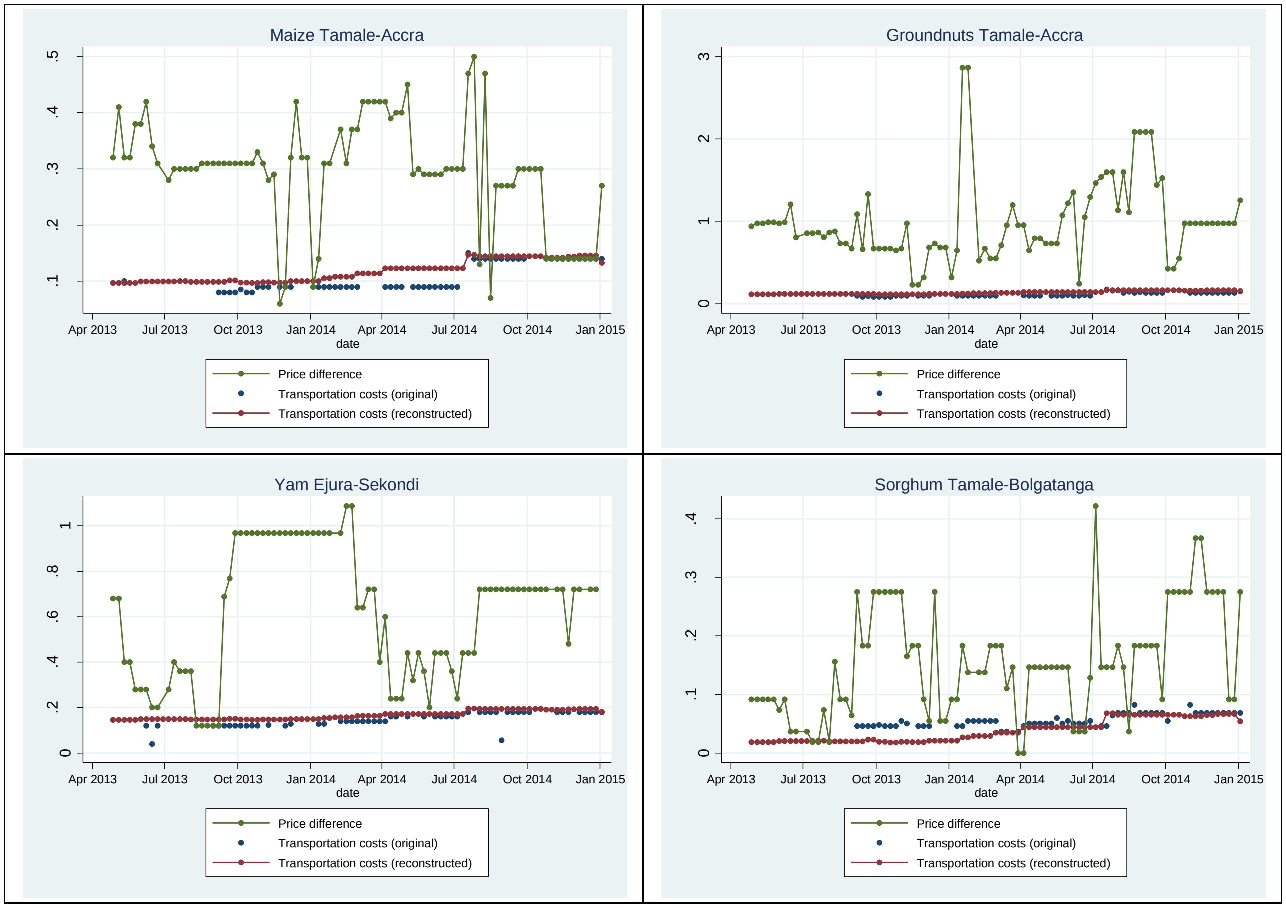
However, these 1.9% of observations are still a significant number (about 9,000). There are two ways to proceed: we can either use the dataset with many missing observations, or we can construct a new dataset based on the existing one. We do so and describe the procedure in Appendix A1.

In the next sections, we first describe the methodology we employ to estimate full transaction costs, then we present the estimates, and, finally, apply the results to analyze commodity specific and location specific premiums.

**5.2 Estimating transaction costs**

As discussed in existing literature (see, for example, Baulch (1997) or Penzhorn and Arndt (2002)), transport costs account only for a fraction of the transaction costs that equate price differences as per the Law of One Price. Indeed, if we look at a plot of price differences and transportation costs (whether original or reconstructed), we see that the price difference is significantly larger: $\left|P_{cj_1,t} - P_{cj_2,t}\right| \gg T_{cj_1j_2,t}$. Figure 2 shows an example of price difference vs transportation costs for four different combinations of commodity, origin, destination.

