Defence of empirical evidence, reply to Shimshon Bichler and Jonathan Nitzan

Paul Cockshott, Allin Cottrell, Alejandro Valle Baeza

December 25, 2010

Bichler and Nitzan argue against the findings of close correlations between labour content and monetary value added on two main grounds:

1. That the empirical studies do not use labour time to estimate labour content but instead use monetary data from the input output tables. They claim that the evidence for the labour theory of value is thus an artefact of circular reasoning. Those of us who have published evidence for close correlations between labour values and prices are guilty of circular reasoning since we presume what we must show: that it is possible to work backwards from money to labour time.

2. That the correlations we observe are essentially supurious since they do not take into account industry scale. They provide an interactive spreadsheet that allegedly shows how such spurious correlations can arise.

We respond to both of these arguments below.

1 Product flows: Quantities versus monetary magnitudes

The claim that all empirical studies of labour value to price correlations have relied on monetary input output tables is false. The Swedish input output tables give labour inputs not in money but in person years. Zachariah[8] has published price/labour value comparisons for a number of economies including Sweden. The Swedish data, which uses actual person years of labour input shows the same strong correlations that have been observed in other studies that started with wage expenditures as their measure of labour input.

Cockshott, Cottrell and Michaelson [7] had earlier used independent data from the New Earnings Survey on average hourly wage rates in each industry to work back from the wage totals given in the UK input output tables to the number of person hours of labour that these represent. When this was done the correlation between labour values and prices remained as strong as when wage expenditure was used as a surrogate for labour input.

The use of monetary magnitudes rather than in-kind product flows is in part a result of the degree of aggregation of the actually available input–output tables for capitalist economies. That is, in order to construct a meaningful input-output table *in natura* it is necessary that the data be fully disaggregated by product, but many of the industries as defined in the actual tables produce a wide range of different products. There can be no meaningful number for the *quantity* of output of "Aircraft and Parts" or "Electronic Components and Accessories", or for the in-kind flow of the product of the latter industry into the former. In a planned economy it would be possible to construct material flows in terms of unique identifiers for each type of product - using bar codes for instance. Since this information is not available to national offices of statistics in capitalist economies the practical solution is to present the aggregate monetary values of flows between.

But this does not create a problem, if one is interested in comparing the aggregate monetary value of the output of the industries with the aggregate labour-value of those same outputs. The point is this: *The vector of aggregate sectoral labour values calculated from a monetary table will agree with the vector calculated from a physical table, up to a scalar, regardless of the price vector and the (common) wage rate used in constructing the monetary table.* Or in other words, the vector of sectoral labour values obtained is independent of the price vector used. One might just as well (if it were practically possible) use an arbitrary vector of accounting prices or weights to construct the "monetary" table. The fact that actual prices are used in the published data does not in any way "contaminate" the value figures one obtains; no spurious goodness of fit between values and prices is induced. We provide a proof of this in section 3.1.

Correlation coefficients between two vectors do not change under scalar multiplication of one of the vectors. A correlation between rainfall and temperature for example is not affected by whether temperature is measured in Farenheit or Centigrade, nor whether the rainfall is measured in centimeters or inches. Thus since the aggregate sectoral values obtained from the monetary data agree—up to a scalar, namely w, the common money wage rate—with those that would be obtained from the data *in natura*, it follows that the correlation coefficients obtained this way will be the same as those that would be obtained with in natura data. The sole source of variation would be the assumption that wage rate was the same accross industries. But as we have said above, tests have already been performed with correcting for differing wage rates accross industries and the correlation coefficient remains very strong. In the UK case [7] found a correlation coefficient of 98% using an assumption of uniform wage rates versus 96% after adjusting for wage rates.

2 Spurious correlation

Bichler and Nitzan[3] argue that the correlations we observe are essentially supurious since they do not take into account industry scale. They provide an interactive spreadsheet that allegedly shows how such spurious correlations can arise. This is continuation of the argument advanced by Kliman [6].

We will respond to this strand of their argument in 3 ways:

1. By focusing on what the hypothesis being tested in our studies was.

2. By showing that their demonstration of spurious correlation is based on what in computer science is called a type error, and what in physics is called a dimensional error. This argument we owe to Valle and Fröhlich [2, 5].

3. By citing additional empirical data that show the observed correlations between labour content and monetary value are not spurious.

