AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

How Spurious is the Relationship Between Food Price and Energy Density?

A Simple Procedure and Statistical Test

George C. Davis[a]

and

Andrea Carlson[b]

Paper submitted to satisfy requirement for oral presentation at the Agricultural and Applied Economics Association Annual Meeting, Seattle Washington, August 2012

a. Department of Agricultural and Applied Economics and Human Nutrition, Foods, and Exercise, Virginia Tech University. b. United States Department of Agriculture, Economic Research Service. The views expressed here are those of the authors and cannot be attributed to the Economic Research Service or the United States Department of Agriculture.

Copyright © 2012 by George C. Davis. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract

An important ongoing debate in the literature is whether or not the relationship between food price per kilocalorie and energy density is real or spurious. No closure has come on this debate because no formal statistical tests have been performed. Rather, the arguments against a real relationship have been more anecdotal or analogy based. The goal of this paper is to develop and demonstrate a simple test for the degree of spurious correlation between price of food per kilocalorie and energy density and apply it to a large number of foods. Whereas previous studies have considered rather small sample sizes (e.g. 300 foods or less), here the test is applied to 4430 different foods and 25 different food sub groups from the National Health and Nutrition Examination Study (NHANES) 2003-04. The results indicate that over all foods the relationship is spurious between price per kilocalorie and energy density. When the analysis is broken down by food groups, 92% of the relationships between price per kilocalorie and energy density are spurious. Because this is such an important issue and has been cast within the context of an economic argument, a brief discussion outlines the more appropriate economic framework for discussing the relationships between price and food attributes, such as energy density.

Key Words: Food price, energy density, spurious correlation, economics

1. Introduction

The diets of most Americans do not meet dietary recommendations (1). The cost of food has become a simple explanation because there is a common (mis)perception that healthy foods cost more than less healthy foods. This perception is based on research that is currently under debate in the literature. The ongoing debate is whether the relationship between food price (measured on a per calorie basis) and energy density is real or spurious. The 'real proponents' claim there is a real and significant inverse relationship between food price and energy density (2-8). The 'spurious proponents' claim the negative relationship between food price and energy density is a mathematical or statistical artifact created by the way the food price is constructed (9-11). If the relationship is real, the higher obesity rate in low income groups has a simple economic story: low income groups eat more high energy dense foods because these foods are cheaper (2-6). If the relationship is spurious, the economic story relating lower income to higher obesity rates, must be more sophisticated and requires more work.

One reason no closure has come on this debate is because no formal statistical tests have been performed. Rather, the arguments against a real relationship have been more anecdotal or analogy based. For example Burns, et al. (10) collected energy content data on 212 foods in Melbourne, Australia and generated random numbers for the price of each food. Then, using the data transformation employed by the real proponents, they plotted the data showing a negative relationship very comparable to that shown by the real proponents. Conclusion: even if the underlying relationship between food price and energy content is completely random, the data transformation will create a figure implying a negative relationship. Similarly, Lipsky (11) generated three variables randomly, representing "kilocalories", "grams", and "total price", applied the data transformation used by the real proponents, and generated figures like those shown by Burns, et al. (10) and the real proponents. Lipsky (11) did go a little further than

Burns et al. (10) by running some regressions on the relationship between serving price (i.e. no transformation of data) and energy density for some observational data and finds only a weak *positive* relationship between serving price and energy density within food categories. Again, the conclusion is that the relationship between food price and energy density, as measured by the real proponents, is spurious. However, though the analyses of the spurious proponents are conceptually and visually compelling, their analyses can be dismissed as suffering the analogy fallacy because they are not direct formal measures and tests of spurious correlation.

The lack of a formal measure of spurious correlation in this debate is surprising. There is a long history in the statistics literature on this issue under the more common heading of 'ratio analysis and spurious correlation' (12-18). In fact, Karl Pearson, a key figure in the development of statistics, wrote of this problem in 1897. Pearson (12) described the situation where there are three component variables (x,y,z) and these three variables are then used to construct other (ratio) variables, such as $u = x \div y$ and $v = y$. This is exactly the type of transformation being used by the real proponents, where x = price of food per gram and y = energy density, so u = price of food per gram divided by energy density, which is equivalent to the price per kilocalorie, and v = energy density. Pearson (12) sought to answer the question: if there is no correlation between the component variables, such as x and y, is it possible for there to be correlation between the ratio variables (u and v)? In answering this question, Pearson's (12) conceptual idea was to decompose the correlation between ratio variables u and v into a function of the correlations between the component variables x and y. Pearson (12) showed that even if there is *no* correlation between the component variables x and y there may be correlation between the constructed variables u and v and he labels this spurious correlation. Though rather straightforward, Pearson's approach relies on a first order approximation that may give misleading results if the approximation is not very good (18).

