Measuring integration in the English wheat market, 1770-1820: new methods, new answers

Appendix 1: List of Counties

Bedfordshire	Lincolnshire
Berkshire	Middlesex
Buckinghamshire	Monmouthshire
Cambridgeshire	Norfolk
Cheshire	Northamptonshire
Cornwall	Northumberland
Cumberland	Nottinghamshire
Derbyshire	Oxfordshire
Devon	Rutlandshire
Dorset	Shropshire
Durham	Somerset
Essex	Staffordshire
Gloucestershire	Suffolk
Hampshire	Surrey
Herefordshire	Sussex
Hertfordshire	Warwickshire
Huntingdonshire	Westmorland
Kent	Wiltshire
Lancashire	Worcestershire
Leicestershire	Yorkshire

Appendix 2: Correlations of prices over time

In the main text, we asked how the pattern of prices changed from year to year and this was illustrated in Figure 3. This showed that cross-sectional variation in prices in one year was similar to the cross-sectional variation of prices in the following year. This raises the obvious question of how cross-sectional variation of prices changed over longer periods.

As in Figure 3, we start by calculating the within-harvest-year average price for each county

(A2.1)
$$\tilde{p}_{i.y} \equiv \sum_{w \in y} p_{i.w} \Big/ 52$$
$$\text{e.g. } \tilde{p}_{\text{Bedfords.1781-2}} \equiv \frac{p_{\text{Bedfords.1781,week 45}} + p_{\text{Bedfords.1781,week 46}} + \cdots + p_{\text{Bedfords.1782,week 44}}}{52}$$

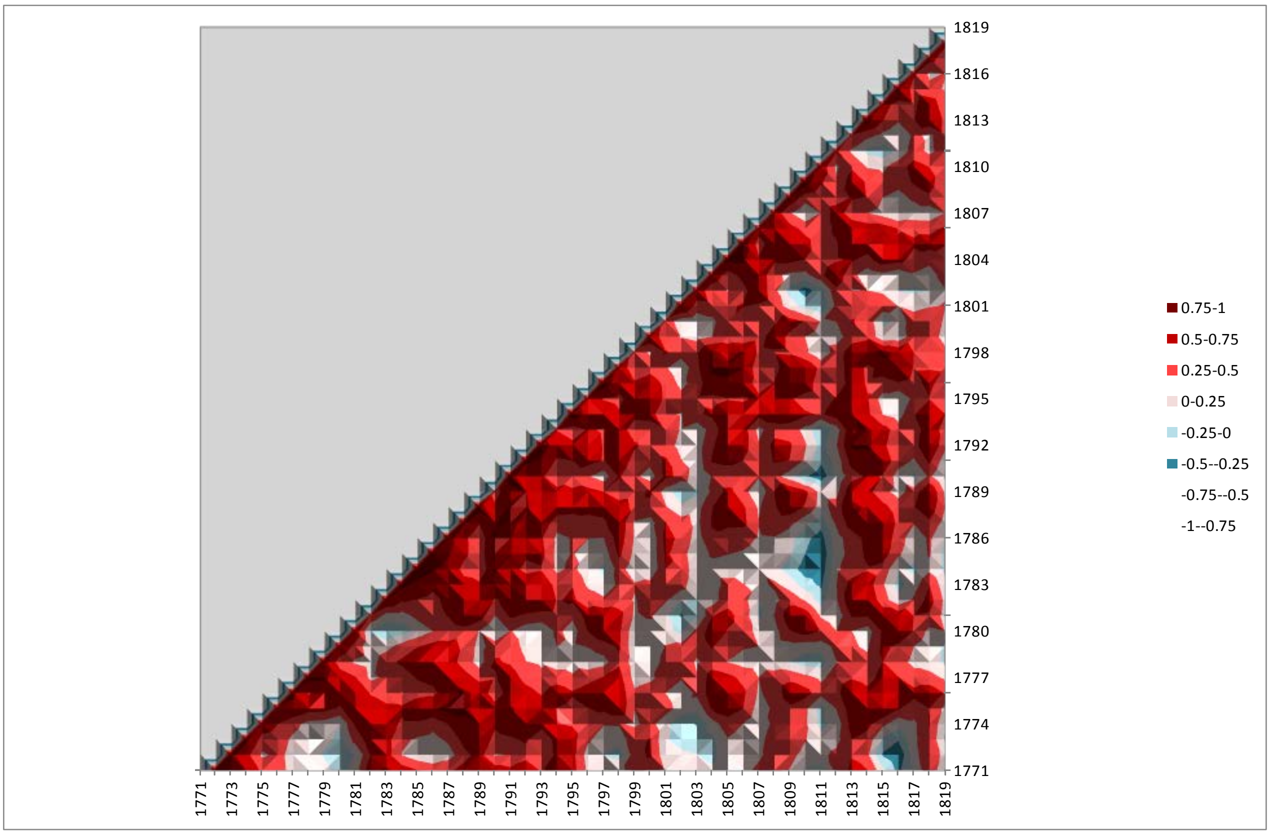
This means that for each harvest year from 1771/2 to 1819/20 we have 43 county prices. For any two harvest years, y and $y+x$, we then can calculate the correlation between the (average) prices for different counties using the conventional correlation coefficient

(A2.2)
$$\text{corr}\left(\tilde{p}_{i.y}, \tilde{p}_{i.y+x}\right) = \frac{\sum_{i=1}^{i=40}\left(\tilde{p}_{i.y} - \bar{\tilde{p}}_{y}\right)\left(\tilde{p}_{i.y+x} - \bar{\tilde{p}}_{y+x}\right)}{\sqrt{\sum_{i=1}^{i=40}\left(\tilde{p}_{i.y} - \bar{\tilde{p}}_{y}\right)^2 \sum_{i=1}^{i=43}\left(\tilde{p}_{i.y+x} - \bar{\tilde{p}}_{y+x}\right)^2}}$$

We illustrate the resulting 946 correlations in the implicitly three-dimensional diagram in Figure A3: the horizontal axis shows the year $y+x$ and the vertical axis year y. The correlation is shown by the colour of the diagram.

For example, if we look at the point corresponding to 1809 on the horizontal axis and 1789 on the vertical axis we see that the area is shaded dark red, so the correlation between county prices in 1789 and 1809 was between 0.50 and 0.75 (in fact it was 0.723). This means that the pattern of prices between different counties in 1789-90 (before the French revolution had really started) was very similar to the pattern of prices in 1809-10 (when France had just defeated Austria for the fifth time and Britain had just embarked on the Peninsular War). In fact, most of the diagram is red or brown, showing that the pattern of prices remained remarkably constant for most of the period.

Figure A2.1: Correlations of cross-sectional price series for all year-pairs



The graph plots the correlation of county prices in each year with all other years.

Appendix 3: Notation

In this section we carefully define our matrix notation. Recall that our most general model in equation (5) is:

(A3.1)
$$\begin{bmatrix}\Delta p_t^i\\ \Delta p_t^j\end{bmatrix}=\begin{bmatrix}\alpha_i\\ \alpha_j\end{bmatrix}\left(p_{t-1}^i-p_{t-1}^j\right)+\begin{bmatrix}\mu_i\\ \mu_j\end{bmatrix}+\begin{bmatrix}\pi_{i.i}^{(1)} & \pi_{i.j}^{(1)}\\ \pi_{j.i}^{(1)} & \pi_{j.j}^{(1)}\end{bmatrix}\begin{bmatrix}\Delta p_{t-1}^i\\ \Delta p_{t-1}^j\end{bmatrix}+\cdots+\begin{bmatrix}\pi_{i.i}^{(K)} & \pi_{i.j}^{(K)}\\ \pi_{j.i}^{(K)} & \pi_{j.j}^{(K)}\end{bmatrix}\begin{bmatrix}\Delta p_{t-K}^i\\ \Delta p_{t-K}^j\end{bmatrix}+\begin{bmatrix}\varepsilon_t^i\\ \varepsilon_t^j\end{bmatrix};$$
$$\begin{bmatrix}\varepsilon_t^i\\ \varepsilon_t^j\end{bmatrix}\sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix},\begin{bmatrix}\omega_{i.i} & \omega_{i.j}\\ \omega_{i.j} & \omega_{j.j}\end{bmatrix}\right);\qquad \alpha_i\le 0;\quad \alpha_j\ge 0;\quad \alpha_j-\alpha_i<1$$

where $\Delta p_t^i \equiv p_t^i - p_{t-1}^i$ which we refer to as the price change and we re-write this in vector notation as

(A3.2) $$\Delta\mathbf{p}_t=\boldsymbol{\alpha\gamma}\mathbf{p}_{t-1}+\boldsymbol{\mu}+\sum_{k=1}^{K}\boldsymbol{\pi}^{(k)}\Delta\mathbf{p}_{t-k}+\boldsymbol{\varepsilon}_t;\quad \boldsymbol{\varepsilon}_t\sim N\left(\mathbf{0},\boldsymbol{\Omega}\right);\quad \boldsymbol{\gamma}\equiv\begin{bmatrix}1 & -1\end{bmatrix}.$$

We use the common (although not universal) practice of denoting vectors and matrices with bold type and scalars in light type. The vectors and matrices are defined formally as

$$\mathbf{p}_t \equiv \begin{bmatrix} p_t^i \\ p_t^j \end{bmatrix}; \quad \Delta\mathbf{p}_t \equiv \begin{bmatrix} \Delta p_{t-s}^i \\ \Delta p_{t-s}^j \end{bmatrix} = \begin{bmatrix} p_t^i - p_{t-1}^i \\ p_t^j - p_{t-1}^j \end{bmatrix}; \quad \boldsymbol{\varepsilon}_t \equiv \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix}$$

(A3.3)
$$\boldsymbol{\alpha} \equiv \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix}; \quad \boldsymbol{\gamma} \equiv \begin{bmatrix} 1 & -1 \end{bmatrix}; \quad \boldsymbol{\mu} \equiv \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}; \quad \boldsymbol{\pi}^{(k)} \equiv \begin{bmatrix} \pi_{i.i}^{(k)} & \pi_{i.j}^{(k)} \\ \pi_{j.i}^{(k)} & \pi_{j.j}^{(k)} \end{bmatrix}$$

$$\boldsymbol{\Omega} \equiv \begin{bmatrix} \omega_{i.i} & \omega_{i.j} \\ \omega_{i.j} & \omega_{j.j} \end{bmatrix}; \quad \mathbf{I} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{0} \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

