DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

CONSUMER DEMAND BASED ESTIMATES OF THE BLACK ECONOMY: PARAMETRIC AND NONPARAMETRIC APPROACHES

Panayiota Lyssiotou,
Panos Pashardes and
Thanasis Stengos

Discussion Paper 99-17

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: ++357-2-892101, Fax: ++357-2-750310
Web site: http://www.econ.ucy.ac.cy

Consumer Demand Based Estimates of the Black Economy: Parametric and Nonparametric Approaches

Panayiota Lyssiotou
University of Cyprus, Nicosia 1678, Cyprus (p.lyssiotou@ucy.ac.cy)

Panos Pashardes
University of Cyprus, Nicosia 1678, Cyprus (p.pashardes@ucy.ac.cy)

Thanasis Stengos[¤]
University of Guelph, Guelph, Ontario, Canada N1G 2W1 (thanasis@css.uoguelph.ca)

October 1999

Abstract

We propose a consumer demand system approach to estimating the size of the black economy where alternative hypotheses a®ecting the empirical results can be tested in a nested framework. This approach allows for the estimation of the under-reporting of household income from various sources, dispensing with the need to use arbitrary criteria to classify households by their main source of income. It also avoids potential bias in black economy estimates arising from mistaking preference heterogeneity (substitution) as income e®ects. We illustrate these arguments by estimating the extent to which self-employment income in the UK is under-reported using parametric and nonparametric techniques.

JEL Classi¯cation: C14, D12

Keywords: Engel curves, consumer behavior, implicit separability, hidden/shadow economy

1 Introduction

The black economy, broadly de¯ned as the economic activities that are hidden from public authorities to avoid taxation, has recently received increased attention in the

[¤]We would like to thank the University of Cyprus for ¯nancial support and the Central Statistical O±ce for making available the UK Family Expenditure Survey data through the ESRC Data Archive. We are solely responsible for the interpretation of the data and all errors.

literature due to arguments that the rising tax burden and state regulation is driving economic activity underground (Tanzi and Schuknecht 1997 and Enste and Schneider, 1998). The usefulness of attempts to estimate the size of the black economy, however, is a controversial issue. The skeptics emphasize the difficulties associated with the precise definition of black economy activities and point to the widely different estimates of the size of these activities resulting from alternative methods (Dixon 1999, Tanzi 1999 and Thomas 1999). Authors approaching the subject in an optimistic frame of mind emphasize positive aspects of estimating the size of the black economy such as the correct measurement of GDP and other macroeconomic variables (e.g. employment) and the design of policy and econometric models (Bhattacharyya 1999 and Giles 1999).

The controversy on estimation issues, however, does not subtract from the widespread concern about the undesired consequences of the black economy. Undeclared economic activities reduce the tax base, thereby undermining the financing of public goods and social protection. A country trying to curtail the loss of tax revenue by raising tax rates can create a vicious cycle, as this reinforces the incentive not to declare economic activities to the public authorities. High tax rates also have a distorting effect resulting in inefficient allocation of resources due to unequal opportunities and willingness to evade in different sectors of the economy and can undermine the ability of the country to compete internationally. Distortions in the labor market can also arise because individuals working in the black economy may have difficulty in moving from one job to another. Furthermore, in countries where health, pension and other social security payments are not universal, individuals working in the black economy and their families are outside the social protection net. In developing economies where the black economy is controlled by the affluent, effective taxation can become regressive.

Most empirical approaches to estimating the size of black economy activities in the literature rely on macroeconomic relationships thought to contain information about such activities. Included among these approaches are the `currency ratio' method pioneered by Cagan (1958) and further developed by Tanzi (1983) and Bhattacharyya (1990), the `transactions' method (Feige 1979) and the `MIMIC' (Multiple Indicators Multiple Causes) method first considered in the context of the black economy by Frey and Weck-Hannemann (1984) and subsequently applied by Schneider (1997) and Giles (1997). The macroeconomic approaches to estimating the size of the black economy are criticized for not being based on theory and for employing flawed econometric techniques (Thomas 1999).

Studies using microeconomic data for the estimation of the black economy include

(i) `direct' methods employing tax audits and data collected from surveys designed to measure the undeclared taxable income (used by the US Inland Revenue Service and discussed in Feinstein and Dixon,1999) and (ii) an `expenditure-based' method proposed by Pissarides and Weber (1989) where an Engel curve (demand for food) estimated from individual household data is used for the calculation of the under-reporting of income by households with a self-employed head.

This paper builds on the Pissarides and Weber (1989) expenditure-based approach and proposes a complete demand system framework for the estimation of the size of the black economy using cross-section individual household data. This framework is argued to have two advantages over the Pissarides and Weber (1989) single equation method: it (i) avoids potential bias from confusing preference heterogeneity (substitution) with under-reporting (income) e®ects in consumer demand and (ii) dispenses with the need to classify households according to their main source of income, i.e. self-employed, wage earners etc. We believe that the latter advantage is important not only because the classi¯cation of households by their main source of income is based on an arbitrary criterion (income share) but also because such classi¯cation can exclude from the analysis income from black economy activities representing a substantial but not the main source of household income (second jobs, fees etc).

Furthermore, the analysis in the paper goes some way towards meeting the criticism expressed against other approaches to estimating the size of the black economy. On the theory side, the proposed method is based on sound principles of consumer behavior and by linking undeclared income to its sources takes account of the point made by Cowell (1990) that tax evasion is tied to labor supply decisions (type of employment). On the estimation side the analysis in the paper not only provides a framework in which alternative hypotheses a®ecting the empirical results can be tested in a nested framework but also proposes a nonparametric measure of the size of the black economy that can be used as a benchmark for the evaluation of parametric results. This measure is based on the distance between two nonparametric regression functions, as suggested in Pinkse and Robinson (1995) and employed by Pendakur (1999) to compute equivalent income scales.

