BANK OF GREECE

EUROSYSTEM

Working Paper

Modelling household expenditure on health care in Greece

Manos Matsaganis
Theodore Mitrakos
Panos Tsakloglou

68

MARCH 2008

BANK OF GREECE
Economic Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Printed in Athens, Greece
at the Bank of Greece Printing Works.

ISSN 1109-6691

MODELLING HOUSEHOLD EXPENDITURE ON HEALTH CARE IN GREECE

Manos Matsaganis
Athens University of Economics and Business

Theodore Mitrakos
Bank of Greece

Panos Tsakloglou
Athens University of Economics and Business

ABSTRACT

Health expenditure data are known to be afflicted by restricted range, zero values, skewness and kurtosis. Several methods for modelling such data have been suggested in the literature to cope with these problems. This paper compares the performance of several alternative estimators, including two-part models and generalized linear models. The dependent variable is household, not individual, expenditure on health care in Greece, a country where out-of-pocket health expenditure is higher than anywhere else in the European Union, whether as a proportion of GDP, as a share of all health spending, or in per capita terms. To facilitate comparison of model performance, household health expenditure is examined in two different specifications: expenditure on all health care (where zero values are rare) and expenditure on hospital services alone (where zero values are common). Three of the estimators performed almost equally well in terms of mean square error and mean absolute prediction error: a modified two-part model with non-linear least squares in the second part, a constant-variance generalized linear model, and a variance-proportional-to-mean generalized linear model. The findings suggest that no estimator is best under all circumstances, while most alternative estimators produce similar results. The paper concludes by discussing implications for further research.

Keywords: Two-part models; generalized linear models; household expenditure; Greece
JEL codes: I10; C52

Acknowledgements: The views expressed are those of the authors and should not be interpreted as those of their respective institutions.

Correspondence:
Manos Matsaganis
Department of International and European Economics
Athens University of Economics and Business
Patission 76, Athens 10434, Greece
Tel.: +30-210-8203380,
fax: +30-210-8214122
e-mail:matsaganis@aueb.gr

1. Introduction

Health expenditure data are typically characterised by restricted range (i.e. no negative values), a number of zero values (that can be large, depending on the type of expenditure considered), as well as skewness (asymmetry) and kurtosis (i.e. heavy right tail). In the presence of these properties, estimation via ordinary least squares is biased and inefficient [Jones 2000].

Alternatives to ordinary least squares include two-part models [Duan et al. 1983; Mullahy 1998; Manning 1998; Manning and Mullahy 2001; Deb and Trivedi 2002; Buntin and Zaslavsky 2004], and generalized linear models [McCullagh and Nelder 1989; Blough et al. 1999; Basu et al. 2004; Buntin and Zaslavsky 2004; Manning et al. 2005; Cantoni and Ronchetti 2006]. Two-part models estimate first the probability of non-zero expenditure, and then its level conditional on non-zero expenditure. The dependent variable is commonly log-transformed before the second-part estimation to accommodate skewness. On the other hand, generalized linear models directly model both the mean and variance functions on the original scale of the dependent variable. These two broad modelling approaches, with their various specifications, can produce a considerable number of alternative estimators. In this paper, we compare six estimators of household expenditure on health care in Greece. As the literature emphasizes that no estimator is "best" under all circumstances, to allow for the possibility that model performance is affected by data characteristics we compare two different dependent variables: expenditure on all health care and expenditure on hospital services. Our objective is to apply recent methodological insights to a different type of dataset than is commonly used, that is household rather than individual expenditure.

The structure of the paper is as follows. The next section briefly describes the data. Section three explains the general methodology and presents the various models. Section four shows the results and compares alternative estimators. The final section concludes.

2. Data

The dataset used in this paper is drawn from the latest Household Budget Survey, conducted by the National Statistical Service of Greece over a 12-month period (February 2004 to January 2005). The survey sample consists of 6,555 households with 17,913 members.

In the survey, expenditure on health care is recorded on a household not an individual basis. Our measure of expenditure on all health care includes spending on hospital services, physician fees, prescription costs, diagnostic tests and therapeutic devices, but excludes dental services.

The 2004-05 Household Budget Survey found that average household expenditure on health care, defined as above, was €1,070 annually, equivalent to 6.3% of all household expenditure. Household expenditure on hospital care in particular was on average €228 a year (1.0% of all household expenditure). These sample means are broadly consistent with information from other sources [Mossialos et al. 2005], and confirm that out-of-pocket payments on health in Greece are very substantial (the highest, in fact, in the European Union).

Table 1 indicates that household expenditure on health varies significantly with private medical insurance status, as privately insured households spend almost twice as much as not privately insured ones. Moreover, while health expenditure does rise with household income, its budget share declines – in view of the fact that high spending on health is common among low-income elderly households. As a matter of fact, with respect to household type, elderly households are shown to register high expenditure on health (budget shares of about 12% to 13%), and the same holds for families with newborn babies and very young children. Finally, as regards sickness fund affiliation, differences in health expenditure (especially in terms of budget shares) are rather pronounced, and probably reflect differences in membership composition as well as differences in health benefits between different funds. The case of the farmers' fund seems to be emblematic of that: a high concentration of elderly members combines with a restricted range of benefits to allow ample scope for high household expenditure on health.

As explained earlier, the paper focuses on two different dependent variables: expenditure on all health care and expenditure on hospital services. The distribution of these two variables differs considerably. As Table 2 shows, only 10.7% of all individuals in the survey have zero expenditure on any health care at all, while the proportion of individuals with zero expenditure on hospital services is as high as

83.2%. Similarly, whereas the top 1% of cases account for 14.0% of all health expenditure, in terms of expenditure on hospital care the top 1% of cases (ranked accordingly) account for 37.6% of the total. While average per capita expenditure is higher for all health care compared to hospital services alone (€392 vs. €83 respectively *per annum*), the opposite is true if zero values are excluded (€438 vs. €495 a year).

The distribution of household expenditure on all health care and on hospital services by centile is presented graphically in Figure 1.

On the whole, our two dependent variables seem different enough to allow us to test the hypothesis that model performance depends on data characteristics. We now turn to the models themselves and the general methodology of the paper.

3. Methodology

Two broad approaches are compared here, two-part models vs. generalized linear models. These are briefly described below.

3.1 Two-part models

As a way to deal with the issue of zero values, two-part models are often used as an alternative to ordinary least squares. The first part predicts the probability of any expenditure and can be specified as probit or logit.

