Knut R. Wangen

An Expenditure Based Estimate of Britain's Black Economy Revisited

Abstract:
The seminal paper by Pissarides and Weber (1989) is one of several previous studies trying to measure the size of the black economy. Pissarides and Weber compared the relationship between food expenditure and income in two groups of workers, self-employed and employees in employment, assuming that employees reported income correctly. For a given level of reported income, the self-employed had a higher food expenditure than employees. Pissarides and Weber concluded that self-employed's actual income was 1.55 times reported income, and that this part of the black economy was about 5.5 percent of GDP in the UK in 1982.

Presumably due to a too informal argumentation, Pissarides and Weber's estimators are not entirely correct and alternative estimators have been overlooked. In all, I suggest three different interval estimators for mean under-reporting. The first is obtained by formally solving optimization problems which Pissarides and Weber tried to solve informally. The other two follows from recognizing, and incorporating, parameter restrictions which were not fully appreciated.

Address: Knut R. Wangen, Statistics Norway, Research Department.
E-mail: knut.reidar.wangen @ssb.no

| **Discussion Papers** | comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc. |

1. Introduction

Knowledge about the size and functioning of the black economy is not only interesting in its own right. The presence of illegal, or informal, activities may distort empirical analyses relying on accurate measurements of economic variables. A non-exhaustive list of affected areas includes revenue predictions and analyses of labor supply, welfare distribution and consumption behavior. Yet, the obviously evasive nature of such activities usually makes direct assessments unfeasible. Several indirect approaches have tried to utilize traces left by black economy activities on measurable variables. In a seminal paper, Pissarides and Weber (1989) (abbreviated as P&W) developed a method for estimating the degree of under-reporting of income among self-employed workers, based on household data.

P&W's basic idea was to compare the relationship between food expenditure and income in two groups of workers, self-employed and employees in employment, assuming that the latter group reported income correctly. For a given level of reported income, the self-employed tended to have a higher food expenditure than employees. Interpreted within P&W's framework, this suggests that actual self-employment income on average was 1.55 times reported income and that this part of the black economy was about 5.5 percent of GDP in the UK in 1982.

Presumably due to a too informal line of reasoning, the estimators on which P&W's conclusions rely are not entirely correct. I suggest a corrected and formally derived estimator. With this estimator at hand two further discoveries appear. First, I show that parameter restrictions in P&W's model framework makes it possible to derive two additional estimators. Secondly, all three of my proposed estimators – which like P&W's estimators are interval estimators for the mean level of under-reporting – are symmetric around mid-points that can be expressed by identifiable parameters. In contrast, P&W's mid-points are derived based on a somewhat arbitrarily chosen procedure, in which the results depend on unidentified parameters.

The proposed estimator that most closely resembles P&W's estimator provides only slightly different results, regardless of which mid-points are used. A second estimator suggests that actual self-employment income is about 8.6 times higher than reported income, if mid-points are calculated according to P&W's specifications, and about 3.5 times higher if my suggested mid-points are used. The latter number yields an estimate of the black economy of 12.4 percent of GDP. This is in better accordance with independent estimates: in an overview given by Lyssiotou, Pashardes and Stengos (2004), P&W's estimate of 5.5 percent is the lowest, while all other estimates lie within the range of 7.2 and 13.2 percent.

Despite any potential weaknesses, one should bear in mind that P&W's method has the benefit of being highly replicable. Household level data on income and expenditure are available in several countries, making international comparison attainable. Already P&W's method has been applied more or less directly to Canadian and Finnish data, confer Schuetze (2002) and Johansson (2000). Further developments of P&W's method escape the problems presented here – for instance Lyssiotou, Pashardes and Stengos (2004), where a demand system with six expenditure types are modelled as functions of income; or the nonparametric single equation approach suggested by Tedds (2004). My proposal is thus simply to remove the estimators suggested by P&W from the practitioner's toolbox and replace them with the modified alternatives presented below.

2. A short review of Pissarides and Weber's method

My points are closely related to P&W original approach. In the following review, I have omitted parts from P&W's study that are not strictly needed for my arguments, for instance their thorough discussion of theoretical assumptions and collateral assumptions in their empirical application. In case the reader wishes to confer P&W's original text, I have used their notation to facilitate parallel reading.

The main structural equation in P&W's analysis is a food expenditure function,

$$(1) \qquad \ln C_i = Z_i\alpha + \beta\ln Y_i^P + \varepsilon_{1i},$$

where C_i is the food expenditure for household i, Z_i is a vector of exogenous variables such as household characteristics, ε_{1i} is a white noise error term, and Y_i^P is "the measure of income that influences consumption decisions, referred to as permanent income". The scalar coefficient β is the marginal propensity to consume food, and α is a vector of parameters.

In addition to permanent income, P&W define two other income variables: true income and reported income. The relationships between the three income definitions are assumed to be

$$(2) \qquad \ln Y_i = \ln p_i + \ln Y_i^P,$$

$$(3) \qquad \ln Y_i = \ln k_i + \ln Y_i',$$

where Y_i and Y_i' are true and reported income, respectively. In (2), the difference between the logs of true and permanent income, $\ln p_i$, is a random variable. Similarly, $\ln k_i$ is a random variable which is supposed to capture deviation between true and reported income. The relationships between

individual values of p_i and k_i and the population means are defined as

$$\ln p_i = \mu_p + u_i, \tag{4}$$

$$\ln k_i = \mu_r + v_i, \tag{5}$$

where μ_p and μ_r are population means, while u_i and v_i are individual deviations from the means with variances σ_u^2 and σ_v^2, respectively.

