Measurement Error in Stylised and Diary Data on Time Use

Man Yee Kan
Department of Sociology
University of Oxford

Stephen Pudney
Institute for Social and Economic Research
University of Essex

ISER Working Paper
2007-03

Addresses for correspondence: Man Yee Kan, Department of Sociology, University of Oxford, Manor Road, Oxford, OX1 3UQ. Email: man-yee.kan@sociology.ox.ac.uk. Stephen Pudney, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ. E-mail: spudney@essex.ac.uk.

Acknowledgement:

We are grateful to participants at the IATUR conference, Copenhagen for useful comments. This work was supported by the Economic and Social Research Council through the Gender Equality Network project "Gender, Time Allocation and the Wage Gap" and MiSoC Research Centre (award nos. RES225252001 and RES518285001).

Readers wishing to cite this document are asked to use the following form of words:

> Kan, Man Yee and Pudney, Stephen (February 2007) 'Measurement Error in Stylised and Diary Data on Time Use', ISER Working Paper 2007-03. Colchester: University of Essex.

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and
- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

We investigate the nature of measurement error in time use data. Analysis of 'stylised' recall questionnaire estimates and diary-based estimates of housework time from the same respondents gives evidence of systematic biases in the stylised estimates and large random errors in both types of data. We examine the effect of these measurement problems on three common types of statistical analyses in which the time use variable is used as: (*i*) a dependent variable, (*ii*) an explanatory variable, and (*iii*) a basis for cross-tabulations. We develop methods to correct the biases induced by these measurement errors.

Keywords: Time use, measurement error, response bias

JEL classifications: C81, D13, J16, J22

1 Introduction

Time use data have played a central role in social science research in the past few decades. They have been used widely in statistical models of market and non-market activities, research into the household division of labour, studies on lifestyle patterns and well-being across different social groups and different countries, and so on. Last but not least, they act as revealing indicators of societal trends. Researchers have used them to understand the changing ways in which different groups of individuals structure their daily life (See, for example, Gershuny, 2000; Robinson, 1987).

1.1 Stylised and diary time use estimates

There are two main types of information on time use. The first, generally referred to as the "stylised" estimate, is based on responses to questions asking about the time devoted to various classes of activity in an "average" or "normal" or "typical" week[1] (Juster and Stafford, 1985). Interpreted literally, it requires respondents to perform two difficult tasks: to recall their activities in the recent past and to carry out an appropriate form of averaging. Doubts about the accuracy of stylised data rest on two concerns: the difficulty of the respondent's task, which might lead to substantial measurement error; and the lack of detail, which allows the respondent scope for choosing responses which correspond to a pre-existing (and possibly inaccurate) self-image.

The second type of data source, usually preferred by researchers, builds up a summary measure from entries in a time-use diary, in which the respondent is required to keep a very detailed record of activities throughout the day, on each of a set of (usually randomly sampled) days throughout the survey period. This approach is certainly not error-free: there may be recording or recall error involved in completing the diaries and the days selected for diary-keeping may, by chance, be unrepresentative of normal activity. Despite this, it is usually felt that there is less scope for systematic bias in diary records, so the analysis of diary-based estimates is likely to be less prone to systematic distortion than is the case for stylised estimates.

However, the use of diary measures imposes other limitations on research. There are very few panel datasets that incorporate time-use diaries and so longitudinal analysis,

[1] See Appendix 1.1 for details of these questions.

involving the study of change over time is not generally possible without resort to stylised time-use variables. A second limitation comes from the limited coverage of the questionnaires used in most specialist time-use surveys. Diary-keeping is a rather burdensome activity for respondents, so it is usually seen as inadvisable to adopt a wide-ranging or intensive general questionnaire to give information on the household context. In particular, it is usually not possible to build up a picture of the household's economic resources in anything like the detail that is possible with most conventional household surveys. Since the household's economic position is likely to have a strong influence on patterns of time use within the household, this is a potentially serious limitation.

The researcher is, therefore, faced with a difficult dilemma. Should we opt for detailed and (presumably) more reliable time-use data, at the cost of severe constraints on the type of research that can be done, or should we accept poorer time-use data, with the attendant risk of measurement error bias, to give us wider research scope? In this paper, we attempt to resolve the dilemma by: (*i*) assessing the nature an extent of measurement error in both diary and stylised time use data, (*ii*) evaluating its impact on various types of research that make use of time use data and (*iii*) exploring methods of adjustment which could reduce the impact of measurement error bias. To do this, we exploit a unique longitudinal dataset, the 1999-2001 Home On-Line Survey (HoL), which contains both stylised and diary estimates of time devoted to household activities and compare it with the widely used British Household Panel Survey (1991 - 2004), which gives only stylised estimates of time use. In this article, we focus on housework time.

In what follows, we review previous studies comparing stylised and diary-based time use estimates. We then go further by using both types of estimates collected from the same respondents to address the statistical and methodological issues arsing from measurement error in the two estimates.

1.2 Comparing stylised and diary estimates

Earlier studies were limited by the lack of data sources containing both stylised and diary-based time use estimates. It is particularly difficult to find comparable estimates because most surveys have collected stylised estimates of *weekly* housework time and paid work time, while many time-diary studies have collected only *one or two days* of diary records (usually one weekday diary and one weekend diary) from respondents.

Studies employing data from separate sources consistently report that stylised estimates of housework time exceed diary estimates. This finding was confirmed in a recent study by Juster, Ono, and Stafford (2003), who provided detailed comparisons between stylised estimates and diary estimates of housework hours and paid work hours from various years of U.S. data (1965–2001). Similar results were reported by other US studies (Bianchi, Milkie, and Sayer, 2000; Marini and Shelton, 1993; Robinson, 1985) and also by Baxter and Bittman (1995), using Australian data, and Niemi (1993) for Finnish data. Some of these studies also suggested that the difference between stylised and diary-based estimates is substantially larger in the case of women than in the case of men (Baxter and Bittman, 1995; Niemi, 1993; Robinson, 1985).

Despite the differences between these estimates, these studies argued that the two methods often reveal roughly similar patterns of variation between different groups. They therefore concluded that, despite being less accurate and reliable than diary-based estimates, stylised estimates provide a useful ordinal scaling of individuals' time use and for multivariate analyses of topics such as the domestic division of labour (Baxter and Bittman, 1985; Marina and Shelton, 1993; Robinson, 1985). Nevertheless, some studies have found the gap between diary-based and stylised estimates to vary systematically with respondents' characteristics (Press and Townsley, 1998).