There are several points to note about $\boldsymbol{\alpha\gamma}$. First,

(A3.4)
$$\boldsymbol{\gamma}\mathbf{p}_t \equiv \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} p_t^i \\ p_t^j \end{bmatrix} = p_t^i - p_t^j$$

which is just the price gap. Second, the matrix $\boldsymbol{\alpha\gamma}$ has imposed three restrictions, two substantive and one an "identifying restriction". To see the first substantive restriction, notice that we could have written

(A3.5)
$$\Delta\mathbf{p}_t = \boldsymbol{\Psi}\mathbf{p}_{t-1} + \cdots; \qquad \boldsymbol{\Psi} \equiv \begin{bmatrix} \psi_{i.i} & \psi_{i.j} \\ \psi_{i.j} & \psi_{j.j} \end{bmatrix}$$

with no restriction on the four parameters. Even with a completely unrestricted version of $\boldsymbol{\gamma}$, which we denote $\boldsymbol{\gamma}^*$, by writing $\boldsymbol{\Psi} = \boldsymbol{\alpha\gamma}^*$ we have imposed the restriction that $\psi_{i.i}\psi_{j.j} = \psi_{i.j}\psi_{i.j}$: to see this note that

(A3.6)
$$\boldsymbol{\alpha\gamma}^* = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \begin{bmatrix} \gamma_i & \gamma_j \end{bmatrix} = \begin{bmatrix} \alpha_i\gamma_i & \alpha_i\gamma_j \\ \alpha_j\gamma_i & \alpha_j\gamma_j \end{bmatrix}.$$

Formally, the restriction consists in restricting the rank of the matrix $\boldsymbol{\Psi}$ to equal one (instead of two). When the individual price series follow unit root processes this restriction corresponds to saying that there is an equilibrium relationship and this could be tested using Johansen's maximum likelihood procedure. (If there is no unit root, then imposing the restriction still makes economic sense and can be tested using conventional t and F tests).

Despite this restriction, the form of the matrix in (A3.6) is underidentified because we could replace $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ with $2\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}/2$ without making any difference to the product $\boldsymbol{\alpha\gamma}$. This means we need a normalising restriction: although any of the parameters could be normalised, it is convenient here to restrict the model to

(A3.7) $$\boldsymbol{\alpha\gamma}^* = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \begin{bmatrix} 1 & \gamma_j \end{bmatrix}$$

The final restriction that we use in this paper is to restrict $\gamma_j = -1$. Again this can be tested within the Johansen maximum likelihood procedure. In our paper we impose this restriction: where we test the restriction it is virtually never rejected.

Appendix 4: The Constant and Seasonals

In the error-correction model the constant can be either restricted or unrestricted. In our specification we follow papers such as Marks (2010) and place no restriction on the constant terms in the vector $\boldsymbol{\mu}$, but in some published articles the constant appears to be constrained to lie in the cointegrating space (for example, in equation 5A.4 in Ejrnæs' appendix in Persson, 1999, p.157 of Ejrnæs, Persson and Rich, 2008); in others, it appears to be omitted altogether (such as in Buyst, Dercon and Van Campenhout, 2006). In this appendix we clarify what we mean by a restricted constant and discusses the consequences of differing modelling strategies.

The constant plays two rôles in the cointegrating model. For notational simplicity our exposition in this section ignores the lagged-dependent variables and we hence re-write equation (5) as

(A4.1) $$\begin{bmatrix} \Delta p_t^i \\ \Delta p_t^j \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \left(p_{t-1}^i - p_{t-1}^j \right) + \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix} + \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix};$$

this model is usually described as having an unrestricted constant. Alternatively it is also possible to restrict the constant to lie in the cointegrating space so that

(A4.2) $$\begin{bmatrix} \Delta p_t^i \\ \Delta p_t^j \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \left(p_{t-1}^i - p_{t-1}^j + \lambda \right) + \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix}.$$

In the restricted version, the equilibrium condition is that $p^i_{t-1} = p^j_{t-1} - \lambda$ (i.e. there is a systematic difference between the price levels). When the market is in equilibrium the expected price change is zero, which means that there is no systematic trend up or down in prices. Notice that this version of the model is the same as the first model with the cross-equation restriction that $\mu_j = \alpha_j \mu_i / \alpha_i$. When the restriction is not imposed, equation (A4.1) can be rewritten as

(A4.3)
$$\begin{bmatrix} \Delta p^i_t \\ \Delta p^j_t \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \left(p^i_{t-1} - p^j_{t-1} + \lambda \right) + \begin{bmatrix} \mu'_i \\ \mu'_j \end{bmatrix} + \begin{bmatrix} \varepsilon^i_t \\ \varepsilon^j_t \end{bmatrix};$$

which emphasises that there can be both a systematic difference between the two prices and a stochastic trend. When the restriction is valid, there is potentially an efficiency gain from imposing the restriction in the estimation; conversely imposing the restriction in the model when it is invalid will bias parameter estimates. It is always possible to test the restriction by using a likelihood ratio test.

In our data, as can be seen from Figure 1, there are often systematic price differences and the overall trend from 1770-1820 is quite small. This has the following consequences, where we summarise analysis that is not reported here or in the main paper.

If we omit the constant altogether and estimate a model for the whole period 1770-1820 then the result is that our other parameter estimates are highly biased, since we are imposing an invalid equilibrium condition that $p^i_{t-1} = p^j_{t-1}$. Typically the estimated half-life is biased up by a factor of as much as two.

If we restrict the constant to lie in the cointegrating space and estimate the model for the whole period 1770-1820, then there is a negligible effect on the estimated half-life. The reason for this is that prices at the end of the period are not much higher than at the beginning of the period and so the unrestricted estimate of the drift term is close to zero anyway: the restriction makes little difference

If we restrict the constant to lie in the cointegrating space and estimate the model for a sub-sample of the data, however, then the restriction can have a big impact on the estimated half-life. The reason for this is that, over various sub-samples, prices do go up or down by substantial amounts (as can be seen in Figure 1, for example 1803-1812) and therefore it is important to include a stochastic drift term in the model.

In principle we could use a sophisticated process by which the constant was sometimes restricted and sometimes unrestricted, using an appropriate test as the criterion for model selection. However, since we would invariably make some Type I errors, this would involve some invalid restrictions: on the other hand the gain in efficiency from imposing the restriction would be reduced whenever we made a Type II error. This might involve making inappropriate choices (as some tests would with the criteria). For this reason we choose never to restrict the constant.

Throughout the main text of the paper we omit seasonal dummies from our formulae for notational compactness (and when we analyse data at an annual frequency the issue of seasonals does not arise). When we include the seasonals for models estimated on weekly or monthly data the model becomes:

(A4.4) $$\begin{bmatrix} \Delta p_t^i \\ \Delta p_t^j \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \left(p_{t-1}^i - p_{t-1}^j \right) + \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix} + \sum_{w=1}^{51} \begin{bmatrix} \delta_w^i \\ \delta_w^j \end{bmatrix} + \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix};$$

(If the data were monthly then there would be eleven, rather than 51, seasonals). Restricting the seasonals to the cointegrating space would imply that there were no seasonal effects on expected price changes, but that the equilibrium relationship between the two prices changed over the year, which is a slightly strange assumption and not borne out by the facts (as prices show a seasonal pattern). For this reason we do not restrict the seasonals, either.

Appendix 5: The Half-life

A5.1: The half-life when there are no lagged price changes

In this section we discuss several technical issues with the half-life. When the VECM model has no lagged price changes, so that it can be written as

(A5.1) $$\Delta \mathbf{p}_t = \boldsymbol{\alpha\gamma} \mathbf{p}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t,$$

then the half-life is a sufficient statistic to describe adjustment back to equilibrium from a position of disequilibrium. This is because the decay in the price gap is geometric. As noted in the main text of the paper, the formula for the half-life in this instance is

(A5.2) $$HL \equiv \frac{\ln\left(0.5\right)}{\ln\left(1+\alpha_i-\alpha_j\right)} = \frac{\ln\left(0.5\right)}{\ln\left(1+\boldsymbol{\gamma\alpha}\right)}$$

If we had a sufficiently large sample then we could rely upon a consistency result that

(A5.3) $$\text{plim}\left[\frac{\ln(0.5)}{\ln(1+\gamma\hat{\alpha})}\right] = \frac{\ln(0.5)}{\ln(1+\gamma \mathsf{E}[\hat{\alpha}])} = \frac{\ln(0.5)}{\ln(1+\gamma\alpha)} = HL\,.$$

But some of our results are based on relatively small sub-samples of the data, so we wish to know the properties of

(A5.4) $$\frac{\ln(0.5)}{\ln(1+\gamma \mathsf{E}[\hat{\alpha}])} = \frac{\ln(0.5)}{\ln(1+\hat{\alpha}_i-\hat{\alpha}_j)}$$

in small samples. Most authors simply substitute $\hat{\alpha}_i-\hat{\alpha}_j$ into this formula to estimate the half-life. Although the half-life is an increasing function of $\alpha_i-\alpha_j$, when this quantity is less than about -0.57 the function is concave; thereafter it is convex. This suggests that the expected value of the half-life will not be the same as the half-life evaluated at the expected value of the parameters. However, in nearly all cases the standard error of $\hat{\alpha}_i-\hat{\alpha}_j$ is sufficiently small that it makes no difference: the reason for this is that $\hat{\alpha}_i$ and $\hat{\alpha}_j$ are highly negatively correlated and the variance of $\alpha_i-\alpha_j$ is correspondingly quite low.