Given (1)-(5) the food expenditure can be written as a function of reported income

$$\ln C_i = Z_i \alpha + \beta \ln Y_i' - \beta(\mu_p - \mu_r) + \eta_i, \tag{6}$$

$$\eta_i = \varepsilon_{1i} - \beta(u_i - v_i). \tag{7}$$

It is assumed that ε_{1i} are uncorrelated with (u_i, v_i). P&W (p. 27) argue that the correlation coefficient $\rho = \mathsf{corr}(u_i, v_i)$ is non-negative. All random variables are assumed to be homoscedastic, but for (u_i, v_i) variances are assumed to differ across occupational groups.

A principal assumption in P&W's study is that self-employed under-report income while employees in employment do not. Let the subscript SE denote the former occupational group, and EE the latter. The accurate reporting of the employees imply that there are no deviation between true and reported income, which imply that $\mu_{rEE} = 1$ and $\sigma_{vEE}^2 = \rho_{EE} = 0$.

The error term in (6) is a composite of three different parts. In order to obtain an independent estimate of the variance of errors in income, P&W introduce a reduced form equation for income:

$$\ln Y_i' = Z_i \delta_1 + X_i \delta_2 + \zeta_i, \tag{8}$$

$$\zeta_i = \varepsilon_{2i} - (u_i - v_i), \tag{9}$$

where δ_1 and δ_2 are coefficient vectors, X_i is a vector of identifying instrument variables. The composite error term ζ_i consists of deviations of actual from permanent income, deviations of actual from reported income, and unexplained variation in permanent income, ε_{2i}.[1]

Equations (6) and (8) are estimated separately. The variances of the composite error terms (η_i, ζ_i), denoted $(\sigma_\eta^2, \sigma_Y^2)$, are allowed to vary between groups of employees, but are assumed constant within each group. Together, the coefficients and composite variances in (6) and (8) forms a *set of basic parameters* (my definition of terms), which are directly identifiable.

---

[1]P&W do not state (9) explicitly. They do state that "[t]he residual ζ_i is again a composite of three errors: unexplained variation in permanent income, deviations of actual from permanent income, u_i, and deviations of actual from reported income, v_i", (P&W p. 24). They also present an expression for $var\zeta_{SE} - var\zeta_{EE}$ which is consistent with (9), (11), and (12). Thus, (9) seems to be the only reasonable interpretation.

Identifiability of other parameters in the model framework is determined by whether or not they can be expressed in terms of the basic parameters.

P&W's main parameter of interest is "the number by which average reported self-employment income has to be multiplied to give average true income", that is $\mathsf{E}(k_i)_{SE}$. This parameter cannot be calculated directly from the set of basic parameters, but P&W derive an expression for it:[2]

$$(10) \qquad \ln \mathsf{E}(k_i)_{SE} = \frac{\gamma}{\beta} + \frac{1}{2} \left(\sigma^2_{vSE} - \sigma^2_{uSE} + \sigma^2_{uEE} \right).$$

P&W pose the question whether the estimate of $\mathsf{E}(k_i)_{SE}$ *"varies within a small range when σ^2_{vSE} and σ^2_{uSE} vary over their feasible range"*. Albeit these two parameters are unidentified, their values must satisfy restrictions implied by the model setup. P&W choose to impose restrictions derived from the variance of the composite error term, ζ_i, for the two occupational groups:

$$(11) \qquad \sigma^2_{YSE} = \sigma^2_{uSE} + \sigma^2_{vSE} - 2cov(u_i, v_i) + var(\varepsilon_{2i}),$$

$$(12) \qquad \sigma^2_{YEE} = \sigma^2_{uEE} + var(\varepsilon_{2i}).$$

In addition there is a definitional relationship between the variances, the covariance, and the correlation coefficient for u_i and v_i, and the variances must be non-negative. P&W perform an informal search for extreme values of $\ln \mathsf{E}(k_i)_{SE}$ for different values of σ^2_{vSE} and σ^2_{uSE}, subject to these restrictions. The extreme values depend on the correlation coefficient ρ. In the case when $\rho = 0$, the supposed minimum and maximum values of $\ln \mathsf{E}(k_i)_{SE}$ are the limits of the interval

$$(13) \qquad \left[\frac{\gamma}{\beta} - \frac{1}{2} \left(\sigma^2_{YSE} - \sigma^2_{YEE} \right), \frac{\gamma}{\beta} + \frac{1}{2} \left(\sigma^2_{YSE} - \sigma^2_{YEE} \right) \right],$$

while in the case $\rho = 1$, the supposed maximum is

$$(14) \qquad \frac{\gamma}{\beta} + \frac{1}{2} \left(\sigma^2_{YSE} + \sigma^2_{YEE} \right) + \sigma_{YSE}\sigma_{YEE}.$$

Since (13) and (14) can be derived from the set of basic parameters, P&W use them as interval estimators for $\ln \mathsf{E}(k_i)_{SE}$. Interval estimators for $\mathsf{E}(k_i)_{SE}$ can then be obtained by simple anti-log transformations.

3. Three interval estimators with natural mid-points

First interval estimator: correcting P&W's derivation. Presumably due to a too informal line of reasoning, P&W's formulae are not completely

---

[2]Their derivation is based on two additional parts which are omitted here; firstly, an expression for the difference between the two occupational groups' intercepts in the food expenditure equation; and secondly, parameter restrictions implied by assuming that u_i and v_i are log-normally distributed.

correct. Corrected intervals can be found by recognizing, and solving, the derivation of bounds as non-linear optimization problems. One should treat $\ln \mathsf{E}(k_i)_{SE}$ in (10) as an objective function, dependent on the three arguments σ^2_{vSE}, σ^2_{uSE}, and σ^2_{ySE}.[3] Simply put, the upper (lower) bound of the interval estimator is the maximum (minimum) value of $\ln \mathsf{E}(k_i)_{SE}$ when $(\sigma^2_{vSE}, \sigma^2_{uSE}, \sigma^2_{ySE})$ is subject to the selfsame restrictions as P&W imposed when deriving their interval estimators. For expositional reasons, the formal statements and solutions of these optimization problems are given in the appendix. For any value of $\rho \in [0, 1]$, the correct interval is

$$(15) \qquad \left(\frac{\gamma}{\beta} + \frac{\sigma^2_{YEE}}{2} \right) \pm \frac{\sigma^2_{YSE}}{2\sqrt{1 - \rho^2}}.$$

Compared to the intervals offered by P&W, the strongest contrast is when $\rho = 1$, in which case the correct interval covers the whole real line, or the whole positive line if we take the anti-log.