Figure 2. Price differences vs Nominal Transportation Costs




In fact, the average ratio of price difference to transportation costs (across all commodities) is about 4.2. In other words, nominal transportation costs account for only 24% of the price difference between markets. This also confirms that the use of nominal transportation costs in the right-hand side of the equation (1) may lead to poor results. The summary statistics for price difference and transportation costs are reported in Table 13.

Table 13. Price difference and nominal transport costs

| Variable | Number of observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Price difference | 87,888 | 0.460 | 0.391 | 0.000 | 3.119 |
| Transport costs (original) | 9,500 | 0.109 | 0.040 | 0.028 | 0.300 |
| Transport costs (reconstructed) | 84,728 | 0.124 | 0.042 | 0.010 | 0.256 |

The previous literature that uses transaction costs to test for market integration (Baulch (1997), Fafchamps, Gavian (1996), Penzhorn and Arndt (2002), and others) adds a reasonable estimate of the opportunity costs of traders. For example, Fafchamps and Gavian (1996), analyzing livestock transportation and trading, compute animal speed to add the trader's opportunity costs to nominal transportation costs. Penzhorn and Arndt (2002) add 32% on top of measurable part of transaction costs to reflect the unmeasurable part.

We believe that the richness of our transportation cost and price datasets enable us to find good estimates of the full transaction cost $\tau$. Since the transaction costs include opportunity costs for traders, we will also use the estimated coefficients to analyze the premia for specific commodities and destinations.

We assume the following structure for transaction costs $\tau_{cj_1j_2t}$ separating measurable transportation costs and unmeasurable parts that depend on 1) trend, 2) seasonality, 3) a component that is specific to origin-destination pair, and 4) a component that is specific to commodity, and an idiosyncratic error. Further we assume that the Law of One Price holds as equality,

$$\left|P_{cj_2t} - P_{cj_1t}\right| = \tau_{cj_1j_2t}. \qquad (8)$$

and that costs are symmetric, i.e., $\tau_{cj_1j_2t} = \tau_{cj_2j_1t}$.

The structural equation for transaction costs can be written as

$$\tau_{cj_1j_2t} = TC_{cj_1j_2t} + \beta t + \gamma_r I(t \in R) + \sum_{\psi=1}^{7} \varphi_\psi I(c = \psi) + \sum_{\mu=1}^{78} m_\mu I((j_1, j_2) = \mu) + \varepsilon_{cj_1j_2t}$$

(9)

where $t$ is the week number, $R$ is the set of dates which belong to the major rainy season (April to July), 7 is the total number of commodities (see table A2.1 in the appendix for the full list), 78 is the total number of market pairs[8] (see the table in the appendix for the full list), and $\varepsilon_{cj_1j_2t}$ is an idiosyncratic shock.

Therefore, we estimate the following regression:

$$Y = \alpha + \beta \cdot t + \gamma_r \cdot I(t \in R) + \sum_{\psi=2}^{7} \varphi_\psi \cdot I(c = \psi) + \sum_{\mu=2}^{78} m_\mu \cdot I((j_1, j_2) = \mu) + \varepsilon_{cj_1j_2t} \quad (10)$$

where $Y = \left|P_{cj_2t} - P_{cj_1t}\right| - TC_{cj_1,j_2t}$. We define our transaction cost estimates as predicted values plus transportation costs:

$$\widehat{\tau_{cj_1j_2t}} = \hat{Y} + TC_{cj_1,j_2t} \quad (11)$$

Results of the regression are shown in Tables 14 and 15.[9] For Table 14 we used the original transportation costs while for Table 15 we used reconstructed transportation costs.

Both tables indicate a significant effect of the time trend: every week the transaction cost increases by 0.0025-0.0033 GHC (25/10,000=1/400) per 1 kg on top of the increase of the transportation costs. This may well be due to inflation.

Seasonality, however, is captured only if we use original transportation costs (Table 14). The seasonal effect is insignificant if we use reconstructed transportation costs (Table 15).

[8] We matched 13 locations, so total number of pairs is $\frac{13 \cdot 12}{2} = 78$, full list is provided in appendix A2

[9] Estimates for location effects are not presented here but available upon request

Table 14. Regression transaction costs I

| | Estimate | Std. Err. | t-stat |
|---|---|---|---|
| Trend (week number) | .0033 | .00028 | 11.55 |
| Rainy season ($\gamma_r$) | -.0502 | .019 | -4.34 |
| Cowpea | 0.00 | -[10] | - |
| Groundnuts | 0.05 | 0.0191 | 2.62 |
| Maize | -0.3644 | 0.0193 | -18.85 |
| Millet | -0.3116 | 0.0216 | -14.46 |
| Rice (local) | -0.1345 | 0.0230 | -5.85 |
| Sorghum | -0.3011 | 0.0236 | -12.75 |
| Yam | -0.2343 | 0.0223 | -10.51 |
| Constant | 0.5309 | 0.0313 | 16.94 |
| Num. obs. | 4404 | | |
| R-squared | .268 | | |