The goal of this article is to utilize Pearson's (12) conceptual idea of decomposing the relationship between constructed variables into a spurious component and a non-spurious component, and develop and demonstrate a simple test that does not require any approximation or randomly generated data. This goal is achieved by using some basic math that leads to a one sided t-test from a simple regression as a test of spurious correlation. The mathematical relationship is intuitively demonstrated with a graph showing how the degree of spurious correlation will vary as the relationship between the price of food and energy density changes. In addition, whereas previous studies have considered rather small sample sizes (e.g. 300 foods or less) here the test is applied to 4430 different foods. In addition to a test of all foods, we grouped the foods by the food groups in the USDA Food Patterns found in the *Dietary Guidelines for Americans 2010* (1) and foods that contained an excessive amount of solid fat, added sugars, and sodium. We also include dishes that contribute to more than one food group, which we refer to as mixed dishes as done in Carlson and Frazao (19). The statistical test results indicate that over all foods the relationship is spurious between price per kilocalorie and energy density. When the analysis is broken down by food groups, 92% of the relationships between price per kilocalorie and energy density are spurious. Because this is such an important issue and has been cast within the context of an economic argument, a brief discussion outlines the more appropriate economic framework for discussing the relationships between food choice, food price and energy density, and consequently income level and high energy density food consumption.

2. Methods

2.1. Analytical and Graphical Representation

Following the lead of Pearson (12), first consider the component variables and their possible relationship. Let p denote the price per gram and d the kilocalories per gram or energy density of a food. Assume the *possible* relationship between the price per gram and energy density is $p = f(d)$. The real proponents focus on the relationship between the price per gram divided by energy density $r = p/d$ and energy density d (e.g., 3-5). Note by definition, the price per gram divided by energy density is equivalent to the price per kilocalorie as the gram units cancel, but for consistency it will be referred to as the price per energy density. Now direct substitution of $p = f(d)$ into $r = p/d$ gives $r = f(d)/d$. The question is then: what is the relationship between r and d? As the spurious proponents correctly point out, even if there is no relationship between p and d (i.e., p remains constant as d changes), r will still decrease if d increases because the denominator of r increases as d increases. A little basic math can be used to show this more formally and generate a simple t-test for a spurious relationship. The arguments put forth in this section can be shown to also apply to the relationship between the price per kilocalorie and kilocalorie.

Because the analysis is working with ratios, it is convenient to use log functions and operators; indeed the graphs in this literature are often reported on a log scale.[1] Let the relationship, $f(d)$, between price p and energy density d be represented by the double logarithm regression model

(1) $\ln p = \alpha + \beta \ln d + \varepsilon,$

[1] All the results generated using the log functional forms and operators can be generated more generally for any functional form, but the testing becomes more involved.

where 'ln' is the natural logarithm operator, α is the unknown intercept, β is the unknown slope, and ε is the disturbance term. Recall in such a double logarithm model, the slope coefficient is a measure of the percentage change in the untransformed dependent variable (price per gram, p) for a one percent change in the untransformed explanatory variable (energy density, d). Importantly, note if there is no relationship between ln p and ln d then $\beta = 0$.

Now by definition, the price per gram divided by energy density (or the price per kilocalorie) is $r = p/d$, so by the rules of logarithms

(2) $\ln r = \ln(p/d) = \ln p - \ln d.$

Substituting (1) into (2) yields

(3) $\ln r = \alpha + (\beta - 1)\ln d + \varepsilon$

$$= \alpha + \lambda \ln d + \varepsilon,$$

where the second line just recognizes in a regression framework only the slope ($\beta - 1$) can be estimated (i.e., β is not identified), so this slope coefficient is defined as $\lambda = (\beta - 1)$.

Figure 1 gives a visual representation of the relationship between equation (2) and (3) and how a spurious negative relationship could be produced. Suppose the average value of ln p is 10 and there is *no* relationship between ln p and ln d (i.e., $\beta = 0$), such that the triangles represent the data plot of ln p against ln d. Alternatively, by construction, the data for the natural log of price per energy density, ln r (equation 3), would look like the dots. Why? Because as ln d increases, that value is being subtracted from ln p. That is, if ln p = 10 and ln d = 1, then ln r = ln p – ln d = 9 and if ln p = 10, ln d = 2, ln r = ln p – ln d = 8, etc. So note if ln p = 10 and ln d = 6, then ln r = 4 as shown in figure 1 by the bracket. Now consider fitting regression lines through these data. If a regression line is fit through the triangles by regressing ln p against ln d, a straight line with no significant slope would result, as represented by the black line (i.e., $\alpha = 10$