The empirical analysis is based on UK data drawn from the 1993 Family Expenditure Survey (FES) and containing information about income from various sources, expenditure on a detailed commodity breakdown and a large number of demographic and other household characteristics found to be signi¯cant in Pissarides and Weber (1989) and other empirical studies of consumer behavior based on individual household data (e.g. Blundell et. al. 1993). Separate estimations are performed for households with blue and white collar heads using the parametric complete demand system and the parametric and nonparametric single equation methods.

The structure of the paper is as follows. Section 2 models the under-reporting of income from various sources in the context of a complete demand system and discusses the empirical advantages of this modelling approach vis-a-vis the single equation approach. Section 3 reports the results obtained from application to the UK data and considers their implications for the size of the black economy. Section 4 concludes the paper.

2 Under-reporting of income and consumer demand

We consider households to have the implicitly separable preference structure (Gorman 1976, Deaton and Muellbauer 1980) defined by the cost function

$$C(p, U) = F\left[c(p, U), d(r, U), U\right], \tag{1}$$

where U is the household utility level, p the price vector of nondurables and r the vector of prices (user costs) for durable goods. The subcost functions $c(.)$ and $d(.)$ reflect the prices (unit cost) of nondurable and durable goods paid by the household, respectively. These are increasing in U and linearly homogeneous in prices.

Household demands obtained in the context of (1) can be thought to result from a standard two-stage budgeting process: first total household expenditure is allocated between nondurable and durable goods according to the efficient (cost minimizing) rule, $\partial F[.]/\partial c(.)$ and $\partial F/\partial d(.)$. Demand for the i^{th} good in the nondurable category is given by

$$q_i = \frac{\partial F[.]}{\partial c(.)}\frac{\partial c(.)}{\partial p_i} \tag{2}$$

and household expenditure on nondurables by

$$y \equiv \Sigma_i q_i p_i = \frac{\partial F[.]}{\partial c(.)}\Sigma_i \frac{\partial c(.)}{\partial p_i} p_i = \frac{\partial F[.]}{\partial c(.)} c(.). \tag{3}$$

Household expenditure on the i^{th} good is given by

$$p_i q_i \equiv p_i \frac{\partial F[.]}{\partial c(.)}\frac{\partial c(.)}{\partial p_i} = p_i y \frac{\partial c(.)}{c(.)\,\partial p_i} \tag{4}$$

and as share in the expenditure on nondurable goods $w_i = \partial \ln c(.)/\partial \ln p_i$.

We assume that the unit cost of nondurables $c(.)$, has the Quadratic Logarithmic form (Lewbel 1990)

$$\ln c(p;U) = a(p) + b(p)\left[\frac{U}{1 - g(p)U}\right]; \tag{5}$$

where $a(p)$, $b(p)$ and $g(p)$ are some functions homogeneous in p. This yields Hicksian shares

$$w_i = a_i(p) + b_i(p)\left[\frac{U}{1 - g(p)U}\right] + \lambda_i(p)\left[\frac{U}{1 - g(p)U}\right]^2; \tag{6}$$

where $a_i(p) = \partial \ln a(p)/\partial \ln p_i$; $b_i(p) = \partial \ln b(p)/\partial \ln p_i$ and $\lambda_i(p) = b(p)\partial \ln g(p)/\partial \ln p_i$. Note that U here is the household utility level obtained from expenditure on nondurable and durable goods. Therefore, to obtain the Marshallian demands for nondurable goods we substitute U in (6) for the indirect utility function $V(p;r;Y)$, where Y is household income. In the present context this is convenient because demand for non-durable goods, assumed to be accurately reported, can be used as a benchmark for the investigation of the degree to which observed income is understated.

At base period prices $p_i = 1$ and $r_i = 1$, all i; utility can be redefined as $u = U/[1 - g_o U]$ where g_o is some constant such that $\partial u/\partial U > 0$. Introducing the h subscript to denote the individual household, the budget shares (6) can then be written as a system of Engel curves,

$$w_{ih} = a_i + \beta_i[\ln Y_h^*] + \lambda_i[\ln Y_h^*]^2; \tag{7}$$

where Y_h^* is the true level of income consisting of K components distinguished by their source, such as income from employment in the private or public sector, self-employment, fees, second job etc. We assume that each true component of income is proportional to its observed counterpart,

$$Y_{kh}^* = \theta_k Y_{kh}; \quad k = 1,...,K; \tag{8}$$

where Y_{kh} is the observed level of the k^{th} component of income of the h^{th} household and $\theta_k \geq 1$ shows the factor by which the observed income must be multiplied in order to become equal to the true income. Replacing (8) in (7) we obtain

$$w_{ih} = a_i + \beta_i[\ln Y_h + \ln(\Sigma_k \theta_k y_{kh})] + \lambda_i[\ln Y_h + \ln(\Sigma_k \theta_k y_{kh})]^2; \tag{9}$$

where $Y_h = \Sigma_k Y_{kh}$ and $y_{kh} = Y_{kh}/Y_h$.

The intuition behind the above modelling of under-reporting is as follows. Suppose that for a given k^{th} income component $\theta_k > 1$.[1] Then $\ln Y_h < \ln Y_h^*$ so that the household in question will behave as if it is on a higher utility level, i.e. allocate more expenditure to luxuries and less expenditure to necessities than otherwise. The parameter θ_k shows the extent to which the observed k^{th} component of income will have to increase in order to compensate for the understatement of the observed household income, i.e. make the allocation of expenditure on nondurable luxuries and necessities as if $\ln Y_{kh} = \ln Y_{kh}^*$.

This is illustrated by Figure 1 where we initially assume that all households have identical preferences and current income is allocated between two goods, food (necessity) and non-food (luxury) according to the continuous Engel curve OM in the diagram. Point A corresponds to the case of a household that reports income OC and has budget shares AC for food and AG for non-food. Under the preferences given by the Engel curve OM in the diagram, however, these budget shares are consistent with the higher level of income OD. In other words the reported level of income must be raised from OC to OD for the budget shares to be as implied by the Engel curve. The θ-parameter in this case is equal to ratio OD/OC.

