Journal of Agricultural and Resource Economics 26(1):75–90

The Food Stamp Benefit Formula: Implications for Empirical Research on Food Demand

Parke E. Wilde

To understand how food stamps affect food spending, nonexperimental research typically requires some source of independent variation in food stamp benefits. Three promising sources are examined: (*a*) variation in household size, (*b*) variation in deductions from gross income, and (*c*) receipt of minimum or maximum food stamp benefits. Based on results of a linear regression model with nationally representative data, 90% of the total variation in food stamp benefits is explained by gross cash income and household size variables alone. This finding raises concern about popular regression approaches to studying the Food Stamp Program.

Key words: food spending, Food Stamp Program, multicollinearity

Introduction

The Food Stamp Program represents a large investment in food resources for poor Americans. Even after five years of caseload declines, the federal government still spent almost $18 billion on the program in 1999, serving over 18 million low-income Americans on average each month [U.S. Department of Agriculture (USDA) 2001]. With considerable effort, the government provides these benefits in a special currency—paper food stamp coupons or special plastic debit cards—so that program participants may not easily spend their benefits on nonfood goods. Policy makers and social scientists are interested in learning whether the use of this special currency increases the program's impact on food spending.

A common approach to measuring the distinct impact of food stamps is to estimate a regression model using cross-sectional survey data, with food spending as the dependent variable and food stamp benefits and cash income as separate regressors. A typical finding is that the marginal propensity to consume food (MPC) out of food stamps is more than double the MPC out of cash income (see Fraker for a review of this literature). This result has sometimes been considered surprising, because most food stamp participants are found to spend some cash income on food. In a simple static theoretical framework, such participants would have identical MPCs out of food stamps and cash income (Southworth; Fraker).

This article explores a concern with the survey data approach, which goes to the heart of interpreting estimates from this class of regression models. The concern is that food

---

The author is an economist with the Economic Research Service, U.S. Department of Agriculture (USDA). Helpful comments from Mark Prell and an anonymous referee are gratefully acknowledged. This research does not necessarily reflect the views of the USDA.

Review coordinated by Gary D. Thompson.

stamp benefits are a particular nonstochastic function of household income after certain deductions, known as the benefit formula. Under some conditions, the benefit formula would make estimation of a regression equation impossible because there would be no variation in food stamp benefits conditional on cash income. In practice, cross-sectional survey data appear to show wide variation in food stamp benefits conditional on cash income; consequently, the regression approach has been widely used without apparent empirical difficulty.

Nevertheless, this study is motivated by two nagging questions: (*a*) What is the source of independent variation in food stamp benefits which permits the distinct impact of food stamp benefits and cash income to be estimated using regression analysis? and (*b*) How does the dependence of food stamp benefits on cash income affect the reliability of such regression analysis?

The literature on food spending by food stamp participants includes surprisingly little comment on the potential problem presented by the benefit formula. A long footnote in Moffitt's analysis of the food stamp cashout in Puerto Rico in the 1980s reports regression estimates for food stamp benefits as a function of cash income, squared cash income, and some demographic variables (Moffitt, p. 387). Moffitt finds the R^2 on this equation varies between 0.33 and 0.35, depending on the specification. He lists variation in deductions across households, nonlinearities in the benefit formula, and errors in the benefit calculation among the possible sources of the remaining benefit variation. A search of the literature revealed no previous study that has attempted to quantify the importance of different sources of variation in food stamp benefits.

The Food Stamp Program's quality control data, used in our empirical analysis, are a unique resource for such a study, because of their large sample size and precise food stamp benefit information. However, these data contain no information on food spending. Conversely, none of the principal data sources on food spending (such as the Consumer Expenditure Survey from the Bureau of Labor Statistics) contain sufficiently precise information about program benefits for our purpose. This article consequently does not estimate a model of food spending. It focuses instead on quantifying the sources of food stamp benefit variation, which are needed for nonexperimental demand analysis.

The Benefit Formula in Program Regulations

When a household applies for food stamps, it must first pass several eligibility "tests." Nonfinancial tests include restrictions against participation by many students, strikers, legal immigrants, and others. Asset tests include a limit on the amount of assets a household may own, not counting the value of a primary residence and some assets in the form of a vehicle. A "gross income test" requires, for households without an elderly or disabled member, that gross cash income may not be greater than 130% of the poverty guideline for a particular household size. A "net income test" requires for all households that net income (defined below) may not be greater than 100% of the relevant poverty guideline (Castner and Rosso).

Six deductions from gross income are used in computing net income (Castner and Rosso):

- a *standard deduction*, equal to $134 in the contiguous United States in fiscal year 1998;

- an *earned income deduction*, equal to 20% of the combined labor market earnings of household members;
- an *excess shelter expense deduction*, equal to shelter costs that exceed 50% of the income remaining after all other deductions are subtracted from gross income; for households without an elderly or disabled member, this deduction is subject to a limit of $250 for fiscal year 1998;
- a *child support payment deduction*, equal to legally obligated child support payments to somebody who is not a member of the household;
- a *medical deduction*, equal to nonreimbursed medical expenses for elderly or disabled members of the household (beyond the first $35 for each elderly or disabled person); and
- a *dependent care deduction*, equal to certain covered expenses for care of children and other dependents while other household members work, seek employment, or go to school.