In the latter case:

$$\Pr ob(y_i > 0) = \frac{e^{xa}}{1+e^{xa}} = \frac{\exp(x\alpha)}{1+\exp(x\alpha)} \tag{1}$$

The second part of the model predicts the level of expenditure, conditional on its taking a non-zero value. Several specifications of this second part are possible, and some are discussed here. Estimates of predicted expenditure can be obtained by multiplying probabilities from the first part of the model by expected levels from the second part:

$$E(y_i \mid x_i) = \Pr ob(y_i > 0 \mid x_i) \times E(y_i \mid x_i, y_i > 0) \tag{2}$$

A common option for the second part of a two-part model is to use an OLS model with a log-transformed dependent variable [Duan et al. 1983; Mullahy 1998; Manning 1998]. The problem with that option is that predictions must then be retransformed back to the original scale (in our case, euros) to draw meaningful conclusions. If the log-scale error term is normally distributed,

$$E(y \mid y > 0; x_i) = \exp(x\beta + \frac{1}{2}\sigma^2) \tag{3}$$

If the error term is not normally distributed, then Duan's non-parametric smearing factor [Duan 1983], developed in the context of the RAND Health Insurance Experiment [Manning et al. 1987], can provide a consistent estimate of its expected value – provided the errors are not heteroskedastic. The smearing factor is the average of the exponential of the residuals from the OLS regression on the log-transformed dependent variable.

$$\phi = \frac{1}{n}\sum_{i=1}^{n}\exp(\varepsilon_i)\text{, where } \varepsilon_i = \log y_i - x_i\hat{\beta} \tag{4}$$

The exponential of the predicted values are then multiplied by the smearing factor to obtain expected values on the original scale.

$$E(y \mid x) = \Pr ob(y > 0 \mid x) \times \exp(x\hat{\beta}) \times \phi \tag{5}$$

However, if the error term is heteroskedastic, use of a single smearing factor is likely to bias the predictions. In the presence of heteroskedasticity, the solution to the retransformation problem would be either to employ multiple, group-specific smearing factors [Mullahy 1998; Manning 1998; Manning and Mullahy 2001], or to model the heteroskedasticity explicitly [Manning et al. 1987; Mullahy 1998; Manning 1998; Manning and Mullahy 2001].

Since multiple smearing factors by sub-group can be difficult if the error term is heteroskedastic in relation to continuous or multiple covariates [Manning et al. 1987], and since computation of the heteroskedastic retransformation can be cumbersome [Manning et al. 2005; Basu et al. 2004], another alternative would be to forego retransformation altogether in favour of a modified two-part model, in which the second part is a non-linear least squares model [Mullahy 1998], equivalent to a constant variance, generalized linear model (see below). In this case,

$$E(y \mid y > 0, x) = \exp(x\beta) \qquad (6)$$

and, substituting from (1) and (6),

$$E(y \mid x) = \Pr ob(y > 0 \mid x) \times E(y \mid y > 0, x) = \frac{\exp(x\alpha) \times \exp(x\beta)}{1 + \exp(x\alpha)} = \frac{\exp(x(\alpha + \beta))}{1 + \exp(x\alpha)} \qquad (7)$$

where *exp*($x\alpha$) is estimated *via* logit, and *exp*($x\beta$) is estimated *via* non-linear least squares. One advantage of this approach is that the elasticities have a very simple form [Mullahy 1998]:

$$\eta_j(x) = (1 - \pi(x))\alpha_j x_j + \beta_j x_j \qquad (8)$$

where $(1-\pi(x))\alpha_j x_j$ corresponds to the logit probability elasticity and βx_j the exponential conditional means elasticity from non-linear least squares.

3.2 Generalized linear models

While a generalized linear model can be used in the second part of a modified two-part model, as above (as zero expenditures in the dataset pose no problem for fitting such a model), it can also be estimated on the entire sample. Generalized linear models accommodate skewness and related issues *via* variance-weighting rather than through transformation and retransformation. More specifically, such models explicitly specify a distribution that reflects the relationship between the mean and the variance, and a link function between the linear part $x\beta$ and the mean μ = E(y|x) on the original scale of the dependent variable. The mean function is represented as:

$$\mathrm{E}(y|x) = \mu(x\beta) \qquad (9)$$

If the link function is the log, as is typically the case in health expenditure applications, then μ is the exponential function.

With respect to the variance function, a commonly used family includes the power functions of the form:

$$v(x) = \kappa(\mu(x\beta))^{\lambda} \qquad (10)$$

If λ = 0, the variance is constant; if λ = 1, the variance is proportional to the mean; and, if λ = 2, the variance is proportional to the mean squared (or the standard deviation is proportional to the mean). Although the residuals need not take these distributional forms (in fact, there is no reason to assume that λ should be an integer at all), in this paper we estimate a family of generalized linear models for $\lambda = 0$, 1 and 2. In the case of all health expenditure, where non-zero values account for almost 90% of all observations, the one-part generalized linear model for $\lambda = 0$ is expected to yield rather similar estimates to the two-part model in which the second part is a non-linear least squares model.

Choice of λ can also be made on a priori grounds with the help of the Park test [Manning and Mullahy 2001; Buntin and Zaslavsky 2004]. The Park test directly estimates the relationship between the mean and the variance by regressing the log-transformed squared residuals from a provisional model (GLM or log-transformed OLS) on the log-transformed predictions ($\hat{y}$) from the same model:

$$ln((yi - \hat{y}i)2) = \lambda 0 + \lambda 1ln(\hat{y}i) + vi \quad (11)$$

The coefficient λ_1 corresponds to λ in (9), indicating which GLM variance function is most appropriate.

3.3 The models under comparison

In view of the preceding analysis, the paper uses six alternative estimators to model household expenditure on health care in Greece:

- 2 two-part models, denoted as 2PM (logit plus OLS with log-transformed y and Duan's non-parametric smearing factor) and M2PM (logit plus NLS) respectively, and
- 3 GLM models, denoted as GLMλ (for $\lambda = 0, 1, 2$).
- 1 standard, one-part, ordinary least squares model, denoted as OLS,

Despite its well-known failings, the latter model is estimated without a transformation of the dependent variable (i.e. on the original scale of y) and is presented alongside the other models for comparison.

To facilitate comparisons between estimators, all models use a common set of regressors. These include household characteristics such as demographic structure, (logarithm of) equivalent income, private medical insurance status, sickness fund affiliation, education and location, as well as a number of interaction terms.