2.1 The basic hypothesis being tested

There seems to be a certain irony to Bichler and Nitzan's opposition to the labour theory of value. The idea, on which they found their work, that capital is power has a respectable classical pedigree, Adam Smith had long ago written that monetary wealth was power. But he was specific, moetary wealth was the *power to command the labour of others.* If Bichler and Nitzan were to seek a measurable correlate to power they would do well to follow Smith. But in doing so they would have to abandon their opposition to the labour theory of value, since in Smith's case the power is a power over labour and there is a direct correlation between amount of money and amount of labour commanded. This correlation, Bichler and Nitzan deny, thus depriving their theory of the realistic foundation that Adam Smith had.

The purpose of the empirical studies that we and others have done on price/labour correlations has been to verify this basic proposition on which classical economics stood: that labour is the source of commodities exchange value. We got into this because we wanted to analyse national income in marxist categories - rate of surplus value, organic composition of capital etc. We were turned down by reviewers on the grounds that we were using monetary quantities to measure what should have been labour value ratios. In order to establish the validity of doing this we responded by showing that even if you break the economy down in much finer detail, using input output tables there was a close correlation between labour value magnitudes and monetary magnitudes, and that it was thus valid to use monetary data to work out ratios like them organic composition of capital or the rate of exploitation.

Marx's analysis of capitalist exploitation rests on the hypothesis that embodied labour is the source of monetary value. To establish the validity of his hypothesis and analysis of exploitation that stems from it, it is sufficient to break down the economy into a large number of sectors and show that the monetary value of the gross output of these sectors correlates closely with the labour expended to produce that gross output. This in turn requires that you compute two vectors

1. A vector of monetary flows of output indexed by sector, each element of this vector is of dimension £Million per year.

2. A vector of the number of people whose annual labour was directly or indirectly embodied in this monetary output, the dimension of each element of this vector is a number of persons since person hours per annum reduces to dimension persons. This, incidentally is exactly the format used by the Swedish input output tables mentioned in section 1.

If a strong correlation exists between the two vectors we can say that the data are consistent with the hypothesis that labour is the source of value. It must be emphasised that this method directly examines what we want to test : whether monetary value is proportional to labour used. The argument by Kliman, Bichler and Nitzan that the correlations observed are spurious depends on the idea that there could be an independent 3rd factor that is the cause of the variation both in the persons vector and the monetary flow vector.

Any correlation observed in science could potentially be spurious, so this is always a possibility. But for an allegation of spurious correlation to be born out, one must both identify this third factor and also show that it actually does induce the correlations observed.

What could this 3rd factor be?

Kliman suggests that it is industry size. Big industries employ more people and also sell more output, so that the correlation arises just because of this fact.

But for a 3rd factor to be the common cause of the variation in our two vectors that third factor must itself be quantifiable. How do you measure industry size?

The obvious measures of an industry's size : how many people it employs, or its turnover are ruled out, since we are looking for something independent. Kliman, Bichler and Nitzan suggest that there is some third form of industry size causes the variations in both employment and turnover.

There certainly are other possible measures : the area of land an industry occupies, the number of tons of output it produces, the number of kilowatts hours of energy it uses. In principle any of these could be the 3rd factor that detemined both the labour used and the turnover of an industry, but we merely have to list them to see how implausible it is that land or tonnage is an appropriate third source of variation.

Agriculture is by far the largest industry in the UK in physical terms. It occupies the most space, but its employment and turnover are far in no way proportional to its 'size' in these terms.

The water supply industry is the largest in terms of kilograms delivered, but again, its postion in terms of turnover and employment falls far short.

Energy input is a more plausible 3rd factor, and we examine this in detail in section 2.3.

Bichler and Nitzan dont propose any of these, instead they suggest that the common cause of variation is the number of units of output produced. Bichler and Nitzan produce a spreadsheet showing that if you have two vectors a, b that are uncorrrelated and that if we multiply these by a 3rd random vector c, then $a \circ c$ will be correlated with $b \circ c$. Mathematically this fine but it has no relevance to the question under dispute unless some economic meaning can be given to the vectors a, b, c. Putting headings at the top of the columns like 'Unit Price', 'Number of units sold' does not give their mathematical example any economic grip unless they can explain what these 'Units' are in the context of the input output tables used in our study.

If we look at the UK industrial subdivisions what are the units of output?

For industry 30, Footware, it is presumably pairs of shoes. But what is the unit of output for industry 47 Rubber Products, or industry 50 Ceramic goods?

For industry 67, Weapons and Amunition, is the unit of output a bullet, a tank or an atom bomb?