Two recent studies compared diary and stylised time use estimates from a single data source. Bonke (2005) compared time-diary estimates and stylised time use from the Danish Time Use Survey, 2001 and found that respondents generally reported less household work time in survey interviews than in diaries. He also found that the difference was significantly larger when the respondents were women, parents or older people. Kitterød and Lyngstad (2005) analyzed data from the Norwegian Time Use Survey 2000 – 01 and found only modest differences between the two types of time use estimates, with the difference associated significantly with age but not gender. However, these studies were based on data collected from one-day diary records, which provide less ideal information for comparison with stylised time use estimates than weekly diary records.

We see that studies using time use data from the same data source and those using data from different sources have yielded different conclusions. It is clearly preferable to compare alternative time use estimates which are derived from a common set of respondents, to ensure comparability. To examine measurement error (and especially systematic error) in time use data, it is also better to use diary-based time use estimates derived from weekly rather than daily records to improve comparability and reduce purely random variation. Our main data source, the HoL survey, is unique in that it provides both weekly diary estimates and stylised estimates from the same respondents. Kan (2006) compared stylised and diary estimates of housework time from the HoL survey and found that the gap between the estimates is associated with gender, presence of dependent children, the amount of housework performed as secondary activities and irregularity in housework hours. She suggests mechanisms, such as social desirability and irregularities in work hours that potentially produce systematic biases in time use data. Although using the same data, in this article we will focus instead on the methodological and technical issues involved in estimating measurement error in time use data.

2 Time Use data

2.1 The Home On-line Survey

The HoL survey consists of three annual waves of household panel data (1999 – 2001). A distinctive advantage of this survey is that it contains both stylised estimates and diary-based estimates of time spent on housework. It also surpasses other time budget studies because it collected seven-day diaries from respondents, while other studies usually collect only one- or two-day diaries. It interviewed about 1,000 households drawn from a national random sample in Great Britain. It was originally intended for the estimation of time-use patterns as a result of the everyday use of information-and-computer technology, and therefore has an over-sample to make sure that 50% of the households have a personal computer. We use sampling weights to correct this non-proportional sampling mechanism.

Individuals aged 16 or over in the selected households were interviewed in all waves. A one-week self-completion diary designed to record what respondents were doing each day of that week every quarter hour of a day was given to the respondents after the interview. They were shown 35 pre-coded activities and were requested to record their primary activity and, if any, their secondary activity, in every time slot of the diaries. (the pre-coded activities are listed in Appendix 1.2). They were advised to fill in the diary at least once each day and then return it at the end of the designated week. Around 2,300 diaries (i.e., 16,100 diary days) were collected from respondents in all the three waves. Two variants of the housework time variable are used to allow for the possibility of simultaneous activities: one counts only time spent on primary activities, the other also includes housework done as a secondary activity.

The sample selected for the present study includes married and cohabiting men and women, to be consistent with the literature on the domestic division of labour of married couples. It is also restricted to cases where both questionnaire and diary records are present and where only fully or nearly completed diaries (missing time less than 3.5 hours) are included. The final sample pooled from the three waves contains 1,422 observations (weighted N = 1,248). In the analyses, standard errors are adjusted to take account of multiple observations for individuals in the sample.

2.2 The British Household Panel Survey

The British Household Panel Survey (BHPS) is the principal source of household- and individual-level panel data in the UK. It interviewed all members of a random sample of households initially in 1991, and subsequently all the original household members, their natural descendents, and all their current household co-residents annually. The first wave interviewed about 5,000 households and 10,300 individuals. Additional samples of 1,500 households in Scotland and the same number in Wales were added to the main sample in 1999, and a sample of 1,000 households in Northern Ireland was added in 2001. The BHPS currently contains 14 waves of data from around 9,000 households. From Wave 2 (1992) onwards, the BHPS has asked about respondents' normal weekly hours of housework. As mentioned earlier, the survey part of the HoL study also collected this variable using the same question wordings as the BHPS.

We create a pooled sample of married and cohabiting individuals from Wave 9 to Wave 11 (1999 – 2001) of the BHPS, i.e. individuals who were interviewed in the same period as the HoL survey. Standard errors are again adjusted to take account of multiple observations for same individuals in the sample.

We do not select cases from the BHPS additional samples of Wales and Scotland in 1999 and of Northern Ireland in 2001, so that the two data samples have similar regional representation of the population in the UK. The BHPS has no detailed time use diary and contains only stylised estimates of housework time. The BHPS and HoL stylised variables are, in principle, comparable.

2.3 Sample characteristics

Tables 1 and 2 summarise the BHPS and HoL samples for the common period 1999 - 2001. To be consistent with the HoL and BHPS stylised questions, we define housework time derived from HoL diary records as time spent on routine housework and cooking; this does not include time spent on DIY, gardening, and care.

Table 1 Characteristics of the HoL and BHPS samples

	HoL		**BHPS**	
	Mean	*SD*	*Mean*	*SD*
Personal characteristics				
Age	50.28	14.97	46.13	15.49
Female	0.52	-	0.52	-
Married	0.90	-	0.81	-
Has a job	0.60	-	0.66	-
Has no job	0.40	-	0.34	-
Education: degree or above	0.18	-	0.40	-
Education: intermediate	0.46	-	0.38	-
Education: O level or below	0.35	-	0.23	-
Household characteristics				
Number of children	0.58	0.92	0.72	1.05
Number of adults	2.34	0.65	2.35	0.75
Partner has a job	0.54	-	0.67	-
Partner has no job	0.46	-	0.33	-
Year				
1999	0.41	-	0.33	-
2000	0.30	-	0.34	-
2001	0.30	-	0.33	-
Weekly housework hours				
Housework hours: diary – primary activities only	10.50	9.18	-	-
Housework hours: diary – primary and secondary activities	11.48	9.74	-	-
Housework hours: stylised	11.75	10.26	11.77	10.78

Notes: For the HoL sample, the values are weighted to adjust the PC bias, weighted N = 1,248; For the BHPS sample, all values are unweighted, N = 19,852; cases where housework hours are missing (4.7% of the original sample) are dropped.