In the context of the budget share system (9) the Pissarides and Weber (1989) approach can be shown to correspond to the case where (i) y_{kh} is replaced by a dummy variable $D_{kh} = 1$ if $y_{kh} > \tilde{y}_{kh}$ and $D_{kh} = 0$ otherwise, where $\tilde{y}_{kh}$ is some ad hoc value of y_{kh} and (ii) $\lambda_i = 0$ all i. Then (9) is written as

$$w_{ih} = a_i + \beta_i \ln Y_h + \Sigma_k \theta_{ik}^* D_{kh}, \qquad (10)$$

where $\theta_{ik}^* = \beta_i \ln \theta_k$. Pissarides and Weber (1989) assume that $\ln \theta_k = \mu_k + \upsilon_h$ and by the log-normality assumption the logarithm of mean θ_k is $\ln \bar{\theta}_k = \mu_k + \frac{1}{2}\sigma_\upsilon^2$.

An important limitation of the single, linear equation approach is that it imposes separability in the components of income defined as

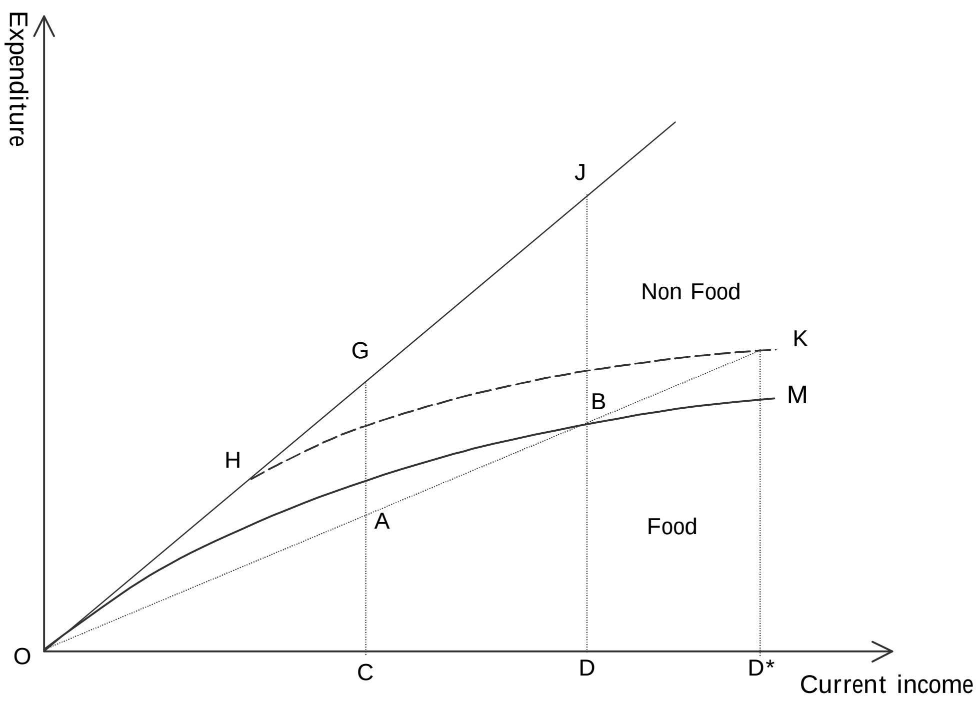
$$\frac{\partial w_{ih}/\partial D_h}{\partial w_{jh}/\partial D_h} = \frac{\partial w_{ih}/\partial \ln Y_h}{\partial w_{jh}/\partial \ln Y_h}, \qquad (11)$$

i.e. income components change the relative demand for two goods in the same way as an income change would affect this relative demand. This implies absence of preference heterogeneity associated with the sources of income: all differences in consumer demand

[1] It should be noted here that the restriction $\Sigma_k x_{kh} = 1$ implies that θ_k must be fixed for at least one income component. In other words the understatement of the various income components can only be estimated relative to the rest. The obvious solution to this problem is to set $\theta_k = 1$ for the k^{th} component of income for which we have reasons to believe that it is accurately reported, e.g. income from salary-paid employment.

are attributed to di®erences in (mis)reported income. Yet, at any given level of income the pattern of demand can vary between consumers due to di®erences in demographic and other characteristics such as the number of children in various age groups, housing tenure and location, economic status, age and other characteristics of head etc. To the extent that the employment position of head, as de¯ned by D_{kh}, is among the household characteristics responsible for this preference heterogeneity, then the estimated $\mu^{¤}_{ik}$ in (10) will pick up the e®ects of both, the heterogeneity of preferences due to the employment position of head and the understatement of the k^{th} component of income. The distinction between the preference heterogeneity (substitution) and under-reporting (income) e®ects requires a system approach to be determined.[2]

Figure 1: Income understatement in a system of Engel curves



This point is illustrated in the diagram of Figure 1 using the discontinuous Engel curve HK, assumed to correspond to households with income from black economy activities. In this case the size of the black economy corresponds to the ratio $OD^{¤}/OC$

---

[2]Note that this problem can also arise in systems of linear Engel curves, e.g. for a given employment dummy D_{1h} the system $w_{ih} = a_i + a_{i1}D_{1h} + ¯_i[\ln Y_h + \mu_1 D_{1h}]$ is observationally equivalent to $w_{ih} = a_i + a^{¤}_{i1}D_{1h} + ¯_i \ln Y_h$ where $a^{¤}_{i1} = a_{i1} + ¯_i\mu_1$ captures both the preference heterogeneity (a_{i1}) and under-reporting (μ_1) e®ects of D_{1h}.

rather than OD/OC, i.e. ignoring the fact that those under-reporting have a different Engel curve (consume more food at any level of income) than other households results in a downward biased black economy estimate.[3]

3 Empirical analysis

This section reports the parameter estimates obtained from a single Engel curve for food (10) and the quadratic system of budget share equations (9), where the preference heterogeneity and under-reporting effects discussed above can be estimated separately. Furthermore, given that both the system and single equation approaches are subject to a potential specification error arising from the choice of functional form, we also estimate the size of the black economy using a nonparametric distance measure. By comparing the results obtained from these alternative approaches we assess the implications of (i) the choice of functional form and (ii) preference heterogeneity on the estimated size of the black economy. The data used are drawn from the 1993 UK Family Expenditure Survey (FES) and consist of 1750 households whose head is in employment and their main source of income (as defined by the FES) is from wages or self-employment.