The first three deductions are the most important in terms of value and the number of households who benefit from the deduction (Castner and Rosso). Net income equals either gross income minus all relevant deductions or zero, whichever is greater.

Only households that pass all eligibility tests may participate in the program and receive benefits. Participant households with net income of zero receive the maximum food stamp benefit, which varies by family size (refer to table 1).[1] Households with positive net income receive either the maximum food stamp value minus 30% of net income or the minimum benefit level, whichever is higher. The minimum benefit level for a household with one or two members was $10 in 1998.[2]

Therefore, for participant household i, the benefit formula may be stated as follows:

$$S_i = \text{Max}\{10, \alpha(k_i) - 0.3\text{Max}[C_i - d(\theta_i), 0]\}, \tag{1}$$

where S_i (for "stamps") is monthly food stamp benefits, C_i (for "cash") is monthly gross income, $\alpha(k_i)$ is the maximum food stamp benefit for a household with k_i members, 0.3 is the benefit reduction rate, and d is a function showing the total amount of deductions a household may take from gross cash income. The argument of d, θ_i, is a vector containing relevant household characteristics, such as shelter expenses and labor market earnings. The term $\text{Max}[C_i - d(\theta_i), 0]$ represents monthly net income.

The literature on the Food Stamp Program has sometimes emphasized the great complexity of the benefit formula (e.g., Fraker and Moffitt). Thus equation (1) may at first appear too spare. However, most of the difficulty with the official benefit formula concerns the order in which deductions are taken, a complication absorbed in this analysis into the general deductions function (d).

---

[1] The maximum benefit schedule differs from table 1 for states and territories outside of the continental United States (see Castner and Rosso).

[2] Because the net income test binds at a sufficiently low level of net income, only rarely do eligible households with more than two members receive more than $10 in food stamp benefits.

Table 1. Maximum Food Stamp Benefits in the Contiguous 48 States, 1998

Number of Members in Food Stamp Unit	Maximum Benefit per Food Stamp Unit ($)	Maximum Benefit per Member ($)
1	122	122
2	224	112
3	321	107
4	408	102
5	485	97
6	582	97
7	643	92
8	735	92

Source: Castner and Rosso
Note: For each additional member of the food stamp unit, the maximum benefit increases by $92.

The Benefit Formula as a Multicollinearity Problem

A large body of food stamp research requires cross-household variation in benefits, conditional on income, in order to estimate regression parameters of interest. In this section a linear regression model for food spending is examined to show very simply why benefit variation is necessary, and in the following section the implications for a broader class of research methods are discussed.

Consider first a linear regression model for food spending by participant household i:

$$F_i = \beta_0 + \beta_1 C_i + \beta_2 S_i + \beta_3' \mathbf{Z}_i + \varepsilon_i, \tag{2}$$

where F_i is monthly food spending, $\mathbf{Z}_i$ is a vector of household characteristics, the βs are parameters to be estimated, and ε_i is an independently and identically distributed error term with variance σ_ε^2. The parameter β_1 represents the marginal propensity to consume food (MPC) out of cash income, and β_2 represents the MPC out of food stamp benefits.

Our primary concern about estimating equation (2) is prompted by a multicollinearity problem. Consider the regression of food stamp benefits on the other independent variables:

$$S_i = \gamma_0 + \gamma_1 C_i + \gamma_2' \mathbf{Z}_i + v_i, \tag{3}$$

where v_i is an independently and identically distributed error term with variance σ_v^2. As is well known [see Greene, p. 421, equation (9-4)], the variance of the OLS estimate of β_2 in equation (2) depends on the goodness of fit of equation (3), as follows:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_\varepsilon^2}{[SS(S_i)(1 - R_S^2)]}, \tag{4}$$

where $SS(S_i)$ is the sum of squares of deviations from the mean for the food stamp variable, and R_S^2 is the coefficient of determination for equation (3).

Equation (3) is not our account of how the food stamp benefit formula *should* be modeled empirically. Rather, it is an auxiliary regression equation derived directly from equation (2) and which is useful for investigating multicollinearity problems. Intuitively, if there is a "large" degree of independent variation of food stamp benefits, then σ_v^2 will be large, R_S^2 will be small, and the OLS estimate of β_2 will be precise. Conversely, if there is not much independent variation in food stamp benefits, then the OLS estimate of β_2 will be imprecise. In the limiting case, where $\sigma_v^2 = 0$ and $R_S^2 = 1$, the denominator of equation (4) is zero and the variance of $\hat{\beta}_2$ is undefined. Thus, in the absence of independent variation in food stamp benefits, one cannot estimate the distinct impact of food stamps as in equation (2).

In some plausible circumstances, this absence of independent variation is just what would be expected from the benefit formula. Suppose we have what might be called a worst-case scenario, a cross-sectional sample of food stamp participants where: (*a*) all households have the same household size $\bar{k}$, (*b*) all households have the same characteristics $\bar{\theta}$, and (*c*) no household receives the maximum or minimum benefits. Under this scenario, equation (1) reduces to a simple intercept term and slope term:

$$S_i = \left[\alpha(\bar{k}) - 0.3d(\bar{\theta})\right] - 0.3C_i. \tag{5}$$

Clearly, equation (5) is equivalent to equation (3), with the restrictions that $\gamma_0 = [\alpha(\bar{k}) - 0.3d(\bar{\theta})]$, $\gamma_1 = -0.3$, $\gamma_2 = \mathbf{0}$, and the variance $\sigma_v^2 = 0$. For this hypothetical example, the auxiliary regression equation shows no independent variation in food stamp benefits, so the main regression equation (2) cannot be estimated.