As a further check, we tried a Monte Carlo procedure to see if the non-linearity made any difference. To do this we assumed that the disturbances had a Normal distribution (which is only approximately correct), so that

(A5.5) $$\widehat{\alpha_i-\alpha_j} \sim \mathsf{N}\left(\alpha_i-\alpha_j, \mathsf{var}\left(\widehat{\alpha_i-\alpha_j}\right)\right).$$

Note that

(A5.6) $$\mathsf{var}\left(\widehat{\alpha_i-\alpha_j}\right) = \mathsf{var}(\hat{\alpha}_i) + \mathsf{var}(\hat{\alpha}_j) - 2\,\mathsf{cov}(\hat{\alpha}_i,\hat{\alpha}_j).$$

Using this as a basis, we simulated 10,000 values $\alpha_i-\alpha_j$ from a Normal distribution $N\left(\widehat{\alpha_i-\alpha_j}, \mathrm{var}\left(\widehat{\alpha_i-\alpha_j}\right)\right)$ and calculated the corresponding 10,000 half-lives (in a very small number of cases the draw of $\alpha_i-\alpha_j$ was negative and these were discarded). We then averaged the 10,000 replications and compared the mean to the conventionally calculated half-life. We found that in nearly all cases the standard error of $\widehat{\alpha_i-\alpha_j}$ was sufficiently small that it made no real difference which method we used.

A5.2: The half-life when there are lagged price changes

In the general case there are lagged price changes and the model can be written as

$$(A5.7) \qquad \Delta \mathbf{p}_t = \boldsymbol{\alpha}\boldsymbol{\gamma}\mathbf{p}_{t-1} + \boldsymbol{\mu} + \sum_{k=1}^{K} \boldsymbol{\pi}^{(k)} \Delta \mathbf{p}_{t-k} + \boldsymbol{\varepsilon}_t.$$

In this case there is no single measure which summarises the speed of adjustment. To understand the difference between this situation and that in (A5.1), consider the hypothetical possibility that prices in period $t-1$ were $p_{t-1}^i - p_{t-1}^j = 0.1$ so that price *i* were approximately ten per cent higher than price *j*. Conceptually, we can distinguish two simple processes that could have resulted in this price gap: either (i) $p_{t-2}^i - p_{t-2}^j = 0$ and there was a shock to the prices in period $t-1$; or (ii) there was no shock in period $t-1$ but $p_{t-1}^i - p_{t-1}^j > 0.1$ and the price gap existed in $t-1$ because prices had not yet fully adjusted back to equilibrium after a shock in period $t-2$ or earlier. With the model in (A5.1), the price adjustment in period t would be identical: it is as if the process generating prices had "forgotten" how the disequilibrium had arisen. With the more general model, the price behaviour in period t would depend upon whether the price gap had arisen from situation (i) or situation (ii). Since the adjustment in period t depends upon the earlier behaviour of prices there cannot be a single measure of the speed of adjustment.

Despite this we wish to summarise the speed of adjustment, even if our measure be imperfect. The method we choose is to plot an impulse response function like that in Figure 6 and then see where this curve crosses the horizontal line $y = 0.5$.

This raises the further issue of how to plot the impulse response function and this is complicated because the impulse response function to a shock to price *i* (i.e. due to a shock ε_t^i) may differ from a shock to price *j* (i.e. due to a shock ε_t^j). In the three-price context of New York, London and Copenhagen, Ejrnæs, Persson and Rich (2006) illustrate impulse response functions to shocks in all three cities to all three price series. In a two-price context, these correspond to the effect on prices of shocks of the form

$$(A5.8) \qquad \text{either } \boldsymbol{\varepsilon}_t \equiv \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ or } \boldsymbol{\varepsilon}_t \equiv \begin{bmatrix} \varepsilon_t^i \\ \varepsilon_t^j \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

i.e. a shock to one price with no effect on the other. The problem with this is that it is rare for price shocks to occur in isolation and we know that ε_t^i and ε_t^j are correlated (as is illustrated in Figure 9). Pesaran and Shin (1996) suggest a method for calculating the impulse response function as follows. First, re-write the VECM of equation (A5.7) in the form of a VAR:

(A5.9)
$$\mathbf{p}_t = \mathbf{\Phi}_1 \mathbf{p}_{t-1} + \mathbf{\Phi}_2 \mathbf{p}_{t-2} + \cdots + \mathbf{\Phi}_{K+1} \mathbf{p}_{t-K-1} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$
$$\mathbf{\Phi}_1 \equiv \left(\mathbf{I} + \boldsymbol{\alpha\gamma} + \boldsymbol{\pi}_1\right); \quad \mathbf{\Phi}_2 \equiv \left(\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1\right); \ldots \quad \mathbf{\Phi}_K \equiv \left(\boldsymbol{\pi}_K - \boldsymbol{\pi}_{K-1}\right); \quad \mathbf{\Phi}_{K+1} \equiv -\boldsymbol{\pi}_K$$

From a hypothetical position of equilibrium $p_{t-2}^i - p_{t-2}^j = 0$, the effect of a shock in period *s* relative to what would have happened if there had been no shock is then calculated iteratively via

(A5.10)
$$\begin{aligned} \mathbf{p}_0 &= \boldsymbol{\varepsilon}_0 \\ \mathbf{p}_1 &= \mathbf{\Phi}_1 \boldsymbol{\varepsilon}_0 \\ \mathbf{p}_2 &= \mathbf{\Phi}_1 \mathbf{\Phi}_1 \boldsymbol{\varepsilon}_0 + \mathbf{\Phi}_2 \boldsymbol{\varepsilon}_0 \\ \mathbf{p}_3 &= \mathbf{\Phi}_1 \left(\mathbf{\Phi}_1 \mathbf{\Phi}_1 \boldsymbol{\varepsilon}_0 + \mathbf{\Phi}_2 \boldsymbol{\varepsilon}_0\right) + \mathbf{\Phi}_2 \mathbf{\Phi}_1 \boldsymbol{\varepsilon}_0 + \mathbf{\Phi}_3 \boldsymbol{\varepsilon}_0 \\ &\vdots \end{aligned}$$

Pesaran and Shin (1996) suggest that a potential and natural shock to use is

(A5.11)
$$\sqrt{\boldsymbol{\gamma\Omega\gamma}'} = \sqrt{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \omega_{i.i} & \omega_{i.j} \\ \omega_{i.j} & \omega_{j.j} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}} = \sqrt{\omega_{i.i} + \omega_{j.j} - 2\omega_{i.j}}$$

and they then calculate the effect of this shock. To implement this, they define

(A5.12)
$$\mathbf{B}_s \equiv \mathbf{0} \quad \text{for} \quad s < 0$$
$$\mathbf{B}_0 \equiv \mathbf{I}$$
$$\mathbf{B}_s \equiv \sum_{k=1}^{K} \Phi_k \mathbf{B}_{s-k} \quad \text{for} \quad s > 0$$

in which case the impulse response function (normalised by adjusting for the variance of the original shock) is

(A5.13)
$$i(s) = \sqrt{\frac{\gamma \mathbf{B}_s \Omega \mathbf{B}_s' \gamma'}{\gamma \Omega \gamma'}}$$

which is the formula used to derive the functions in Figure 6.

Appendix 6: Decomposing the RMS price difference when there are lagged dependent variables.

In this section we derive the decomposition of equation (9) more formally. Recall the general VECM from equation (6)

(A6.1) $$\Delta\mathbf{p}_t = \boldsymbol{\alpha\gamma}\mathbf{p}_{t-1} + \boldsymbol{\mu} + \sum_{k=1}^{K} \boldsymbol{\pi}^{(k)} \Delta\mathbf{p}_{t-k} + \boldsymbol{\varepsilon}_t$$

which, by adding $\mathbf{p}_{t-1}$ to both sides, can be re-written as

(A6.2) $$\mathbf{p}_t = \left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\mathbf{p}_{t-1} + \boldsymbol{\mu} + \sum_{k=1}^{K} \boldsymbol{\pi}^{(k)} \Delta\mathbf{p}_{t-k} + \boldsymbol{\varepsilon}_t .$$

Multiplying by γ gives

(A6.3) $$\boldsymbol{\gamma}\mathbf{p}_t = \boldsymbol{\gamma}\left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\mathbf{p}_{t-1} + \boldsymbol{\gamma\mu} + \sum_{k=1}^{K} \boldsymbol{\gamma\pi}^{(k)} \Delta\mathbf{p}_{t-k} + \boldsymbol{\gamma\varepsilon}_t .$$

Consider first the simplest case where there are no lagged differences so that $\boldsymbol{\pi}^{(k)} = \mathbf{0}$. Note also that $\boldsymbol{\gamma}\left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right) = \boldsymbol{\gamma} + \boldsymbol{\gamma\alpha\gamma} = \left(1 + \boldsymbol{\gamma\alpha}\right)\boldsymbol{\gamma}$, so that the simple case becomes

(A6.4) $$\boldsymbol{\gamma}\mathbf{p}_t = \left(1 + \boldsymbol{\alpha\gamma}\right)\boldsymbol{\gamma}\mathbf{p}_{t-1} + \boldsymbol{\gamma\mu} + \boldsymbol{\gamma\varepsilon}_t$$

or, since the formula consists of scalars,

(A6.5) $$\left(p_t^i - p_t^j\right) = \left(1 + \alpha_i - \alpha_j\right)\left(p_{t-1}^i - p_{t-1}^j\right) + \left(\mu_i + \mu_j\right) + \left(\varepsilon_t^i + \varepsilon_t^j\right).$$

Squaring this formula and taking expectations yields (in matrix and scalar notation respectively)

(A6.6) $$\begin{aligned} \mathsf{E}\left[\boldsymbol{\gamma}\mathbf{p}_t\mathbf{p}_t'\boldsymbol{\gamma}'\right] &= \left(1 + \boldsymbol{\gamma\alpha}\right)^2 \mathsf{E}\left[\boldsymbol{\gamma}\mathbf{p}_{t-1}\mathbf{p}_{t-1}'\boldsymbol{\gamma}'\right] + \boldsymbol{\gamma\mu\mu}'\boldsymbol{\gamma}' + \boldsymbol{\gamma\Omega\gamma}' \\ \mathsf{E}\left[\left(p_t^i - p_t^j\right)^2\right] &= \left(1 + \alpha_i - \alpha_j\right)^2 \mathsf{E}\left[\left(p_{t-1}^i - p_{t-1}^j\right)^2\right] + \left(\mu_i - \mu_j\right)^2 + \omega_{i.i} + \omega_{j.j} + 2\omega_{i.j} \end{aligned},$$

which is the same as equation (10) as discussed in the text: the squared price difference *this period* depends upon the shocks *this period*; the constants; and the price dispersion *in the previous period* multiplied by a term showing the speed of adjustment.