3.1 Single equation approaches

First we consider the results obtained from the single equation approach. We estimate the quadratic Engel curve for food

$$\ln Y_{Fh} = a_F + \delta_F S_h + \beta_F \ln Y_h + \lambda_F (\ln Y_h)^2 + \Sigma_j a_{Fj} z_{jh} + u_{Fh}; \qquad (12)$$

where $S_h = 1$ if the main source of household income comes from self-employment and $S_h = 0$ otherwise; $z_{jh};\ j = 1;\ldots;J$ are household characteristics and υ_{Fh} an error term. The vector of household characteristics, z_h; includes: the number of children in various age categories; the age, occupation, economic position, profession etc of the household head and spouse; housing tenure and geographical location; numbers of rooms in the house, cars and ownership of other durable goods; and a large number of other variables found to be significant in Pissarides and Weber (1989) and in other empirical studies of consumer behavior based on individual household data (e.g. Blundell et. al. 1993).

[3] The problem here is analogous to the substitution bias in the construction of price indices, except for the fact that here the direction of the bias cannot be a priori determined because the substitution is due to the change in the sources of income, not the relative prices.

The Engel curve for food (12) is estimated separately for households with a head in blue and white collar occupation by 2SLS.[4] The top part of Table 1 reports the parameter estimates corresponding to the self-employed dummy and log income for the two household types, white and blue collar. Under the heading `Quadratic model' are the results obtained from the unrestricted estimation of (12); whereas under the heading `Linear model' are the results obtained subject to the restriction $\lambda_F = 0$. Using the empirical results obtained from the linear model and the method proposed by Pissarides and Weber (1989) we have calculated the upper and lower bounds of the under-reporting parameter for blue and white collar households. These bounds are shown in section (a) in the bottom part of Table 1 and, on average, imply an under-reporting coefficient of 1.39 for blue collar and 1.18 for white collar households. The estimates obtained here are comparable to those reported in Pissarides and Weber (1989).

The parameter corresponding to the log income squared in Table 1 is significant for both blue and white collar households. It then follows that the estimates of under-reporting based on the linear model may suffer from misspecification bias. Given that the quadratic model is also subject to a specification error, the implications of using a linear model to estimate the black economy coefficient are assessed through comparison to nonparametric results, where the black economy parameter is measured as the distance between two nonparametric regression functions described as follows.[5]

Suppose we are concerned with the estimation of the following regressions:

$$Y_{Fh}^{w} = c_i^{w} z_h + m(Y_h^{w}) + u_h^{w}; \quad h = 1; \ldots; H^{w} \tag{13}$$

$$Y_{Fh}^{s} = c_i^{s} z_h + m(Y_h^{s}) + u_h^{s}; \quad h = 1; \ldots; H^{s}; \tag{14}$$

where z_h is the vector of household characteristics defined earlier; the superscripts w and s denote households whose main source of income is from wages and self-employment, respectively.

To remove heterogeneity (influence of z_h; excluding the main source of income)

[4] More precisely, $\ln Y_h$ and S_h are instrumented with age and years of schooling of husband and wife and their squares, number of children in different age groups and their squares, number of rooms in the house and a large number of dummies (head self-employed, car, season, housing tenure, washing machine, central heating, wife self-employed, wife working full time, wife working part time,wife unemployed etc.) and interactions of the dummy for self employed head with the variables above and dummies for head in subsidiary job as self-employed, dummy for head in subsidiary job as employee, dummy for wife in subsidiary job as self employed and wife in subsidiary job as employee. In Pissarides and Weber (1989), households are classified as self-employed if more than 25% of their income is from self-employment. We have found that using this definition (rather than the FES one) does not affect the results.

[5] Also see Pinkse and Robinson (1995). A similar distance measure was proposed by Härdle and Marron (1990) for the fixed design case.

from Y_{Fh}^{ℓ}; for $\ell = w; s$; we consistently estimate c_i using a nearest neighbor estimator proposed by Estes and Honore (1995) and Yatchew (1997) and used in Lyssiotou et al (1999)[6]: given that Y_h^{ℓ} is a continuous variable we first sort the data by Y_h^{ℓ} and compute the first differences $\Delta Y_{Fh}^{\ell} = Y_{Fh}^{\ell} - Y_{Fh-1}^{\ell}$ and $\Delta z_h = z_h - z_{h-1}$ on the sorted data; we then run the regressions $\Delta Y_{Fh}^{\ell} = c_i^{\ell} \Delta z_h$ + errors, to obtain a consistent estimate of c_i^{ℓ}, say $\hat{c}_i^{\ell}$ and compute $\tilde{Y}_{Fh}^{\ell} = Y_{Fh}^{\ell} - \hat{c}_i^{\ell} z_h$.

The regression function $m(\tilde{Y}_j^{\ell})$ is estimated for $\ell = w; s$ at a given point $\tilde{Y}_o^{\ell}$ using kernel methods as

$$\hat{m}(\tilde{Y}_j^{\ell} = \tilde{Y}_o^{\ell}) = \frac{\sum_{h=1}^{H} K\left(\frac{\tilde{Y}_h^{\ell} - \tilde{Y}_j^{\ell}}{b}\right) Y_{Fh}^{\ell}}{\sum_{h=1}^{H} K\left(\frac{\tilde{Y}_h^{\ell} - \tilde{Y}_j^{\ell}}{b}\right)} = \frac{r(\tilde{Y}_j^{\ell} = \tilde{Y}_o^{\ell})}{f(\tilde{Y}_j^{\ell} = \tilde{Y}_o^{\ell})}; \qquad (15)$$

where $K(:)$ and b are the kernel and the bandwidth respectively. The choice of the kernel function is the Gaussian kernel and the bandwidth is chosen by cross-validation, see Hardle and Marron (1990).