This worst-case scenario suggests where to look for possible sources of useful variation in food stamp benefits, which would permit measurement of the impact of food stamps through regression analysis: (*a*) variation in household size, (*b*) variation in deductions from gross income, and (*c*) corner solutions with receipt of minimum or maximum benefits. Each source could in principle induce variation in the error term of equation (3), which is necessary for using linear regressions to analyze food stamp impacts on food spending.[3]

To understand how household size provides a useful source of benefit variation, suppose the assumption of constant household size in equation (5) were relaxed. Variation in household size would induce variation in the benefit level of food stamps on the left-hand side of this equation. In equation (3), this variation due to household size would not be explained by the constant intercept term (γ_0) or the cash income term ($\gamma_1 C_i$), so it could instead be reflected in the error term. Similarly, if the assumption of constant characteristics θ_i were relaxed, the resulting variation in deductions would induce variation in food stamp benefits, which again would not be explained by γ_0 or $\gamma_1 C_i$. Likewise, the minimum and maximum benefit regulations introduce piecewise-linear kinks in the functional relationship between cash income and food stamps, which would not be captured by the first two terms in equation (3).

---

[3] These sources of independent benefit variation are the most promising, in principle, for providing a sound basis for estimating the food spending model in equation (2). Another possible source is administrative errors in the assignment of food stamp benefits, which would produce variation in food stamp benefits conditional on cash income. But administrative errors are unlikely to provide a substantial basis for accurate estimation of a food spending equation. There are many other sources of variation in food stamp benefits which do not provide a sound basis for estimating food spending equations. In real-world, cross-sectional microdata, food stamp benefits may appear to vary across households because of different reference periods for benefits and cash income, measurement error in either variable or both variables, or many other reasons.

The usefulness of all three sources of variation may be reduced to the extent they can be explained using control variables which appear in the vector $\mathbf{Z}_i$ of equation (3). The most serious example is household size, widely recognized as an important determinant of food demand. Household size variables will in practice always appear in $\mathbf{Z}_i$ in some form. Suppose $\mathbf{Z}_i$ includes an indicator variable for each possible household size (with one omitted category). In this case, the variation in food stamp benefits induced by variation in household size would be entirely explained by the third term of the auxiliary regression equation (3), so this source of variation would be of no use in estimating the food demand equation (2). This problem may be somewhat less severe if $\mathbf{Z}_i$ includes a simple scalar rather than a complete vector of indicator variables for household size; but the restrictive specification is usually justified by convenience, not because formal tests or other knowledge show it to be correct.

A similar argument can be made for variation in deductions. If the control variables in $\mathbf{Z}_i$ are good predictors of the deductions, then the remaining benefit variation captured by the error term in equation (3) is reduced. For example, in empirical practice, $\mathbf{Z}_i$ often includes demographic or geographic variables reflecting the same types of characteristics that determine deductions. This potential problem may be systematic, because in designing the regulations for deductions, policy makers explicitly sought to compensate for factors which might make it difficult for poor families to buy enough food. Factors affecting food demand therefore commonly appear in both vectors θ_i and $\mathbf{Z}_i$, which could make deductions less useful as a source of residual variation in equation (3).

Finally, the receipt of minimum or maximum benefits could serve as a useful source of variation, because the linear cash income term ($\gamma_1 C_i$) in equation (3) does not capture the piecewise-linear kinks contained in the true benefit formula in equation (1). However, if the functional form for the effect of cash income were made more flexible (for example, if it permitted kinks at the income levels where households become just barely eligible to receive exactly the minimum or maximum benefit), then this source of variation would also fail to provide useful variation in the error term of equation (3).

To summarize the preceding discussion, the precision of parameter estimates in the main food demand equation (2) depends on having a sufficiently large variance in the error term of the auxiliary equation (3). The concern in this study is that this variance may be comparatively small if cash income and $\mathbf{Z}_i$ suffice to explain most of the variation in food stamp benefits.

To evaluate this concern empirically, the analysis below estimates four specifications of equation (3) with different specifications for $\mathbf{Z}_i$:

- The first specification regresses food stamp benefits on just an intercept and cash income, so benefit variation due to household size, deductions, and corner solutions is captured in the error term.

- The second specification regresses food stamp benefits on an intercept, cash income, and a set of indicator variables for household size, so benefit variation due only to deductions and corner solutions is captured in the error term.

- The third specification regresses food stamp benefits on all the preceding variables and a variable representing the deduction amount. The deterministic portion of this model controls for household size and all variation in deductions, so only variation due to corner solutions is captured in the error term.

- The fourth specification includes all the preceding variables and a set of additional interaction terms to account for receipt of minimum or maximum benefits. This model is simply a data check to confirm the benefits do conform to the official benefit formula. We expect to find no residual variance in this model.

These four specifications correspond to four food demand models that might be estimated as in equation (2). The first may be expected to have ample residual variance in equation (3), indicating no multicollinearity problem, but it corresponds to a food demand model that implausibly includes no control variables in $\mathbf{Z}_i$. In the remaining three specifications, as more variables are added to $\mathbf{Z}_i$, the implicit food demand model becomes more plausible, but the difficulty with multicollinearity increases.