The more general case is messier but has the same underlying intuition. From equation (A6.3) we obtain

$$\text{(A6.7)}\quad \begin{aligned} \mathrm{E}\left[\gamma \mathbf{p}_t \mathbf{p}_t' \gamma'\right] &= \left(1+\gamma\alpha\right)^2 \mathrm{E}\left[\gamma \mathbf{p}_{t-1}\mathbf{p}_{t-1}'\gamma'\right] + \gamma\mu\mu'\gamma' + \gamma\Omega\gamma' \\ &+\mathrm{E}\left[\sum_{k=1}^{K}\gamma\pi^{(k)}\Delta\mathbf{p}_{t-k}\mathbf{p}_{t-1}'\gamma' + \gamma\mathbf{p}_{t-1}\sum_{l=1}^{K}\Delta\mathbf{p}_{t-l}'\pi^{(l)\prime}\gamma'\right]. \\ &+\mathrm{E}\left[\sum_{l=1}^{K}\sum_{k=1}^{K}\gamma\pi^{(k)}\Delta\mathbf{p}_{t-k}\Delta\mathbf{p}_{t-l}'\pi^{(l)\prime}\gamma'\right] \end{aligned}$$

The first row of the right-hand side of this formula is the same as (A6.6): the difference lies in the complicated set of terms in the second and third rows. What these terms denote are the adjustments to price dispersion in previous periods. Recall that the presence of lagged price changes corresponds to a complicated adjustment process to price dispersion in previous periods. Therefore, to describe perfectly the adjustment back towards equilibrium requires a full knowledge of the behaviour of prices over the previous k periods. In our analysis in section 5 we summarise this with a single statistic, namely the half-life, calculated as described in the text.

Appendix 7: Within-period price adjustment and the interpretation of the parameter $\omega_{i,j}$

Our data in this paper are weekly data and many of the markets we analysed opened only one or two days of the week. This means that there is no or very little temporal aggregation of the form discussed by Taylor (2001). However, even without temporal aggregation, infrequent sampling affects our interpretation of some of the parameters. In this appendix we consider the effect on the parameters in which we are most interested in this paper. For expositional purposes we consider the simplest version of our model, namely

$$\text{(A7.1)}\quad \begin{bmatrix}\Delta p_t^i \\ \Delta p_t^j\end{bmatrix} = \begin{bmatrix}\alpha_i \\ \alpha_j\end{bmatrix}\left(p_{t-1}^i - p_{t-1}^j\right) + \begin{bmatrix}\mu_i \\ \mu_j\end{bmatrix} + \begin{bmatrix}\varepsilon_t^i \\ \varepsilon_t^j\end{bmatrix};$$

which can more conveniently be written as

$$\text{(A7.2)}\quad \mathbf{p}_t = \left(\mathbf{I}+\alpha\gamma\right)\mathbf{p}_{t-1} + \mu + \varepsilon_t.$$

To separate the effects of infrequent sampling and time aggregation, let us assume that markets traded twice per week and suppose (counter-factually) that we observed end-of-week prices: this would mean that we would observe prices only for $t = \{2, 4, 6, \ldots\}$.

Then the relationship between one end-of-week price and the previous end-of-week price would be

$$\mathbf{p}_t = \left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\underbrace{\left\{\left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\mathbf{p}_{t-2} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t-1}\right\}}_{=\mathbf{p}_{t-1}} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

$$= \left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)^2 \mathbf{p}_{t-2} + \left(2\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\boldsymbol{\mu} + \left\{\left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t\right\}$$

(A7.3) $$\mathbf{p}_t - \mathbf{p}_{t-2} = \left(2\boldsymbol{\alpha} + \boldsymbol{\alpha\gamma\alpha}\right)\boldsymbol{\gamma}\mathbf{p}_{t-2} + \left(2\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\boldsymbol{\mu} + \left\{\left(\mathbf{I} + \boldsymbol{\alpha\gamma}\right)\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t\right\}$$

$$\left(2\boldsymbol{\alpha} + \boldsymbol{\alpha\gamma\alpha}\right) = \begin{bmatrix} \alpha_i\left(2 + \alpha_i - \alpha_j\right) \\ \alpha_j\left(2 + \alpha_i - \alpha_j\right) \end{bmatrix}$$

If we change the dating convention this can be re-written as

(A7.4) $$\Delta\mathbf{p}_w \equiv \mathbf{p}_w - \mathbf{p}_{w-1} = \boldsymbol{\alpha}^*\boldsymbol{\gamma}\mathbf{p}_{w-1} + \boldsymbol{\mu}^* + \boldsymbol{\varepsilon}_w^*$$

where the stars indicates the parameters from the weekly data. We can now ask what parameters we shall estimate. The loadings will be

(A7.5) $$\boldsymbol{\alpha}^* \equiv \left(2\boldsymbol{\alpha} + \boldsymbol{\alpha\gamma\alpha}\right) = \begin{bmatrix} \alpha_i\left(2 + \alpha_i - \alpha_j\right) \\ \alpha_j\left(2 + \alpha_i - \alpha_j\right) \end{bmatrix}$$

Our estimated speed of adjustment will be based on

(A7.6) $$1 + \alpha_i^* - \alpha_j^* = 1 + \alpha_i\left(2 + \alpha_i - \alpha_j\right) - \alpha_j\left(2 + \alpha_i - \alpha_j\right) = \left(1 + \alpha_i - \alpha_j\right)^2$$

which confirms that we shall estimate exactly the same speed of adjustment: the difference between $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}$ is entirely due to the different units of measurement (weekly versus half-weekly respectively).

When we turn to the disturbances, whose covariance matrix can be derived as follows:

$$\begin{aligned}\text{var}\left[\varepsilon_w^*\right] &= \text{E}\left[\left\{\left(\mathbf{I}+\boldsymbol{\alpha\gamma}\right)\varepsilon_{t-1}+\varepsilon_t\right\}\left\{\left(\mathbf{I}+\boldsymbol{\alpha\gamma}\right)\varepsilon_{t-1}+\varepsilon_t\right\}'\right] \\ &= \left(\mathbf{I}+\boldsymbol{\alpha\gamma}\right)\boldsymbol{\Omega}\left(\mathbf{I}+\boldsymbol{\alpha\gamma}\right)'+\boldsymbol{\Omega} \\ &= 2\boldsymbol{\Omega}+\left\{\boldsymbol{\alpha\gamma\Omega}+\boldsymbol{\Omega\gamma}'\boldsymbol{\alpha}'+\boldsymbol{\alpha\gamma\Omega\gamma}'\boldsymbol{\alpha}'\right\}\end{aligned}$$

(A7.7)

$$\begin{aligned} &= 2\begin{bmatrix}\omega_{i.i} & \omega_{i.j} \\ \omega_{i.j} & \omega_{j.j}\end{bmatrix} \\ &\quad+\begin{bmatrix}2\alpha_i\left(\omega_{i.i}-\omega_{i.j}\right) & \alpha_i\left(\omega_{i.j}-\omega_{j.j}\right)+\alpha_j\left(\omega_{i.i}-\omega_{i.j}\right) \\ \alpha_i\left(\omega_{i.j}-\omega_{j.j}\right)+\alpha_j\left(\omega_{i.i}-\omega_{i.j}\right) & 2\alpha_j\left(\omega_{i.j}-\omega_{j.j}\right)\end{bmatrix} \\ &\quad+\left(\omega_{i.i}-2\omega_{i.j}+\omega_{j.j}\right)\begin{bmatrix}\alpha_i^2 & \alpha_i\alpha_j \\ \alpha_i\alpha_j & \alpha_j^2\end{bmatrix}\end{aligned}$$

The covariance between the disturbances from weekly data consists of the actual covariance of the underlying (half-weekly) disturbances (the parameter $\omega_{i.j}$) and the adjustment which takes place within the week, which is

(A7.8)

$$\alpha_i\left(\omega_{i.j}-\omega_{j.j}\right)+\alpha_j\left(\omega_{i.i}-\omega_{i.j}\right)+\alpha_i\alpha_j\left(\omega_{i.i}-2\omega_{i.j}+\omega_{j.j}\right)$$

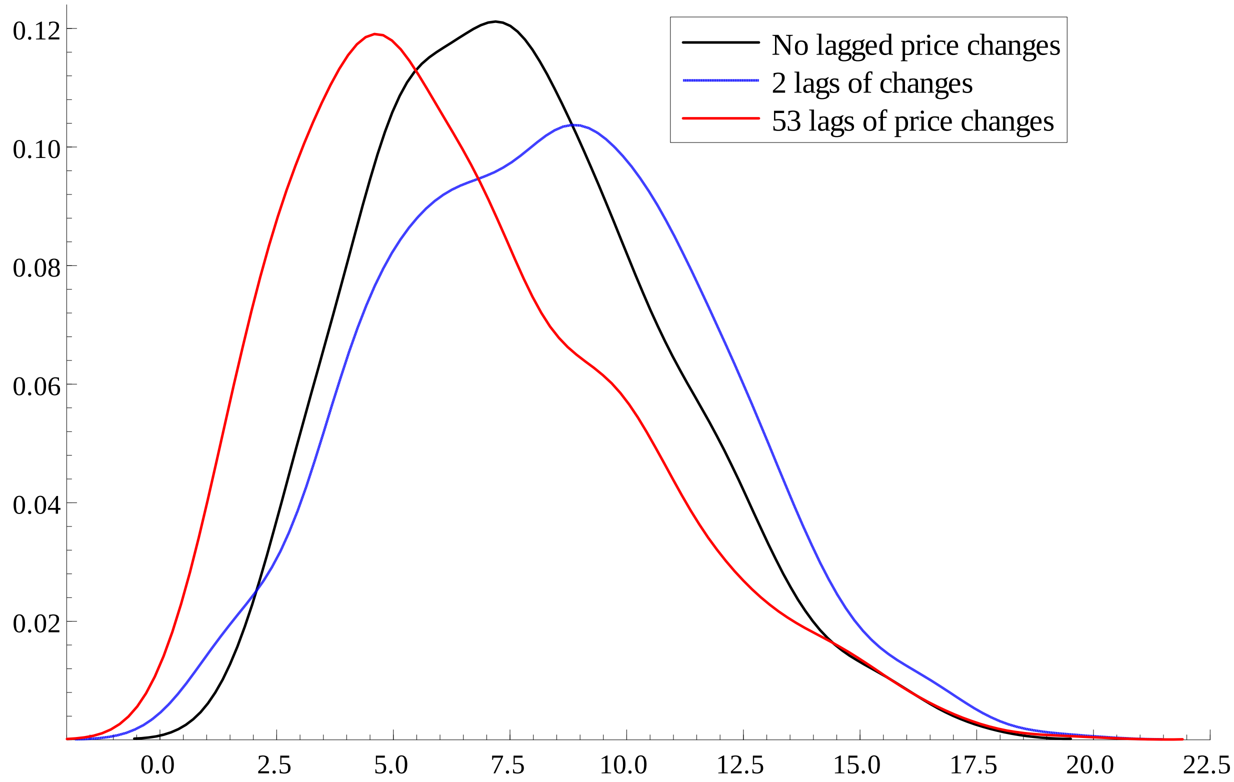
Appendix 8: Additional Tables and Figures