Model selection was based on the procedure described in the following section.

4. Results

A histogram of the log-transformed non-zero data seemed to approach near symmetry, much more than was the case with the non-transformed data. This suggests that 2PM (the two-part log-transformed OLS model) could be a good estimator. Our estimate of Duan's smearing factor was 1.87 for all health expenditure and 1.93 for hospital expenditure only, falling within the expected range of 1.5 and 4.0 [Duan 1983].

Nonetheless, as explained in the previous section, the standard 2PM will yield biased estimates unless the homoskedasticity hypothesis holds (i.e. the variance of the errors is unrelated to the predictions). Testing for heteroskedasticity rejected the null hypothesis, and at a very high level of statistical significance (0.0001 and 0.0020 for all health expenditure and hospital expenditure respectively). Repeating the test after substituting standardized or studentized residuals, as has been recently suggested [Manning et al. 2005], did not substantially alter the basic finding that homoskedasticity is rejected. In view of that, results for the 2PM are not shown here (though are available on request).

Turning to GLM, testing for kurtosis produced scores of 3.08 for all health expenditure and 2.86 for hospital expenditure, both log transformed. The fact that these scores are both reasonably close to 3, the value for the normal distribution, suggests that the common practice of assuming a log link function between the linear part $x\beta$ and the mean $\mu = E(y|x)$ can be justified.

To determine the variance function *a priori* we performed the Park test by actually estimating both provisional models described in the previous section. In the case of all health expenditure, the log-transformed OLS provisional model produced an estimate of $\lambda = 1.67$, while the provisional GLM model estimated $\lambda = 1.52$. In the case of hospital expenditure, the estimates for λ were 1.89 and 1.73 respectively. A modified Park test, using Gamma regression as has also been recently suggested [Manning et al. 2005], gave very similar λ values. These results seem to suggest that the constant variance model ($\lambda = 0$) would not be a good candidate, and that the best model fall somewhere between the variance proportional to the mean (λ = 1) and the variance proportional to the mean squared (λ = 2) models.

In order formally to evaluate model performance we computed the mean square error, the mean prediction error and the mean absolute prediction error for the remaining five models and for both dependent variables. The results are shown in Table 3.

In terms of mean square error, GLM_0 (the constant variance model) seemed to perform best, closely followed by M2PM. That was also the case with respect to the mean absolute prediction error for hospital expenditure – though not for all health expenditure, where it was GLM_1 (i.e. the variance proportional to mean model) that seemed to perform best, closely followed in its turn by M2PM and GLM_0. In terms of mean prediction error, a weaker criterion of fit, GLM_1 appeared again to do better than M2PM and GLM_0. Overall, in stark contrast to what might have been expected on the basis of the Park test, GLM_2 (the variance proportional to mean squared model) was clearly outperformed by all other models shown here.

On the whole, the different models estimated here seem hard to distinguish. This is particularly evident in Figure 2, where ratios of actual to predicted expenditure on all health care by centile of actual expenditure are plotted for the alternative estimators. As is usually the case with exercises of this kind, all models over-predict actual expenditure where the latter is zero (centiles 1-11) or low (centiles 12-72), and under-predict it where it is high (centiles 73-100) – and especially where actual expenditure is very high (centiles 92-100), in which case the ratio of actual to predicted expenditure rises above 2.0 to reach 6.0 (or 6.7 in the case of the OLS model).

Figure 3 presents ratios of actual to predicted expenditure on hospital care by centile of actual expenditure. In this case too, although the predictions of alternative estimators differ more clearly, the same pattern of the models over-predicting actual expenditure where it is zero or low, and under-predicting it where it is high, is clearly manifest. In the case of hospital care the ratio of actual to predicted expenditure reaches 5.0 by the 99th centile, and rises sharply higher still in the 100th (top) centile of actual expenditure.

If the objective of the paper had been to draw inferences and test hypotheses about the effects of household characteristics on health expenditure, the estimated coefficients from the various models would also have to be compared. While this is straightforward for generalized linear models, direct comparisons are not possible for different models (such as the M2PM estimated here).

As Table 4 indicates, the coefficient estimates are indeed fairly similar across the three generalized linear models. As might have been expected on *a priori* grounds, the coefficients of the income, private insurance and demographic variables (for the age groups 0-6 and 55+) are positive and mostly significant, especially for all health expenditure. In contrast, the coefficients had opposite signs or were less significant in the case of the sickness fund, education and location variables. More research is needed to establish if an adjustment of the link or the variance function might improve fit.

5. Conclusion

The objective of the paper was to draw on recent methodological insights to model household – rather than individual – expenditure on health care in Greece. Six alternative estimators were tested, including two-part models and generalized linear models. To reflect the fact that no estimator is "best" under all circumstances, and to allow for the possibility that model performance can be affected by data characteristics, two different dependent variables were employed: expenditure on all health care (where only 11% of cases in the sample had zero expenditure) and expenditure on hospital services alone (where only 17% of cases had *non*-zero expenditure). Tests for heteroskedasticity, skewness and kurtosis, as well as for the variance function of the generalized linear models (Park test) were also carried out.

Our results appear to support the finding that most alternative estimators often produce very similar results in practice [Buntin and Zaslavsky 2004]. More specifically, three of our estimators, namely M2PM (a modified two-part model in which the second part is a non-linear least squares model), GLM_0 and GLM_1 (the generalized linear models for $\lambda = 0$ and $\lambda = 1$ respectively), seemed on the whole to perform equally well in terms of our main criteria of fit, mean square error and mean absolute prediction error. With expenditure on all health care as the dependent variable, GLM_1 did slightly better in terms of mean absolute prediction error, while M2PM and GLM_0 did (again, slightly) better in terms of mean square error. In proportional terms, the score differences between the three estimators were below 0.3% in the case of mean absolute prediction error and less than 2.1% in the case of mean square error. In contrast, the performance of the generalized linear model *a priori* favoured by the Park test, GLM_2 (the variance proportional to mean squared model, for $\lambda = 2$) appeared to be distinctly inferior to that of all other models.

Several extensions to this work are possible. Cross-validation, fitting the models to one part of the sample to assess predictive accuracy on the remaining part, is a useful antidote to over-fitting. An interest in response to covariates would require a stronger focus on the interpretation of regression coefficients, and might lead to a reassessment of the relative performance of different models. Furthermore, alternative estimators can also be compared in terms of their accuracy in predicting average expected expenditure of meaningful and policy relevant sub-samples (e.g. population groups differentiated by sickness fund affiliation). Such issues lie well beyond the scope of this paper, but we do intend to tackle many, if not all, of these in future research.