Table 1: Empirical results based on demand for food

Effect on logarithmic food expenditure	Quadratic model Parameter	Quadratic model t-ratio	Linear model Parameter	Linear model t-ratio
Self-employment dummy	0.16463	4.37	0.11360	3.05
Self-employment x white collar	-0.09055	-1.71	-0.02758	-0.53
Log income	0.28814	5.99	0.37503	7.94
Log income x white collar	0.13841	0.51	0.15922	6.48
Log income square	0.02355	9.22	-	-
Log income square x white collar	-0.01935	-0.90	-	-
Root Mean Square Error	0.39596		0.39624	
R-square	0.25780		0.23460	
Undereporting parameters				
(a) Linear model:	Lower bound		Upper bound	
Blue collar households	1.37		1.41	
White collar households	1.09		1.26	
(b) Nonparametric model (t-ratio):				
Blue collar households	1.64 (13.8)			
White collar households	1.35 (14.3)			

[6] An alternative way of obtaining consistent estimates of c_i is to use Robinson's (1988) semiparametric estimator. The nearest neighbor estimator proposed by Estes and Honore (1995) produces less efficient estimates of c_i than Robinson's (1988) estimator, but has computational advantages and is easier to implement.

Following Pendakur (1999), we fit the estimated nonparametric expenditure equation of each household type, the reference w and the non-reference s type. Then the distance between the estimated expenditure function of the two household types is found by minimizing a Loss Function that was suggested in this context by Pinkse and Robinson (1995). The distance measure is based on the following relationship between the expenditures of the w and s household types:

$$m(\Upsilon_h^s) = m(\Upsilon_h^w - \theta^s) + \delta^s. \tag{16}$$

Then using the result that $m(\Upsilon_h^l) = r(\Upsilon_h^l)/f(\Upsilon_h^l)$ for $l = s, w$ we obtain

$$f(\Upsilon_h^w - \theta^s)r(\Upsilon_h^s) = f(\Upsilon_h^s)r(\Upsilon_h^w - \theta^s) + \delta^s f(\Upsilon_h^s)f(\Upsilon_h^w - \theta^s). \tag{17}$$

Pinkse and Robinson (1995) suggest the minimization of the integrated squared difference between the two sides of equation (17) with respect to the parameters θ^s and δ^s using a simple gridsearch over a wide span of values.

The nonparametric estimate of the black economy coefficient obtained from the model above is 1.64 for blue and 1.35 for white collar households (section b in the lower part of Table 1). On the basis of these results one can argue that the choice of a linear functional form for the food expenditure equation causes a downward bias to the black economy coefficient, especially for households with head in blue collar occupation.

3.2 Demand system approach

In the context of the demand system approach we consider six categories of non-durable goods: food, alcohol, fuel, clothing, personal goods/services and leisure goods/services.[7] For comparability with the results obtained from the single equation approach, here we also consider two sources of income: wages and self employment income. Thus, the estimated system of budget shares is

$$\begin{aligned} w_{ih} &= a_i + \Sigma_j a_{ij} z_{jhi} + \delta_i y_h^s + \beta_i [\ln Y_h + \ln(\theta_0 y_h^\omega + \theta_1 y_h^s)] \\ &\quad + \lambda_i [\ln Y_h + \ln(\theta_0 y_h^\omega + \theta_1 y_h^s)]^2 + \upsilon_{ih}, \end{aligned} \tag{18}$$

where y_h^ω; and y_h^s are the wage and self-employment proportions of Y_h, respectively, z_{jh}; $j = 1, ..., J$ is the vector of households characteristics described earlier and υ_{ih} an error term. The parameter θ_0 is set equal to unity on the assumption that wage income

[7] The assumption concerning durable goods is that they act as conditioning variables and this is modelled by including dummies for housing tenure, the size and value of the house, car ownership, the presence of smokers in the household etc in the vector of household characteristics.

is correctly reported. Therefore the parameter θ_1 is interpreted as the black economy coefficient, i.e. the proportion by which reported self-employment income needs to be raised to reach its correct level, in the sense explained in section 2.

The system of equations (18) is estimated by nonlinear three-stage least squares, where the components of income are instrumented to avoid measurement error due to transitory elements. In addition to all the exogenous variables, their squares and cross-products, the years of schooling (including their squares and cross-products) of both adults in the household are used as instruments. The estimation is performed using all the observations in the sample allowing for differences in the parameters corresponding to blue and white collar households through dummies.

Table 2 reports results of interest obtained from the estimation of (18): the parameters and t-ratios reflecting under-reporting (θ_1), preference heterogeneity (δ_i), log income (β_i) and log income square (λ_i) effects. Under the heading `blue collar households' are estimates corresponding to households with head in blue collar occupation and under the heading `white collar dummies' the estimated differences between these households and households with head in white collar occupation. Only the parameters estimates of five budget share equations are reported in the table since the parameters of the sixth equation (leisure goods/services) are redundant due to adding up.[8]

The under-reporting parameters for self-employment income is 1.96 for blue collar and 1.61 (i.e. 1.96-0.35) for white collar households. These estimates are well above those obtained from the single equation approach above. Regarding the preference heterogeneity effects, the results in Table 2 suggest that households at a given level of income with a higher proportion from self-employment tend to have a higher share of necessities (food and fuel) in consumer expenditure at the expense of other goods. Possibly this is because these households tend to use their home as workplace, thereby spending relatively more on food (rather than, for example, eating out) and fuel for heating.