Table A8.1: Summary of half-lives estimated for 1770-1820

	All county pairs			Adjacent-county pairs		
frequency of data	weekly	monthly	annual	weekly	monthly	annual
mean	8.0	10.8	22.1	4.1	6.9	22.9
median	7.7	10.4	20.7	3.7	6.4	21.2
st.dev.	3.2	3.7	9.0	1.4	2.2	10.9
minimum	1.5	3.3	5.5	1.5	3.3	5.5
maximum	18.7	25.8	65.0	8.0	14.0	60.9

This table describes the same econometric analysis as that illustrated in Figure 5. The first three columns summarise the distribution of 780 half-lives (slightly fewer for annual data, where some half-lives could not be calculated). Each half-life is estimated from a regression of the form reported in equation (9) using data from the entire period 1770-1820, except where one of the prices is from London, when it is 1770-1793. The final three columns report analogous statistics for the 103 pairs where the counties are adjacent.

Figure A8.2: Distributions of half-life estimates



These half-lives correspond to the estimation in section 4.1, where a single model is estimated for the whole period (ignoring issues of parameter instability over the period).

Each of these distributions summarises the half-lives from 780 regressions, each of which is estimated on weekly data for a county pair for the entire period 1770-1820. Every county pair is estimated, not just adjacent-county pairs.

The only difference between the distributions is the number of lagged price changes used in the regression. These distributions are based on the same information as the third row of Table 1. Note that all the half-lives were positive but an artefact of the kernel smoothing method used to estimate the density was that the curves appear to extend to the left of the origin.

Table A8.3: Regressions using different measures of market integration

Lags in 1st-stage VAR:	0	1	2	3
Dependent variable: Average standard deviation of disturbances				
Roads	-0.030 (2.517)	-0.024 (1.941)	-0.025 (1.965)	-0.026 (2.056)
Canals	-0.171 (2.440)	-0.196 (2.641)	-0.205 (2.720)	-0.204 (2.687)
Newspapers	-0.073 (0.264)	-0.034 (-0.117)	-0.124 (0.418)	-0.111 (0.366)
N × T	4642	4642	4642	4642
R-squared	0.722	0.688	0.680	0.674
Dependent variable: Ratio of standard deviations of disturbances				
Roads	-0.015 (2.526)	-0.014 (2.250)	-0.013 (2.079)	-0.013 (2.087)
Canals	0.011 (0.305)	0.015 (0.465)	0.005 (0.142)	0.005 (0.132)
Newspapers	-0.705 (3.874)	-0.709 (3.760)	-0.674 (3.524)	-0.601 (3.216)
N × T	4642	4642	4642	4642
R-squared	0.090	0.085	0.084	0.086
Dependent variable: Correlation of disturbances				
Roads	0.002 (0.640)	0.004 (1.236)	0.003 (1.039)	0.003 (1.122)
Canals	0.032 (1.987)	0.029 (1.717)	0.032 (1.851)	0.036 (2.060)
Newspapers	0.202 (3.044)	0.213 (3.320)	0.188 (2.939)	0.187 (2.953)
N × T	4642	4642	4642	4642
R-squared	0.340	0.316	0.313	0.302
Dependent variable: Half-life				
Roads	0.067 (2.316)	0.063 (1.673)	0.039 (1.044)	0.064 (1.508)
Canals	0.307 (2.131)	0.199 (0.852)	0.397 (1.771)	0.281 (1.181)
Newspapers	-0.296 (0.332)	1.132 (1.084)	0.928 (1.051)	-0.243 (0.195)
N × T	4564	4384	4330	4308
R-squared	0.051	0.032	0.030	0.032

Results are for sixteen separate regressions. The explained variables are themselves estimated from regressions on weekly data for each county pair: the column headings refer to the number of lags in the first-stage time-series regressions.

Appendix 9: Figures in Colour

The figures in the main body of the paper are re-produced in black and white: this section contains the colour analogues.

Figure 1: Wheat prices 1770-1820

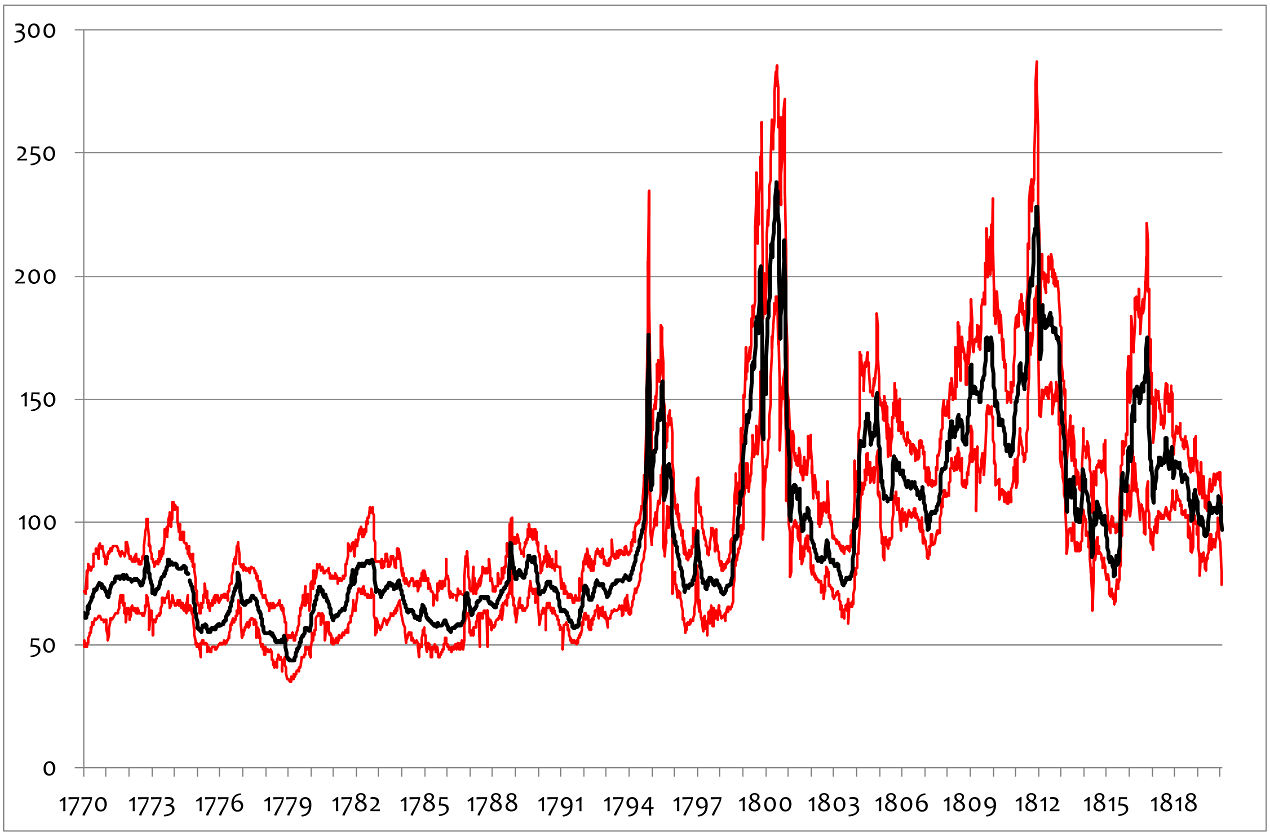
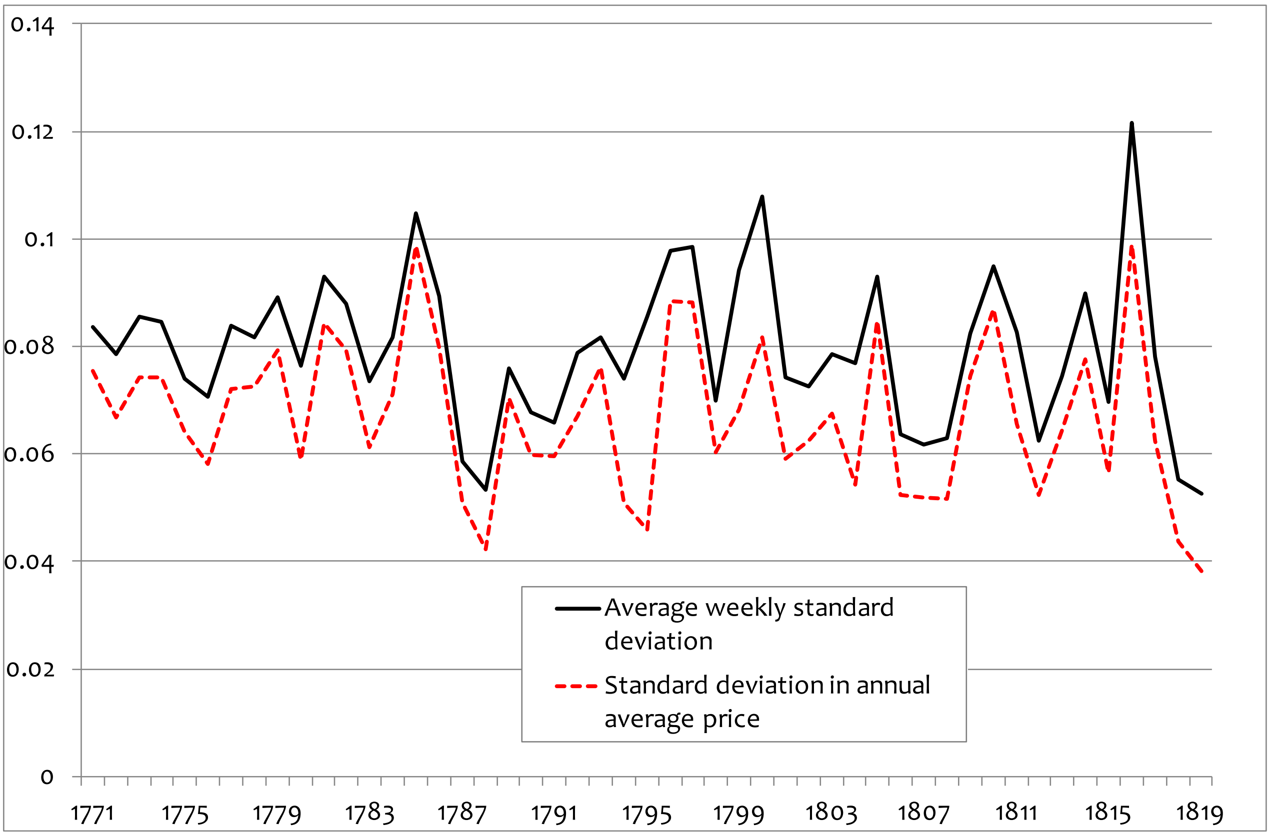