References

Basu A., Manning W. G. and J. Mullahy (2004). Comparing alternative models: log vs Cox proportional hazard? Health Economics 13: 749–765.

Blough D. K., Madden C. W. and M. C. Hornbrook (1999). Modeling risk using generalized linear models. Journal of Health Economics 18: 153–171.

Buntin M. B. and A. M. Zaslavsky (2004). Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures, Journal of Health Economics 23: 525–542.

Cantoni E. and E. A. Ronchetti (2006). Robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. Journal of Health Economics 25: 198–213.

Deb P. and P. K. Trivedi (2002). The structure of demand for health care: latent class versus two-part models. Journal of Health Economics 21: 601–625

Duan N. (1983). Smearing estimate: a nonparametric retransformation method. Journal of American Statistical Association 78: 605–610.

Duan N., Manning W. G., Morris C. N. and J. P. Newhouse (1983). A comparison of alternative models of the demand for medical care. Journal of Business Economics and Statistics 1: 115-126.

Jones A. M. (2000). Health Econometrics, In Handbook of Health Economics, Culyer A. J. and J. P. Newhouse (eds), North-Holland: Amsterdam.

Manning W. G., Basu A. and J. Mullahy (2005). Generalized modelling approaches to risk adjustment of skewed outcomes data. Journal of Health Economics 24: 465–488.

Manning W. G. and J. Mullahy (2001). Estimating log models: to transform or not to transform? Journal of Health Economics 20: 461–494.

Manning W. G., Newhouse J. P., Duan N., Keeler E. B., Leibowitz A. and M. S. Marquis (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. American Economic Review 77: 251-277.

Manning W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics 17: 283–295.

McCullagh P. and J. A. Nelder (1989). Generalized linear models. (2nd edn), Chapman and Hall: London.

Mossialos E., Allin S. and K. Davaki (2005). Analysing the Greek health system: A tale of fragmentation and inertia. Health Economics 14: S151–S168.

Mullahy J. (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. Journal of Health Economics 17: 247–281.

Table 1: Household characteristics and health expenditure

	Population share		All health expenditure		Expenditure on hospital care	
	% of	% of	mean (€ p.a.)	budget share	mean (€ p.a.)	budget share
household type						
couple with 1 child aged 0-1	1.3	1.4	1,268	12.7	722	7.3
couple with 1 child aged 2-6	2.6	2.8	365	4.0	60	0.7
couple with 1 child aged 7-13	2.1	2.3	301	3.2	72	0.8
couple with 2 children aged 0-1, 2-6	1.2	1.8	891	12.0	457	6.2
couple with 2 children both aged 2-6	1.1	1.6	271	3.7	31	0.4
couple with 2 children both aged 7-13	2.1	3.1	159	2.2	18	0.3
single adult aged under 65	9.3	3.4	624	4.3	125	0.9
couple both aged under 65	7.9	5.8	515	4.6	74	0.7
single adult aged 65+	11.6	4.2	884	11.4	104	1.3
elderly couple both aged 65+	10.6	7.8	669	11.3	126	2.1
other household types	50.2	65.8	315	4.1	64	0.8
household equivalent income						
low	25.0	24.5	400	7.7	55	1.1
middle	50.0	51.1	438	5.8	87	1.1
high	25.0	24.4	641	4.7	145	1.1
private medical insurance status						
privately insured	7.1	8.3	751	5.9	235	1.8
not privately insured	92.9	91.7	459	5.6	83	1.0
sickness fund affiliation						
employee	34.0	31.3	475	5.8	84	1.0
farmer	13.6	11.2	498	9.3	57	1.1
government	8.7	7.6	522	4.8	123	1.1
self-employed	6.2	6.2	578	5.9	131	1.3
other	37.5	43.6	609	5.4	145	1.3
all	100.0	100.0	479	5.6	93	1.1

Note: Health expenditure figures are household member averages (i.e. per capita).

Table 2: Distribution of dependent variables

	y_1 = all health expenditure		y_2 = expenditure on hospital care	
	% of all expenditure	mean (€ p.a.)	mean (€ p.a.)	mean (€ p.a.)
top 0.1% of cases	4.0%	19,420	13.9%	13,499
top 1% of cases	14.8%	7,096	43.5%	4,032
top 10% of cases	49.8%	2,386	95.4%	892
top 25% of cases	73.9%	1,417	100.0%	374
top 50% of cases	92.2%	884	100.0%	187
all cases	100.0%	479	100.0%	93
cases with zero expenditure (% of all cases)	11.1%		84.8%	
mean expenditure of non-zero cases (€ p.a.)	539		614	

Notes: Cases were ranked separately by the value of each dependent variable.

Health expenditure figures are household member averages (i.e. per capita).

Table 3: Comparison of model performance

y₁ = all health expenditure	Mean square error (×10^6)	Mean absolute prediction error	R^2
log OLS	3.53	818.00	0.13
M2PM	3.12	916.17	0.16
GLM_0	3.12	916.63	0.16
GLM_1	3.19	913.93	0.14
GLM_2	3.34	923.87	0.10
y₂ = expenditure on hospital care	Mean square error (×10^6)	Mean absolute prediction error	R^2
log OLS	1.48	749.91	0.15
M2PM	1.08	303.75	0.16
GLM_0	1.02	285.49	0.20
GLM_1	1.13	332.79	0.12
GLM_2	5.44	437.23	0.03