A well-specified food demand model would certainly control carefully for household size, as in the second specification. It would typically include demographic and geographic variables which are correlated to some extent with deductions, but these variables probably would not explain quite all variation in deductions as the third specification does. Therefore, we consider the second and third specifications to provide plausible upper and lower bounds, respectively, on the likely amount of independent variation in food stamp benefits available for food demand estimation in practice.[4]

The Generality of the Multicollinearity Problem

The multicollinearity issues raised here do not arise in cashout demonstrations, such as two random-assignment cashout experiments and two comparison-site demonstrations sponsored by the U.S. Department of Agriculture in the early 1990s. The experiments found no effect of cashout in Alabama and a cashout effect equal to 6.9% of food spending in San Diego. The comparison-site demonstrations found stronger apparent cashout effects (Fraker, Martini, and Ohls).

The multicollinearity concerns also do not arise in nonexperimental studies using a binary food stamp participation variable rather than a continuous variable for food stamp benefits. Such studies vary widely in how they control for other confounding differences between participants and nonparticipants. Some studies use a program participation dummy along with other variables to control for observed household or individual characteristics (e.g., Wilde, McNamara, and Ranney). Others control further for selection on unobservables using selection-bias correction methods (e.g., Butler and Raymond). These investigations require data on both participants and nonparticipants, and they also generally require a restriction by assumption that certain behavioral parameters are identical for participants and nonparticipants. Moreover, they do not attempt to provide an estimate of the impact of *marginal* changes in food stamp benefits and cash income.

---

[4] Estimating an auxiliary regression equation is just one common approach to collinearity diagnostics. An alternative approach, based on the "condition number" of the matrix of independent variables, is described by Belsley, Kuh, and Welsch. We pursue this approach for the second specification of $\mathbf{Z}_i$ as a way of corroborating our conclusions. Because this approach requires substantial additional notation and terminology, and reaches the same diagnosis as our auxiliary regression approach, these results are reported in an appendix available from the author upon request. In support of their approach, Belsley, Kuh, and Welsch (p. 112) argue that in many regression applications there is no obvious choice for which regressor to use on the left-hand side of an auxiliary regression equation. In the present application, however, the benefit formula itself makes food stamp benefits the obvious choice for this role. Thus the auxiliary regression approach is quite illuminating.

A number of studies that do use a scalar benefit variable have emphasized the importance of functional form in regression models such as equation (2). Some of these analyses have compared multiple functional forms (Moffitt; Levedahl; Wilde and Ranney). The collinearity problem discussed in the previous section arises under some of the specifications examined, but not others. Any functional form such as the quadratic that takes the linear form in equation (2) as a special case is naturally subject to the same multicollinearity concerns. In contrast, there is no obvious multicollinearity problem in a double-logarithmic specification:

$$\ln(F_i) = \beta_0 + \beta_1 \ln(C_i) + \beta_2 \ln(S_i) + \beta_3' \mathbf{Z}_i + \varepsilon_i. \tag{6}$$

For this specification, there is nothing in the official benefit formula that demonstrates a problematic collinearity between $\ln(C_i)$ and $\ln(S_i)$, although we still consider it worthwhile to address this specification in the empirical analysis below.

A popular class of models expresses food spending as a general Engel function which depends on "effective income," EY_i:

$$g(F_i) = f(EY_i, \mathbf{Z}_i) + \varepsilon_i, \tag{7}$$

where g is a continuous function monotonically increasing in F_i, f is a continuous function monotonically increasing in EY_i, $EY_i = C_i + \delta S_i$, and δ is a parameter to be estimated.[5] The parameter δ represents the amount of cash income having the same effect on food spending as does $1 of food stamp benefits. Because effective income is a linear function of benefits and cash income, it might seem the collinearity implied by the benefit formula would make estimation difficult for all models in this class. But this is not true in general. For example, household size provides a more useful source of benefit variation for many models in this class than it does for the linear model, because effective income and the household size variables in $\mathbf{Z}_i$ may enter function f in a form that presents no multicollinearity problem.

In sum, the multicollinearity problem explored here is limited to nonexperimental food demand specifications where food stamp benefits and cash income enter linearly. While other popular specifications do not face the same multicollinearity concern, they rely on the assumption that the functional form is correct.

Quality Control Data

The Food Stamp Program quality control data are derived from monthly quality control reviews conducted by state agencies to verify the accuracy of their benefit payments (Brinkley). For a sample of participating food stamp units,[6] quality control reviewers gather selected information from case files and visits to the selected households. The data set is stratified by state and month, and in some cases within states. It is nationally representative with the use of sampling weights to compensate for the stratification.

---

[5] The three "simple" models estimated by Moffitt (pp. 393–98) fall into this class, although other models in the same article do not. All of the specifications in Wilde and Ranney are in this class, as are four of the seven specifications compared in Levedahl. Of course, the linear specification in equation (2) is a member of this class.

[6] In program terminology, a food stamp unit is defined as individuals who live in the same residence and who purchase and prepare food together. In this analysis, the term "household" is used as a synonym for food stamp unit.

After removing identifying information, the USDA's Food and Nutrition Service has released public-use Quality Control (QC) files for 1997 and 1998. Similar files are used to produce the annual report series on characteristics of participants (most recently Castner and Rosso).