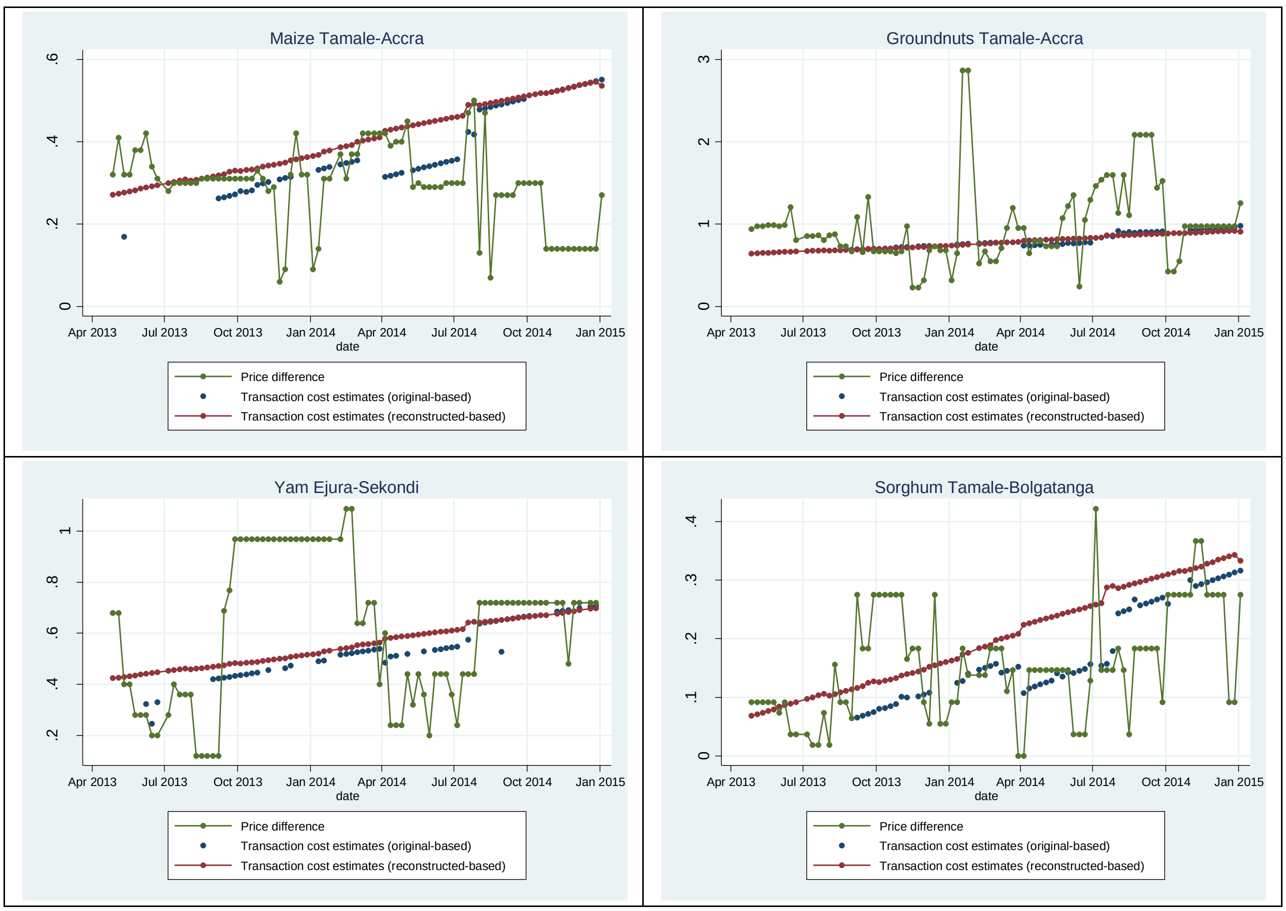
Table 15. Regression transaction costs II

| | Estimate | Std. Err. | t-stat |
|---|---|---|---|
| Trend (week number) | 0.0026 | 0.0001 | 38.07 |
| Rainy season ($\gamma_r$) | 0.0039 | 0.0037 | 1.05 |
| Cowpea | 0.00 | - | - |
| Groundnuts | 0.0120 | 0.0062 | 1.93 |
| Maize | -0.3433 | 0.0062 | -55.66 |
| Millet | -0.2515 | 0.0062 | -40.78 |
| Rice (local) | -0.1702 | 0.0067 | -25.53 |
| Sorghum | -0.2358 | 0.0064 | -36.59 |
| Yam | -0.2835 | 0.0064 | -44.06 |
| Constant | | | |
| Num. obs. | 38,154 | | |
| R-squared | .2795 | | |

[10] Cowpea dummy is excluded from regression, and all commodity coefficients are in the reference to cowpea

The law of one price assumes no arbitrage, and, therefore, no profit apart from the required return on capital and opportunity costs. However, since regression (10) essentially takes an average across all commodities and markets, we may see that for some directions there is a significant difference between price differentials and estimated transaction costs. This difference should not be interpreted other than a random deviation from an average. Examples for a few combinations of commodity, market1 and market2 are demonstrated in figure 3.