Table 4a: Coefficient estimates

y_1 = all health expenditure														
	OLS of y on x's		Modified two-part model				One-part generalized linear models							
Explanatory variable	β_i's from OLS		β_i's from logit		β_i's from NLS		β_i's from GLM_0		β_i's from GLM_1		β_i's from GLM_2			
Demographics														
no. of female members aged 0-1	2,521.90	**	3.19	**	1.02	**	1.11	**	1.23	**	1.40	**		
no. of female members aged 2-6	140.58		0.23		0.07		0.08		0.15		0.18	*		
no. of female members aged 7-13	-127.77		0.09		-0.49	**	-0.41		-0.10		-0.04	*		
no. of female members aged 14-18	24.48		0.29	*	-0.07		-0.01		0.00		0.05	*		
no. of female members aged 19-24	-39.00		0.30	*	-0.29	**	-0.22		-0.03		0.02			
no. of female members aged 25-34	160.45	*	0.37	**	0.22	**	0.24		0.15	*	0.19	**		
no. of female members aged 35-44	176.84	*	0.18		0.12		0.10		0.14		0.15	*		
no. of female members aged 45-54	207.15	**	0.28	*	0.15	*	0.10		0.17		0.29	**		
no. of female members aged 55-64	221.67	**	0.56	**	0.10		0.09		0.19	*	0.29	**		
no. of female members aged 65-74	402.44	**	0.85	**	0.35	**	0.40		0.36	**	0.42	**		
no. of female members aged 75-84	573.85	**	1.34	**	0.47	**	0.50		0.50	**	0.55	**		
no. of female members aged 85+	360.96	**	1.20	**	0.29	*	0.37		0.31	**	0.41	**		
no. of male members aged 0-1	2,277.21	**	3.55	**	0.90	**	0.96	**	1.17	**	1.35	**		
no. of male members aged 2-6	181.54	*	0.35	*	0.22	**	0.26		0.16	*	0.19	**		
no. of male members aged 7-13	62.45		0.18		0.08		0.04		0.06		0.07	*		
no. of male members aged 14-18	28.50		0.19		-0.07		-0.11		0.02		0.09	*		
no. of male members aged 19-24	-88.98		0.13		-0.07		0.00		-0.10		-0.09	*		
no. of male members aged 25-34	27.73		-0.25	*	0.00		-0.04		0.03		0.01			
no. of male members aged 35-44	16.89		-0.20		0.03		0.03		0.01		0.01			
no. of male members aged 45-54	116.51		-0.12		0.17	*	0.13		0.11		0.04			
no. of male members aged 55-64	74.52		-0.08		-0.12		-0.22		0.06		0.12	*		
no. of male members aged 65-74	243.36	**	0.18		0.06		-0.03		0.23	**	0.32	**		
no. of male members aged 75-84	436.41	**	0.39		0.31	**	0.30		0.38	**	0.43	**		
no. of male members aged 85+	309.35		1.23	*	0.18		0.18		0.28	**	0.35	**		

Continue in the next page

y_1 = all health expenditure												
	OLS of y on x's		Modified two-part model				One-part generalized linear models					
Explanatory variable	β_i's from OLS		β_i's from logit		β_i's from NLS		β_i's from GLM_0		β_i's from GLM_1		β_i's from GLM_2	
household income												
log equivalent income	439.11	**	0.16		0.57	**	0.62		0.40	**	0.38	**
private medical insurance status												
privately insured	850.98	**	1.18	**	0.54	**	0.61		0.54	**	0.51	**
sickness fund affiliation												
farmer	-20.51		-0.50		-0.07		-0.25		-0.24		-0.14	
government	-162.49		-0.11		-0.19		-0.21		-0.15		-0.16	
self-employed	423.39		0.08		0.29		0.31		0.35		0.36	*
other	-136.40		-0.30		-0.26		-0.30		-0.15		-0.06	
uninsured	215.41		-0.84		0.51		0.53		0.19		0.02	
employee + farmer	-201.46		0.12		-0.27		-0.17		-0.17		-0.24	*
employee + government	-102.50		-0.65		0.00		-0.08		-0.09		-0.17	*
employee + self-employed	-152.60		-0.57		0.16		0.21		-0.14		-0.30	*
employee + other	-533.71		-0.98		-0.43		-0.39		-0.77	*	-1.00	**
Education												
no. of schooling years 12+	235.08		-0.40		0.17		0.05		0.24		0.14	*
no. of schooling years 10-12	358.82	**	-0.11		0.53	**	0.54		0.39	**	0.28	**
no. of schooling years 6-8	249.38	*	-0.10		0.26		0.29		0.26	**	0.23	*
no. of schooling years 0-5	344.36	*	0.25		0.30	*	0.36		0.37	**	0.35	**
currently in education	173.57		-0.43		-0.21		-0.37		-0.49		-0.56	*
Location												
urban	73.29		0.10		-0.14	*	-0.15		0.07		0.08	*
rural	53.54		-0.07		0.07		0.01		0.05		0.07	*
constant	-2,866.06	**	0.32		2.46	**	2.04		3.31	**	3.37	**

Notes: Coefficient estimates for interaction terms suppressed. (**) and (*) indicate statistical significance at the 0.01 and 0.05 level respectively.

Table 4b: Coefficient estimates

y*$_2$ = *expenditure on hospital care												
	OLS of *y* on *x's*		Modified two-part model				One-part generalized linear models					
Explanatory variable	β_i's from OLS		β_i's from logit		β_i's from NLS		β_i's from GLM_0		β_i's from GLM_1		β_i's from GLM_2	
Demographics												
no. of female members aged 0-1	1,585.55	**	3.97	**	0.56	**	0.05		2.11	**	1.98	**
no. of female members aged 2-6	-41.86		0.11		0.02		-0.44	*	0.15		-0.02	
no. of female members aged 7-13	-43.56		0.24	*	-0.14		-1.34	**	-0.02		-0.12	
no. of female members aged 14-18	-47.23		0.15		0.03		-1.60	*	-1.22		-0.37	*
no. of female members aged 19-24	-45.79		-0.25		-0.04		-0.35		-0.73		-0.23	
no. of female members aged 25-34	47.72		0.07		0.41	**	0.51	*	-0.34		0.28	
no. of female members aged 35-44	51.58		-0.08		0.43	**	0.83	**	-0.72		0.19	
no. of female members aged 45-54	1.43		0.22		0.00		-0.61	*	-2.07		-0.06	
no. of female members aged 55-64	47.62		0.34	*	0.31	*	-0.05		-0.73		0.27	
no. of female members aged 65-74	81.91		0.54	**	0.26		0.29		0.15		0.39	
no. of female members aged 75-84	100.56	*	0.79	**	0.12		0.04		-0.36		0.50	*
no. of female members aged 85+	-17.27		0.18		-0.19		-0.75		-0.76		-0.32	
no. of male members aged 0-1	1,482.62	**	4.24	**	0.48	**	-0.45	**	1.94		1.96	**
no. of male members aged 2-6	-31.62		-0.04		-0.14		0.19		-0.66		-0.01	
no. of male members aged 7-13	-2.15		0.11		-0.07		-0.82	**	0.22		0.00	
no. of male members aged 14-18	-2.09		0.19		-0.22		0.30		-0.87		-0.03	
no. of male members aged 19-24	-49.80		-0.03		-0.13		-2.87	*	-0.66		-0.40	
no. of male members aged 25-34	-20.40		-0.08		0.16		0.22		0.01		0.00	
no. of male members aged 35-44	34.01		0.12		0.22		0.14		-0.55		0.12	
no. of male members aged 45-54	50.04		0.11		-0.05		0.60	**	0.61		0.29	
no. of male members aged 55-64	1.87		0.20		-0.18		-0.13		-0.77		0.00	
no. of male members aged 65-74	28.58		0.25		0.05		-0.13		-0.91		0.16	
no. of male members aged 75-84	109.26	*	0.35	*	0.29		0.99	**	0.44		0.49	*
no. of male members aged 85+	-39.26		0.11		-0.01		-0.50		-1.30		-0.40	