This analysis uses the 1998 QC public-use files (USDA). Unlike many cross-sectional data sources which report cash income on an annual basis and food stamp benefits for the past month, the QC files report benefit and income variables for the same monthly reference period. While many cross-sectional data sources rely on income and benefits as recalled and reported by the household, the QC files use administrative records supplemented with household visits.

The data editing and production of the QC public-use files are described in Brinkley. For the most part, the data editing algorithm makes choices that are well-advised for our research purpose. For example, the algorithm selects from two possible sources of household-level income values in the state-provided data files: (*a*) household-level income values reported directly, and (*b*) the sum of individual-level values for the same households. If the two sources disagree, but one of them appears consistent with the official benefit formula, then the consistent data source is selected. In occasional cases where neither source appears precisely consistent, the algorithm nevertheless chooses one of the two data sources and essentially "corrects" the inconsistent values so they appear to adhere to the benefit formula. For purposes of this research, we would have preferred these latter inconsistencies had been left in the edited data source. Nevertheless, the "reconciled" QC data are better suited for our task than the "raw" QC data from either the household-based or individual-based income series separately.

Household units living outside the contiguous 48 states are deleted from this analysis because of their different maximum benefit levels. Units with more than eight members are also deleted because we must control for unit size, and the sample size falls off for larger units. Fewer than 1% of participating households in the United States live in Alaska, Hawaii, Guam, and the Virgin Islands, and fewer than 1% have more than eight members. The final sample for this study consists of 45,263 observations, representing 8.1 million food stamp units nationally.

Descriptive Analysis: Sources of Benefit Variation

In the section that follows, regression estimates are presented for benefit equations derived from equation (3). To interpret those regression results, one must first understand how dramatically deductions and benefits differ for households of different types and sizes. Three broad and mutually exclusive categories of food stamp households are used, all of which face different eligibility rules and opportunities for deductions:

- *Elderly/Disabled:* Households with an elderly or disabled member (40% of the sample);
- *Working:* Non-elderly, non-disabled households with labor market earnings (23% of the sample);
- *Nonworking:* Non-elderly, non-disabled households without earnings (37% of the sample).

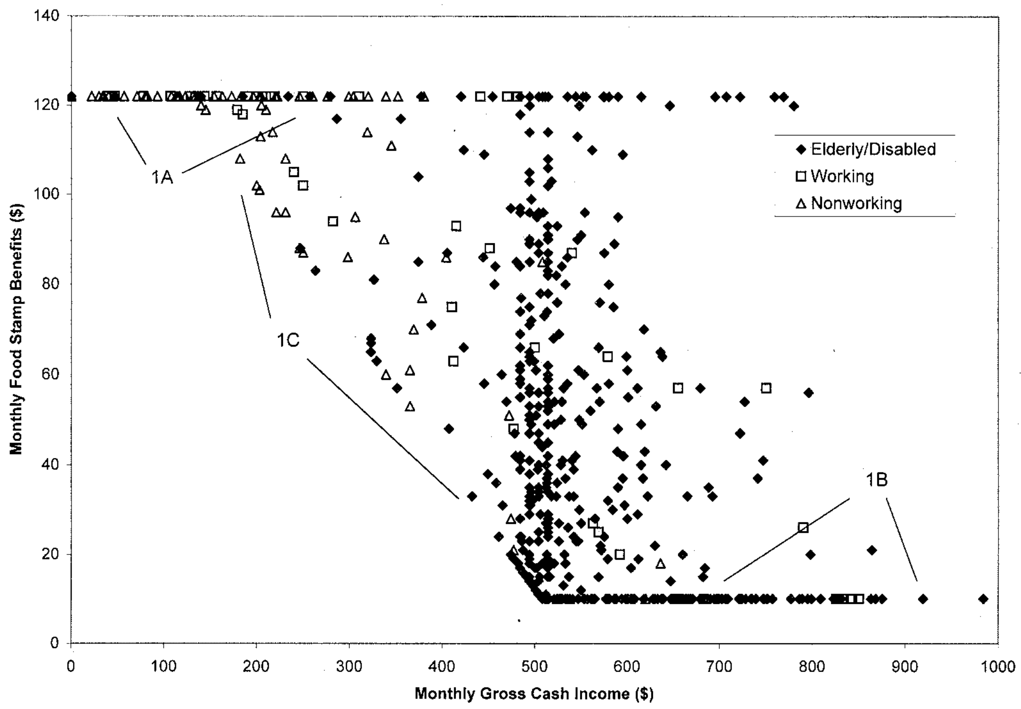

Figure 1. Variation in benefits for households with one member

About two-thirds of all *elderly/disabled* households have one member, and about two-thirds of all one-person households are *elderly/disabled*. The remaining two categories of households both tend to have larger household sizes, but they differ in their economic conditions. *Working* households have far higher cash income and lower food stamp benefits than *nonworking* households, most of which are cash welfare recipients. Because of these major qualitative differences, the deductions available to households in different categories and with different household sizes are examined separately.