Second interval estimator: a single equation only. The importance of the structural similarity between the composite variances of the food expenditure function and the income equation is not fully recognized by P&W. By rewriting (7) we obtain

$$(16) \qquad \frac{\eta_i}{\beta} = \frac{\varepsilon_{1i}}{\beta} - (u_i - v_i),$$

which mathematically corresponds to (9). This observation leads to two alternative estimators, which I refer to as the second and third.

The second alternative estimator is the simplest obtainable of all three alternatives and can be derived from the food expenditure function only. Replacing (11) and (12) with corresponding equations derived from (7) yields

$$(17) \qquad \left(\frac{\gamma}{\beta} + \frac{\sigma^2_{\eta EE}}{2\beta^2} \right) \pm \frac{\sigma^2_{\eta SE}}{2\beta^2 \sqrt{1 - \rho^2}},$$

confer the appendix for details.[4]

The set of basic parameter estimates needed to calculate (17) is included in the set needed to obtain estimates based on (15). In this sense, there are two competing interval estimators even if P&W's approach is followed exactly.

Third interval estimator: simultaneous equations with restrictions. The third interval estimator can be found if P&W's approach is

---

[3]Also P&W's solutions depend on σ^2_{ySE}, even though the quote after (10) above suggests otherwise.

[4]For future studies it may be worth noting that the width of the interval depends on the income elasticity.

slightly modified. Instead of dismissing the income equation one can in-corporate the structural similarity between (7) and (9) into the estimation strategy. It is straightforward to show that the composite variances satisfy the restriction

$$(18) \qquad \frac{\sigma^2_{\eta SE} - \sigma^2_{\eta EE}}{\beta^2} = \sigma^2_{Y SE} - \sigma^2_{Y EE}.$$

This restriction can be implemented in estimations: the food and income equation for the two groups of employees – four in total – should be estimated simultaneously, restricting the diagonal residual covariance matrix according to (18).[5]

A simultaneous estimation along these lines would yield an unique in-terval estimator, for which the formulae will depend on the relationship between (estimates of) $\sigma^2_{Y EE}$ and $\sigma^2_{\eta EE}/\beta^2$, confer the appendix. If $\sigma^2_{Y EE} > \sigma^2_{\eta EE}/\beta^2$, the formula is (15), otherwise it is (17). Note that even though the formulae are the same, estimates of the basic parameters will usually differ between a restricted simultaneous estimation scheme and the equation-by-equation approach applied by P&W.

Natural mid-point estimators. P&W use their interval estimates to con-struct a point estimate of mean under-reporting. The value mentioned in the introduction, 1.55, is obtained in two steps: first they calculate intervals for $\mathsf{E}(k_i)_{SE}$ for two groups of workers (white and blue-collared), assuming that $\rho = 0.5$; second, they find the mid-points for each interval estimate, and calculate a rough average of these midpoints. This procedure seems to be somewhat arbitrarily chosen. Why is the assumption $\rho = 0.5$ a good choice? And why is the mid-point in the interval estimator more interesting than any other point within the same interval?

Such questions could also be taken into consideration in the alternative interval estimators. However, each of the two intervals for $\ln \mathsf{E}(k_i)_{SE}$, (15) and (17), are symmetric around mid-points that are independent of $\rho \in [0, 1]$. In my view, this property makes them natural mid-point estimators, or at least important points of reference. Of course, the 'natural mid-point' estimators for $\ln \mathsf{E}(k_i)_{SE}$ are below the mid-point of the interval estimators for $\mathsf{E}(k_i)_{SE}$, due to the convexity of the anti-log transformation.

---

[5]A similar relationship between the constant terms in the two equations may also apply. Let Γ_1 and Γ_2 denote the composite constant terms in the food expenditure equation and the income equation, respectively. If the two groups of employees share the same genuine constant term for each equation, the composite constant terms must satisfy $(\Gamma_{1SE} - \Gamma_{1EE})/\beta^2 = \Gamma_{2SE} - \Gamma_{2EE}$.

4. Discussion

In all, I have suggested three different interval estimators for mean under-reporting of income: first, a corrected version of P&W's original one; second, an alternative based on the food expenditure function only; and third, an alternative based on simultaneous estimation of the food and income equations – subject to certain parameter constraints.

The latter two are derived based on an observation of structural similarity between compound error terms in the food expenditure and income equations. Both of these estimators utilize all information contained in their respective model frameworks. In this respect, the corrected version of P&W's original estimator is theoretically inferior, since it only uses parts of the available information.

Choosing between the food expenditure only and the simultaneous estimation approach is harder. If correctly specified, multi-equation models are asymptotically most efficiently estimated by full-information models. Yet, this merit is often less clear in finite samples, see for instance Cragg (1967) or Phillips (1983), and full-information estimators are more susceptible to specification errors, since specification errors in one equation can distort estimates in others.

We will now turn to the question of whether my points have empirical relevance. P&W reported estimates of the basic parameters needed to calculate the first and second interval estimators. The original estimates of basic parameters are given in Table 1, for white and blue-collared workers separately.

In the upper part of Table 2, P&W's original interval estimates for $\mathsf{E}(k_i)_{SE}$ are reproduced for comparison. Comparing them to estimates based on the first alternative estimator, in the middle part of Table 2, reveals only minor differences except for the upper bounds when $\rho = 0$. For both white and blue-collar workers, the upper bound is then roughly 0.1 higher in the first alternative intervals than in the original intervals.