and $\beta = 0$). Alternatively, if a regression line is fit through the dots by regressing ln *r* against ln *d*, a straight line with a negative significant slope would result, as represented by the gray line (i.e., $\alpha = 10$ and $\lambda = -1$). In fact, the triangle data in figure 1 was generated by drawing randomly 40 values (4 for each value of ln *d*) within the range of – 1 to 1 and adding them to 10. The dot data (ln *r*) was then generated by subtracting the corresponding value of ln *d* from the value of ln *p*. Fitting the regressions generated the following results: ln *p* = 10.0 + 0.0 ln *d* with an $R^2 = 0.00$ and ln *r* = 10.0 – 1.0ln *d* with an $R^2 = 0.97$. Note the relationship between ln *r* and ln *d* is negative and highly significant while there is no significant relationship between ln *p* and ln *d* and the coefficient estimates of α, β, and λ are 10, 0, and – 1, respectively, as indicated by the mathematical derivations. More generally, as the relationship between ln *p* and ln *d* becomes more positive the entire interior of figure 1 would rotate counter-clockwise around the y axis value of 10 implying a less negative relationship between ln *r* and ln *d.* Alternatively, as the relationship between ln *p* and ln *d* becomes more negative, then the entire interior of figure 1 would rotate clockwise around the y axis value of 10, implying a more negative relationship between ln *r* and ln *d.*

A spurious test can be deduced from evaluating all the possible ways $\lambda = (\beta - 1)$ in equation (3) can be negative. First, as shown in figure 1, even if there is *no* relationship between *p* and *d* (i.e., $\beta = 0$), there is still a negative relationship between *r* and *d*, because in this case the value of the slope coefficient will be $\lambda = -1$. So a one percent increase in energy density *d* will lead to a one percent decrease in the price per energy density, *r*. This is the completely spurious case. Alternatively, if the relationship between *p* and *d* is such that a one percent increase in *d* leads to a less than one percent increase in *p* (i.e., $0 < \beta < 1$), then there will still be a negative relationship between the price per energy density *r* and energy density *d* but it will be within the negative unit interval ($-1 < \lambda < 0$), thus a one percent increase in *d* leads to a less than one percent decrease in *r*. In this case the negative relationship occurs simply because the spurious component (the -1) is greater in absolute value than the positive β. This is the spurious dominating case. Finally, if the relationship between *p* and *d* is such that a one percent increase in *d* leads to a greater than one percent decrease in *p* (i.e., $\beta < -1$), then there will be a negative relationship between the price per energy density *r* and energy density *d* ($\lambda < \beta < -1$), thus a one percent increase in *d* leads to a greater than one percent decrease in *r*. This is the non-spurious case. So in summary, these three cases reveal that if the coefficient in the regression model (3) falls in the range $-1 \leq \lambda < 0$, then the negative relationship is spurious.[2]

2.2. Statistical Testing Framework for Spurious Correlation

The previous section suggests that a simple test for a spurious relationship between ln *r* and ln *d* is a one sided t-test using equation (3) of the form

(4) $H_0 : \lambda \geq -1 \qquad H_a : \lambda < -1.$

If the null hypothesis is rejected, the relationship is statistically not spurious. If the null hypothesis is not rejected, the relationship is spurious.

---

[2] Note that redefining the dependent variable to be the ratio of kilocalories to price ($r = k/p$) does not "put to rest the argument" that the relationship is spurious as claimed by Drewnowski and Monsivais (7). This does nothing but change the relationship from a possible spurious negative to possible spurious positive relationship. That is, equation (2) becomes $\ln r = \ln d - \ln p$ and equation (3) to be $\ln r = -\alpha + (1-\beta)\ln d - \varepsilon$. Working through the math leads to the not surprising summary result that the relationship will be spurious if $0 \leq \lambda < 1$. Redefining the dependent variable to be kilocalories per price does nothing to change the spurious nature of the relationship, except change it from being a negative relationship to a positive relationship. See Chayes (13 p.13-14)]

3. An Application

3.1. Data Description

We merged data from the National Health and Nutrition Examination Survey (NHANES) 2003-04 (20), the MyPyramid Equivalent Database 2.0 (MPED: 21), and the CNPP Food Prices Database 2003-04 (FPD: 22-23). We used 2003-04 because it is the most recent data available for both prices and cup and ounce equivalents. However, ERS research suggests that newer price data is not likely to affect most of our conclusions (24).

NHANES: This is a well known multi-stage probability sample of non-institutionalized individuals living in the United States. The study includes two 24-hour dietary recalls for most subjects. The dietary recall data include the quantity of food reported consumed, as well as the nutrient information on the number of calories, grams of saturated fat, teaspoons of added sugars, and the mg of sodium consumed in each food. We use the dietary recall data to generate a list of foods reported consumed by survey respondents, and calculate the average amount consumed by adults age 19 and older who report consuming that food. The 2003-04 sample includes 4,578 adult participants with at least one complete dietary recall, and for the data used in this analysis 4,430 individual foods.

MPED: The MPED gives the number of cup- and ounce- equivalents in 100 edible grams of each food item. Vegetables, fruits and dairy products are measured in cup-equivalents, and grains and protein foods in ounce-equivalents (25).