Figure 3. Price Differences vs Estimated Transaction Costs




Even though we use Law of One Price to estimate the transaction costs, there is still a significant difference between price differentials and transaction cost estimates. This can only be explained by the volatility of the food prices. Despite earlier results (3.1 – 3.3) indicating that price behavior supports the LOP, the excess variability of price differences that we observe contradicts that, and suggests that further research is required to understand price variability in staple food markets in Ghana.

The difference between transaction costs and transportation costs can proxy for several things at the same time: opportunity costs, insurance, marketing fees, etc. In the analysis below we concentrate on the opportunity costs, as this is the main component of the difference. We believe that opportunity costs of trading commodity C between locations A and B can be separated to a part that depends only on commodity, a part that depends on the location pair A-B mostly through distance, and a part that is related to specific market B. A location pair should affect opportunity costs since the time required to travel between various locations varies with distance. Commodities have different intrinsic risk due to perishability and differences in how actively they are traded.

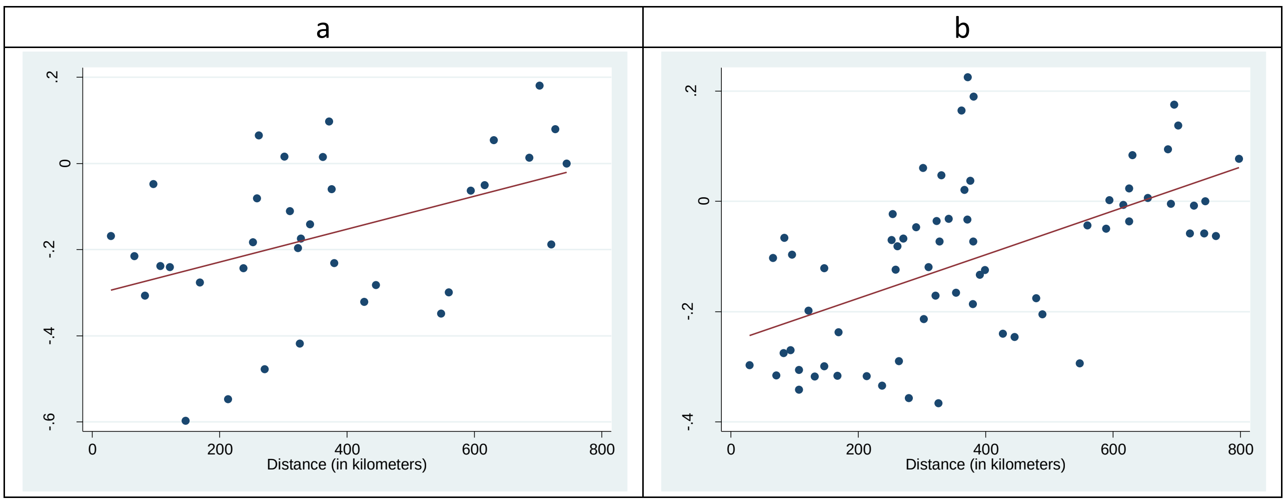
Marketing fees (loading-unloading fees and entry fees) and insurance are also a part of the difference between transaction costs and transportation costs. In the analysis that follows, we should keep in mind that we cannot disentangle between those fees and some of the premium estimates. Tables 14 and 15 allows us to rank commodities by premia involved in trading them. The commodity with the highest premia has the largest coefficient estimate. The safest commodity (ie lowest premium) is maize, the riskiest are cowpea and groundnuts according to both criteria.[11] We know that maize and yam are among most actively traded

[11]The complete ranking is very similar using each criterion: maize, millet, sorghum, yam, rice, cowpea, groundnuts if we use original transportation costs and maize, yam, millet, sorghum, rice, cowpea, groundnuts if we use reconstructed costs.

commodities, therefore, traders that are mainly involved in trading those commodities risk less than traders who are trading other commodities.

Next, we can take a closer look at the fixed effect of the locations pair. We observe a significant correlation between location fixed effects and distances: the correlation is 0.44 for original costs (scatter plot is presented in Figure 4a) and 0.56 for reconstructed costs (scatter plot is presented in Figure 4b).

Figure 4. Location effect and distance




We believe that for the most part distance between locations captures the opportunity cost of traders' time, however, there is a significant variation in transaction costs that is not captured either by nominal transportation costs or by distance. We think that a major part of this uncaptured variation is the premium related to trading at a specific market. We can compare the magnitude of this premium across different locations, as follows.

We take the estimates of the location effect from regression 10, $m_\mu$, that corresponds to some market pair $(j_1, j_2)$.[12] Define $m_{j_1 j_2} = m_\mu$ – an estimate from regression 10.