Estimates based on the second alternative estimators are given in the lower part of Table 2. They are substantially wider than the corrected P&W-like intervals. Three out of four lower bounds are below unity, allowing the interpretation that self-employed may *over-report* income. Even more disturbing, the upper bounds are in order of magnitude ten times higher than P&W's original intervals. This leads to unreasonably high mid-point estimates, confer the lower part of Table 3: if P&W's procedure is followed as strictly as otherwise possible, the mid-points estimates are 8.26 for white-collar workers and 8.95 for blue-collar workers, with 8.6 as a rough average. By construction my suggested mid-points are the lowest, suggesting that

actual income is roughly 3.5 times higher than reported income. This may seem unlikely high, but at least it implies a reasonable estimate of the size of the black economy: if P&W's factor of 1.55 implied that the black economy constituted 5.5 percent of GDP, a factor of 3.5 should correspond to about 12.4 percent of GDP.

We should keep in mind that the above results depend on P&W's specific estimates of basic parameters, and that re-estimation on other data could yield less discrepancy between the estimators. Also the third alternative estimator, based on simultaneous estimation of the food expenditure and income equations, could yield lower results than the alternative based on food expenditure only. Such estimates are not provided here, since that would require a full re-estimation without adding theoretical insight. Nevertheless, the included results illustrate that the choice of method is highly influential on the outcome. Given the genuine difficulties associated with measurements of under-reported income, the best strategy for future studies is perhaps not to pick just one of the three alternative estimators, but apply at least the second and third.

References

Cragg, J.G. (1967): On the Relative Small-Sample Properties of Several Structural-Equation Estimators. *Econometrica*, 35, 89–110.

Johansson, E. (2000): *An Expenditure-Based Estimation of Self-Employment Income Underreporting in Finland*. Working paper no. 433, Swedish School of Economics and Business Administration, Helsinki, Finland.

Lyssiotou, P., P. Pashardes, and T. Stengos (2004): Estimates of the Black Economy Based on Consumer Demand Approaches, *The Economic Journal*, 114, 622–640.

Phillips, P.C.B. (1983): Exact Small Sample Theory. In Griliches, Z. and Intriligator, M.D. (Eds.) *Handbook of Econometrics*, Volume 1, North Holland, Amsterdam, The Netherlands.

Pissarides, A. and G. Weber (1989): An Expenditure-Based Estimate of Britain's Black Economy, *Journal of Public Economics*, 39, 17–32.

Schuetze, H.J. (2002): Profiles of Tax Non-compliance Among the Self-Employed in Canada: 1969 to 1992, *Canadian Public Policy – Analyse De Politiques*, 18, 219–238.

Tedds, L.M. (2004): *Nonparametric Expenditure-Based Estimation of Income Under-Reporting and the Underground Economy*. Working paper no. 2004-17, Department of Economics, McMaster University, Ontario, Canada.

TABLE 1. Estimates of relevant basic parameters[a]

	β	γ	$\sigma^2_{\eta SE}$	$\sigma^2_{\eta EE}$	σ^2_{YSE}	σ^2_{YEE}
White-collar	0.270	0.092	0.185	0.138	0.250	0.065
Blue-collar	0.235	0.107	0.157	0.083	0.146	0.060

[a]Extract from Pissarides and Weber's Table 2.

TABLE 2. Interval estimates for mean under-reporting. All based on Pissarides and Weber's original results

	P&W's intervals		
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$
White-collar	$[1.28, 1.54]$	$[1.28, 1.66]$	$[1.28, 1.87]$
Blue-collar	$[1.51, 1.64]$	$[1.51, 1.74]$	$[1.51, 1.92]$
	First alternative estimator (P&W-like intervals)		
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$
White-collar	$[1.28, 1.65]$	$[1.26, 1.68]$	$(0, \infty)$
Blue-collar	$[1.51, 1.75]$	$[1.49, 1.77]$	$(0, \infty)$
	Second alternative estimator (food expenditure only)		
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$
White-collar	$[1.02, 12.89]$	$[0.84, 15.68]$	$(0, \infty)$
Blue-collar	$[0.81, 13.85]$	$[0.65, 17.26]$	$(0, \infty)$

TABLE 3. Mid-points of interval estimates of mean under-reporting

	P&W's intervals			
	Plain mid-point			Natural mid-point
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$	$\rho \in [0, 1]$
White-collar	1.41	1.47	1.58	–
Blue-collar	1.58	1.63	1.71	–
	First alternative (P&W-like intervals)			
	Plain mid-point			Natural mid-point
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$	$\rho \in [0, 1]$
White-collar	1.46	1.47	∞	1.45
Blue-collar	1.63	1.63	∞	1.62
	Second alternative (food expenditure only)			
	Plain mid-point			Natural mid-point
	$\rho = 0$	$\rho = 0.5$	$\rho = 1$	$\rho \in [0, 1]$
White-collar	6.95	8.26	∞	3.62
Blue-collar	7.33	8.95	∞	3.34

5. APPENDIX: CORRECTED DERIVATIONS OF BOUNDS

Three different methods for deriving bounds for mean under-reporting are discussed in the main text. Mathematically, these are closely related and will benefit from a common notation. Following the exposition in the main text, I will first derive the first interval estimator, then the second and third.

To simplify notation, define $(a_1, a_2, a_3) = (\sigma_{vSE}^2, \sigma_{uSE}^2, \sigma_{uEE}^2)$ and

$$(A.1) \qquad f(a_1, a_2, a_3) = a_1 - a_2 + a_3,$$

and observe that $\ln \mathsf{E}(k_i) = \frac{\gamma}{\beta} + \frac{1}{2} f(a_1, a_2, a_3)$. Obviously, the optima for the function f are the same as for $\ln \mathsf{E}(k_i)$.