Continue in the next page

y_2 = *expenditure on hospital care*												
	OLS of *y* on *x's*		Modified two-part model				One-part generalized linear models					
Explanatory variable	β_i's from OLS		β_i's from logit		β_i's from NLS		β_i's from GLM_0		β_i's from GLM_1		β_i's from GLM_2	
household income												
log equivalent income	97.93	**	0.06		0.23	*	1.37	**	0.74		0.37	**
private medical insurance status												
privately insured	249.35	**	0.57	**	0.38	*	0.29		0.22		0.62	**
sickness fund affiliation												
farmer	-28.29		-0.02		-0.17		-0.13		-0.67		-0.34	
government	32.27		0.38	*	0.01		-0.48		0.05		0.21	
self-employed	37.72		0.02		0.03		-0.26		0.19		0.21	
other	-64.85		0.10		-0.17		-1.56		-0.82		-0.48	
uninsured	-0.96		-0.27		0.22		-1.30		-0.98		-0.49	
employee + farmer	2.87		0.29		-0.13		-1.31	*	-0.25		-0.02	
employee + government	-7.20		-0.40		0.12		-0.45		0.56		0.02	
employee + self-employed	40.58		0.49	*	-0.34		-0.23		0.90		0.21	
employee + other	72.31		0.15		-0.04		-0.60		-0.86		0.18	
Education												
no. of schooling years 12+	125.84	*	0.26		0.44	**	-0.28		1.65		0.47	
no. of schooling years 10-12	69.09		0.08		0.29	*	-0.62	**	1.88		0.35	
no. of schooling years 6-8	89.11		0.26		0.23		-0.50	*	1.60		0.35	
no. of schooling years 0-5	90.60		0.32		0.17		-0.38		1.26		0.34	
currently in education	122.05		-1.80	*	2.15	*	2.83	**	1.12		-0.20	
Location												
urban	34.27		0.12		0.14		-0.55	**	0.22		0.10	
rural	40.25		-0.03		0.05		-1.13	**	-0.25		0.20	
constant	-706.67	**	-3.28	**	4.16	**	-1.98		-1.49		1.67	

<u>Notes</u>: Coefficient estimates for interaction terms suppressed. (**) and (*) indicate statistical significance at the 0.01 and 0.05 level respectively.