There is simply no practical way of measuring the scale of these industries in terms of 'units'. Unless Bichler and Nitzan can:

- say how to measure the number of units produced by these, and around 100 hundred other industries in the I/O tables;
- show that the vector of numbers of units produced is in fact highly correlated with both the monetary turnover vector and the embodied labour vector;

their claim that a spurious correlation is induced by the number of units produced will remain no more than idle speculation.

We will now go on to argue both that the very idea of correlating price vectors with labour content vectors, as they purport to do in their example spreadshee (section 2.2), and to show that a number of plausible potential sources of spurious correlation have already been tested and found not to operate (section 2.3).

2.2 A type error.

The spreadheet provided by Bichler and Nitzan purports to show an example in which prices and values are uncorrelated but, once multiplied by the output vector, they become correlated. We have made some criticisms of this above, but now we will focus on another part of their argument : the part that shows the initial correlations between prices and values is insignificant. This part relies on the library correlation function built into Excel. Now Excel is an untyped computer maths package. It does not check that the mathematical operations one is performing make sense since it knows nothing about what the numbers in a spreadsheet represent. More rigourous programming systems like Fortress [1] or Vector Pascal [4] allow the user to specify the units being used for variables so that dimensional analysis can be applied. Had this been done the computer would have warned Bichler and Nitzan of the mistake that they were making.

Dimensional analysis is a set of rules to verify basic aspects of mathematical models, it contains necessary but not sufficient conditions for the validity of a model. Variables, in general, are ordered pairs: a magnitude x and a unit of measurement $[m]$, by example oil price is 90 $[USdollar/barrel]$. Basic rules of dimensional analysis are:

1.Any mathematical expression must be dimensionally consistent, i.e. units of left side of the expression must be equal to units of the right side of the expression.

2.Addition or subtraction of magnitudes with same units is allowed:
$x[m] + y[m] = (x + y)[m]$

3.Addition or subtraction of magnitudes with different units is not allowed: $x[m] + y[t]$ is impossible.

4.Multiplication of variables with different units is allowed $x[m]y[t] = (xy)[mt]$

5.Division of variables with different units is aceptable$x[m]/y[t] = (x/y)[m/t]$

The value of commodity i is$\lambda_i[tl/u_i]$ and it's price $p_i[\$/u_i]$ where tl means labor time and u_i the physical unit of merchandise i.

Does the average price of oil and a pencil means anything?

Of course do not.

In the same vein, average price of all commodities is nonsense. Analogously,the average value of different commodities is not definable. Dimensional analysis corroborate this because averaging oil price $p_o[\$/barrel]$ with $p_p[\$/pencil]$ is forbidden by rule 3.

The correlation coefficient between two vectors $\mathbf{a}, \mathbf{b}$ is the inner or dot product of the normalised vectors

$$\rho(\mathbf{a}, \mathbf{b}) = N(\mathbf{a}).N(\mathbf{b})$$

Where the normalisation $N(\mathbf{x})$ function for a vector $\mathbf{x}$ subtracts $\mu(\mathbf{x})$ the mean of $\mathbf{x}$ and divides by $\sigma(\mathbf{x})$ the standard deviation of $\mathbf{x}$:

$$N(\mathbf{x}) = \frac{(\mathbf{x} - \mu(\mathbf{x}))}{\sigma(\mathbf{x})}$$

Since normalisation depends on computing the mean of a vector and since computing the mean depends on addition, it is clear that normalisaation is only defined on vectors of homogenous dimension. Hence correlation likewise only applies to vectors of homogenous dimension.

Value price correlation would be defined by:
$\rho(\lambda, \mathbf{p}) = N(\lambda).N(\mathbf{p}) = \left(\frac{(\lambda - \mu(\lambda))}{\sigma(\lambda)}\right) \cdot \left(\frac{(\mathbf{p} - \mu(\mathbf{p}))}{\sigma(\mathbf{p})}\right)$

where $\mu(\lambda)$ and $\mu(\mathbf{p})$ are simple means of values and prices; and $\sigma(\lambda)$,$\sigma(\mathbf{p})$ are standard deviations of λ and $\mathbf{p}$. However, averaging prices and values is an invalid operation as is taking the standard deviation since this in turn depends on the mean. Hence the attempt to compute $\rho(p, \lambda)$ is a big mistake, but why?

It is because vectors λ and $\mathbf{p}$ are not dimensionally homogenous, they are not n elements of two variables but two vectors each of n different variables. Sustainers of the spurious correlation criticism have confused the problem: there is no correlation of two variables with n observations complicated by introducing a third variable. The correlation of prices multiplied by gross production $p_i x_i$ and labor values $\lambda_i x_i$ multiplied by gross production is well defined because each

$p_i x_i$ has dimension [$] and each $\lambda_i x_i$ has dimension $[tl]$. Each is of homogenous dimension and thus correlation is well defined on them.