FPD: Because NHANES does not include food prices, USDA's Center for Nutrition Policy and Promotion (CNPP) developed a Food Prices Database for all foods reported consumed in the NHANES 2003-04. The database estimates the price per edible 100 grams, that is, the price of the food after it is prepared. These prices take into account the inedible parts that are included in the purchase weight. The retail price data for the FPD comes from Nielsen's 2004 Homescan Panel data, a national panel of consumers who record their retail food purchases. Prices are national average prices, and include all package sizes and brands that were recorded by panel participants.

After merging the three data sets, foods were placed in groups following the method outlined by Carlson and Frazao (19). Appendix Tables 11-15 of *the Dietary Guidelines for Americans 2010* (1) suggest a standard portion size for vegetables, fruits, dairy, protein foods and grains. If the average amount of the food reported as having been consumed by adults in the sample was at least half of this amount, the food was placed in the respective group. For example, vegetable portions are listed as a half of a cup, so in order to count in one of the vegetable groups, the mean amount consumed must contain at least one quarter cup –equivalent of vegetables. Foods that met this standard for more than one group (say vegetables and grain) were classified as mixtures. Mixtures were assigned to sub-groups such as vegetable-based mixtures, based on which food group was more predominant. Foods that did not contain sufficient amounts of any group were classified as "non-food group based foods". Once classified into a food group (vegetables, fruit, grain, dairy, and protein), the foods were divided into the same sub-groups used in the USDA Food Patterns (see Appendix 7, 8, and 9 of (1)). We depart slightly from Carlson and Frazao (19), by dividing grains into breads, crackers and snacks, rice, pasta and cooked cereal, and ready to eat cereal. The final cut was to define any food with at least 480 mg sodium, 1 teaspoon of added sugars and/or 3 grams saturated fat as "moderation foods". Protein and mixed dishes foods were allowed 4 grams of saturated fat, while mixed dishes were allowed 600 mg of sodium, 1.25 teaspoons of added sugars, and 4 grams of saturated fats. All of these partitions then produced a set of 25 food groups.

Table 1 shows summary statistics for both the edible gram price ($/100 g) and energy density (kcal/100 g) for each of the 25 groups. All except two groups (Lean Red Meat and Fish) have a price per 100 gram less than a dollar but, perhaps not surprisingly, there is a greater range of energy densities per 100 grams (from 42.01 for the Dark Green Vegetables group to 312.14 for the Grain Moderation group). The standard deviations within each group (in parentheses) indicate there is a large degree of variability even within each group in terms of price and energy density.

3.2. Results of Tests

Table 2 gives the food categories (column one), the estimate of the slope coefficient β from the regression of the price per gram on energy density (equation 1) with its corresponding *p*-value (column 2), the estimate of the slope coefficient λ from the regression of price per energy density on energy density (equation 3) with its corresponding *p*-value (column 3), and the *p*-value for the one-sided t-test of the null hypothesis that the relationship is spurious. Focusing on the test results (last column *p*-value) indicates that of the 25 food groups, the null hypothesis of spurious correlation cannot be rejected (i.e., p-value ≥ 0.05 in column four) for 23 of the groups or 92% show spurious correlation. The last row in the table also shows when all foods are considered as one group, the relationship is again spurious.

4. Discussion

The analytical results from section 2 can be used to explain the results. As shown analytically, the slope coefficient on the relationship between the price per energy density and energy density (λ column three table 2) is equal to the slope coefficient on the relationship between the price per gram and energy density (β column two table 2) minus one or simply $\lambda = \beta$

– 1, and this is verified in table 2. Consequently, as discussed, even if $0 < \beta < 1$, indicating a *positive* relationship between the price per gram and energy density, this relationship will be dominated by the spurious component (the – 1 in λ) such that $\lambda < 0$, suggesting the relationship between price per energy density and energy density is negative. For example, consider the Red Orange Vegetables group (second row table 2). The second column indicates that a one percent increase in the energy density is associated with a 0.05 percent increase in the price per gram of Red Orange Vegetable products but this relationship is not significant ($p \geq 0.05$). However, the third column indicates that a one percent increase in energy density is associated with a 0.95 percent *decrease* in the price per energy density of the Red Orange Vegetable products and this relationship is highly significant ($p \leq 0.01$). Note this third column slope coefficient is then equal to the second column slope coefficient minus 1, as indicated above. Yet the test for a spurious relationship (fourth column) indicates one fails to reject to null hypothesis of a spurious relationship at even a generous significance level (i.e. $p = 0.75$).

Of the 25 food categories, 20 (80%) show a negative relationship and five a positive relationship between the price per energy density and energy density (column three). Focusing on the 20 cases that show a negative relationship between the price per energy density and energy density, 16 of these negative relationships appear to be significantly different from zero (i.e., p-value ≤ 0.05 in column three) as would be claimed by the real proponents. However, of these 16 cases, the relationship is spurious in 14 of the cases (i.e., p-value ≥ 0.05 in column four). Consequently, there are only two cases where the negative relationship between price per energy density and energy density is real: poultry and fish. So in sum, of the 25 cases, 23 show a spurious relationship. Finally, when all food groups are analyzed together (as one group), the relationship between price per energy density and energy density appears to be significantly negative (bottom of table column 3), but again this result is spurious (column 4).