As we have argued using the diagram of Figure 1, a positive correlation between the income elasticity and the preferences of the self-employed can result in the a downward biased under-reporting parameter. To test this hypothesis we re-estimate the system of budget share equations (18) imposing no preference heterogeneity ($\delta_i = 0$; all i). As seen in Table 2, section b, this hypothesis is rejected and causes the under-

[8]Following the acceptance of a hypothesis test (based on the χ^2 distribution) the substitution parameters δ_i are set to be the same for blue and white collar households to obtain better determined under-reporting parameters.

reporting parameters to be reduced to 1.70 for blue collar and 1.29 for white collar households. Notably, these estimates are close to those obtained from the nonparametric single-equation method which also imposes the same restriction on the demand for food equation. This is not surprising, given that (i) the most sizeable preference heterogeneity e®ect comes from the food equation and (ii) the quadratic logarithmic model is generally found to be an adequate speci¯cation for individual household data when tested against nonparametric alternatives (e.g. Banks et al 1997 and Lyssiotou et al 1999). Given (ii) and the fact that preference heterogeneity cannot be rejected, we conclude that the demand system approach yields more reliable estimates of the black economy than the single-equation (parametric and nonparametric) approach.

Table 2: Empirical results based on the complete demand system

a. Budget share effects:		Blue collar		White collar dummies	
		Parameter	t-ratio	Parameter	t-ratio
Under-reporting		1.96041	7.66	-0.35439	-1.38
Self-employment:	Food	0.03809	4.08	0.03809	4.08
	Alcohol	-0.00179	-0.30	-0.00179	-0.30
	Fuel	0.01266	2.26	0.01266	2.26
	Clothing	-0.01026	-1.16	-0.01026	-1.16
	Personal	-0.01178	-2.07	-0.01178	-2.07
Log income:	Food	-0.18525	-2.68	0.27926	2.95
	Alcohol	0.08830	2.01	-0.09185	-1.76
	Fuel	-0.19090	-4.84	0.12719	2.47
	Clothing	0.21547	3.24	-0.16398	-2.00
	Personal	0.02057	0.48	0.02305	0.43
Log income square:	Food	0.04476	1.97	-0.06286	-2.61
	Alcohol	-0.01855	-1.28	0.01995	1.35
	Fuel	0.04750	3.58	-0.04227	-3.12
	Clothing	-0.05743	-2.59	0.05397	2.38
	Personal	-0.00478	-0.34	-0.00175	-0.12
b. Functional form effects:		Blue collar		White collar dummies	
		Parameter	t-ratio	Parameter	t-ratio
Separability		1.69569	10.5	-0.48602	2.60
Chi-square value (d.o.f.)		24(5), rejected at 5% signifjcance level			
Linearity		1.43273	7.2	0.10784	0.45
Chi-square value (d.o.f.)		54(15), rejected at 5% signifjcance level			

The second hypothesis tested is linearity ($¸_i$ = 0; all i): This involves ¯fteen parameter restrictions: ¯ve for blue collar households, ¯ve for white collar households and

¯ve for the interaction terms between the log income squared and the dummy for the presence of children in the household. The interesting result here is that, imposing linearity reduces the under-reporting parameter from 1.96 to 1.43 for the blue households whereas for the white collar households the reduction of the same parameter is much smaller, from 1.61 to 1.54 (or to 1.43, if we consider that in this case the white collar dummy in insigni¯cant). Therefore, the linearity restriction (i) has a more profound downward e®ect on the under-reporting parameter estimate than the separability restriction and (ii) this is particularly so in the case of blue collar households. The latter result re-iterates the conclusion reached from the comparison between the parametric and nonparametric estimates of the demand for food equation above.

4 Conclusions

This paper considers the problem of black economy in the context of consumer theory and proposes a method of estimating the under-reporting of income using parameters of a complete demand system. It compares the results obtained from this method with those obtained from two single equation approaches: a log linear Engel curve and a measure of distance between two nonparametric Engel curves.

The log linear Engel curve approach is argued to rely on restrictive functional form assumptions and cannot distinguish between the preference heterogeneity (substitution) and under-reporting (income) e®ects attributed to the sources of income. Furthermore, it requires households to be classi¯ed by their main source of income, thereby introducing an element of arbitrariness in empirical application. The nonparametric approach improves on the log linear Engel approach insofar as it makes no assumptions about the functional form of the Engel curve and the distribution of errors, including measurement errors in observed right hand side variables (transitory income).

The complete demand system approach needs no functional form assumptions other than those required by fundamental principles of consumer theory and generally found to hold for the quadratic logarithmic demand system used for the empirical analysis in the paper. Furthermore, this approach not only allows for separate preference heterogeneity and under-reporting e®ects, but also allows for all income sources to be included in the analysis (i) regardless of their relative contribution to total household income and (ii) without the need to classify households by their main source of income.

The empirical analysis, based on individual household data drawn from the UK Family Expenditure Survey (1993), suggests that the log linear Engel curve approach is statistically inferior to the nonparametric approach and results in lower estimates of the black economy coe±cients. The estimates obtained from the complete demand system approach con¯rm that restricting the functional form of Engel curves to be log linear result in lower black economy estimates. Furthermore, the complete demand system results suggest that failing to allow for preference heterogeneity also leads to understated black economy parameters, i.e. we ¯nd that households with higher income from self-employment tend to have higher budget shares of necessities (food and fuel). The hypothesis of preference homogeneity is rejected by the data.

According to the results obtained from the complete demand system, self-employment income reported by households with head in blue collar occupation needs to be scaled up by a factor of 1.96 to correct for under-reporting, whereas the corresponding ¯gure for self-employment income reported by households with head in white collar occupation is 1.61. Considering that households with head in blue collar occupation account for 45.6% and households with head in white collar occupation for 54.4% of the reported self-employment income, our estimates suggest that self-employment related black economy activities in the UK amount to 9.24% of GDP.

References

Banks J., R. Blundell and A. Lewbel (1997), Quadratic Engel Curves and Consumer Demand, Review of Economics and Statistics, 4,527-539.