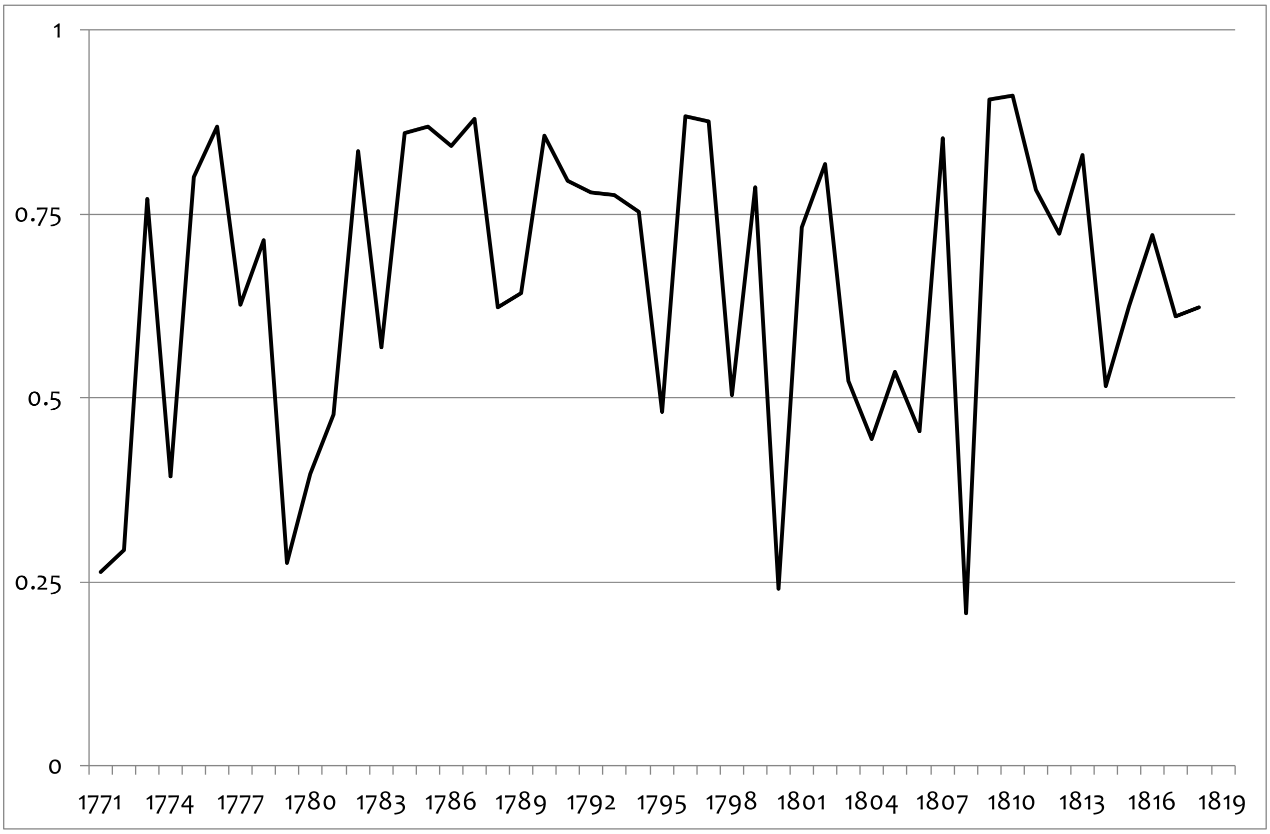
Figure shows the minimum, maximum and average London Gazette wheat price in each week from November 1770 to September 1820.

Figure 2: Dispersion of prices between counties



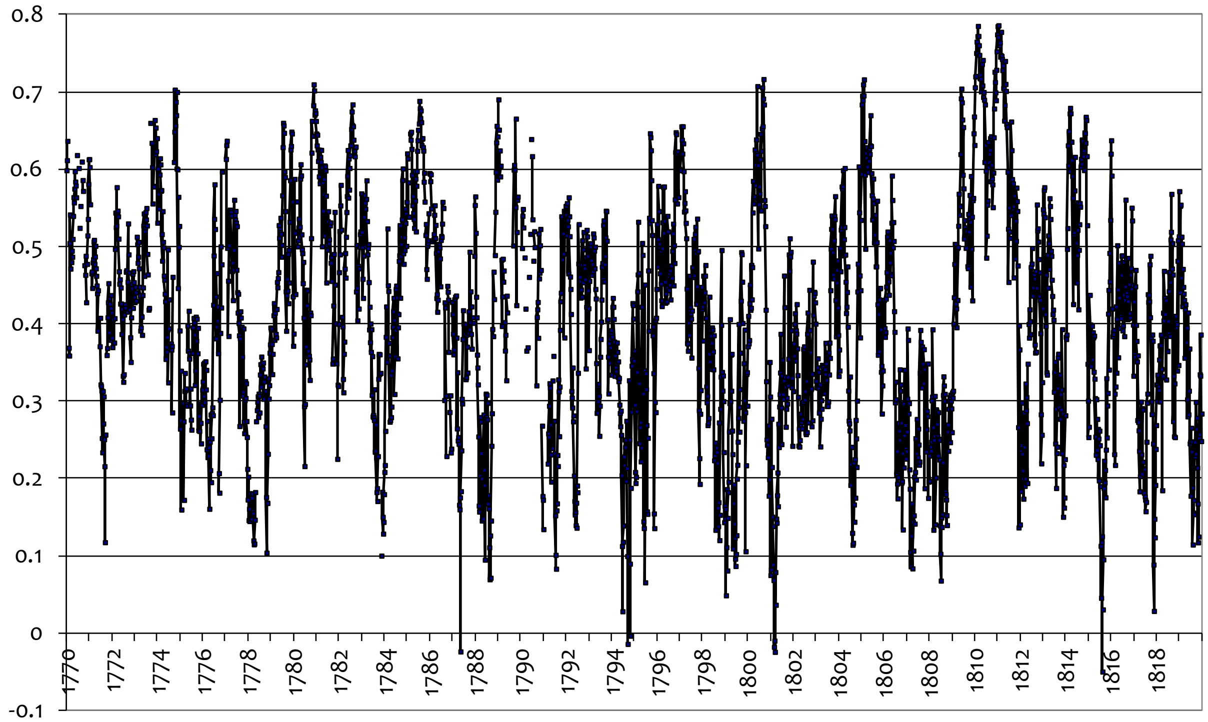
Average weekly standard deviation: the standard deviation of log prices is calculated for each week of the sample and then the 52 standard deviations are averaged for a harvest year (October-September). Standard deviation in annual average: the harvest-year mean price is calculated for each county and then the standard deviation is calculated of the forty mean prices.

Figure 3: Year-on-year correlations of cross-sections of prices



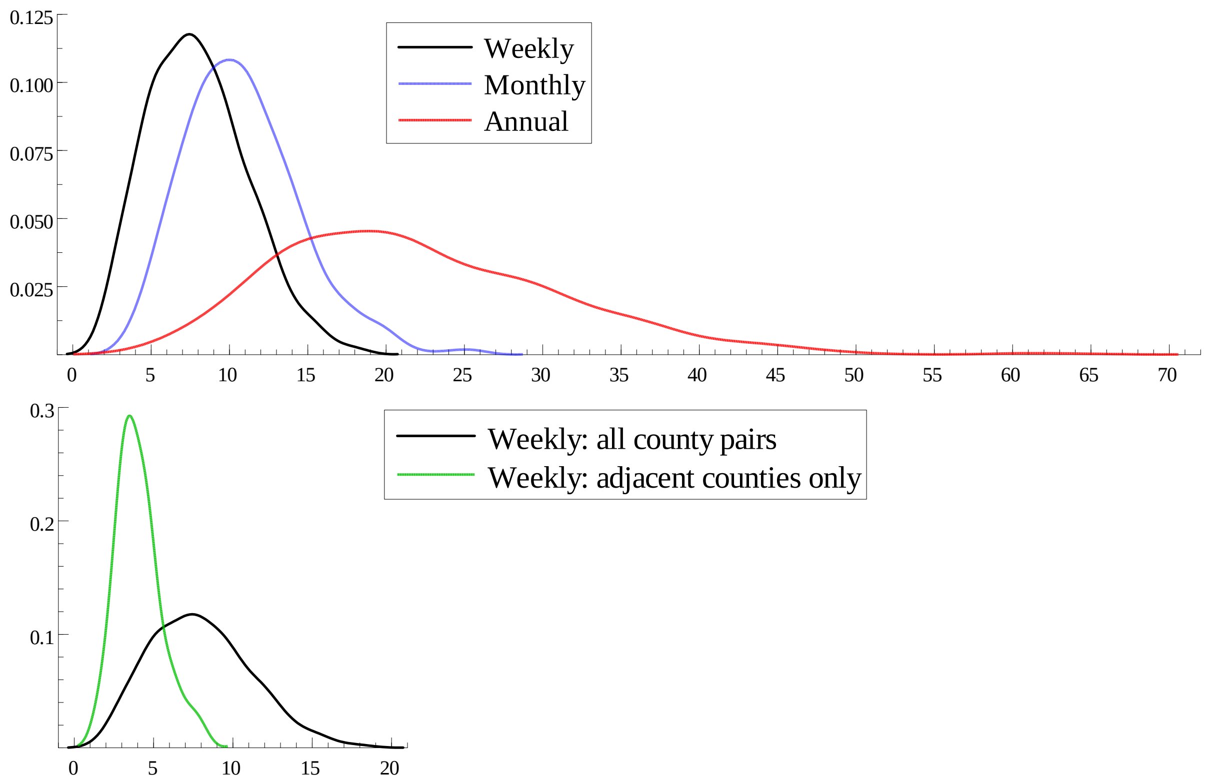
The graph plots the correlations of county prices in each year with prices in the following year (equation 2).