According to above arguments the incorectly named spurious correlation is actually the right way to measure value price deviations and Bichler and Nitzan's assumed non spurious correlation is a total mistake.

Also the correct approach allows us to see that if the basket used is changed the results cannot be compared. Exactly as when inflation is measured: inflation is the variation of the monetary value of a specific basket.

2.3 Poor correlations with other value bases

3 APPENDICES

3.1 Proof

Consider an economy characterized by the following arrays:

- U An $n \times n$ matrix of intersectoral product flows in kind, such that u_{ij} represents the amount of industry j's output used as input in industry i.
- q An $n \times 1$ vector of gross outputs of the industries, in their natural units.
- l An $n \times 1$ vector of direct labour-hours performed in each industry.

It will be useful also to define an $n \times n$ diagonal matrix Q such that

$$Q_{ij} = \begin{cases} q_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

The standard calculation of labour-values proceeds as follows. First calculate the $n \times n$ matrix of technical coefficients as $A = Q^{-1}U$ and the n-vector of direct labour input per unit of physical output as $\lambda = Q^{-1}l$. The n-vector of unit values (vertically integrated labour coefficients) is then given by

$$v = (I - Q^{-1}U)^{-1}Q^{-1}l = (I - A)^{-1}\lambda$$

and the n-vector of *aggregate* values of the sectoral outputs is

$$V = Qv = Q(I - A)^{-1}\lambda \qquad (1)$$

We now construct the monetary counterpart to the above arrays. Let the n-vector p represent the prices of the commodities and the scalar w denote the (common) money wage rate.[1] Let us also define an $n \times n$ diagonal matrix P such that

[1] Having addressed the issue of intersectoral wage differentials above, we abstract from it here.

$$P_{ij} = \begin{cases} p_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

Corresponding to each of the initial "real" arrays above there is a monetary version as follows:

$\hat{U} = UP$	Matrix of money-values of inter-sectoral product flows
$\hat{q} = Pq$	Vector of money-values of gross outputs
$\hat{l} = wl$	Vector of industry wage-bills

From these we can construct counterparts to the derived "real" arrays. First the $n \times n$ diagonal matrix $\hat{Q}$, whose diagonal elements are $p_i q_i$, is given by

$$\hat{Q}_{ij} = \begin{cases} \hat{q}_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} = QP \tag{2}$$

The counterpart to the matrix of technical coefficients is

$$\begin{aligned} \hat{A} = \hat{Q}^{-1}\hat{U} = (QP)^{-1}UP \\ = P^{-1}Q^{-1}UP = P^{-1}AP \end{aligned} \tag{3}$$

The elements of $\hat{A}$ represent the pounds' worth of input from sector j required to produce a pound's worth of output in sector i. Finally, the counterpart to λ is the n-vector $\hat{\lambda}$

$$\begin{aligned} \hat{\lambda} = \hat{Q}^{-1}\hat{l} = (QP)^{-1}wl \\ = wP^{-1}Q^{-1}l = P^{-1}w\lambda \end{aligned} \tag{4}$$

whose elements represent the direct labour cost per pound's worth of output in each sector.

Now here is the issue: suppose we are not privy to the information on product flows in kind and labour-hours, and have at our disposal only the information given in the monetary tables. On this basis we can calculate a vector $\hat{v}$,

$$\hat{v} = (I - \hat{A})^{-1}\hat{\lambda}$$

While v_i represented the vertically integrated labour hours per physical unit of output of commodity i, the $\hat{v}_i$ that we are able to obtain from the monetary tables represents the vertically integrated labour cost per pound's worth of output of commodity i. If we then multiply up by the money-value of the gross outputs of the industries we obtain the vector of vertically integrated labour costs for the industries.

$$\hat{V} = \hat{Q}\hat{v} = \hat{Q}(I - \hat{A})^{-1}\hat{\lambda} \tag{5}$$

We are interested in the relationship between (1), the aggregate sectoral values that could be obtained in principle from the data *in natura*, and (5), the corresponding figures obtained by using the monetary data.