To find the effect of the destination market risk premium we estimate the following regression:.

$$m_{j_1 j_2} = \alpha + \beta \cdot DST_{j_1 j_2} + \sum_{k=2}^{13} \kappa_k \cdot I(j_1 = k \; or \; j_2 = k) + \varepsilon_{j_1 j_2} \tag{12}$$

where $DST_{j_1 j_2}$ – distance between market pairs, 13 – total number of markets, and $\varepsilon_{j_1 j_2}$ is an idiosyncratic error.

The results of the regression are in Tables 16 and 17.

[12] Note that $\mu$ indexes market pair from table A2.2, and each $j_1$ and $j_2$ indexes an individual market from table $A2.3$.

Table 16. Regression and Destination Market Risk Premium

| | Estimate | Std. Err. | T-stat |
|---|---|---|---|
| Constant | -0.3747 | 0.0768 | -4.88 |
| Distance | 0.0575 | 0.0116 | 4.95 |
| Ejura | 0.1581 | 0.0547 | 2.89 |
| Techiman | 0.1133 | 0.0554 | 2.04 |
| Mankessim | 0.0750 | 0.1249 | 0.6 |
| Tema | 0.0428 | 0.0552 | 0.78 |
| Sunyani | 0.0080 | 0.1161 | 0.07 |
| Accra | 0 | | |
| Koforidua | -0.0140 | 0.0631 | -0.22 |
| Tamale | -0.0188 | 0.0545 | -0.34 |
| Sekondi | -0.0344 | 0.0635 | -0.54 |
| Kumasi | -0.0692 | 0.0559 | -1.24 |
| Wa | -0.0842 | 0.0757 | -1.11 |
| Bolgatanga | -0.1301 | 0.0658 | -1.98 |
| Cape Coast | -0.2266 | 0.0915 | -2.48 |
| Num. obs. | 72 | | |
| R-squared | 0.48 | | |
| Adj. R-square | 0.37 | | |

With reconstructed costs

Table 17. Regression and Destination Market Risk Premium

| | Estimate | Std. Err. | T-stat |
|---|---|---|---|
| Constant | -0.3134 | 0.0398 | -7.87 |
| Distance | 0.0681 | 0.0047 | 14.61 |
| Techiman | 0.1276 | 0.0270 | 4.73 |
| Ejura | 0.0753 | 0.0270 | 2.79 |
| Accra | 0 | | |
| Sekondi | -0.0017 | 0.0270 | -0.06 |
| Sunyani | -0.0067 | 0.0310 | -0.22 |
| Tema | -0.0097 | 0.0269 | -0.36 |
| Kumasi | -0.0162 | 0.0310 | -0.52 |
| Koforidua | -0.0412 | 0.0309 | -1.33 |
| Mankessim | -0.0517 | 0.0309 | -1.67 |
| Tamale | -0.0585 | 0.0275 | -2.12 |
| Cape Coast | -0.0716 | 0.0310 | -2.31 |
| Wa | -0.0960 | 0.0284 | -3.38 |
| Bolgatanga | -0.1933 | 0.0296 | -6.53 |
| Num. obs. | 136 | | |
| R-squared | 0.67 | | |
| Adj. R-square | 0.64 | | |

The results indicate that Techiman and Ejura are the markets with highest risk premium (as we have define it), probably, reflecting both barriers to entry and low trading volumes. Sending goods to these markets is risky for traders, therefore, they require higher risk premium. Cape Coast and Mankessim are coastal markets, which naturally results in a low risk premium. Surprisingly, Accra, the capital market, is very high in the list. However, since this is a capital market, it may be very specific and access to it may be limited by local market queens, which results in high entry costs for traders, that would be one of the unobservable variables in our regression.

**6. Conclusion**

An important consideration in evaluating the extent to which market frictions exist in development contexts is whether simple relationships like market integration (MI) and the law of one price (LOP) hold. Empirical investigation of even such basic relationships is hampered by the fact that high quality data are generally not available, especially on transportation costs. In this paper, we collected such data and used existing methods to evaluate the extent of MI in Ghana as implied by the LOP. We found that generally a case can be made for MI, conditional on the distance between markets, which implies that the LOP can be weakly supported. We also show that we can estimate normally unobserved transaction costs, assuming the LOP that has been supported by other tests and using 3-dimensional panel data on prices and transportation costs. We conclude by applying our method to the premia analysis for specific commodities and markets.

## References

Abdoulai A. (2000). “Spatial Price Transmission and Asymmetry in the Ghanaian Maize Market”, *Journal of Development Economics,* 63, pp. 327–349.

Alexander C. and Wyeth J. (1994) “Cointegration and Market Integration: an Application to the Indonesian Rice Market”, *Journal of Development Studies*, 30 (2), pp. 303-328.

Alexander and Wyeth (1994). “Cointegration and market integration: An application to the Indonesian rice market”, *The Journal of Development Studies*, 30:2, 303-334.

Alderman and Shively. 1996. “Economic Reform and Food Prices: Evidence from Markets in Ghana,” *World Development*, v 24, 521-534.