5. Conclusions

The goal of this paper has been to provide a method for helping to settle the important debate on whether the relationship between food price per energy density and energy density is real or spurious. Following the lead of Pearson (12), an analytical relationship between price per gram and energy density and price per energy density and energy density was developed leading to a simple one sided t-test of a spurious relationship. Applying the method to a total of 4430 foods grouped into 25 categories indicated that the relationship is spurious for 23 out of the 25 categories.

Though the results here support the position of the spurious proponents, one should not dismiss the larger contribution that has been made by the real proponents: the obesity crisis clearly has an economic component as there is an income-obesity gradient. However, this gradient varies by income group and other factors and is more complex and subtle than being portrayed by the real proponents (e.g., 26, 27). The flaw in the economic argument is a flaw in application, not a flaw in economics. The real proponents are attempting to couch the explanation of the relationship between income and obesity within the context of an unrealistically simplistic two-dimensional theory of demand which is the inappropriate economic framework.

Economists have long recognized that consumers - and producers - make choices and tradeoffs within a multidimensional environment where consumers and producers interact. Economic models are much more sophisticated than they are being characterized by the real proponents. If there is interest in the relationship between price of a food and its attributes, as is the case for the real proponents, then the more appropriate economic framework is the hedonic model (28). The hedonic model consists of a demand side and a supply side. On the demand side consumers evaluate and choose products not only based on price but also on the attributes of the products and their income. On the supply side, producers (retailers) evaluate and offer products also based on price and attributes as those affect profitability. Consequently the interaction of consumers and producers in the market equilibrium process determines the actual price and product attributes observed in the market. Some of these attributes may be easily observed and measured (e.g., calories, fat content) but others may be more latent, contextual, and difficult to measure, such as those considered in the behavioral economics literature (e.g., lighting, background noise, or even shape; see 29 for an overview of such factors). The key point however is that the demand (and supply!) for a product will be determined by the collection of attributes, not a single attribute and the relationship between the price and an attribute does *not* represent a demand curve but rather represents a collection of intersections of demand and supply for the attribute.

There is only one analysis we are aware of that is more consistent with a hedonic price approach and that is the work by Brooks, Simpson, and Raubenheimer (30), though they do not mention or reference the extensive hedonic price literature in their article. They analyze 106 foods from the US and Australia and run a multiple regression of the price per gram on the macronutrient contents. They find that as protein content increased the price of food increased but as carbohydrate content increased the price of food decreased.

The Brooks, Simpson, and Raubenheimer (30) analysis is clearly a more sophisticated economic argument than the real proponents have put forth and is a step in the right direction, though it still may be ignoring some other difficult to quantify attributes, such as taste and convenience (31). However, more importantly it completely ignores the supply side of the market, and the policies they recommend should be cautiously evaluated. For example, assuming their biological argument of the protein leveraging hypothesis is correct, incorporating the supply side in the analysis suggests that a tax on energy density, as advocated by the real proponents, could actually lead to more overeating. Why? As Barzel (32) explains, producers will reallocate inputs so as to minimize the impact of the tax. If carbohydrates are cheaper, as indicated by the hedonic price analysis of Brooks, Simpson, and Raubenheimer (30), then producers may substitute more carbohydrates in food production, while reducing protein content (as carbohydrates and protein have the same caloric value per gram), thus consumers would have to eat even more food to achieve the protein target. Stated, alternatively, the point of Brooks, Simpson, and Raubenheimer (30) is that all calories are not equal. It is the composition of calories that may matter when addressing overeating and satiation and how producers respond to policies is just as important as to how consumers respond.

References

1. U. S. Department of Agriculture, U. S. Department of Health and Human Services, 2010. Dietary Guidelines for Americans 2010. U.S. Government Printing Office, Washington, DC.
2. Drewnowski, A. Fat and sugar: an economic analysis. J Nutr 2003;133:838S-40S.
3. Drewnowski, A, Specter SE. Poverty and obesity: the role of energy density and energy costs. Am J Clin Nutr 2004;79:6-16.
4. Drewnowski A, Darmon N. The economics of obesity: dietary energy density and energy cost. Am J Clin Nutr 2005;82:265S-73S.
5. Drewnowski A, Darmon N. Food choices and diet costs: an economic analysis. J Nutr 2005;135:900-4.
6. Townsend MS, Aaron GJ, Monsivais P, Klein NL, and Drewnowski A. Less-energy-dense diets of low income women in California are associated with higher energy adjusted costs. Am J Clin Nutr 2009;89:1220-6.
7. Drewnowski A, Monsivais P. Reply to E Frazao. Am J Clin Nutr 2009;89:701.
8. Drewnowski A. Authors Reply: Healthier foods cost more. Nutr Rev 2010;68(3):184-5.
9. Frazao E. Less-energy-dense diets of low income women in California are associated with higher energy-adjusted costs but not with higher daily costs. Am J Clin Nutr 2009;89:701.
10. Burns C, Sacks G, Bilenkij G, Swinburn B. Correctly calculating the cost of food. Nutr Rev 2010;68(3):182-3.
11. Lipsky LM. Are energy-dense foods really cheaper? reexamining the relation between food price and energy density. Am J Clin Nutr 2009;90:1397-401.
12. Pearson K. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc Roy Soc 1896-97;60:489-502.