Figure 1

Distribution of household expenditure on health by centile

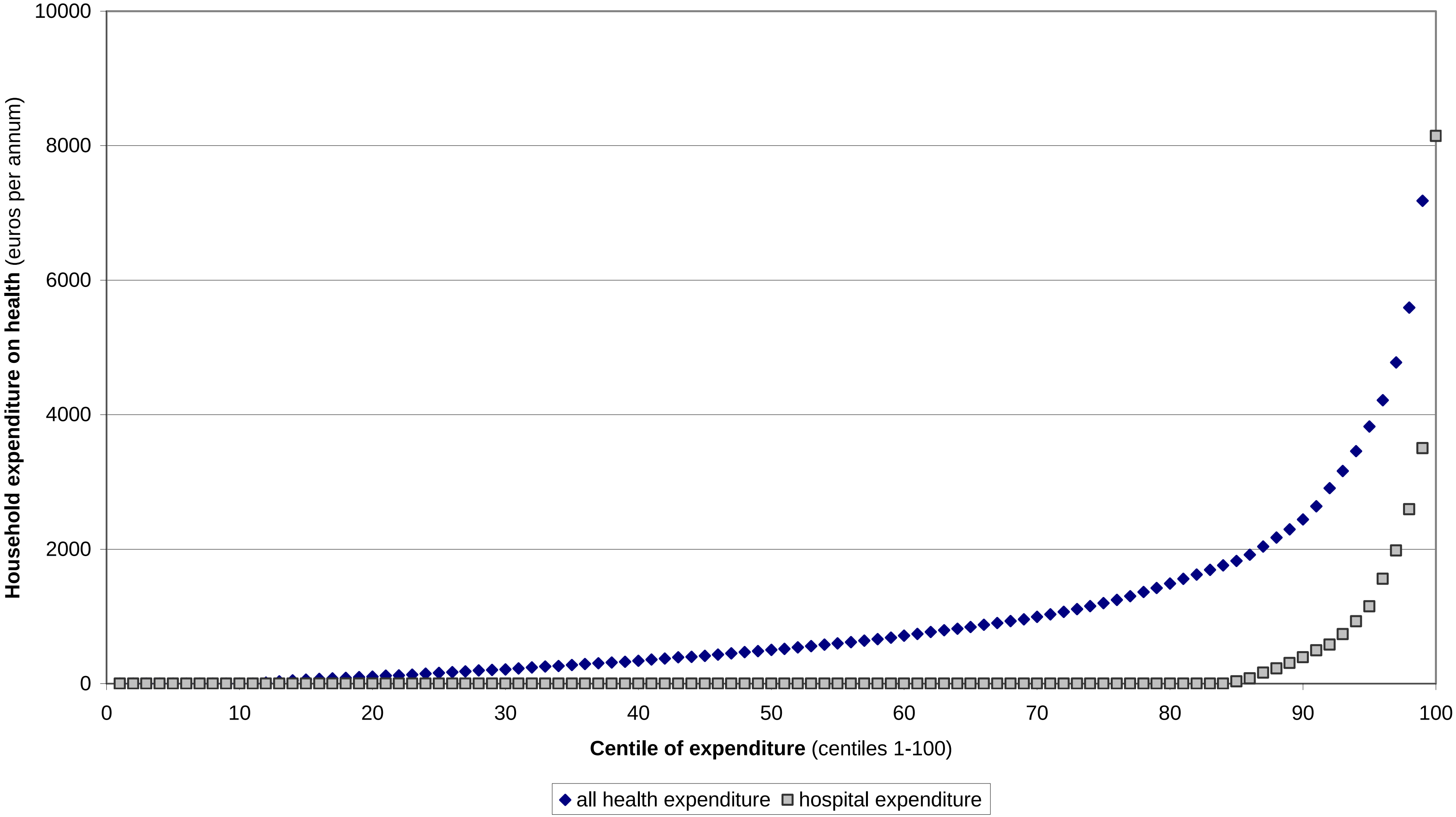


Figure 2

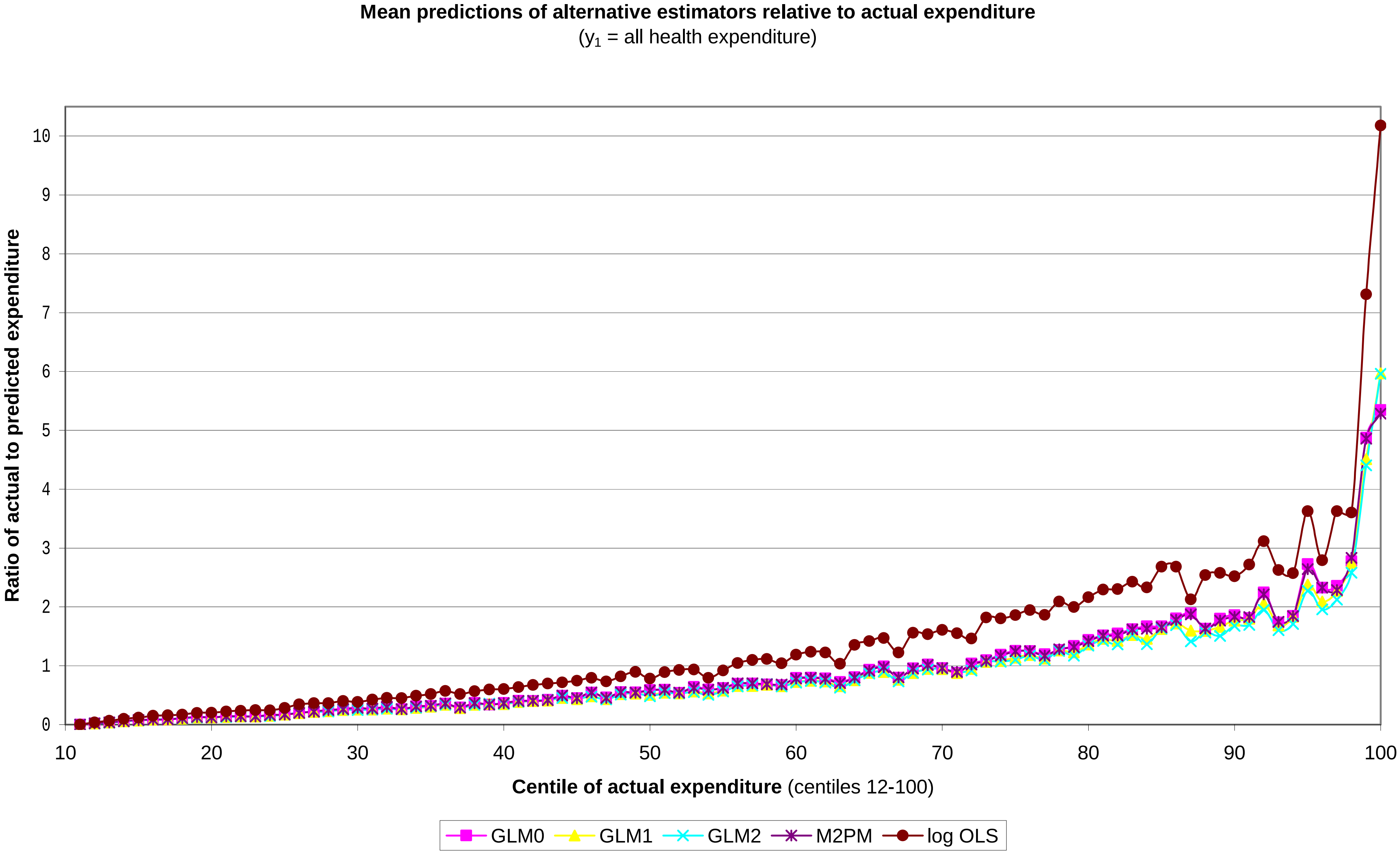
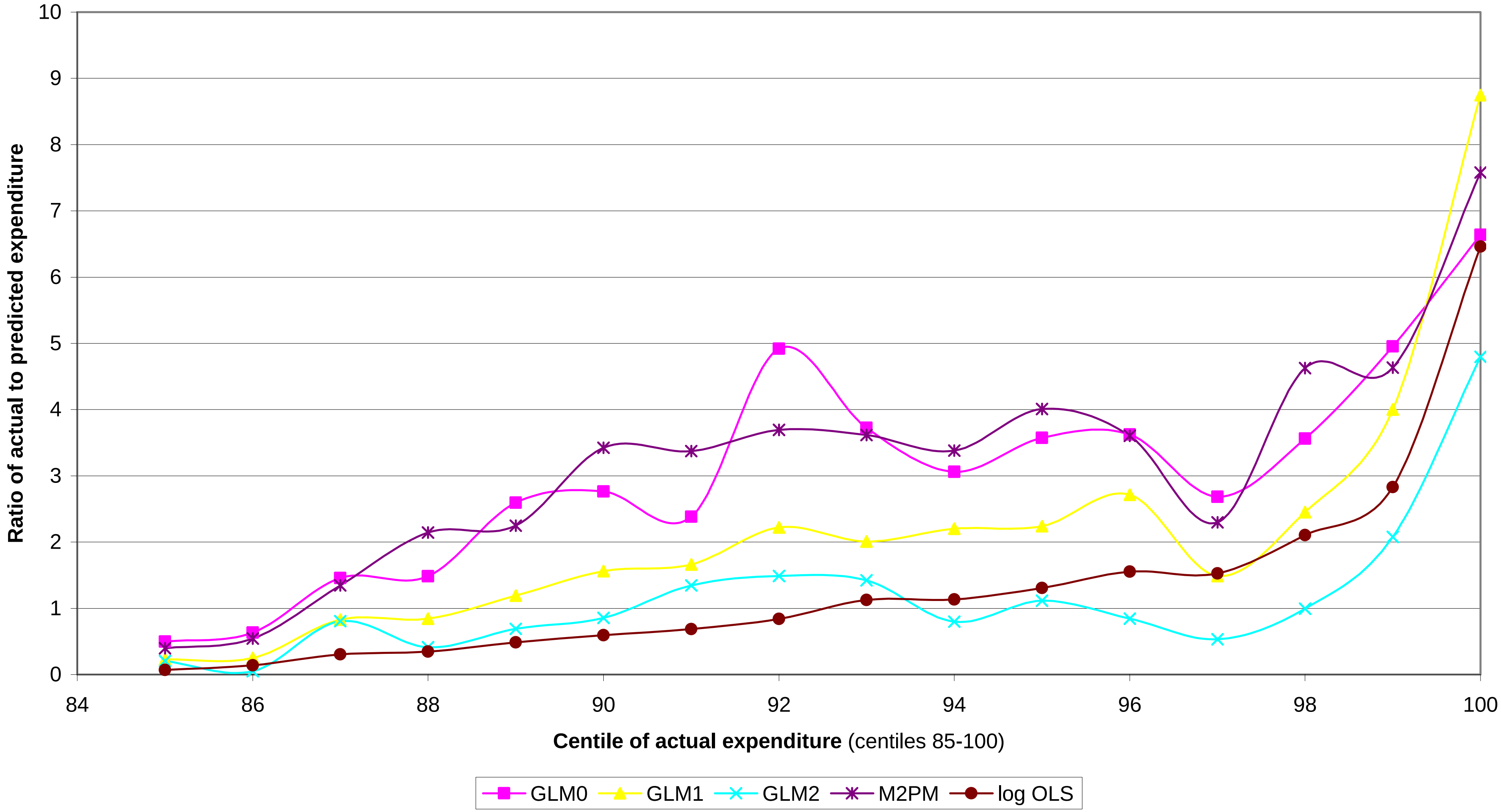

