In his response to our note "Robust Correlations between Sectoral Prices and Values", Andrew Kliman argues that if two vectors of economic data **p** and **v** exhibit a strong positive correlation, such that they can be modelled using the equation

$$\log \mathbf{p} = \log \mathbf{v} + \epsilon \tag{1}$$

then they will remain positively correlated after division by a third vector **c**, even if this third vector is strongly correlated with both **p** and **v**. He argues that since, for any **c**,

$$(\log \mathbf{p} - \log \mathbf{c}) = (\log \mathbf{v} - \log \mathbf{c}) + \epsilon \tag{2}$$

the error term, ϵ, in the original equation (1) is unchanged in (2) and thus

> Regressing ($\ln P - \ln C$) on ($\ln V - \ln C$), we would again obtain a positive correlation, a zero intercept and a unit slope (p. 318)

But this can be shown to be wrong. Any pair of correlated vectors **p** and **v** can be split into correlated and uncorrelated components as shown in Figure 1. If the correlated component, **c**, is subtracted from both **p** and **v**, we are left with the pair of orthogonal vectors $\mathbf{u}_p$ and $\mathbf{u}_v$.

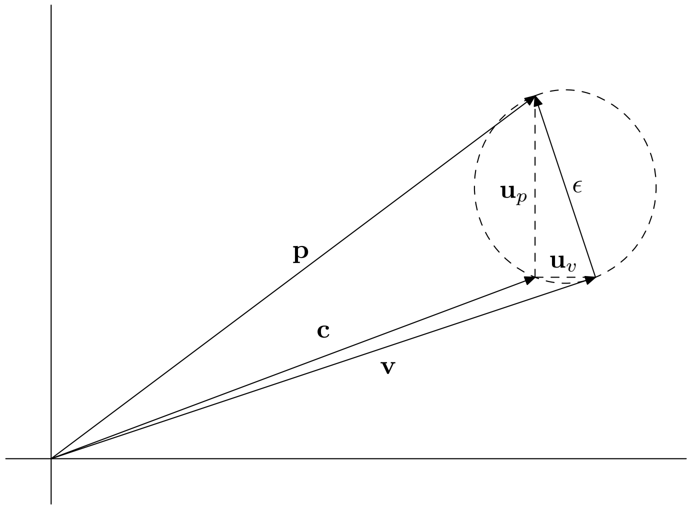

Figure 1: Decomposition of **p** and **v** into correlated and orthogonal components

The vector **c** in Figure 1 is one example of a 'perfect decorrelator' for **p** and **v**. It is constructed as $\mathbf{c} = \mathbf{p} \min \mathbf{v}$, the vector each of whose elements, c_i, equals the lesser of p_i and v_i. If two vectors have a zero inner product they are orthogonal and hence uncorrelated. The inner product of $(\mathbf{p} - \mathbf{c})$ and $(\mathbf{v} - \mathbf{c})$, with $\mathbf{c} = \mathbf{p} \min \mathbf{v}$, is

$$\sum_{i=1}^{n} (\mathbf{p}_i - (\mathbf{p}_i \min \mathbf{v}_i))(\mathbf{v}_i - (\mathbf{p}_i \min \mathbf{v}_i)) = 0$$

In this summation, clearly each term will be zero, since one or other of the two multiplicands is zero. Note that we get the same effect if we select either the minimum or the maximum of the corresponding elements from the two correlated vectors.

Consider the circle centred on the midpoint of the vector ϵ: any vector from the origin to a point on the circumference of this circle will be a perfect decorrelator.[1] Further, any vector from the origin into the interior of the circle will convert a positive correlation between **p** and **v** into a negative correlation, when subtracted from each of them. In geometric terms, the 'remainder' vectors $\mathbf{p} - \mathbf{c}$ and $\mathbf{v} - \mathbf{c}$ form an obtuse angle in this case.

The above account describes the two-dimensional case. In the n-dimensional case, the set of perfect decorrelators forms a hypersphere with radius $|\epsilon|/2$ centred on the midpoint of the error vector, ϵ.

[1] This follows from Proposition 31 of Book III of Euclid's *Elements*, to the effect that a triangle one of whose sides is the diameter of a circle and all of whose vertices lie on the circle is right.

This shows that Kliman's point is incorrect: a positive correlation between P and V will not, in general, survive division by an arbitrary common deflator.

There is obviously an infinite number of vectors in the neighbourhood of the perfect decorrelators that are nearly perfect and will reduce the correlation between $\mathbf{p}$ and $\mathbf{v}$ to statistical insignificance. Kliman's result—a loss of correlation between US log-price and log-value vectors after subtraction of a log-cost vector—is consistent with his log-cost vector approximating to a decorrelator for log-price and log-value. The fact that such a decorrelator exists has no bearing on whether the original correlation was spurious. Similar manipulations can remove correlation from any pair of correlated data series.

A secondary argument advanced by Kliman is that the correlations arise from aggregation effects in the collection of data for input–output tables.

> Although C&C deny that inter-industry variations in costs and physical output are sources of spurious correlation, I do not think they can deny aggregation is such a source (p. 320).

We don't deny that aggregation will tend to enhance correlation. There are two points at issue here.

1. The dispersion of the ratio of price to value—which Farjoun and Machover (1983) refer to as ψ—will tend to be narrower at the level of industries than at the level of individual firms. This follows from the Central Limit Theorem.

2. A correlation is likely to be induced at the aggregate level if the aggregate sectors contain widely differing numbers of firms.

Both of these are possible factors; the issue is how serious they are—whether they induce such a large effect as to account for the strong price–value correlations that have been observed.

Were that true, then one would encounter equally strong correlations between prices and 'oil values', 'electricity values', etc., but this is not the case.

The correlations Kliman generated in his simulation (page ref) example are not spurious, but arise from the fact that in his simulation the randomised firm output price and output value are drawn from the same distribution for all his simulated industries. In consequence each firm in every industry has the same expected price and the same expected value as every other. This is both a strong and unrealistic assumption. It is unsurprising in these circumstances that he gets a strong correlation between aggregate industry prices and values.