Table 2 Mean housework hours by sex and other characteristics

	BHPS (stylised estimates)		HoL (stylised estimates)		HoL (diary estimates - primary activities only)		HoL (diary estimates - primary and secondary activities)	
	Men	Women	Men	Women	Men	Women	Men	Women
Overall	5.82	17.29	6.72	16.41	4.66	15.90	5.02	17.46
	(6.07)	(11.26)	(7.05)	(10.55)	(4.96)	(8.86)	(5.21)	(9.11)
Marital Status								
Married	5.82	18.19	6.58	16.94	4.73	16.54	5.06	18.11
	(6.28)	(11.39)	(7.01)	(10.65)	(4.97)	(8.82)	(5.20)	(9.07)
Cohabitating	5.81	13.57	7.96	11.59	3.97	10.12	4.63	11.52
	(5.09)	(9.86)	(7.36)	(8.29)	(4.86)	(7.07)	(5.30)	(7.25)
Age								
≤ 30	5.46	13.43	6.41	11.06	4.27	8.91	5.00	10.71
	(4.93)	(10.20)	(4.38)	(11.73)	(4.05)	(6.04)	(5.20)	(6.96)
> 30 and ≤ 45	5.38	17.88	6.73	16.40	4.05	14.27	4.67	16.76
	(5.23)	(11.76)	(6.93)	(10.75)	(4.27)	(8.48)	(4.66)	(8.75)
> 45	6.25	18.44	6.74	17.48	4.99	18.24	5.18	19.19
	(6.87)	(10.90)	(7.32)	(9.88)	(5.33)	(8.63)	(5.45)	(9.04)
Number of children								
0	5.96	15.74	6.60	15.78	4.89	16.59	5.09	17.48
	(6.24)	(10.32)	(7.15)	(9.79)	(5.16)	(8.63)	(5.34)	(9.05)
1	5.45	16.51	7.02	14.07	3.63	13.55	4.13	16.14
	(5.56)	(10.00)	(6.29)	(8.65)	(4.30)	(8.55)	(4.59)	(8.71)
2	5.43	20.00	6.03	19.45	4.80	15.19	5.69	18.00
	(5.54)	(11.16)	(4.83)	(13.24)	(4.92)	(9.69)	(5.54)	(9.87)
≥ 3	6.26	25.70	8.84	21.85	4.77	17.02	5.38	19.41
	(6.75)	(15.70)	(11.40)	(12.71)	(4.13)	(9.16)	(4.28)	(9.07)
Educational level								
Degree or	5.46	15.05	6.86	14.14	5.66	13.36	6.00	15.02
above	(5.02)	(9.83)	(8.40)	(9.99)	(5.33)	(8.62)	(5.39)	(9.07)
Secondary	5.74	17.42	6.09	16.14	4.52	14.36	4.83	16.40
	(6.23)	(11.09)	(5.89)	(10.31)	(4.87)	(8.01)	(5.06)	(8.45)
O level or	6.69	20.37	7.56	18.06	4.40	19.21	4.83	20.12
below	(7.52)	(12.64)	(7.82)	(10.92)	(4.87)	(9.02)	(5.30)	(9.28)
Work status								
Have a job	4.81	14.87	5.65	14.29	3.73	13.03	4.13	14.69
	(4.54)	(9.40)	(5.50)	(9.65)	(4.11)	(7.78)	(4.47)	(8.25)
Have no job	8.37	20.92	8.53	19.24	6.24	19.72	6.51	21.13
	(8.31)	(12.73)	(8.83)	(11.05)	(5.82)	(8.79)	(5.99)	(8.92)
Partner's work status								
Have a job	5.46	16.90	6.20	15.47	4.25	14.52	4.69	16.26
		(10.73)	(5.60)	(10.92)	(4.29)	(8.88)	(4.68)	(9.14)
Have no job	6.38	18.33	7.23	17.72	5.06	17.82	5.34	19.12
	(7.27)	(12.47)	(8.22)	(10.92)	(5.22)	(8.49)	(5.67)	(8.82)

Notes: For the HoL sample, the values are weighted to adjust the PC bias, weighted *N* = 1,248; For the BHPS sample, all values are unweighted, *N* = 19,852; cases where housework hours are missing (4.7% of the original sample) are dropped. Figures in parentheses are standard deviations.

3 Measurement error

3.1 Modelling the measurement error process in HoL data

We first establish some notation. Let D_i be housework time as recorded by the time-use diary, for individual i. S_i is the corresponding stylised observation from responses to the recall questions. Both variables come from the same HoL survey and both are regarded as error-prone measures of the same underlying "true" current average level of housework time. Let us call this unobserved true measure H_i, and make the following assumptions.

Assumption 1 The diary variable D_i is an unbiased estimate of the true housework time H_i but deviates randomly from it:

$$D_i = H_i + U_i \tag{1}$$

where:

$$\mathrm{E}(U_i) = 0 \tag{2}$$

$$\mathrm{Var}(U_i) = \sigma_u^2 \tag{3}$$

Assumption 2 The stylised variable S_i is a potentially biased estimate of true housework time H_i, deviating systematically and randomly from it:

$$S_i - H_i = \alpha_0 + \alpha_1 X_{1i} + \ \ldots \ + \alpha_k X_{ki} + V_i \tag{4}$$

where:

$$\mathrm{E}(V_i) = 0 \tag{5}$$

$$\mathrm{Var}(V_i) = \sigma_v^2 \tag{6}$$

and $X_{1i} \ \ldots \ X_{ki}$ are a set of variables believed to describe the nature of the bias in the stylised estimate.

Assumption 3 The measurement errors, U_i and V_i, embedded in the stylised and diary variables are uncorrelated:

$$\mathrm{corr}(U_i, V_i) = 0 \tag{7}$$

The idea underlying Assumption 1 is that diary responses are essentially reliable in terms of the actual days covered by the diary, but that those days are randomly sampled, giving some purely random differences from "average" days. Assumption 2 captures the idea that the recall questions on which the stylised estimate is based may confuse respondents about the notion of an "average" week; it also gives respondents scope to portray their activity in a favourable light. These considerations introduce the possibility of systematic bias, as well as randomness. Assumption 3 reflects the presumption that the misreporting process distorting the stylised measure S_i is unrelated to the sampling process which generates random departures of the diary measure from the underlying “normal” level. If the diary-keeping process is also distorted by misreporting, then Assumption 3 is dangerous and we explore the robustness of our results to violations of this assumption in section 6 below.