The arguments in f cannot vary freely. Rewriting (11) and (12), by letting $(A, B, w, c) = \left(\sigma_{YSE}^2, \sigma_{YEE}^2, cov(u_i, v_i), var(\varepsilon_{2i})\right)$, we obtain

$$(A.2) \qquad A = a_1 + a_2 - 2w + c,$$

$$(A.3) \qquad B = a_3 + c,$$

where $A > B > 0$.[6] In addition, the inequality constraint

$$(A.4) \qquad w^2 \le a_1 a_2,$$

must be satisfied due to the definitional relationship between the variances and the covariance of two random variables. Finally, the variances a_1, a_2, a_3, and c must be non-negative.

Combining (A.2) and (A.3), by eliminating w and c, we obtain

$$(A.5) \qquad (a_1 + a_2 - a_3 - A + B)^2 \le 4a_1 a_2,$$

$$(A.6) \qquad 0 \le a_3 \le B.$$

For a given allowed value of a_3, (A.5) constitute a filled parabola, symmetric around the line $a_1 = a_2$. The non-negativity restrictions for a_1 and a_2 are binding in two points where the border is tangent to the axes; $(a_1, a_2, a_3) = (0, a_3 + A - B, a_3)$ and $(a_1, a_2, a_3) = (a_3 + A - B, 0, a_3)$.

The general form of the problem can now be expressed clearly: the lower (upper) bound is the minimum (maximum) of f, subject to the inequality constraints (A.5) and (A.6). Due to the linearity of the objective function, the optima must lie on the border of the admitted set, but without further restrictions there are no solutions. For any allowed a_3, it can be shown that the objective function is not parallel to the border in any point, not even asymptotically. It follows that arbitrarily large, positive or negative, values of the objective function can be obtained.

---

[6]P&W seem to refer to the restriction $A > B > 0$ as a theoretical restriction derived from their other assumptions, but it is a separate assumption. If $B \le A$, bounds can be found along the lines suggested here, with slightly more complicated formulae.

Making assumptions on the correlation coefficient for u_i and v_i, defined as

$$(A.7) \qquad \rho^2 = \frac{w^2}{a_1 a_2}, \quad a_1 a_2 > 0,$$

amounts to restricting the admitted set. Following P&W, I will only consider cases when $0 \le \rho \le 1$.

In the case when $\rho = 1$, (A.4) and (A.5) become equalities. This imply that only the border of the general admitted set is allowed and that the interior is disregarded. Even though the admitted set is reduced, the conclusion is the same as for the general case: no optima exist, and the objective function can take arbitrarily high positive or negative values.

Let us redefine the other extreme, $\rho = 0$ to mean $w = 0$, since ρ is not well defined when a_1 or a_2 equals zero. Then (A.5) becomes an equality where the right hand side is zero, and the admitted set forms a plane in (a_1, a_2, a_3) that meets the axes for a_1 and a_2 in the same points as the tangency points in general case. The linearity of f makes it sufficient to check the corners of this plane for optimal values. The minimum and maximum values are $B - A$ and $B + A$.

In intermediate cases, with a fixed $\rho \in (0, 1)$, (A.5) becomes and equality with $4\rho^2 a_1 a_2$ on the right hand side. For any allowed a_3, the admitted area now forms an ellipse in (a_1, a_2). This ellipse has the same tangency points with the axes of a_1 and a_2 as in the unrestricted case, regardless the value of ρ. Given the geometric situation, it is obvious that the optima must be tangency points between the ellipse and contours of f. There are two such tangency points, $(a_1, a_2) = (a^*, a^{**})$ and $(a_1, a_2) = (a^{**}, a^*)$, where $a^* = \frac{(a_3 + A - B)(1 + \sqrt{1 - \rho^2})}{2(1 - \rho^2)}$ and $a^{**} = \frac{(a_3 + A - B)(1 - \sqrt{1 - \rho^2})}{2(1 - \rho^2)}$. Since $f(a^*, a^{**}, a_3)$ and $f(a^{**}, a^*, a_3)$ are linear in a_3, the optima are obtained with either $a_3 = 0$ or $a_3 = B$. The minimum and maximum values are $f(a^*, a^{**}, B)$ and $f(a^{**}, a^*, B)$, respectively, and define the interval of interest:

$$(A.8) \qquad \left(B - \frac{A}{\sqrt{1 - \rho^2}}, B + \frac{A}{\sqrt{1 - \rho^2}} \right).$$

A few things are worth noting. Firstly, the interval in the intermediate case can be used to describe the intervals in the extreme cases, $\rho = 0$ and $\rho = 1$. Secondly, the mid-point of the interval is always B, regardless the value of ρ. This is because for any ρ, the graph of the admitted set is symmetric around the plane where $a_2 = a_1$, and because f is linear with contours parallel to this plane.

With minor adjustments, the formula above applies to the case where only the food expenditure equation is taken into account. The whole idea is to replace the equations (11) and (12) with corresponding equations derived from (16). Defining $(A', B', w, c') = \left(\sigma^2_{\eta SE}/\beta^2, \sigma^2_{\eta EE}/\beta^2, cov(u_i, v_i), var(\varepsilon_{1i}/\beta^2) \right)$, this amounts to replacing (A, B, c) with (A', B', c') above. Otherwise the arguments are exactly the same, and the second interval estimator follows directly.
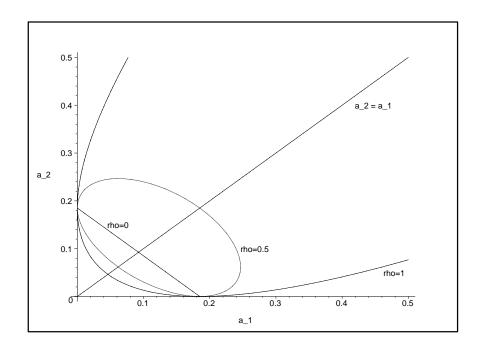
Some additional modifications apply in the case where the food expenditure and income equations are estimated simultaneously with restrictions. When $A - B = A' - B'$, (A.5) will be the same regardless on whether one chooses to use (A, B), or (A', B'). The only real difference is (A.6), which should be replaced with

(A.9)
$$0 \leq a_3 \leq \min(B, B').$$

Whichever of B and B' is binding should be used in (A.8), together with its respective mate (A or A'), in order to obtain the third interval estimator.