Ankamah-Yeboah, I. (2012). “Spatial Price Transmission in the Regional Maize Markets in Ghana”, MPRA Paper No. 49720

Baulch B. (1997)., “Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration”, *American Journal of Agricultural Economics*, Vol. 79, No. 2, pp. 477-487.

Barrett, B., and Li J. (2002) “Distinguishing between Equilibrium and Integration in Spatial Price Analysis”, *American Journal of Agricultural Economics*, Vol. 84, No. 2, pp. 292-307.

Blyn G. (1973) “Price Series Correlation as a Measure of Market Integration”, *Indian journal of Agricultural Economics*, 28(1): 56-59.

Cournot A. A., and Fisher, I. (1897). *Researches into the Mathematical Principles of the Theory of Wealth.* New York: The Macmillan Company.

Said S. and Dickey D. (1984). “Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order”, *Biometrica,* 71, 3, pp. 599-607.

Goodwin and Piggot (2001). “Spatial Market Integration in the Presence of Threshold Effects” *American Journal of Agricultural Economics* 83(2): 302–317.

Fafchamps M. and Gavian S., (1996). “The Spatial Integration of Livestock Markets in Niger”, *Journal of African Economies*, vol. 5(3), pp: 366 – 405.

Headey, D. 2008. “Anatomy of a crisis: the causes and consequences of surging food prices”, Agricultural Economics, v. 39.

Harriss, B. (1979). “There is Method in My Madness: Or Is It Vice Versa? Measuring Agricultural Market Performance”, *Food Research Institute Studies*, XVH (2): 197-218.

Quaye, W. 2008. “Food Security Situation in Northern Ghana, Coping Strategies and Related Constraints,” African Journal of Agricultural Research, v 3(5), 334-342.

Kaminski, J., Christiaensen, L. and Christopher L. Gilbert (2016). “Seasonality in local food markets and consumption: evidence from Tanzania,” *Oxford Economic Papers*, 68(3), 736-757.

Ravallion M. (1986). Testing Market Integration", *American Journal of Agricultural Economics*, Vol. 68, No. 1, pp. 102-109.

Wang D. and Tomek W. (2007). Commodity Prices and Unit Root Tests, American Journal of Agricultural Economics 89(4): 873–889.

Penzhorn N. & Arndt C. (2002). Maize Markets in Mozambique: Testing For Market Integration

Devereux, S. and Maxwell, S. (2001). Food Security in sub-saharan Africa. Devereux, S.; Maxwell, S. (eds.). Wiley & Sons.

**Appendix A1: reconstructing transportation cost dataset.**

Step 1: we identify the most important factors that can affect the transportation cost. There are: distance, oil price, season, commodity factor, origin, and destination.

Step 2: We run a regression with transportation costs on the left-hand side and all factors we identified on the right-hand side

Step 3: we use the regression estimates to compute out-of-sample prediction for transportation costs.

We run the following regression:

$$TC_{cj_1j_2t} = \alpha + \beta_1 DST_{j_1j_2} + \beta_2 DST_{j_1j_2}^2 + \gamma Oil_t + S_r \cdot I(t \in R) + \sum_{\psi=2}^{7} \varphi_\psi \cdot I(c = \psi) + \sum_{\nu=2}^{29} l_\nu \cdot I(j_1 = \nu) + \sum_{\mu=2}^{29} m_\mu \cdot I(j_2 = \mu) + \varepsilon_{cj_1j_2t}$$

where $DST_{j_1j_2}$ is the distance between the markets, $Oil_t$ –price for gasoline in Ghana at date $t$,[13] $R$ is a set of dates that belong to major rainy season (from April to July of each year), and $S_r$ – seasonal effect, $\varphi_\psi$ – commodity effect, $l_\nu$ – is the origin effect, $m_\mu$ – is the destination effect, and $\varepsilon_{cj_1j_2t}$ is the idiosyncratic error. 7 is the total number of commodities, and 29 is the total number of locations. The full lists of commodities and locations are in the tables A1.1 and A1.2 below.

Table A1.1

| Index | Commodity |
|---|---|
| 1 | Cowpea |
| 2 | Groundnuts |
| 3 | Maize |
| 4 | Millet |
| 5 | Rice (local) |
| 6 | Sorghum |
| 7 | Yam |

Table A1.2

| Index | Location | Index | Location |
|---|---|---|---|
| 1 | Accra | 16 | Koforidua |
| 2 | Bawku | 17 | Kpassa |
| 3 | Bimbilla | 18 | Kumasi |
| 4 | Bole | 19 | Mankessim |
| 5 | Bolgatanga | 20 | Sefwi wiaso |
| 6 | Borae | 21 | Sekondi |
| 7 | Cape Coast | 22 | Sunyani |
| 8 | Damango | 23 | Tamale |
| 9 | Donkorkrom | 24 | Techiman |
| 10 | Ejura | 25 | Tema |