On the basis of the correspondences (2), (3) and (4) we can rewrite (5) as

$$\hat{V} = QP(I - P^{-1}AP)^{-1}P^{-1}w\lambda \tag{6}$$

Recall that (1) specified $V = Q(I - A)^{-1}\lambda$. Comparing these two equations we see that $\hat{V} = wV$ on condition that

$$(I - A)^{-1} = P(I - P^{-1}AP)^{-1}P^{-1} \tag{7}$$

That this condition is indeed satisfied may be seen by taking inverses on both sides of (7). On the left, we simply get $(I - A)$; on the right we get

$$[P(I - P^{-1}AP)^{-1}P^{-1}]^{-1}$$
$$= P(I - P^{-1}AP)P^{-1}$$
$$= (P - AP)P^{-1} = I - A$$

This means we have proved that $\hat{V} = wV$, which is to say that the aggregate sectoral values obtained from the monetary data agree—up to a scalar, namely w, the common money wage rate—with those that would be obtained from the data *in natura*, if these were available. The aggregate value vector is independent of the price vector used in forming the monetary tables.

This proof was originally presented in a conference paper at the Easter Economics Association in 1997.

3.2 Formal treatment of dimensioned data

Dimensional analysis is familiar to scientists and engineers and provides a routine check on the sanity of mathematical expressions. Dimensions can not be expressed in the otherwise rigourous type system of standard Pascal, but they are a useful protection against the sort of programming confusion between imperial and metric units that caused the demise of a recent Mars probe. They provide a means by which floating point types can be specialised to represent dimensioned numbers as is required in physics calculations. For example:

```
kms =(mass,distance,time);
meter=real of distance;
kilo=real of mass;
second=real of time;
newton=real of mass * distance * time POW -2;
meterpersecond = real of distance *time POW -1;
```

The identifier must be a member of a scalar type, and that scalar type is then refered to as the *basis space* of the dimensioned type. The identifiers of the basis space are refered to as the dimensions of the dimensioned type. Associated with each dimension of a dimensioned type there is an integer number refered to as the power of that dimension. This is either introduced explicitly at type declaration time, or determined implicitly for the dimensional type of expressions.

A value of a dimensioned type is a dimensioned value. Let $\log_d t$ of a dimensioned type t be the power to which the dimension d of type t is raised. Thus for t =newton in the example above, and d =time, $\log_d t = -2$

If x and y are values of dimensioned types $t_x$and $t_y$respectively, then the following operators are only permissible if $t_x = t_y$: +, - ,<, >, =, <=, >=. For + and -, the dimensional type of the result is the same as that of the arguments. The operations. The operations *, / are permited if the types $t_x$and t_y share the same basis space, or if the basis space of one of the types is a subrange of the basis space of the other.

The operation **POW** is permited between dimensioned types and integers.

Dimension deduction rules

1. If $x = y * z$ for $x : t_1, y : t_2, z : t_3$ with basis space B then $\forall_{d \in B} \log_d t_1 = \log_d t_2 + \log_d t_3$.
2. If $x = y/z$ for $x : t_1, y : t_2, z : t_3$ with basis space B then $\forall_{d \in B} \log_d t_1 = \log_d t_2 - \log_d t_3$.
3. If $x = y$ **POW** z for $x : t_1, y : t_2, z : integer$ with basis space for t_2, B then $\forall_{d \in B} \log_d t_1 = \log_d t_2 \times z$.

References

[1] E. Allen, D. Chase, J. Hallett, V. Luchangco, J.W. Maessen, S. Ryu, G.L. Steele Jr, S. Tobin-Hochstadt, J. Dias, C. Eastlund, et al. The Fortress language specification. *Sun Microsystems*, 139:140, 2005.

[2] Alejandro Valle Baeza. Dimensional analysis of price-value correspondence: a spurious case of spurious correlation. *Investigacion economica*, LXIX:274, 2010.

[3] P. Cockshott, S. Bichler, and J. Nitzan. Testing the Labour Theory of Value: An Exchange. *bnarchives.yorku.ca*, 2010.

[4] Paul Cockshott. Vector pascal reference manual. *SIGPLAN Not.*, 37(6):59–81, 2002.

[5] N. Fröhlich. Dimensional Analysis of price-value deviations. 2010.

[6] Andrew J. Kliman. The law of value and laws of statistics: sectoral values and prices in th us economy, 1977-97. *Camb. J. Econ.*, 26(3):299–311, 2002.

[7] G. Michaelson, W. P. Cockshott, and A. F. Cottrell. Testing marx: some new results from uk data. *Capital and Class*, pages 103–129, 1995.

[8] David Zachariah. Labour value and equalisation of profit rates. *Indian Development Review*, 4(1):1–21, 2006.

v