FIGURE 1. Graphs of admitted area for three values of ρ and symmetry-line when $a_3 = 0$, $A = 0.250$ and $B = 0.065$

Tangency points between admitted set and axes when $0 < \rho \leq 1$.
When $0 < \rho \leq 1$, the admitted set can be derived from (A.5):

$$(\text{B.1}) \qquad (a_1 + a_2 - a_3 - A + B)^2 = 4\rho^2 a_1 a_2.$$

Differentiating with respect to a_1 and a_2, keeping a_3 fixed, we obtain

$$(\text{B.2}) \quad da_1 \left(a_1 + (1 - 2\rho^2)a_2 - a_3 - A + B \right)$$
$$= -da_2 \left((1 - 2\rho^2)a_1 + a_2 - a_3 - A + B \right)$$

The tangency point between the admitted set an the a_1-axis satisfy the restrictions

$$(\text{B.3}) \qquad \frac{da_1}{da_2} = 0 \quad \text{and} \quad a_2 = 0,$$

which gives the solution

$$(\text{B.4}) \qquad a_1 = a_3 + A - B.$$

The tangency point between the admitted set and the a_2-axis is found by a corresponding argument.

Extreme values of f when $0 < \rho < 1$. For a_3 fixed, the admitted set is an ellipse when $0 < \rho < 1$. An extreme value is a point on the ellipse where a contour of f is tangent to the ellipse.

Define $C = -a_3 - A + B$, then the slope of the ellipse is

(B.5)
$$\frac{da_2}{da_1} = -\frac{a_1 + (1 - 2\rho^2)a_2 + C}{(1 - 2\rho^2)a_1 + a_2 + C}.$$

The slope of the contours of f is

(B.6)
$$\frac{da_2}{da_1} = 1.$$

If a contour is tangent to the ellipse these two slopes must equal, that is the requirement

(B.7)
$$-\frac{a_1 + (1 - 2\rho^2)a_2 + C}{(1 - 2\rho^2)a_1 + a_2 + C} = 1$$

must be satisfied. Solving this expression with respect to a_2 yields

(B.8)
$$a_2 = -a_1 - \frac{C}{1 - \rho^2}.$$

Inserting into the formula of the ellipse, (B.1), yields

(B.9)
$$\left(C - \frac{C}{1 - \rho^2}\right)^2 = -4\rho^2 \left(a_1^2 + a_1 \frac{C}{1 - \rho^2}\right).$$

This is a quadratic equation in a_1 which can be rewritten as

(B.10)
$$a_1^2 + \frac{C}{1 - \rho^2}a_1 + \frac{\rho^2 C^2}{4(1 - \rho^2)^2} = 0$$

and has two solutions:
(B.11)
$$a_1 = \frac{-\left(\frac{C}{1-\rho^2}\right) \pm \sqrt{\left(\frac{C}{1-\rho^2}\right)^2 - 4\left(\frac{\rho^2 C^2}{4(1-\rho^2)^2}\right)}}{2} = -\frac{C}{2(1-\rho^2)}\left(1 \pm \sqrt{1 - \rho^2}\right).$$

For each solution, a corresponding solution for a_2 is found from (B.8). Defining $a^* = -\frac{C}{2(1-\rho^2)}\left(1 - \sqrt{1 - \rho^2}\right)$ and $a^{**} = -\frac{C}{2(1-\rho^2)}\left(1 + \sqrt{1 - \rho^2}\right)$ it is easily verified that the pairs of solutions are $(a_1, a_2) = (a^*, a^{**})$ and $(a_1, a_2) = (a^{**}, a^*)$.

For both pairs of solutions, f is linear in a_3:

$$(\text{B.12}) \quad f(a^*, a^{**}, a_3) = a_3 + \frac{C}{\sqrt{1 - \rho^2}} = a_3\left(1 - \frac{1}{\sqrt{1 - \rho^2}}\right) - \frac{A - B}{\sqrt{1 - \rho^2}},$$

$$(\text{B.13}) \quad f(a^{**}, a^*, a_3) = a_3 - \frac{C}{\sqrt{1 - \rho^2}} = a_3\left(1 + \frac{1}{\sqrt{1 - \rho^2}}\right) + \frac{A - B}{\sqrt{1 - \rho^2}}.$$
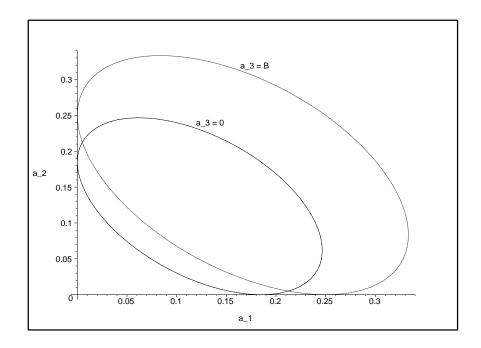
Since $C = -a_3 - A + B < 0$, the expressions after the first equalities in (B.12) and (B.13) imply that $f(a^*, a^{**}, a_3) < f(a^{**}, a^*, a_3)$, for any given a_3. The fact that $f(a^{**}, a^*, a_3)$ is (linearly) increasing in a_3 implies that

$$(\text{B.14}) \qquad\qquad f(a^{**}, a^*, B) = B + \frac{A}{\sqrt{1 - \rho^2}}$$

is the maximum value in the admitted set. Correspondingly, $f(a^*, a^{**}, a_3)$ is decreasing in a_3, implying that the minimum value in the admitted set is

$$(\text{B.15}) \qquad\qquad f(a^*, a^{**}, B) = B - \frac{A}{\sqrt{1 - \rho^2}}.$$