One way of proceeding with the analysis would be to analyse the difference between the diary and stylised variables, i.e., to use D_i - S_i as a dependent variable in a regression on X_{1i} ... X_{ki}. This is done by Kan (2006), for example. However, a disadvantage of that approach is that it is impossible to estimate the measurement error variances σ_u^2 and σ_v^2 separately. These variances are important parameters which determine the magnitude of measurement error biases that are encountered in certain types of empirical analysis.

To understand the measurement error process in full detail, we need to estimate separate models for the two measures D_i and S_i. This, in turn, requires a regression model to explain the unobserved true variable H_i. Let us write this model as follows:

$$H_i = \beta_0 + \beta_1 Z_{1i} + \ldots + \beta_r Z_{ri} + \varepsilon_i \tag{8}$$

where Z_{1i} ... Z_{ki} are the variables believed to influence the individual’s use of time and ε_i is a random unobservable component of behaviour. Note that Z_{1i} ... Z_{ki} may overlap with the variables X_{1i} ... X_{ki}. However, we assume that the random components of measurement error, U_i and V_i, are uncorrelated with each other and with underlying behaviour, ε_i.

Substitute equation (5) into (1) and (4) to give:

$$D_i = \beta_0 + \beta_1 Z_{1i} + \ldots + \beta_r Z_{ri} + \varepsilon_i + U_i \tag{9}$$

$$\begin{aligned} S_i = {} & \beta_0 + \beta_1 Z_{1i} + \ldots + \beta_r Z_{ri} \\ & + \alpha_0 + \alpha_1 X_{1i} + \ldots + \alpha_k X_{ki} + \varepsilon_i + V_i \end{aligned} \tag{10}$$

Note that, if any variable appears both among the Xs and the Zs, then the corresponding terms can be combined in (10). For example, if X_{i1} and Z_{i1} are the same variable, then (10) will involve a single term $(\beta_1 + \alpha_1) Z_{1i}$. For the same reason, (10) has a singe combined intercept $(\beta_0 + \alpha_0)$.

If we estimate (9) and (10) as a pair of regression equations, the coefficients of $X_{1i} \ldots X_{ki}$ will give estimates of the impact of each X variable on the reporting bias in S_i. For a variable that appears among the Xs and the Zs, the difference between its coefficient in (10) and its coefficient in (9) is the required bias estimate.

Define the residuals in equations (9) and (10) as $R_{1i} = \varepsilon_i + U_i$ and $R_{2i} = \varepsilon_i + V_i$. Their variances and covariance are:

$$\mathrm{Var}(R_{1i}) = \sigma_\varepsilon^2 + \sigma_u^2 \tag{11}$$

$$\mathrm{Var}(R_{2i}) = \sigma_\varepsilon^2 + \sigma_v^2 \tag{12}$$

$$\mathrm{Cov}(R_{1i}, R_{2i}) = \sigma_\varepsilon^2 \tag{13}$$

Therefore, under our assumptions, the measurement error variances can be estimated very simply as:

$$\sigma_u^2 = \mathrm{Var}(R_{1i}) - \mathrm{Cov}(R_{1i}, R_{2i}) \tag{14}$$

$$\sigma_v^2 = \mathrm{Var}(R_{2i}) - \mathrm{Cov}(R_{1i}, R_{2i}) \tag{15}$$

We estimate the two regression equations (9) and (10) efficiently by the method of “seemingly unrelated regressions” (Zellner, 1962). In the application reported below, we use the same set of variables for the Zs and Xs (thus we allow for the possibility that all factors affecting the allocation of housework time also affect misreporting behaviour). In this special case, SUR is equivalent to using two separate OLS regressions of D_i and of S_i on the same set of variables $Z_{1i} \ldots Z_{ri}$.

4 Results for HoL data

Table 3 gives SUR estimates of the regression coefficients for HoL data, using two variants of the diary variable: one which includes only primary activities and another including both primary and secondary activities. We find highly significant differences between the regression results for the stylised and diary measures, implying the existence of systematic measurement error bias. In particular, the biases affect estimated gender differences strongly, since most of the coefficients of gender interacted with age, education employment and the number of children show significant differences between the equations for S_i and D_i at the 5% level.

Table 3 SUR results for HoL data

Explanatory variable	S_i	Primary activity only: D_i	Primary activity only: $\chi^2(1)$ Wald test for equality of coefficients	Primary & secondary activities: D'_i	Primary & secondary activities: $\chi^2(1)$ Wald test for equality of coefficients
Year 2000[a]	0.893	-0.358	5.02*	-0.609	3.00
Year 2001[a]	-0.331	-1.033*	1.56	-1.159	0.18
Female	-8.871	-18.159**	3.01	-17.108***	2.13
Age	-0.119	-0.202	0.26	-0.236*	1.11
Age-squared	0.001	0.002*	0.84	0.003*	1.83
Female*Age	0.584**	1.132***	5.94*	1.134***	6.48*
Female*Age^2	-0.004	-0.010***	7.25**	-0.010***	7.91**
Employed	-4.279***	-2.501**	2.67	-2.717**	1.96
Female*employed	0.060	-2.966**	4.97*	-2.803**	5.29*
Spouse employed	0.337	0.760	0.27	0.650	0.27
Female*spouse employed	-0.040	-0.226	0.03	0.004	0.04
High education[e]	0.651	1.005	0.14	1.202	0.13
Low education[e]	0.962	-0.745	5.13*	-0.551	4.13*
Female*high education	-2.980*	-2.373*	0.23	-2.865**	0.04
Female*low education	-0.567	3.095***	10.91**	2.352*	6.19*
No. of adults	1.207*	0.931*	0.24	0.983*	0.03
Female*no. of adults	-0.276	0.291	0.55	0.161	0.15
No. of children under 16	1.152**	0.676*	1.21	1.040**	0.33
Female*number of children	2.312***	0.935*	5.61*	0.568**	2.64
Intercept	8.848*	7.554**	0.11	8.963**	0.05
Pseudo-R^2	0.323	0.530		0.537	
Overall Wald test for systematic measurement error bias		$\chi^2(20)$ = 120.69 (P = 0.0000)		$\chi^2(20)$ = 81.18 (P = 0.0000)	

Notes: Data from the Home On-line Survey, 1999 - 2001. The sample is weighted to adjust the PC bias. Weighted N = 1,248. S_i is the stylised weekly housework hours. D_i is the diary-based weekly hours on housework, where housework is undertaken as a primary activity. D'_i is the diary-based weekly housework hours, where housework is undertaken as a primary or a secondary activity. [a]Omitted category = Year 1999. [e]High education = degree or above; medium education = A-level or O-level (omitted category); low education = below O-level. * p < .05, ** p < .01, *** p < .001.