[13] Measured in Ghana Cedis, source: National Petroleum Authority

| | | | |
|---|---|---|---|
| 11 | Fumbisi | 26 | Tumu |
| 12 | Gushiegu | 27 | Wa |
| 13 | Hohoe | 28 | Yeji |
| 14 | Jirapa | 29 | Yendi |
| 15 | Kintampo | | |

The regression results are presented in the tables A1.3[14]
Table A1.3
Dependent variable: transportation cost

| | Estimate | Std. Err. | T-stat |
|---|---|---|---|
| Constant | -0.05424 | 0.00405 | -13.36 |
| Distance (100 km) | 0.02704 | 0.00092 | 29.31 |
| Distance square | -0.00211 | 0.00009 | -21.47 |
| Oil price | 0.0378 | 0.00052 | 71.82 |
| Rainy season effect ($S_R$) | 0.0023 | 0.00057 | 4.00 |
| Num. obs. | 9,820 | | |
| R-squared | .7308 | | |

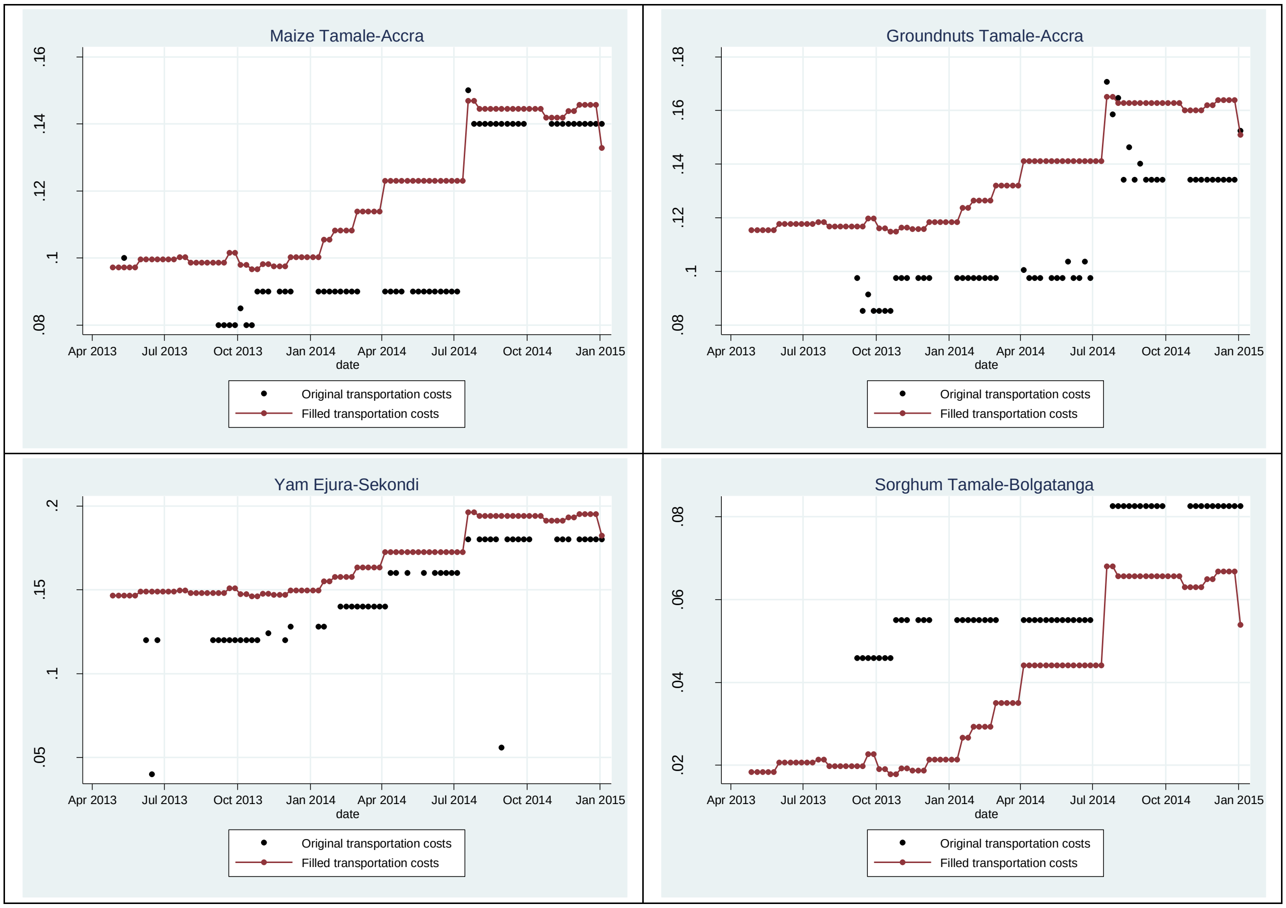
The reconstructed dataset is much more complete than the original one. It has a total of 461,020 observations, which is about 90% of all possible observations (for all commodities, dates, and all pairs of origin-destination). We were still not able to identify location effects and for some combinations of origin and destination and could not find the predicted values. They are recorded as 'missing'.