Consider one-person households first. Figure 1 is a scatter plot of household food stamp benefits against household gross income, for a random sample of one-person households. For ease of graphical display, 800 households were randomly chosen for this illustration. Note that most observations are *elderly/disabled*. The observations clustered in a horizontal line labeled "1A" have a net income of $0, and therefore food stamp benefits equal to the maximum of $122. The observations clustered in a horizontal line labeled "1B" have a net income greater than $407, and food stamp benefits equal to the minimum of $10. The observations clustered in a downward-sloped line labeled "1C" are those with no deductions beyond the standard deduction. The dense vertical cloud of observations in the center of figure 1 is comprised primarily of elderly and disabled Social Security and Supplemental Security Income (SSI) recipients, who typically receive just over $500 in gross cash income.[7] Most of these households do not receive any

[7] Interestingly, neither Social Security nor Supplemental Security Income (SSI) independently is as strongly clustered at $500. The sum of Social Security and SSI is more strongly clustered. This pattern is due to the benefit reduction rate for Social Security in response to SSI income, which is almost exactly -1.0—a phenomenon beyond the scope of this investigation.

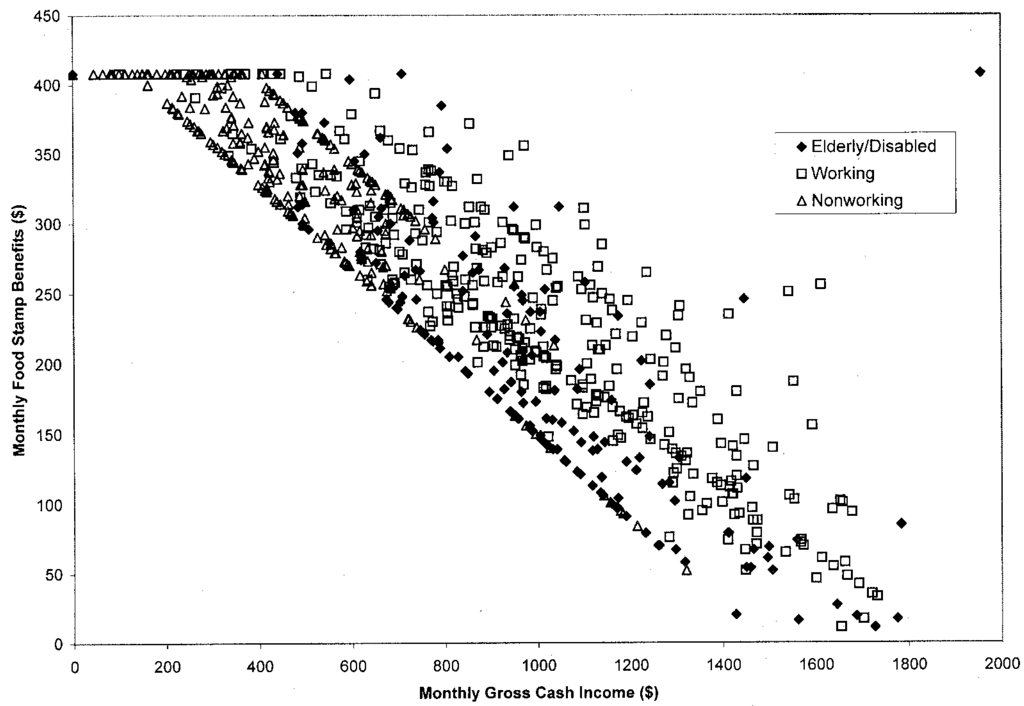

Figure 2. Variation in benefits for households with four members

earnings deductions, so the vertical dispersion of this cloud arises mainly because of variation in excess shelter deductions, which leads to variation in food stamp benefit amounts.

Now consider the contrasting case of four-person households. Figure 2 is a scatter plot for a random sample of 800 four-person households. Fewer four-person households receive the maximum benefit, and none receive the minimum benefit. Only 18% of four-person households are *elderly/disabled*, and the remainder are evenly split between *working* households and *nonworking* households. In comparison with one-person households (figure 1), the scatter plot for four-person households has proportionately less vertical dispersion away from the main downward-sloping trend line.

For purposes of the main research question, the descriptive analysis suggests one-person households differ greatly from four-person households with regard to two potential sources of independent variation in food stamp benefits: deductions and corner solutions with the minimum or maximum benefit. The following regression analysis confirms that the proportion of independent variation in food stamp benefits due to these two sources decreases with household size.

Regression Analysis

The vertical benefit dispersion visible in figures 1 and 2 may be measured by estimating regression models for food stamp benefits. As discussed earlier, the four specifications of the auxiliary regression model (3) are arranged in order from the least to the greatest

Table 2. Regression Estimates for Household Food Stamp Benefits

	Model Variant			
Independent Variable	[1] No Controls	[2] Control for HH Size Only	[3] Control for HH Size and Deductions	[4] Official Benefit Formula
Gross Income	-0.05	-0.20	-0.23	-0.30
Intercept	186.94	150.08	114.44	121.97
(*HH Size* = 2)		101.72	97.52	102.00
(*HH Size* = 3)		197.74	191.31	199.00
(*HH Size* = 4)		275.65	271.00	286.01
(*HH Size* = 5)		342.90	340.34	363.01
(*HH Size* = 6)		429.27	429.80	460.03
(*HH Size* = 7)		484.71	486.02	521.01
(*HH Size* = 8)		557.84	566.78	613.02
Total Deductions			0.186	0.30
(*Minimum Benefit*)				-111.97
(*Minimum Benefit*) × (*HH Size* = 2)				-102.00
Gross Income × (*Minimum Benefit*)				0.30
Gross Income × (*Maximum Benefit*)				0.30
Total Deductions × (*Minimum Benefit*)				-0.30
Total Deductions × (*Maximum Benefit*)				-0.30
R^2	0.02	0.90	0.96	1.00

Notes: Sample includes 45,263 food stamp units with eight or fewer members, living in the contiguous 48 states. Variables described in parentheses are dummy variables with the given characteristic. Standard errors are omitted because every parameter estimate reported is statistically significant at the 0.01 level.

number of control variables in $\mathbf{Z}_i$ (see table 2), which is equivalent to arrangement in order from the greatest to the least amount of useful residual variation in food stamp benefits.