13. Chayes F. Ratio correlation: a manual for students of petrology and geochemistry. university of Chicago press 1971.

14. Schuessler K. Analysis of ratio variables: opportunities and pitfalls. Amer J Soc 1974;80:379-96.

15. Kenney BA. Beware of spurious self-correlations! Water Resources Research 1982;18:1041-48.

16. Kronmal RA. Spurious correlation and the fallacy of the ratio revisited. J Roy Stat Soc. Series A 1993;156:379-92.

17. Raubenheimer D. Problems with ratio analysis in nutritional studies. Func Ecol 1995;9:21-9.

18. Brett MT. When is a correlation between non-independent variables “spurious”? Oikos 2004;105:647-656.

19. Carlson, A., and E. Frazao. "Are Healthy Foods Really More Expensive? It depends on How You Measure the Price." EIB- 96 U.S. Department of Agriculture Economic Research Service, Washington, DC , May 2012.

20. Centers for Disease Control and Prevention (CDC), 2003-04. National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

21. Bowman, S.A., Friday, J., Moshfegh, A., 2008. MyPyramid Equivalents Database 2.0 for USDA Survey Foods, 2003-2004 [Online]. Food Surveys Research Group. Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture,Beltsville, MD.

22. Carlson, A., Lino, M., Juan, W., Marcoe, K., Bente, L., Hiza, H.A.B., Guenther, P.M., Leibtag, E., 2008. Development of the CNPP Prices Database U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, Alexandria, VA.

23. Center for Nutrition Policy and Promotion USDA, 2009. Food Prices Database, 2003-04 User's Guide, June 2009 ed. USDA/CNPP.

24. Todd, J.E., Leibtag, E., Penberthy, C., 2011. Geographic Differeneces in the Relative Price of Healthy Foods, EIB-78. U.S. Department of Agriculture, Economic Research Service, Washington, DC. June 2011

25. U. S. Department of Health and Human Services, U. S. Department of Agriculture, 2005. Dietary Guidelines for Americans 2005. U.S. Government Printing Office, Washington, DC.

26. Chang, V. W.; Lauderdale, D. S. Income Disparities in Body Mass Index and Obesity in the United States, 1971-2002 *Arch Intern Med*. 2005;165(18):2122-2128.

27. Braveman, P. A., Cubbin, C., Egeter, S., Williams, D. R., and Pamuk, E. "Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us." *American Journal of Public Health*: April 2010, Vol. 100, No. S1, pp. S186-S196.

28. Rosen S. Hedonic markets and implicit markets: product differentiation in pure competition. J Pol Econ 1974;82:34-55.

29. Just, D.R., L. Mancino, and B. Wansink. "Could Behavioral Econoimics Help Imporve Diet Quality for Nutrition Assistance Program Participants?" ERR 43 U.S. Department of Agriculutre, Economic Research Service, June 2007.

30. Brooks RC, Simpson SJ, Raubenheimer D. The price of protein: combining evolutionary and economic analysis to understand excessive energy consumption. Obesity Rev 2010;11:887-894.

31. Glanz, K., M. Basil, E. Maibach, J. Goldberg, and D. Snyder. 1998. "Why Americans Eat What They Do: Taste, Nutrition, Cost, Convenience, and Weight Control Concerns as Influences on Food Consumption." *Journal of the American Dietetic Association* 98(10):1118-1126.