Bhattacharyya D. (1990), An econometric method of estimating the `hidden' economy, United Kingdom (1960-1984): estimates and tests, Economic Journal, 100, 703-717

Bhattacharyya D. (1999), On the economic rationale of estimating the hidden economy, Economic Journal, 109, F348-F359.

Blundell R., P. Pashardes and G. Weber (1993), What do we learn about consumer demand patterns from micro-data, American Economic Review, 83, 570-597.

Cagan P. (1958), The demand for currency relative to the total money supply, Journal of Political Economy, 66, 303-329.

Cowell F. (1990), Cheating the Government, Cambridge, Massachusetts: The MIT Press.

Deaton A. and J. Muellbauer (1980), An Almost Ideal Demand System, American Economic Review, 70, 312-326.

Dixon H. (1999), Controversy: on the use of the hidden economy's estimates, Economic Journal, 109, F335-F337.

Estes, E., and Honore, B. (1995), Partially Linear Regression using one Nearest Neighbor, Princeton University, Department of Economics manuscript.

Enste D. and S. Friedrich (1998), Increasing shadow economies all over the world - ¯ction or reality? A survey of the global evidence of their size and of their impact from 1970-1995, Discussion Paper No. 26, IZA.

Feige E. (1979), (ed.) How big is the irrecular economy, Challenge, 22, 5-13.

Feinstein J. and.H. Dixon (1999), Approaches for estimating noncompliance: examples from Federal taxation in the United States, Economic Journal, 109, F360-F369.

Frey B. and H. Weck-Hannemann (1984), The hidden economy as an unobserved variable, European Economic Review, 26, 33-53.

Giles D. (1997), The hidden economy and the tax gap in New Zealand: a latent variable analysis, Discussion Paper 97-8, Department of Economics, University of Victoria.

Giles D. (1999), Measuring the hidden economy: implications for econometric modelling, Economic Journal, 109, F370-F389.

Gorman T. (1976), Tricks with utility functions, in M. Artis and R. Nobay (eds) Essays in Economic Analysis, Cambridge University Press.

Hardle W. and J. Marron (1990), Semiparametric comparison of regression curves, Annals of Statistics 18, 63-89.

Lewbel A. (1990), Full Rank Demand Systems, International Economic Review, 31, 289-300.

Lyssiotou P., P. Pashardes and T. Stengos (1999), Preference Heterogeneity and the Rank of Demand Systems, Journal of Business and Economic Statistics, 17, 248-252.

Pissarides C. and G. Weber (1989), An expenditure-based estimate of Britain's black economy, Journal of Public Economics, 39, 17-32.

Pinkse C. and P. Robinson (1995), Pooling nonparametric estimates of regression functions with a similar shape: in G. Maddala, P. Phillips and T. Srinivisan, (eds), Advances in Econometrics and Quantitative Economics, 172-195.

Pendakur K. (1999), Semiparametric estimates and tests of base-independence equivalence scales, Journal of Econometrics, 88 1-40.

Robinson, P. (1988), Root-N Consistent Semiparametric Regression, Econometrica, 56, 931-954.

Schneider F. (1997), Empirical results for the size of the shadow economy of Western European countries over time, Working Paper 9710, Linz University.

Smith S. (1986), Britain's Shadow Economy, Oxford: Oxford University Press.

Tanzi V. (1983), The underground economy in the United States: annual estimates 1930-1980, IMF Sta® Papers, 30, 283-305.

Tanzi V. (1999), Uses and abuses of estimates of the underground economy, Economic Journal, 109 (June), F338-F347.

Tanzi V. and L. Schuknecht (1997), Reconsidering the ¯scal role of government: the international prospective, American Economic Review, 87, 164-168.

Thomas J. (1999), Quantifying the black economy: `measurement without theory' yet again, Economic Journal, 109, F381-F337.

Thomas J. (1992), Informal economic activity, LSE Handbooks in Economics, London, Harvester Wheatsheaf.

Yatchew J. (1997), An elementary estimator for the partial linear model, Economics Letters, 57, 135-143.

SELECTED RECENT PUBLICATIONS

Bertaut, C. and Haliassos, M,. "Precautionary Portfolio Behavior from a Life - Cycle Perspective", *Journal of Economic Dynamics and Control,* 21, 1511-1542, 1997).

Caporale, G. and Pittis, N. "Causality and Forecasting in Incomplete Systems", *Journal of Forecasting*, 1997, 16, 6, 425-437.

Caporale, G. and Pittis, N. "Efficient estimation of cointegrated vectors and testing for causality in vector autoregressions: A survey of the theoretical literature", *Journal of Economic Surveys*, forthcoming.

Caporale, G. and Pittis, N. "Unit root testing using covariates: Some theory and evidence", *Oxford Bulletin of Economics and Statistics*, forthcoming.

Caporale, W., Hassapis, C. and Pittis, N. “Unit Roots and Long Run Causality: Investigating the Relationship between Output, Money and Interest Rates”. *Economic Modeling*, 15(1), 91-112, January 1998.

Clerides, K., S. "Is Learning-by-Exporting Important? Micro-Dynamic Evidence from Colombia, Morocco, and Mexico." *Quarterly Journal of Economics* 113(3), pp. 903-947, August 1998, (with Lach and J.R. Tybout).

Cukierman, A., Kalaitzidakis, P., Summers, L. and Webb, S. "Central Bank Independence, Growth, Investment, and Real Rates", Reprinted in Sylvester Eijffinger (ed), *Independent Central Banks and Economic Performance,* Edward Elgar, 1997, 416-461.

Eicher, Th. and Kalaitzidakis, P. “The Human Capital Dimension to Foreign Direct Investment: Training, Adverse Selection and Firm Location”. In Bjarne Jensen and Kar-yiu Wong (eds), *Dynamics,Economic Growth, and International Trade,* The University of Michigan Press, 1997, 337-364.