Table 4 shows the estimated residual covariance matrix, the implied estimates of the variances σ_ε^2, σ_u^2 and σ_v^2, and the mean estimated bias (defined as $\hat{\alpha}_0 + \hat{\alpha}_1\bar{x}_1 + ... + \hat{\alpha}_r\bar{x}_r$) of the stylised estimate, implied by the SUR results. The residual variances and covariances can be used to solve for the underlying values of the variance parameters σ_ε^2, σ_u^2 and σ_v^2. The result is as we might expect: there is evidence of much more error in the stylised estimate, with σ_v^2 being about twice the size of σ_u^2. The residual variance, σ_ε^2, of the underlying model of "true" time use is estimated to be a little smaller than the variance of measurement error in the diary estimate. We therefore have the important conclusion that both the diary and (particularly) the stylised measure are subject to measurement errors which are large in comparison with the "intrinsic" unobservable variability of the correctly-measured time use variable. The measurement error variances are large compared to the average bias in the stylised measure, so that mismeasurement appears to be dominated by random error rather than biased reporting. However, taken as a whole, these findings suggest strongly that measurement problems cannot be ignored when analysing time use data.

Table 4 Estimated residual covariance matrix and the mean estimated bias

	Stylised estimates	Diary estimates	
		Primary activities only	Primary & secondary activities
Residual variance	71.00	39.45	43.81
Covariance of residuals		18.44	20.37
Variance estimates:			
$\hat{\sigma}_\varepsilon{}^2$		18.44	20.37
$\hat{\sigma}_u{}^2$		21.00	23.42
$\hat{\sigma}_v{}^2$		52.56	50.63
Mean estimated bias $= \hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}_1 + ... + \hat{\alpha}_r \bar{x}_r$		1.33 (*SD* = 1.96)	0.30 (*SD* = 1.61)

Note: Data from the Home On-line Survey, 1999 - 2001. Weighted *N* = 1,248.

5 Using the results to understand and reduce bias in other datasets

The HoL survey gives valuable evidence on the nature of measurement error in time use data but other datasets are of interest in their own right. To illustrate the implications of measurement error, we use stylised housework data from the BHPS, as our main focus. Our earlier results of measurement error properties based on the HoL estimates allow us to examine the problems arising from use of BHPS stylised data. There are three cases to consider: (*i*) analyses using a "stylised" time use variable as a dependent variable; (*ii*) analyses using the stylised variable as an explanatory factor. Our findings from HoL; and (*iii*) a predictive analysis attempting to estimate the "true" level of time use for an individual or group, given observations on the corresponding stylised and/or diary variables.

5.1 The impact of measurement error in models with time use as a dependent variable

Assume here that our aim is to understand the process determining the pattern of time use. To this end, we want to estimate the regression model (8), which describes "true" behaviour. However, assume also that we are using a dataset containing only the stylised variable S_i rather than the true variable H_i. An analyst who was unaware of the systematic error in S_i would mistakenly regress S_i on Z_{1i} ... Z_{ki} , ignoring the fact that average size of measurement errors also depends on the characteristics X_{1i} ... X_{ki}. Rewriting (10):

$$S_i = \beta_0 + \beta_1 Z_{1i} + \ ... \ + \beta_r Z_{ri} + \xi_i \tag{16}$$

where the "residual" is:

$$\xi_i = \alpha_0 + \alpha_1 X_{1i} + \ ... \ + \alpha_k X_{ki} + \varepsilon_i + V_i \tag{17}$$

The purely random element, V_i , causes no bias here: by contributing additional random variation, it merely reduces the R^2 of the regression. However, bias is caused by the systematic part of the measurement error, $\alpha_0 + \alpha_1 X_{1i} + \ ... \ + \alpha_k X_{ki}$, which will generally be correlated with the explanatory variables Z_{1i} ... Z_{ki}.

Table 5 illustrates this, by comparing the BHPS regression of S_i on Z_{1i} ... Z_{ri} with the regressions of S_i and D_i on Z_{1i} ... Z_{ri} for HoL data. Columns (1) and (2) of the table are biased estimates and column (3) is expected to be unbiased (or, at least, less so). Under our assumptions about measurement error, we would expect to find no significant differences between the coefficients in the BHPS stylised regression and the HoL stylised regression. This is the case with one exception involving the year for 2000. We conjecture that this is the consequence of the different patterns of sample attrition in the BHPS (a long established panel) and the HoL (a newer panel with higher inter-wave attrition). Comparing the coefficients in the models of Table 5 in which stylised estimates and diary estimates respectively are used as the dependent variable, we see the effects of systematic measurement biases on the regression coefficients, which are biased to a certain extent. However, the biases do not alter our major conclusions about the influences on housework hours. For example, all models predict that women undertake more housework than men. The coefficient of the dummy variable for females is negative in all models, but it should be interpreted together with its positive and significant interactions with other major variables, such as age and the number of dependent children. Individuals who are engaged in paid work tend to have shorter housework hours, but the diary-estimate models further predict that the effects on employed women are greater, since the coefficients of the interaction between gender and employment status are negative and significant. Moreover, women with high educational qualification tend to undertake less housework and those with low educational qualification tend to undertake more; the coefficients are however larger in the diary-estimate models. There is a positive association between housework hours and the number of dependent children, but the coefficients are greater in the stylised-estimate models.

In sum, we find that measurement error will bias the coefficients of explanatory variables when the time-use estimate is adopted as a dependent variable in the model. Nevertheless, the bias is not substantial because, as we found earlier, most of the measurement difference is random rather than systematic.