FIGURE 2. Graphs of admitted area when $a_3 = 0$ and $a_3 = B$, for $\rho = 0.5$, $A = 0.250$ and $B = 0.065$

Recent publications in the series Discussion Papers

322 E. Røed Larsen (2002): The Political Economy of Global Warming: From Data to Decisions

323 E. Røed Larsen (2002): Searching for Basic Consumption Patterns: Is the Engel Elasticity of Housing Unity?

324 E. Røed Larsen (2002): Estimating Latent Total Consumption in a Household.

325 E. Røed Larsen (2002): Consumption Inequality in Norway in the 80s and 90s.

326 H.C. Bjørnland and H. Hungnes (2002): Fundamental determinants of the long run real exchange rate:The case of Norway.

327 M. Søberg (2002): A laboratory stress-test of bid, double and offer auctions.

328 M. Søberg (2002): Voting rules and endogenous trading institutions: An experimental study.

329 M. Søberg (2002): The Duhem-Quine thesis and experimental economics: A reinterpretation.

330 A. Raknerud (2002): Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption

331 M.W. Arneberg, J.K. Dagsvik and Z. Jia (2002): Labor Market Modeling Recognizing Latent Job Attributes and Opportunity Constraints. An Empirical Analysis of Labor Market Behavior of Eritrean Women

332 M. Greaker (2002): Eco-labels, Production Related Externalities and Trade

333 J. T. Lind (2002): Small continuous surveys and the Kalman filter

334 B. Halvorsen and T. Willumsen (2002): Willingness to Pay for Dental Fear Treatment. Is Supplying Fear Treatment Social Beneficial?

335 T. O. Thoresen (2002): Reduced Tax Progressivity in Norway in the Nineties. The Effect from Tax Changes

336 M. Søberg (2002): Price formation in monopolistic markets with endogenous diffusion of trading information: An experimental approach

337 A. Bruvoll og B.M. Larsen (2002): Greenhouse gas emissions in Norway. Do carbon taxes work?

338 B. Halvorsen and R. Nesbakken (2002): A conflict of interests in electricity taxation? A micro econometric analysis of household behaviour

339 R. Aaberge and A. Langørgen (2003): Measuring the Benefits from Public Services: The Effects of Local Government Spending on the Distribution of Income in Norway

340 H. C. Bjørnland and H. Hungnes (2003): The importance of interest rates for forecasting the exchange rate

341 A. Bruvoll, T.Fæhn and Birger Strøm (2003): Quantifying Central Hypotheses on Environmental Kuznets Curves for a Rich Economy: A Computable General Equilibrium Study

342 E. Biørn, T. Skjerpen and K.R. Wangen (2003): Parametric Aggregation of Random Coefficient Cobb-Douglas Production Functions: Evidence from Manufacturing Industries

343 B. Bye, B. Strøm and T. Åvitsland (2003): Welfare effects of VAT reforms: A general equilibrium analysis

344 J.K. Dagsvik and S. Strøm (2003): Analyzing Labor Supply Behavior with Latent Job Opportunity Sets and Institutional Choice Constraints

345 A. Raknerud, T. Skjerpen and A. Rygh Swensen (2003): A linear demand system within a Seemingly Unrelated Time Series Equation framework

346 B.M. Larsen and R.Nesbakken (2003): How to quantify household electricity end-use consumption

347 B. Halvorsen, B. M. Larsen and R. Nesbakken (2003): Possibility for hedging from price increases in residential energy demand

348 S. Johansen and A. R. Swensen (2003): More on Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models: Restricted Drift Terms

349 B. Holtsmark (2003): The Kyoto Protocol without USA and Australia - with the Russian Federation as a strategic permit seller

350 J. Larsson (2003): Testing the Multiproduct Hypothesis on Norwegian Aluminium Industry Plants

351 T. Bye (2003): On the Price and Volume Effects from Green Certificates in the Energy Market

352 E. Holmøy (2003): Aggregate Industry Behaviour in a Monopolistic Competition Model with Heterogeneous Firms

353 A. O. Ervik, E.Holmøy and T. Hægeland (2003): A Theory-Based Measure of the Output of the Education Sector

354 E. Halvorsen (2003): A Cohort Analysis of Household Saving in Norway

355 I. Aslaksen and T. Synnestvedt (2003): Corporate environmental protection under uncertainty

356 S. Glomsrød and W. Taoyuan (2003): Coal cleaning: A viable strategy for reduced carbon emissions and improved environment in China?

357 A. Bruvoll T. Bye, J. Larsson og K. Telle (2003): Technological changes in the pulp and paper industry and the role of uniform versus selective environmental policy.

358 J.K. Dagsvik, S. Strøm and Z. Jia (2003): A Stochastic Model for the Utility of Income.

359 M. Rege and K. Telle (2003): Indirect Social Sanctions from Monetarily Unaffected Strangers in a Public Good Game.

360 R. Aaberge (2003): Mean-Spread-Preserving Transformation.

361 E. Halvorsen (2003): Financial Deregulation and Household Saving. The Norwegian Experience Revisited

362 E. Røed Larsen (2003): Are Rich Countries Immune to the Resource Curse? Evidence from Norway's Management of Its Oil Riches

363 E. Røed Larsen and Dag Einar Sommervoll (2003): Rising Inequality of Housing? Evidence from Segmented Housing Price Indices

364 R. Bjørnstad and T. Skjerpen (2003): Technology, Trade and Inequality

365 A. Raknerud, D. Rønningen and T. Skjerpen (2003): A method for improved capital measurement by combining accounts and firm investment data