In the first specification, with just gross cash income and an intercept on the right-hand side, the R^2 is 0.02, indicating very little of the variation in household food stamp benefits is explained. In the second specification, simply adding dichotomous variables for household size on the right-hand side raises the R^2 to 0.90. The third specification includes total deductions as a regressor, so that deductions do not provide a source of variation in the error term of equation (3). The R^2 in this specification is 0.96. In the fourth specification, the third and final source of independent variation in food stamp benefits is controlled by adding a set of interaction terms to account for the kinks in the budget formula due to receipt of minimum or maximum benefit amounts. The R^2 for this specification is, of course, 1.00, indicating there is no remaining variation in food stamp benefits after controlling for household size, variation in deductions, and receipt of minimum and maximum benefits.

Because the second specification provides an upper bound on the amount of useful benefit variation assumed to be available in practice, the matrix of regressors for this specification is further investigated using an alternative approach suggested by Belsley, Kuh, and Welsch. The authors advise that a troublesome degree of collinearity may be diagnosed from the following symptom: a "large" condition index for the matrix of

Table 3. Regression Estimates for Per Person Food Stamp Benefits and the Natural Log of Household Benefits

	Dependent Variable			
	Per Person Benefits		Natural Log of Benefits	
	----- Model Variant -----		----- Model Variant -----	
Independent Variable	[1] No Controls	[2] Control for HH Size Only	[1] No Controls	[2] Control for HH Size Only
Per Person Gross Income	-0.13	-0.16		
Natural Log of Gross Income			-0.12	-0.21
Intercept	104.40	130.36	5.28	4.83
(*HH Size* = 2)		-16.31		1.11
(*HH Size* = 3)		-23.83		1.71
(*HH Size* = 4)		-32.70		1.96
(*HH Size* = 5)		-40.22		2.14
(*HH Size* = 6)		-41.86		2.37
(*HH Size* = 7)		-47.10		2.50
(*HH Size* = 8)		-49.26		2.62
R^2	0.49	0.61	0.04	0.61

regressors, which is associated with a high proportion of variation for at least two regressors. As suggested by Belsley, Kuh, and Welsch, "15 or 30 seems a good start" as a cutoff for determining a "large" condition index (p. 157), and 50% is a useful rule of thumb for determining a high proportion of variation for a particular associated regressor.

Using food stamp benefits, cash income, and household size variables as the regressors, the largest condition index found is 16.4. The proportion of variation associated with this condition index exceeds 90% for the food stamp benefit and cash income variables.[8] These results indicate the presence of one main collinear relationship in the matrix of regressors, and the collinearity is strongly associated with the food stamp and cash income variables.

Results are somewhat different when the same four auxiliary regression specifications are estimated using per person food stamp benefits and per person income variables (table 3). These specifications correspond to a food demand model estimated on a per person basis rather than a household basis. Controlling only for household size results in an R^2 of just 0.61 in the per capita model version, in contrast with 0.90 in the household version (see table 2), suggesting deductions and corner solutions may provide a more promising source of benefit variation in per person food demand specifications than in per household specifications.

Likewise, there is less evidence of multicollinearity between the logarithms of the benefit and income variables than there is between the variables themselves (table 3). When the natural log of household food stamp benefits is regressed on an intercept, the log of gross cash income, and the household size dummies, the R^2 is 0.61, in contrast with an R^2 of 0.90 in table 2.

[8] An appendix with further detail is available from the author upon request.

Table 4. Selected Regression Parameters for Food Stamp Benefits, in Households with One to Four Members

Household Size		Model Variant [2] Control for HH Size Only	[3] Control for HH Size and Deductions	[4] Official Benefit Formula
One-Person Households:	(n = 17,849)			
Gross Income		-0.14	-0.16	-0.30
R^2		0.53	0.81	1.00
Two-Person Households:	(n = 9,297)			
Gross Income		-0.20	-0.22	-0.30
R^2		0.73	0.87	1.00
Three-Person Households:	(n = 7,958)			
Gross Income		-0.21	-0.25	-0.30
R^2		0.79	0.94	1.00
Four-Person Households:	(n = 5,601)			
Gross Income		-0.23	-0.26	-0.30
R^2		0.84	0.95	1.00

Notes: Household size variables are not included in the specifications because all models are estimated with data in which household size is invariant. Complete tables of parameter estimates are available from the author upon request.

Because the graphical descriptive analysis above revealed the benefit variation may differ in important ways for households of different sizes, the second, third, and fourth model specifications in table 2 were estimated separately for samples of one to four household members (see table 4). These models do not include household size variables as regressors, of course, because household size is controlled automatically in the selected samples. For one-person households, model specification 2, which controls for household size only, had an R^2 of 0.53, indicating a substantial proportion of total benefit variation is due to deductions and corner solutions. The R^2 for model specification 2 increases steadily as household size increases. For four-person households, model specification 2 has an R^2 of 0.84; thus, variation in gross cash income alone explains 84% of all variation in food stamp benefits for four-person households.