Figure 4: Moran's I Statistics



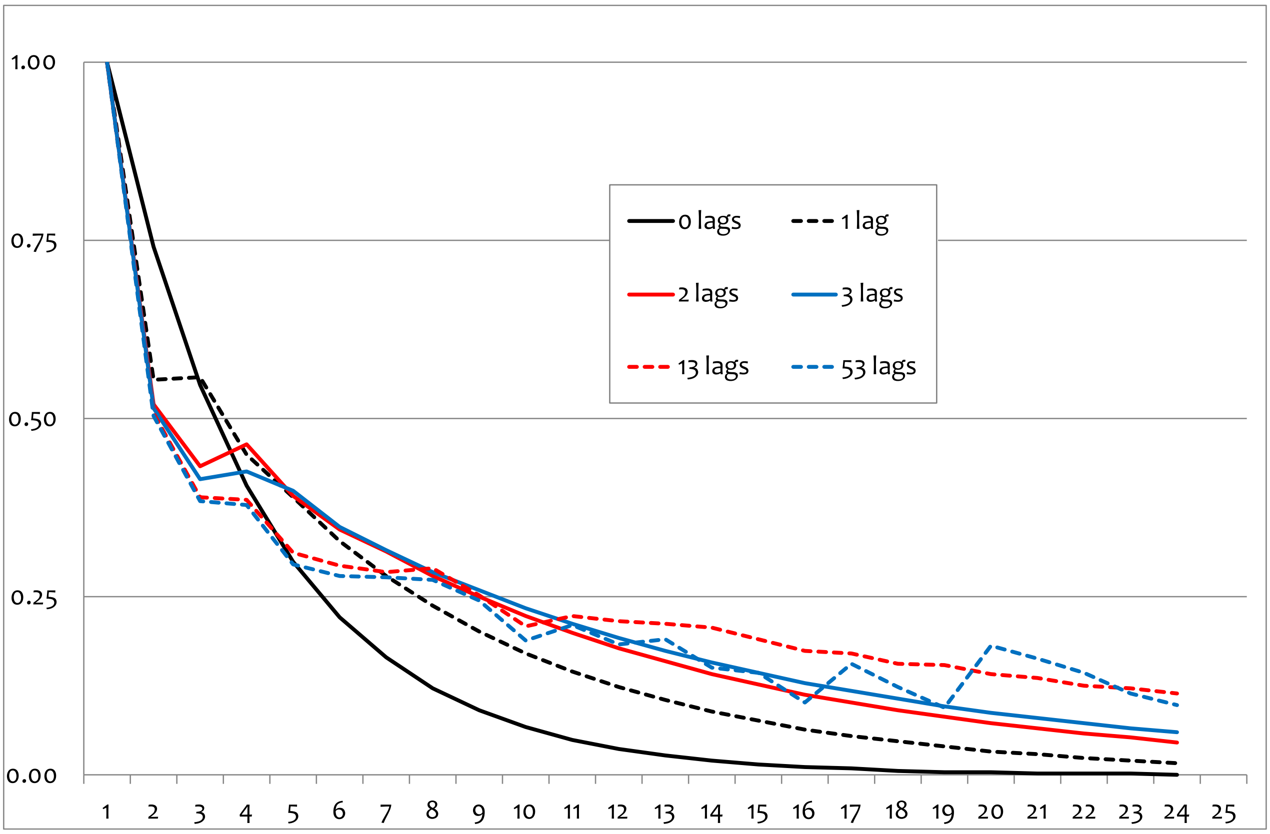
Each point plotted in the figure is a Moran I statistic calculated from a separate cross-section of weekly wheat prices using the formula in equation (3).

Figure 5: Distribution of half-lives from models estimated on 1770-1820 data



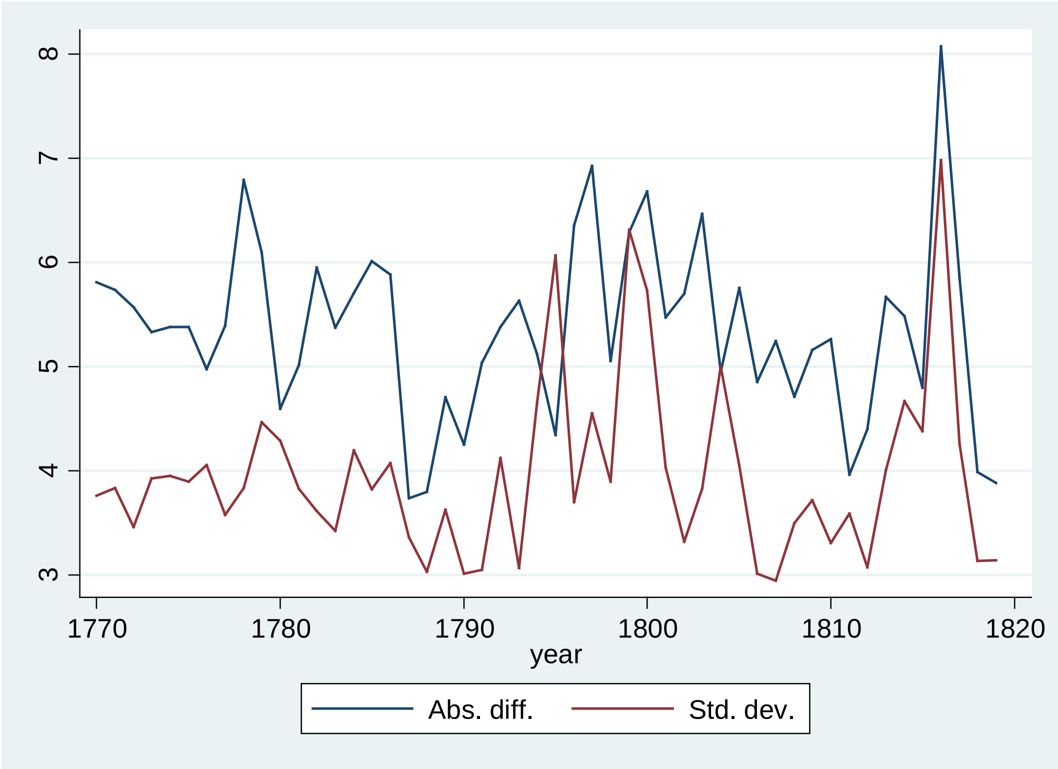
Each distribution in the top panel is based on 780 half-lives (slightly fewer for annual data, where some half-lives could not be calculated). Each half-life is estimated using a model of the form reported in equation (9) using data from the entire period 1770-1820, except where one of the prices is from London, when it is 1770-1793. The bottom panel reproduces the distribution of all 780 half-lives from the top panel and compares it to the distribution of the 103 half-lives where the counties are adjacent.

Figure 6: Impulse response functions for Bedfordshire-Buckinghamshire



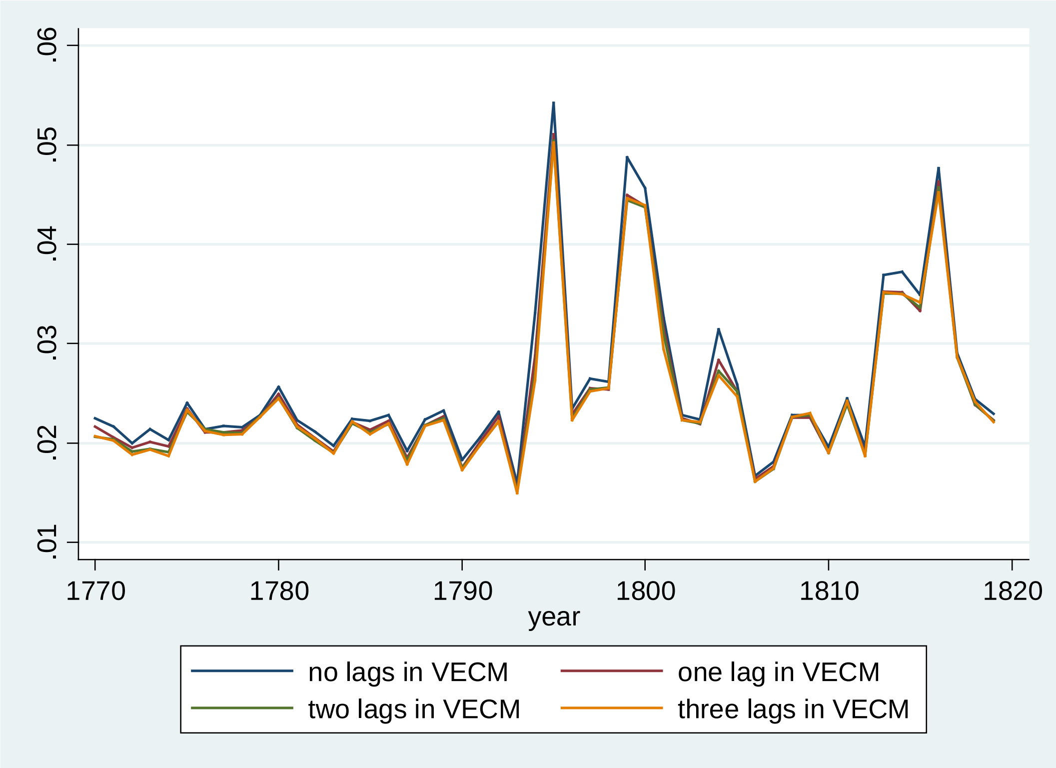
The graph illustrates the speed with which a log-price difference dies away over time (the horizontal axis is measured in weeks). Each impulse response function is estimated using equations (10) to (12). The underlying models are estimated on the full sample of weekly data from 1770-1820 and differ only in the number of lagged dependent variables (the parameter K).