The regression results show that the fit is very decent: distance, oil price, commodity and location dummies explain 72% of the variation.

We also impose symmetry by assuming that $TC_{cj_1j_2t} = TC_{cj_2j_1t}$ for all $j_1$, $j_2$, $c$ and $t$. Technically, if predicted value is available for $c$, $j_1$, $j_2$, $t$ but not available for $c$, $j_2$, $j_1$, $t$ we make $TC_{cj_2j_1t} = TC_{cj_1j_2t}$. If it is available for both pairs, we take the average value.

Below we show four examples of original transportation cost series together with reconstructed for a few popular combinations of commodity-origin-destination. Note that in the reconstructed series the predicted value is taken for all dates, not only for those where we do not have original observation.

[14] To save the space the commodity and location coefficients are not reported but available upon request




**Appendix A2**

Table A2.1: full list of commodities

| Number | Commodity |
|---|---|
| 1 | Cowpea |
| 2 | Groundnuts |
| 3 | Maize |
| 4 | Millet |
| 5 | Rice (local) |
| 6 | Sorghum |
| 7 | Yam |

Table A2.2: full list of pairs of markets

| Number | Market 1 | Market 2 | Number | Market 1 | Market 2 |
|---|---|---|---|---|---|
| 1 | Accra | Bolgatanga | 40 | Ejura | Techiman |
| 2 | Accra | Cape Coast | 41 | Ejura | Tema |
| 3 | Accra | Ejura | 42 | Ejura | Wa |
| 4 | Accra | Koforidua | 43 | Koforidua | Kumasi |
| 5 | Accra | Kumasi | 44 | Koforidua | Mankessim |
| 6 | Accra | Mankessim | 45 | Koforidua | Sekondi |
| 7 | Accra | Sekondi | 46 | Koforidua | Sunyani |
| 8 | Accra | Sunyani | 47 | Koforidua | Tamale |
| 9 | Accra | Tamale | 48 | Koforidua | Techiman |
| 10 | Accra | Techiman | 49 | Koforidua | Tema |
| 11 | Accra | Tema | 50 | Koforidua | Wa |
| 12 | Accra | Wa | 51 | Kumasi | Mankessim |
| 13 | Bolgatanga | Cape Coast | 52 | Kumasi | Sekondi |
| 14 | Bolgatanga | Ejura | 53 | Kumasi | Sunyani |
| 15 | Bolgatanga | Koforidua | 54 | Kumasi | Tamale |
| 16 | Bolgatanga | Kumasi | 55 | Kumasi | Techiman |
| 17 | Bolgatanga | Mankessim | 56 | Kumasi | Tema |
| 18 | Bolgatanga | Sekondi | 57 | Kumasi | Wa |
| 19 | Bolgatanga | Sunyani | 58 | Mankessim | Sekondi |
| 20 | Bolgatanga | Tamale | 59 | Mankessim | Sunyani |
| 21 | Bolgatanga | Techiman | 60 | Mankessim | Tamale |
| 22 | Bolgatanga | Tema | 61 | Mankessim | Techiman |
| 23 | Bolgatanga | Wa | 62 | Mankessim | Tema |
| 24 | Cape Coast | Ejura | 63 | Mankessim | Wa |
| 25 | Cape Coast | Koforidua | 64 | Sekondi | Sunyani |
| 26 | Cape Coast | Kumasi | 65 | Sekondi | Tamale |
| 27 | Cape Coast | Mankessim | 66 | Sekondi | Techiman |
| 28 | Cape Coast | Sekondi | 67 | Sekondi | Tema |
| 29 | Cape Coast | Sunyani | 68 | Sekondi | Wa |
| 30 | Cape Coast | Tamale | 69 | Sunyani | Tamale |
| 31 | Cape Coast | Techiman | 70 | Sunyani | Techiman |
| 32 | Cape Coast | Tema | 71 | Sunyani | Tema |
| 33 | Cape Coast | Wa | 72 | Sunyani | Wa |
| 34 | Ejura | Koforidua | 73 | Tamale | Techiman |
| 35 | Ejura | Kumasi | 74 | Tamale | Tema |
| 36 | Ejura | Mankessim | 75 | Tamale | Wa |
| 37 | Ejura | Sekondi | 76 | Techiman | Tema |
| 38 | Ejura | Sunyani | 77 | Techiman | Wa |
| 39 | Ejura | Tamale | 78 | Tema | Wa |

Table A2.3: full list of markets

| Number | Commodity |
|---|---|
| 1 | Accra |
| 2 | Bolgatanga |
| 3 | Cape Coast |
| 4 | Ejura |
| 5 | Koforidua |
| 6 | Kumasi |
| 7 | Mankessim |
| 8 | Sekondi |
| 9 | Sunyani |
| 10 | Tamale |
| 11 | Techiman |
| 12 | Tema |
| 13 | Wa |