Mean predictions of alternative estimators relative to actual expenditure
(y_1 = all health expenditure)
Ratio of actual to predicted expenditure
0
1
2
3
4
5
6
7
8
9
10
10
20
30
40
50
60
70
80
90
100
Centile of actual expenditure (centiles 12-100)
GLM0
GLM1
GLM2
M2PM
log OLS

Figure 3

Mean predictions of alternative estimators relative to actual expenditure
(y_2 = hospital expenditure)



BANK OF GREECE WORKING PAPERS

41. Genberg, H., “Exchange-Rate Arrangements and Financial Integration in East Asia: On a Collision Course?”, including comments by James A. Dorn and Eiji Ogawa, May 2006.

42. Christl, J., “Regional Currency Arrangements: Insights from Europe”, including comments by Lars Jonung and the concluding remarks and main findings of the workshop by Eduard Hochreiter and George Tavlas, June 2006.

43. Edwards, S., “Monetary Unions, External Shocks and Economic Performance: A Latin American Perspective”, including comments by Enrique Alberola, June 2006.

44. Cooper, R. N., “Proposal for a Common Currency Among Rich Democracies” and Bordo, M. and H. James, “One World Money, Then and Now”, including comments by Sergio L. Schmukler, June 2006.

45. Williamson, J., “A Worldwide System of Reference Rates”, including comments by Marc Flandreau, August 2006.

46. Brissimis, S. N., M. D. Delis and E. G. Tsionas, “Technical and Allocative Efficiency in European Banking”, September 2006.

47. Athanasoglou, P. P., M. D. Delis and C. K. Staikouras, “Determinants of Bank Profitability in the South Eastern European Region”, September 2006.

48. Petroulas, P., “The Effect of the Euro on Foreign Direct Investment”, October 2006.

49. Andreou, A. S. and G. A. Zombanakis, “Computational Intelligence in Exchange-Rate Forecasting”, November 2006.

50. Milionis, A. E., “An Alternative Definition of Market Efficiency and Some Comments on its Empirical Testing”, November 2006.

51. Brissimis, S. N. and T. S. Kosma, “Market Conduct, Price Interdependence and Exchange Rate Pass-Through”, December 2006.

52. Anastasatos, T. G. and I. R. Davidson, “How Homogenous are Currency Crises? A Panel Study Using Multiple Response Models”, December, 2006.

53. Angelopoulou, E. and H. D. Gibson, “The Balance Sheet Channel of Monetary Policy Transmission: Evidence from the UK”, January, 2007.

54. Brissimis, S. N. and M. D. Delis, “Identification of a Loan Supply Function: A Cross-Country Test for the Existence of a Bank Lending Channel”, January, 2007.

55. Angelopoulou, E., "The Narrative Approach for the Identification of Monetary Policy Shocks in a Small Open Economy", February, 2007.

56. Sideris, D. A., "Foreign Exchange Intervention and Equilibrium Real Exchange Rates", February 2007.

57. Hondroyiannis, G., P.A.V.B. Swamy and G. S. Tavlas, "The New Keynesian Phillips Curve and Lagged Inflation: A Case of Spurious Correlation?", March 2007.

58. Brissimis, S. N. and T. Vlassopoulos, "The Interaction between Mortgage Financing and Housing Prices in Greece", March 2007.

59. Mylonidis, N. and D. Sideris, "Home Bias and Purchasing Power Parity: Evidence from the G-7 Countries", April 2007.

60. Petroulas, P., "Short-Term Capital Flows and Growth in Developed and Emerging Markets", May 2007.

61. Hall, S. G., G. Hondroyiannis, P.A.V.B. Swamy and G. S. Tavlas, "A Portfolio Balance Approach to Euro-area Money Demand in a Time-Varying Environment", October 2007.

62. Brissimis, S. N. and I. Skotida, "Optimal Monetary Policy in the Euro Area in the Presence of Heterogeneity", November 2007.

63. Gibson, H. D. and J. Malley, "The Contribution of Sector Productivity Differentials to Inflation in Greece", November 2007.

64. Sideris, D., "Wagner's Law in 19th Century Greece: A Cointegration and Causality Analysis", November 2007.

65. Kalyvitis, S. and I. Skotida, "Some Empirical Evidence on the Effects of U.S. Monetary Policy Shocks on Cross Exchange Rates", January 2008.

66. Sideris, D., "Real Exchange Rates over a Century: The Case of Drachma/Sterling Rate", January 2008.

67. Petroulas, P., "Spatial Interdependencies of FDI Locations: A Lessening of the Tyranny of Distance?", February 2008.