Table 5 OLS regression results for HoL and BHPS data

Explanatory variable	*BHPS* S_i	*HoL* S_i	*$\chi^2(1)$ Wald test for equality of coefficients of BHPS S_i & HOL S_i*	*HoL* D_i	*HoL* D_i^*
Year 2000	-0.081	0.893	*4.08**	-0.358	-0.609
Year 2001	-0.381***	-0.353	*0.01*	-1.033*	-1.159
Female	-11.023***	-9.265	*0.15*	-18.159***	-17.108***
Age	-0.046	-0.123	*0.24*	-0.202	-0.236*
Age^2	0.000	0.001	*0.12*	0.002*	0.003*
Female*Age	0.545***	0.598*	*0.02*	1.132***	1.134***
Female*Age^2	-.004***	-0.004	*0.01*	-0.010***	-0.010***
In work	-4.544***	-4.275***	*0.07*	-2.501**	-2.717**
Female*in work	-0.903	0.030	*0.38*	-2.966**	-2.803**
Spouse in work	1.070***	0.342	*1.32*	0.760	0.650
Female*spouse in work	1.622**	0.002	*1.79*	-0.226	0.004
High education[a]	0.028	0.650	*0.43*	1.005	1.202
Low education[a]	0.085	0.975	*1.41*	-0.745	-0.551
Female*high education	-1.428***	-2.952	*1.24*	-2.373*	-2.865**
Female*low education	1.547**	-0.568	*2.30*	3.095***	2.352*
No. of adults in household	0.079	1.207	*3.47*	0.931*	0.983*
Female*no. of adults	1.290***	-0.275	*3.01*	0.291	0.161
No. of children age < 16	0.342***	1.144*	*2.82*	0.676*	1.040**
Female*no. of children	3.265***	2.314**	*1.78*	0.935*	0.568**
Intercept	9.477***	8.965**	0.07	7.554**	8.963**
R^2	0.410	0.323	$\chi^2(20) = 41.26$ (P=0.0000)	0.530	0.537

Note: Data from the British Household Panel Survey, 1999 – 2001 (N = 19,852) and Home On-line Survey, 1999 – 2001 (weighted N = 1,248). S_i is stylised weekly housework hours. D_i is the diary-based weekly hours on housework, where housework is undertaken as a primary activity; D_i^* is the diary-based weekly hours on housework, where housework is undertaken as a primary or a secondary activity.

[a]High education = degree or above; medium education = A-level or O-level; low education = below O-level (omitted category). *p < .05. **p < .01. ***p < .001.

5.2 The impact of measurement error in models with time use as an explanatory variable

We are often interested in time use variables as explanatory factors in other types of analysis. An example is the study of women's satisfaction with housework arrangements. We use three alternative dependent variables, based on the responses to BHPS questions about satisfaction with life in general, with the spouse and with the amount of leisure time. Responses are coded as categorical variables taking the values 1 to 7 (with higher values indicating a higher level of satisfaction and 4 being neutral), but we simplify the analysis by treating them as continuous and continuing with linear regression methods.

In this case, using a regression framework again, our model is:

$$Y_i = \gamma H_i + \delta_0 + \delta_1 W_{1i} + \ \dots \ + \delta_q W_{qi} + \eta_i \tag{18}$$

where Y_i is the dependent variable, W_{1i} ... W_{ri} are additional explanatory variables and η_i is a random residual. The variable H_i is not available, so we use the diary estimate, D_i, as a proxy. This corresponds to the following model:

$$Y_i = \gamma S_i + \delta_0 + \delta_1 W_{1i} + \ \dots \ + \delta_q W_{qi} + \xi_i \tag{19}$$

Since $S_i = H_i + \alpha_0 + \alpha_1 X_{1i} + \ \dots \ + \alpha_k X_{ki} + V_i$, the residual in (19) is:

$$\xi_i = \varepsilon_i - \gamma \{ \alpha_0 + \alpha_1 X_{1i} + \ \dots \ + \alpha_k X_{ki} + V_i \} \tag{20}$$

This residual structure causes two problems. Firstly, the term V_i is a component of the proxy variable S_i and thus - γV_i is negatively correlated with the regressor S_i. This generally causes an attenuation bias - the estimate of γ is biased towards zero. However, further biases are contributed by the residual term $-\gamma\{\alpha_0 + \alpha_1 X_{1i} + \ \dots \ + \alpha_k X_{ki}\}$ which may be correlated with some or all of the explanatory variables S_i, W_{1i} ... W_{qi}. This systematic part of the measurement error bias is complex since it involves many variables and its sign is consequently difficult to predict.

It is possible to correct this measurement error bias, with a sufficiently large sample, at least. The method is implemented as follows:

Step 1 Using the HoL data, use SUR to estimate the regression models for S_i and D_i. Construct an estimate of the measurement error coefficients, α_0 ... α_k as the difference in the coefficients in the S_i and D_i equations. (Equivalently, compute a regression of (S_i - D_i) on (X_{1i} ... X_{ki})). Calculate an estimate of the measurement error variance σ_v^2 as the difference between the residual variance in the equation for S_i and the covariance between the residuals of the equations for S_i and D_i.

Step 2 Using BHPS data, construct a new variable $S_i^* = [S_i - \alpha_0 - \alpha_1 X_{1i} - \ldots - \alpha_k X_{ki}]$ and regress Y_i on S_i^* and (W_{1i} ... W_{qi}), using the errors-in-variables regression method,[2] with the estimated value of σ_v^2 specified as the known value of the measurement error variance for S_i^*.

Tables 6 - 8 show the result of applying this corrected estimator, for each of the three alternative dependent variables. (We restrict our analyses the 1999 wave of BHPS data here. It is because the EIV regression models cannot adjust the standard errors due to multiple observations of some respondents across waves.) For most coefficients, the effect of adjusting for measurement error bias is modest, with no changes in the sign of those coefficients which are significant and few changes in significance. However, the attenuation bias is clear: in all three models, the coefficient of housework hours is at least twice as large in the EIV estimates as it is in the OLS estimate. Thus a researcher using stylised BHPS data would greatly underestimate the role of housework as an influence on satisfaction, although not to the extent of wrongly concluding that its effect is zero.

[2] Errors-in-variables regression is a simple modification of the regression estimator to correct the measurement error bias. In the regression formula $(X'X)^{-1}X'y$, we simply subtract n times the (known) variance of the measurement error on variable X_j from the jth diagonal element of the matrix $X'X$. The EIV estimator is implemented in Stata. A more efficient estimator would involve simultaneous estimation of equations (8), (9) for HoL data and (9) and (15) for BHPS data, with cross-equation restriction; this would be much more complex. The two-step procedure proposed here does not take account of the fact that $\alpha_0 \ldots \alpha_k$ and σ_v^2 are estimated, so the quoted standard errors should be treated with caution.