366 B.J. Holtsmark and K.H. Alfsen (2004): PPP-correction of the IPCC emission scenarios - does it matter?

367 R. Aaberge, U. Colombino, E. Holmøy, B. Strøm and T. Wennemo (2004): Population ageing and fiscal sustainability: An integrated micro-macro analysis of required tax changes

368 E. Røed Larsen (2004): Does the CPI Mirror Costs.of.Living? Engel's Law Suggests Not in Norway

369 T. Skjerpen (2004): The dynamic factor model revisited: the identification problem remains

370 J.K. Dagsvik and A.L. Mathiassen (2004): Agricultural Production with Uncertain Water Supply

371 M. Greaker (2004): Industrial Competitiveness and Diffusion of New Pollution Abatement Technology – a new look at the Porter-hypothesis

372 G. Børnes Ringlund, K.E. Rosendahl and T. Skjerpen (2004): Does oilrig activity react to oil price changes? An empirical investigation

373 G. Liu (2004) Estimating Energy Demand Elasticities for OECD Countries. A Dynamic Panel Data Approach

374 K. Telle and J. Larsson (2004): Do environmental regulations hamper productivity growth? How accounting for improvements of firms' environmental performance can change the conclusion

375 K.R. Wangen (2004): Some Fundamental Problems in Becker, Grossman and Murphy's Implementation of Rational Addiction Theory

376 B.J. Holtsmark and K.H. Alfsen (2004): Implementation of the Kyoto Protocol without Russian participation

377 E. Røed Larsen (2004): Escaping the Resource Curse and the Dutch Disease? When and Why Norway Caught up with and Forged ahead of Its Neughbors

378 L. Andreassen (2004): Mortality, fertility and old age care in a two-sex growth model

379 E. Lund Sagen and F. R. Aune (2004): The Future European Natural Gas Market - are lower gas prices attainable?

380 A. Langørgen and D. Rønningen (2004): Local government preferences, individual needs, and the allocation of social assistance

381 K. Telle (2004): Effects of inspections on plants' regulatory and environmental performance - evidence from Norwegian manufacturing industries

382 T. A. Galloway (2004): To What Extent Is a Transition into Employment Associated with an Exit from Poverty

383 J. F. Bjørnstad and E.Ytterstad (2004): Two-Stage Sampling from a Prediction Point of View

384 A. Bruvoll and T. Fæhn (2004): Transboundary environmental policy effects: Markets and emission leakages

385 P.V. Hansen and L. Lindholt (2004): The market power of OPEC 1973-2001

386 N. Keilman and D. Q. Pham (2004): Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area

387 G. H. Bjertnæs and T. Fæhn (2004): Energy Taxation in a Small, Open Economy: Efficiency Gains under Political Restraints

388 J.K. Dagsvik and S. Strøm (2004): Sectoral Labor Supply, Choice Restrictions and Functional Form

389 B. Halvorsen (2004): Effects of norms, warm-glow and time use on household recycling

390 I. Aslaksen and T. Synnestvedt (2004): Are the Dixit-Pindyck and the Arrow-Fisher-Henry-Hanemann Option Values Equivalent?

391 G. H. Bjønnes, D. Rime and H. O.Aa. Solheim (2004): Liquidity provision in the overnight foreign exchange market

392 T. Åvitsland and J. Aasness (2004): Combining CGE and microsimulation models: Effects on equality of VAT reforms

393 M. Greaker and Eirik. Sagen (2004): Explaining experience curves for LNG liquefaction costs: Competition matter more than learning

394 K. Telle, I. Aslaksen and T. Synnestvedt (2004): "It pays to be green" - a premature conclusion?

395 T. Harding, H. O. Aa. Solheim and A. Benedictow (2004). House ownership and taxes

396 E. Holmøy and B. Strøm (2004): The Social Cost of Government Spending in an Economy with Large Tax Distortions: A CGE Decomposition for Norway

397 T. Hægeland, O. Raaum and K.G. Salvanes (2004): Pupil achievement, school resources and family background

398 I. Aslaksen, B. Natvig and I. Nordal (2004): Environmental risk and the precautionary principle: "Late lessons from early warnings" applied to genetically modified plants

399 J. Møen (2004): When subsidized R&D-firms fail, do they still stimulate growth? Tracing knowledge by following employees across firms

400 B. Halvorsen and Runa Nesbakken (2004): Accounting for differences in choice opportunities in analyses of energy expenditure data

401 T.J. Klette and A. Raknerud (2004): Heterogeneity, productivity and selection: An empirical study of Norwegian manufacturing firms

402 R. Aaberge (2005): Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality

403 F.R. Aune, S. Kverndokk, L. Lindholt and K.E. Rosendahl (2005): Profitability of different instruments in international climate policies

404 Z. Jia (2005): Labor Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure

405 Z. Jia (2005): Retirement Behavior of Working Couples in Norway. A Dynamic Programming Approch

406 Z. Jia (2005): Spousal Influence on Early Retirement Behavior

407 P. Frenger (2005): The elasticity of substitution of superlative price indices

408 M. Mogstad, A. Langørgen and R. Aaberge (2005): Region-specific versus Country-specific Poverty Lines in Analysis of Poverty

409 J.K. Dagsvik (2005) Choice under Uncertainty and Bounded Rationality

410 T. Fæhn, A.G. Gómez-Plana and S. Kverndokk (2005): Can a carbon permit system reduce Spanish unemployment?

411 J. Larsson and K. Telle (2005): Consequences of the IPPC-directive's BAT requirements for abatement costs and emissions

412 R. Aaberge, S. Bjerve and K. Doksum (2005): Modeling Concentration and Dispersion in Multiple Regression

413 E. Holmøy and K.M. Heide (2005): Is Norway immune to Dutch Disease? CGE Estimates of Sustainable Wage Growth and De-industrialisation

414 K.R. Wangen (2005): An Expenditure Based Estimate of Britain's Black Economy Revisited