32. Barzel Y. An alternative approach to the analysis of taxation. J Pol Econ 1976;84:1177-1197

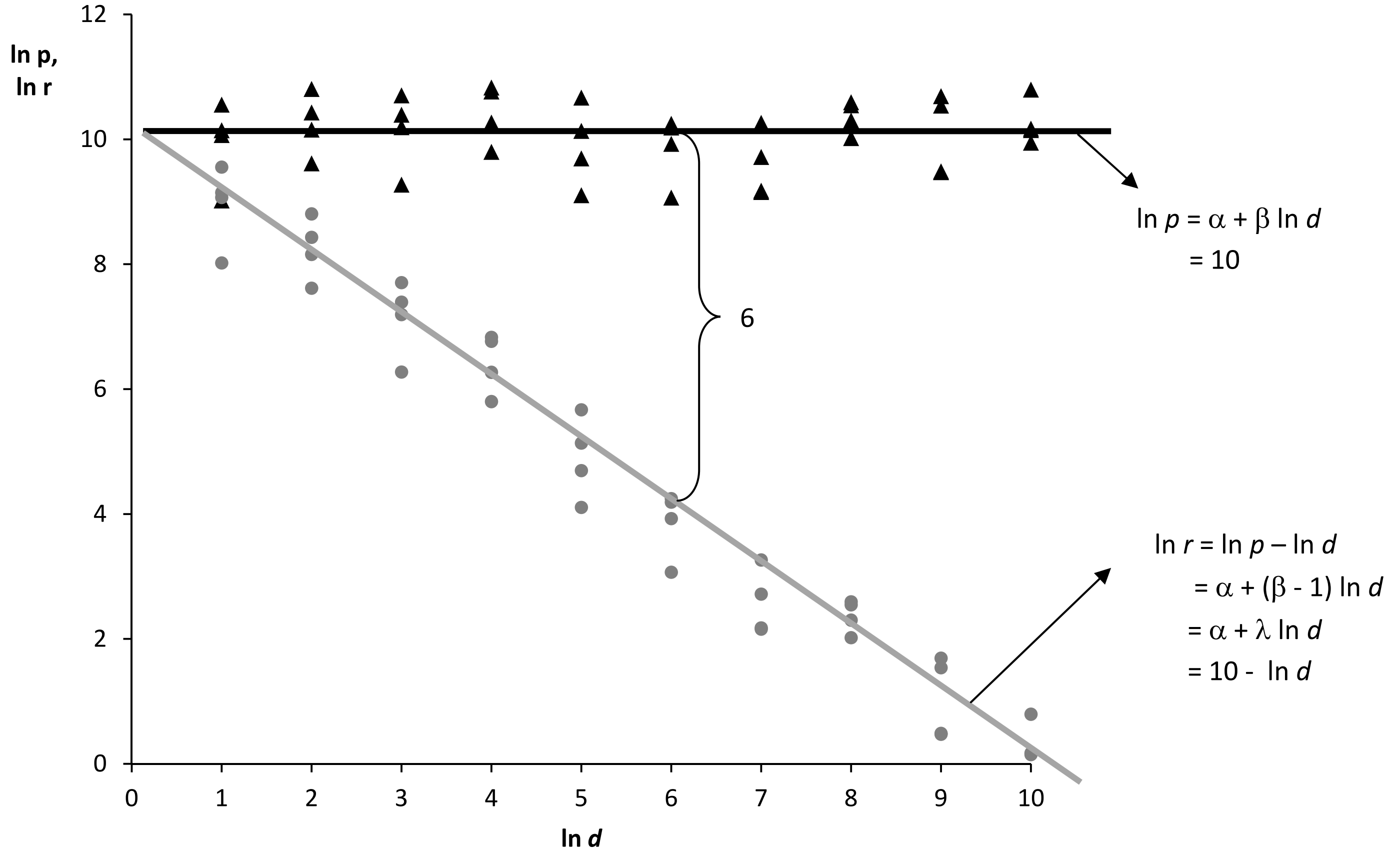

Figure 1. Generating a Spurious Relationship

Caption: In the figure, the triangles represent a situation where there is no relationship between the price of food in grams (ln p), and energy density (ln d). Note that the slope of this line, β, is 0. The dots represent the data for the natural log of price per energy density (r), calculated from $\ln r = \ln p - \ln d$. The slope of the line through the dots, $\lambda = \beta - 1$, is negative one. Thus if we regress energy density on price of energy density, ($\ln r = \alpha + \lambda \ln d + \varepsilon$) we can consider the relationship between energy density and energy density price to be spurious if λ is between -1 and 0.

Table 1. Means and Standard Deviations for Price and Energy Density by Food Group

Food Group (CODE) (observations)	Mean Price per 100 grams (Std. Dev.)	Mean Energy Density per 100 grams (Std. Dev.)
Categories		
Dark Green Vegetables (*n* = 83)	0.53 (0.74)	42.01 (26.60)
Red Orange Vegetables (*n* = 58)	0.24 (0.09)	54.62 (34.84)
Legumes from Vegetable Group (*n* = 48)	0.37 (0.38)	185.44 (65.00)
Other Vegetables (*n* = 236)	0.38 (0.25)	45.24 (26.04)
Starchy Vegetables (*n* = 141)	0.45 (0.41)	123.28 (81.63)
Mixed Vegetables (*n* = 66)	0.34 (0.23)	64.94 (31.51)
Whole Fruit (*n* = 98)	0.56 (0.56)	95.38 (86.82)
100% Fruit Juice (*n* = 41)	0.21 (0.18)	62.07 (68.28)
Whole Grains (*n* = 49)	0.47 (0.45)	271.26 (155.07)
Non-Whole Grains (*n* = 210)	0.39 (0.36)	272.65 (127.14)
Whole and Non-Whole Grain Mixtures (*n* = 63)	0.42 (0.28)	296.52 (71.05)
Low Fat Fluid Milk and Yogurt (unsweetned) (*n* = 21)	0.19 (0.21)	63.05 (69.72)
Lean Red Meat (*n* = 43)	1.26 (0.86)	192.09 (35.88)