Discussion

Even before considering the empirical analysis, our discussion of the three sources of benefit variation—household size, deductions, and corner solutions—suggests some caution. The usefulness of a particular source of benefit variation depends on other assumptions about specification in ways perhaps not recognized in the literature. For example, if household size variables are properly included in the food spending equation, then benefit variation due to household size will not be useful in estimating the distinct impact of food stamp benefits. Likewise, if household or locational characteristics which are highly correlated with deductions from gross income are included in the food spending equation, then the usefulness of deductions as a source of benefit variation is reduced. There are many opportunities for solving the collinearity problem discussed here through assumptions about functional form or through exclusion restrictions on the variables included in the food spending equation, but these assumptions must be correct to be helpful.

Based on our empirical analysis, the sources and magnitudes of benefit variation depend substantially on household size. For one-person households, a comparatively high share of benefit variation is due to deductions and corner solutions where households receive the minimum or maximum benefit. By contrast, for four-person households, gross cash income alone explains fully 84% of the total variation in food stamp benefits.

These auxiliary regression estimates suggest greater concern about using linear regressions with nonexperimental data, in comparison to earlier regression estimates of food stamp benefits discussed in the introduction (Moffitt). The earlier research found a substantially lower R^2 on the benefit equation. This difference is not necessarily surprising, because the earlier research examined a different geographical location (Puerto Rico), a different time period (the early 1980s), and it used survey data on food stamp benefits and cash income. Because the QC data used here are the best source for measuring variation in food stamp benefits in the United States at present, it appears there is less potentially useful variation in food stamp benefits than previously may have been surmised.

The present study employed a diagnostic equation useful for assessing a multicollinearity problem that may be faced in estimating linear food demand equations. Due to data limitations, however, food spending equations were not estimated. Whether the collinearity between food stamp benefits and cash income is so strong as to render unreliable the estimation of regression equations for food spending remains an empirical question, and depends on the quality of the data and sample size used.

In circumstances where food stamp benefits are highly collinear with cash income, the good news is there may be no need to include both variables in the food spending model. One variable or the other may be dropped without much loss in the explanatory power of the regression model. In this case, recovering the distinct impact of food stamp benefits per se, as contrasted with the impact of having more cash to spend, is impossible. Experimental research methods or suitable participant/nonparticipant comparisons would be required. However, even without such experimental methods, one could still discover how food spending is affected by the different bundles of food stamp benefits and cash income that arise in practice, ranging from low benefits with high cash income to high benefits with low cash income.

Future nonexperimental analyses should consider more explicitly the sources of benefit variation and the assumptions about specification which make it possible to estimate the distinct effects of food stamps and cash income. Further study of the benefit formula as it works in practice for different types and sizes of households is useful for better understanding the distribution of food stamp benefits and also for evaluating more reliably the program's impact on food spending.

[*Received September 2000; final revision received March 2001.*]

References

Belsley, D. A., E. Kuh, and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: John Wiley and Sons, 1980.

Brinkley, M. "Technical Documentation for the Fiscal Year 1998 FSP QC Database and QC Minimodel." Mathematica Policy Research, Inc., Washington DC, 1999.

Butler, J. S., and J. E. Raymond. "The Effect of the Food Stamp Program on Nutrient Intake." *Econ. Inquiry* 34(October 1996):781–98.

Castner, L., and R. Rosso. "Characteristics of Food Stamp Households, Fiscal Year 1998." Mathematica Policy Research, Inc., Washington DC, 2000.

Fraker, T. M. "The Effects of Food Stamps on Food Consumption: A Review of the Literature." Mathematica Policy Research, Inc., Washington DC, 1990.

Fraker, T. M., A. Martini, and J. Ohls. "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations." *J. Human Resour.* 30,4(1995):633–49.

Fraker, T. M., and R. Moffitt. "The Effect of Food Stamps on Labor Supply: A Bivariate Selection Model." *J. Public Econ.* 35,1(1988):25–56.

Greene, W. H. *Econometric Analysis*, 3rd ed. Upper Saddle River NJ: Prentice-Hall, 1997.

Levedahl, J. W. "A Theoretical and Empirical Evaluation of the Functional Forms Used to Estimate the Food Expenditure Equation of Food Stamp Recipients." *Amer. J. Agr. Econ.* 77,4(1995):960–68.

Moffitt, R. "Estimating the Value of an In-Kind Transfer: The Case of Food Stamps." *Econometrica* 57,2(1989):385–409.

Southworth, H. M. "The Economies of Public Measures to Subsidize Food Consumption." *J. Farm Econ.* 27,1(1945):38–66.

U.S. Department of Agriculture, Food and Nutrition Service. Public-Use Quality Control Files. USDA/FNS, Washington DC, 1997 and 1998.

———. "Program Data." Online. Available at http://www.fns.usda.gov/pd/fspmain.htm. (Accessed April 5, 2001.)

Wilde, P., P. McNamara, and C. Ranney. "The Effect of Income and Food Programs on Dietary Quality: A Seemingly Unrelated Regression Analysis with Error Components." *Amer. J. Agr. Econ.* 81,4(2000): 959–71.

Wilde, P., and C. Ranney. "The Distinct Impact of Food Stamps on Food Spending." *J. Agr. and Resour. Econ.* 21,1(1996):174–85.