Table 6 OLS and Errors-in-variables regression models of overall life satisfaction (N =6,515)

Explanatory variable	OLS	EIV [a]	EIV [b]
Weekly housework hours S_i [a,b]	-0.002	-0.004	-0.004
Female	0.391	0.409	0.417
Age	-0.074***	-0.074***	-0.074***
Age^2	0.001***	0.001***	0.001***
Female*age	0.025	0.027	0.027*
Female*age^2	0.0003*	0.0003*	0.0003*
In work	0.611***	0.601***	0.601***
Female*in work	-0.460***	-0.469***	-0.472***
Spouse in work	0.035	0.038	0.038
Female*spouse in work	0.120	0.125	0.126
Log(household income)	0.164**	0.166***	0.167***
Female*log(household income)	-0.021	-0.025	-0.026
Log(wage)	0.002	0.001	0.001
Female*log(wage)	-0.029	-0.029	-0.029
Cohabitating[c]	-0.048	-0.048	-0.048
Female*cohabitating	-0.084	-0.085	-0.087
Intercept	4.340***	4.349***	4.351***
R^2	0.047	0.047	0.047

Note: Data from the British Household Panel Survey, 1999. [a,b] In the EIV cases, the housework variable is corrected by removing estimated measurement error bias, where: (a) only primary; or (b) both primary and secondary activities are taken account of in the calculation of diary-based housework hours [c]Omitted category = Married. *p < .05. **p < .01. ***p < .001.

Table 7 OLS and Errors-in-variables regression models of satisfaction with spouse (N =6,507)

Explanatory variable	OLS	EIV [a]	EIV [b]
Weekly housework hours S_i [a,b]	-0.004*	-0.008	-0.009
Female	0.984	1.012	1.028
Age	0.012	0.012	0.012
Age^2	0.0001	0.0001	0.0001
Female*age	-0.064***	-0.060***	-0.060***
Female*age^2	0.001***	0.001***	0.001***
In work	-0.021	-0.039	-0.040
Female*in work	0.007	-0.009	-0.015
Spouse in work	0.036	0.040	0.040
Female*spouse in work	0.278**	0.287**	0.289**
Log(household income)	0.064	0.068	0.069
Female*log(household income)	0.027	0.020	0.019
Log(wage)	-0.003	-0.005	-0.006
Female*log(wage)	-0.019	-0.020	-0.020
Cohabitating [c]	-0.172**	-0.171**	-0.171**
Female*cohabitating	-0.145	-0.145	-0.149
Intercept	5.339***	5.353***	5.357***
R^2	0.021	0.021	0.021

Notes: see Table 6.

Table 8 OLS and Errors-in-variables regression models of satisfaction with leisure time ($N = 6{,}515$)

Explanatory variable	OLS	EIV [a]	EIV [b]
Weekly housework hours S_i [a,b]	-0.009***	-0.019**	-0.020**
Female	0.018	0.087	0.129
Age	-0.028*	-0.028*	-0.028*
Age^2	0.001***	0.001***	0.001***
Female*age	0.015	0.025	0.027
Female*age^2	0.0001	0.0002	0.0002
In work	-0.695***	-0.737***	-0.742***
Female*in work	0.474***	0.437***	0.421***
Spouse in work	0.167*	0.177*	0.178*
Female*spouse in work	-0.264*	-0.243*	-0.238*
Log(household income)	0.047	0.056	0.057
Female*log(household income)	0.032	0.016	0.012
Log(wage)	0.009	0.003	0.002
Female*log(wage)	-0.090*	-0.094*	-0.092*
Cohabitating [c]	0.059	0.061	0.061
Female*cohabitating	-0.035	-0.037	-0.046
Intercept	4.510***	4.543***	4.556***
R^2	0.140	0.141	0.142

Notes: see Table 6.

5.3 Estimating the average level of time use for population groups

A third type of analysis involves the use of the error-prone stylised time use variable to generate estimates of average time use levels for different population groups. The difficulty here is that the stylised variable S_i contains a systematic bias element which may contaminate group average measures. Even in the case of surveys like HoL, where we believe the detailed diary estimate D_i to be unbiased, it is possible in principle to use the additional information in S_i to improve statistical precision, although we do not pursue that idea here.

Suppose we are interested in a set of population sub-groups, labelled $g = 1...G$. Our aim is to estimate the mean of H within each of these subpopulations. Let us call these group means μ_1 ... μ_G. Given knowledge of the distribution of S_i , Z_{1i} ... Z_{ri} , X_{1i} ... X_{ki} within group g, the group mean μ_g can be written:

$$\begin{aligned}\mu_g &= \beta_0 + \beta_1\mu_{1g}^Z + \ldots + \beta_r\mu_{rg}^Z \\ &= \mu_g^S - \left(\alpha_0 + \alpha_1\mu_{1g}^X + \ldots + \beta_r\mu_{kg}^X\right)\end{aligned} \tag{21}$$

where μ_g^S, $\mu_{1g}^X ... \mu_{kg}^X$, $\mu_{1g}^Z ... \mu_{rg}^Z$ are respectively the means of S, $X_1 ... X_k$ and $Z_1 ... Z_r$ for population group g. The sample analogue of (21) gives the following correction for the sub-sample mean of S:

$$\hat{\mu}_g = \bar{s}_g - \left(\hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}_{1g} + ... + \hat{\alpha}_r \bar{x}_{rg}\right) \tag{22}$$

where $\hat{\alpha}_0 ... \hat{\alpha}_r$ are the relevant coefficients from the HoL regression model for S_i.

Table 9 shows the impact of these bias corrections on tabulations of time use by sex, marital status, age, self and partner's employment statuses, educational level and the number of children. In most of cases, the corrected estimates $\hat{\mu}_g{}^{a}$ and $\hat{\mu}_g{}^{b}$ are smaller than the stylised estimate $\bar{s}_g$. Nevertheless, when both primary and secondary activities are taken into account the corrected estimate $\hat{\mu}_g{}^{b}$ is on average greater than the stylised estimate $\bar{s}_g$ for women. This result confirms the view that women often undertake housework simultaneously with other activities. We can also observe that, where women have a low educational level (O-level or below) and where women have no child, the corrected estimates are larger. These findings are in line with the conjecture that systematic biases exist in stylised estimates.