Gatsios, K., Hatzipanayotou, P. and Michael, M. S. "International Migration, the Provision of Public Good and Welfare", *Journal of Development Economics*, 60/2, 561-577, 1999.

Haliassos, M. "On Perfect Foresight Models of a Stochastic World", *Economic Journal*, 104, 477-491, 1994.

Haliassos, M. and Bertaut, C., "Why Do So Few Hold Stocks?", *The Economic Journal,* 105, 1110- 1129, 1995.

Haliassos, M. and Tobin, J. "The Macroeconomics of Government Finance", reprinted in J. Tobin, *Essays in Economics,* Vol. 4, Cambridge: MIT Press, 1996.

Hassapis, C., Pittis, N. and Prodromidis, K. "Unit Roots and Granger Causality in the EMS Interest Rates: The German Dominance Hypothesis Revisited”, *Journal of International Money and Finance,* pp. 47-73, 1999.

Hassapis, C., Kalyvitis, S., and Pittis, N. “Cointegration and Joint Efficiency of International Commodity Markets”, *The Quarterly Review of Economics and Finance,* Vol 39, pp. 213-231, 1999.

Hassapis, C., Pittis, N., and Prodromides, K. “EMS Interest Rates: The German Dominance Hypothesis or Else?” in *European Union at the Crossroads: A Critical Analysis of Monetary Union and Enlargement*, Aldershot, UK., Chapter 3, pp. 32-54, 1998. Edward Elgar Publishing Limited.

Hatzipanayotou, P. and Michael, M. S. "General Equilibrium Effects of Import Constraints Under Variable Labor Supply, Public Goods and Income Taxes", *Economica*, 66, 389-401, 1999.

Hatzipanayotou, P. and Michael, M. S. “Public Good Production, Nontraded Goods and Trade Restriction”, *Southern Economic Journal,* 63, 4, 1100-1107, 1997.

Hatzipanayotou, P. and Michael, M. S. "Real Exchange Rate Effects of Fiscal Expansion Under Trade Restrictions", *Canadian Journal of Economics,* 30-1, 42-56, 1997.

Kalaitzidakis, P. "On-the-job Training Under Firm-Specific Innovations and Worker Heterogeneity", *Industrial Relations,* 36, 371-390, July 1997.

Kalaitzidakis, P., Mamuneas, Th. and Stengos, Th. “European Economics: An Analysis Based on Publications in Core Journals*.” European Economic Review*,1999.

Lyssiotou, P., Pashardes, P. and Stengos, Th. “Testing the Rank of Engel Curves with Endogenous Expenditure”, *Economics Letters,* 64, 1999, 61-65.

Lyssiotou, P., Pashardes, P. and Stengos, Th. "Preference Heterogeneity and the Rank of Demand Systems", *Journal of Business and Economic Statistics*, Vol 17, No 2, April 1999, 248-252.

Lyssiotou, Panayiota, "Comparison of Alternative Tax and Transfer Treatment of Children using Adult Equivalence Scales", *Review of Income and Wealth*, Series 43, No. 1 March 1997, 105-117.

Mamuneas, T.P. (with Demetriades P.). “Intertemporal Output and Employment Effects of Public Infrastructure Capital: Evidence from 12 OECD Economies”, *Economic Journal,* forthcoming.

Mamuneas T.P. (with Kalaitzidakis P. and Stengos T.). “A Nonlinear Sensitivity Analysis of Cross-Country Growth Regressions”, *Canadian Journal of Economics,* forthcoming.

Mamouneas T.P. (with Bougheas S. and Demetriades P.).”I´nfrastructure, Specialization and Economic Growth”, *Canadian Journal of Economics*, forthcoming.

Mamuneas, Theofanis P. “Spillovers from Publicly – Financed R&D Capital in High-Tech Industries”, *International Journal of Industrial Organization,* 17(2), 215-239, 1999.

Mamuneas, T. P. (with Nadiri, M.I.). “R&D Tax Incentives and Manufacturing-Sector R&D Expenditures”, in *Borderline Case: Interntational Tax Policy, Corporate Research and Development, and Investment*, James Poterba (ed.), National Academy Press, Washington D.C., 1997. Reprinted in *Chemtech*, 28(9), 1998.

Michaelides, A. and Ng, S. “Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of three Simulation Estimators”, *Journal of Econometrics,* forthcoming.

Pashardes, Panos. “Equivalence Scales in a Rank-3 Demand System”, *Journal of Public Economics*, 58, 143-158, 1995.

Pashardes, Panos. “Bias in Estimating Equivalence Scales from Grouped Data”, *Journal of Income Distribution*, Special Issue: Symposium on Equivalence Scales, 4, 253-264, 1995.

Pashardes, Panos. “Abstention and Aggregation in Consumer Demand”, *Oxford Economic Papers*, 46, 502-518, 1994 (with V. Fry).

Pashardes, Panos. “Bias in Estimation of the Almost Ideal Demand System with the Stone Index
Approximation”, *Economic Journal*, 103, 908-916, 1993.

Pashardes, Panos. “What Do We Learn About Consumer Demand Patterns From Micro-Data?”, *American Economic Review*, 83, 570-597, 1993 (with R. Blundell and G. Weber).

Pashardes, Panos. “Non-Linearities and Equivalence Scales”, *The Economic Journal*, 103, 359-368, 1993 (with R. Dickens and V. Fry).

Spanos, Aris " Revisiting Date Mining: ´hunting´ with or without a license”, forthcoming *Journal of Methodology,* July 2000.

Spanos, Aris. "On Modeling Heteroscedasticity: The Student's *t* and Elliptical Linear Regression Models", *Econometric Theory*, 10, 286-315, 1994.

Spanos, Aris. "On Normality and the Linear Regression Model", *Econometric Reviews,* 14, 195-203, 1995.

Spanos, Aris. "On Theory Testing in Econometrics: Modeling with nonexperimental Data", *Journal of Econometrics,* 67, 189-226, 1995.