Table 1. Con't

Poultry (*n* = 71)	0.56 (0.22)	193.51 (27.07)
Fish (*n* = 34)	1.34 (0.67)	129.94 (34.53)
Eggs (*n* = 20)	0.29 (0.10)	142.50 (45.75)
Nuts and Seeds (*n* = 47)	0.98 (0.97)	563.43 (108.78)
Mixed Dishes (*n* = 206)	0.63 (0.45)	165.65 (82.10)
Vegetable Moderation (*n* = 143)	0.37 (0.61)	133.50 (128.56)
Fruit Moderation (*n* = 88)	0.30 (0.21)	102.81 (79.42)
Grain Moderation (*n* = 637)	0.45 (0.34)	312.14 (121.88)
Dairy Moderation (*n* = 132)	0.49 (0.51)	186.22 (133.35)
Protein Moderation (*n* = 382)	0.90 (0.59)	243.23 (99.99)
Mixed Dishes Moderation (*n* = 837)	0.46 (0.28)	188.22 (83.53)
Not In Any Food Group Moderation (*n* = 677)	0.54 (1.58)	241.14 (202.40)
All Foods (*n* = 4430)	0.52 (0.75)	204.29 (146.23)

Table 2. Results For By Food Group and All Food Groups

Food Group (CODE) (observations)	**Slope coefficient β estimate from ln *p* on ln *d*[a]**	**Slope coefficient λ estimate from ln *r* on ln *d*[b]**	**Spurious Test P-Value[c]**
Categories			
Dark Green Vegetables (*n* = 83)	-.07 (0.56)	-1.07 (0.00)	0.28
Red Orange Vegetables (*n* = 58)	0.05 (0.50)	-0.95 (0.00)	0.75
Legumes from Vegetable Group (*n* = 48)	0.80 (0.09)	-0.20 (0.67)	0.95
Other Vegetables (*n* = 236)	0.03 (0.66)	-0.97 (0.00)	0.67
Starchy Vegetables (*n* = 141)	0.76 (0.00)	-0.24 (0.16)	1.00
Mixed Vegetables (*n* = 66)	0.30 (0.03)	-0.70 (0.00)	0.98
Whole Fruit (*n* = 98)	0.42 (0.00)	-0.58 (0.00)	1.00
100% Fruit Juice (*n* = 41)	0.31 (0.04)	-0.69 (0.00)	0.98
Whole Grains (*n* = 49)	1.07 (0.00)	0.07 (0.50)	1.00
Non-Whole Grains (*n* = 210)	1.11 (0.00)	0.11 (0.18)	1.00
Whole and Non-Whole Grain Mixtures (*n* = 63)	1.11 (0.00)	0.11 (0.74)	1.00
Low Fat Fluid Milk and Yogurt (unsweetned) (*n* = 21)	1.24 (0.00)	0.24 (0.31)	1.00
Lean Red Meat (*n* = 43)	0.42 (0.41)	-0.57 (0.26)	0.80

Table 2. Con't

Poultry (*n* = 71)	-0.67 (0.01)	-1.67 (0.00)	0.01
Fish (*n* = 34)	-0.91 (0.01)	-1.91 (0.00)	0.01
Eggs (*n* = 20)	-0.20 (0.20)	-1.20 (0.00)	0.10
Nuts and Seeds (*n* = 47)	0.43 (0.02)	-0.56 (0.00)	0.99
Mixed Dishes (*n* = 206)	0.67 (0.00)	-0.32 (0.00)	1.00
Vegetable Moderation (*n* = 143)	0.46 (0.00)	-0.53 (0.00)	1.00
Fruit Moderation (*n* = 88)	0.83 (0.23)	-.17 (0.08)	1.00
Grain Moderation (*n* = 637)	0.71 (0.00)	-0.29 (0.00)	1.00
Dairy Moderation (*n* = 132)	1.08 (0.00)	0.08 (0.15)	1.00
Protein Moderation (*n* = 382)	0.003 (0.96)	-0.997 (0.00)	0.52
Mixed Dishes Moderation (*n* = 837)	0.39 (0.00)	-0.61 (0.00)	1.00
Not In Any Food Group Moderation (*n* = 677)	0.51 (0.00)	-0.49 (0.00)	1.00
All Foods (*n* = 4430)	0.43 (0.00)	-0.57 (0.00)	1.00

a. Slope parameter estimate and p-value corresponding to equation (1). b. Slope parameter estimate and p-value corresponding to equation (3). c. p-value for null hypothesis of spurious correlation (equation 4: H_0: $\lambda \geq -1$).