Figure 7: Dispersion of prices



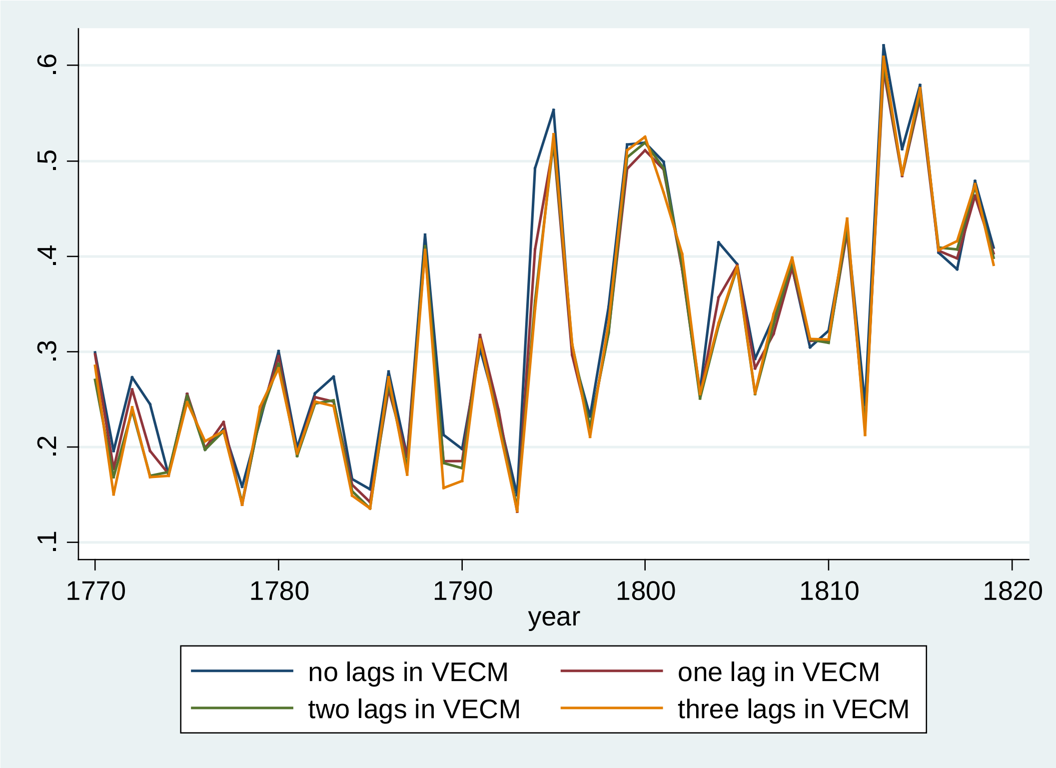
In each year the average for the 93 adjacent-county pairs is plotted of two variables: (i) Abs. diff. is the absolute value of the difference between the mean prices; (ii) Std. dev. is the standard deviation of the price gap. These variables are defined formally in equation (16).

Figure 8: Average magnitude of shocks for each year (per cent)



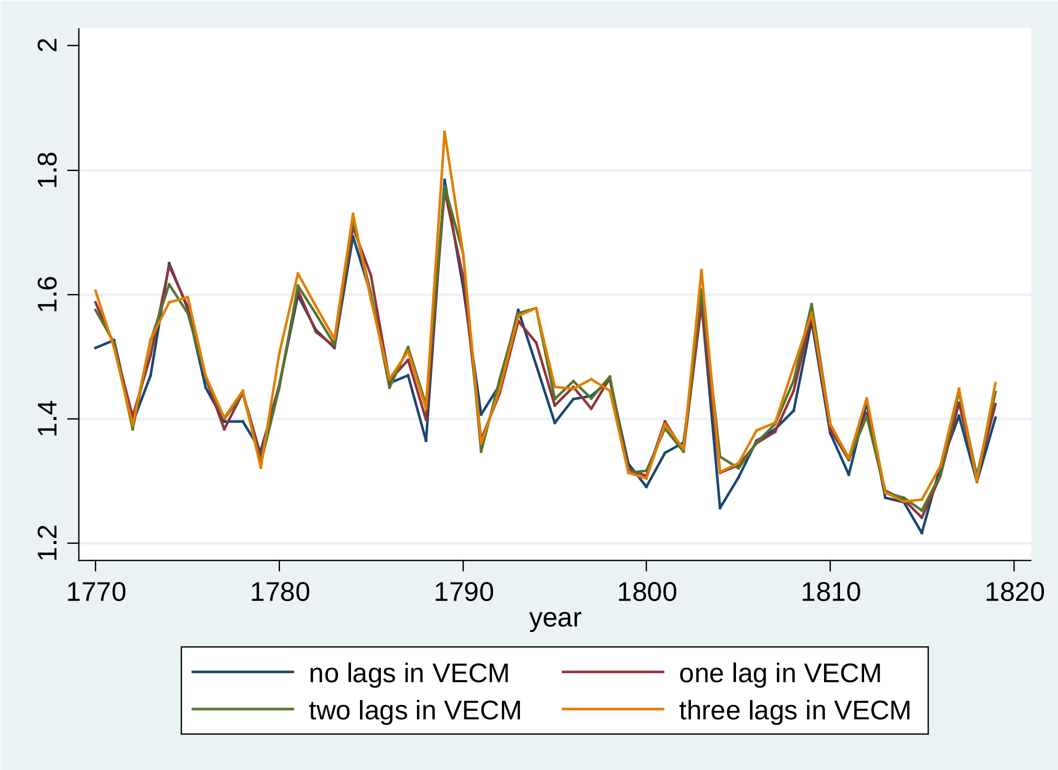
For each year this shows the standard deviation of the shocks (averaged across all 93 adjacent-county pairs) as defined in equation (14) from a model estimated of the form in equation (6).

Figure 9: Average correlation of shocks for each year



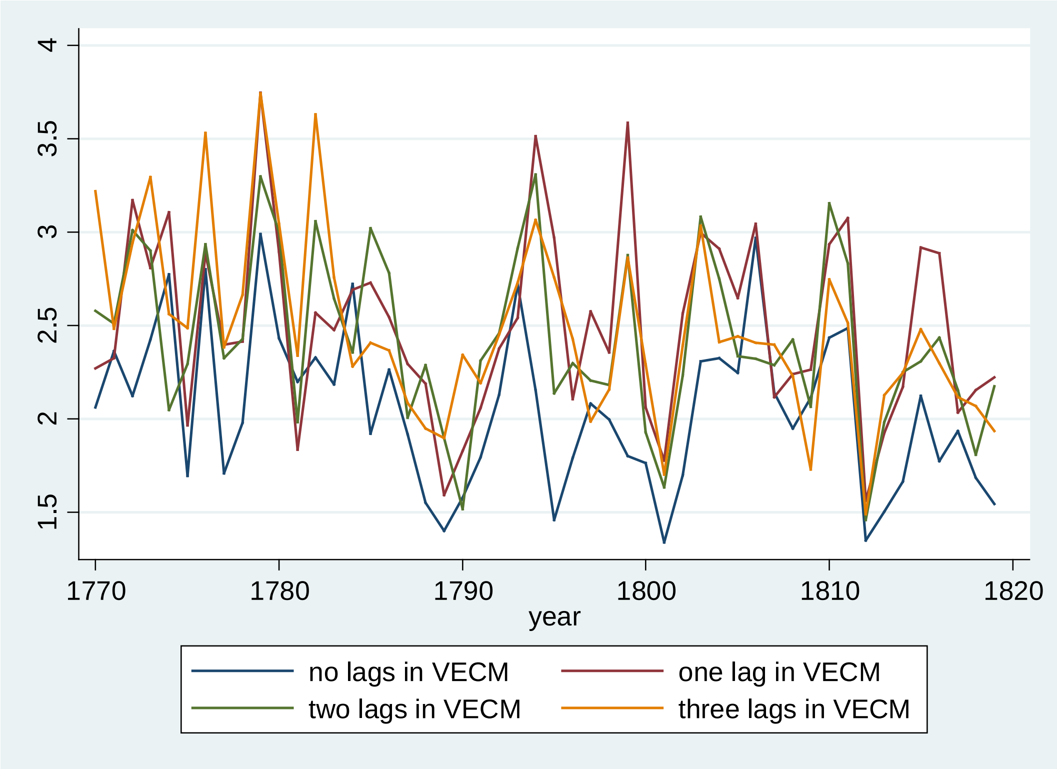
For each year this shows the correlation of the shocks (averaged across all 93 adjacent-county pairs) as defined in equation (15) from a model estimated of the form in equation (6).

Figure 10: Average relative size of shocks (ratio)



For each year this shows the ratio of the larger to the smaller standard deviations of the shocks (averaged across all 93 adjacent-county pairs) as defined in equation (16) from models estimated of the form in equation (6).

Figure 11: Average half-lives for each year (weeks)



For each year this shows the half-life of the response function to a disequilibrium between two prices (averaged across all 93 adjacent-county pairs), estimated on models with differing lags in the VECM. When there are no lagged price changes in the half-life is the measure defined in equation (8). When there are lags in the first-stage model the half-life is calculated using the method described in Appendix 5.2.