Table 9 Mean housework hours by sex and other characteristics

	Men			Women		
	$\bar{s}_g$	$\hat{\mu}_g$ [a]	$\hat{\mu}_g$ [b]	$\bar{s}_g$	$\hat{\mu}_g$ [a]	$\hat{\mu}_g$ [b]
Overall	5.82	3.86	4.22	17.29	16.44	18.14
Married	5.81	3.83	4.20	18.19	17.41	19.10
Cohabitating	5.82	3.98	4.34	13.57	12.42	14.14
Age ≤ 30	5.46	3.69	4.05	13.43	11.92	13.69
Age > 30 and ≤ 45	5.38	3.58	3.94	17.88	15.51	17.85
Age > 45	6.25	4.12	4.48	18.43	19.12	20.26
In work	4.81	2.98	3.34	14.87	13.43	15.16
Not in work	8.37	6.10	6.46	20.91	20.94	22.59
Partner in work	5.46	3.62	3.98	16.90	15.58	17.45
Partner not in work	6.38	4.25	4.61	18.33	18.73	19.96
Education: degree or above	5.46	3.84	4.20	15.05	13.66	15.53
Education: intermediate	5.74	4.13	4.49	17.42	15.84	17.81
Education: O level or below	6.69	3.46	3.82	20.37	21.47	22.48
No. of children aged 0 - 15:						
0	5.96	3.93	4.29	15.74	16.39	17.49
1	5.45	3.62	3.98	16.51	15.00	16.97
2	5.43	3.63	3.99	20.00	16.50	19.24
≥ 3	6.26	4.38	4.74	25.71	19.90	23.55

Note: Data from the British Household Panel Survey, 1999 – 2001 (N = 19,852). $\bar{s}_g$ is the mean housework hours derived from the stylised data. $\hat{\mu}_g$ [a] and $\hat{\mu}_g$ [b] are the corrected estimates of mean housework hours, where the measurement bias is calculated from the HoL data. For $\hat{\mu}_g$ [a], only housework undertaken as primary activities in the diary are considered when calculating the measurement bias; for $\hat{\mu}_g$ [b], housework undertaken as both primary and secondary activities are taken into account.

6 Robustness

It is quite likely that survey respondents who make errors in their "stylised" reporting of time use will make similar errors in their diary entries. This suggests an alternative to assumption 3:

Assumption 3* The measurement errors, U_i and V_i, embedded in the stylised and diary variables are positively correlated:

$$\text{corr}(U_i, V_i) = \rho > 0 \tag{7}$$

It is not possible to estimate the correlation ρ together with σ_ε^2, σ_u^2 and σ_v^2, but we can investigate the robustness of our results by assuming a range of alternative positive values for ρ. Appendix 2 shows how we can solve for the values of σ_ε^2, σ_u^2 and σ_v^2 from the variances and covariance of the residuals from the regressions for S_i and D_i. It turns out that we are not completely free to choose values for ρ, since a large value of ρ may require a larger value for $\text{Cov}(R_{1i}, R_{2i})$ than we observe in the data. In practice, we find that values of ρ greater than 0.3 permit no solution (see Appendix A2.1). Therefore, the data are only consistent with only limited correlation between the measurement errors in S_i and D_i.

Tables 10 and 11 give results for a range of assumptions, where $\rho = 0, 0.1, 0.2$ and 0.3 respectively[3]. We find that the coefficients of significant explanatory variables are not greatly affected by changing the assumed value of ρ. The one exception to this is the coefficient of housework hours, which is more than doubled as we increase ρ from 0 to 0.3. The significance of variables is not affected in any important way by variations in the assumed value of ρ.

[3] For simplicity sake, we only present the models of satisfaction with leisure time in this section. Findings concerning the models of the general life satisfaction and the models of the satisfaction with spouse can be obtained from the authors upon request.

Table 10 OLS and Errors-in-variables regression models of satisfaction with leisure time – only primary activities in diary data are taken into account

Explanatory variable	OLS	Correlation ρ			
		0	**0.1**	**0.2**	**0.3**
Weekly housework hours S_i [a]	-0.009***	-0.019**	-0.023**	-0.030**	-0.049**
Female	0.018	0.087	0.118	0.179	0.349
Age	-0.028*	-0.028*	-0.028*	-0.028*	-0.028*
Age^2	0.001***	0.001***	0.001***	0.001***	0.001***
Female*age	0.015	0.025	0.027	0.033	0.048*
Female*age^2	0.0001	0.0002	0.0002	0.0003	-0.0004
In work	-0.695***	-0.737***	-0.751***	-0.779***	-0.856***
Female*in work	0.474***	0.437***	0.426***	0.405***	0.347**
Spouse in work	0.167*	0.177*	0.180**	0.186**	0.202**
Female*spouse in work	-0.264*	-0.243*	-0.235*	-0.219*	-0.175
Log(household income)	0.047	0.056	0.059	0.064	0.078
Female*log(household income)	0.032	0.016	0.010	-0.001	-0.031
Log(wage)	0.009	0.003	0.001	-0.003	-0.015
Female*log(wage)	-0.090*	-0.094*	-0.094*	-0.095**	-0.098**
Cohabitating [b]	0.059	0.061	0.062	0.063	0.067
Female*cohabitating	-0.035	-0.037	-0.041	-0.048	-0.070
Intercept	4.510***	4.543***	4.562***	4.599***	4.701***
R^2	0.140	0.141	0.142	0.143	0.146

Note: Data from the British Household Panel Survey, 1999. N = 6,515. [a] In the EIV cases, the housework variable is corrected by removing estimated measurement error bias. [b]Omitted category = Married. *p < .05. **p < .01. ***p < .001.

Table 11 OLS and Errors-in-variables regression models of satisfaction with leisure time – both primary and secondary activities in diary data are taken into account

Explanatory variable	OLS	Correlation ρ			
		0	**0.1**	**0.2**	**0.3**
Weekly housework hours S_i [a]	-0.009***	-0.020**	-0.024**	-0.030**	-0.046**
Female	0.018	0.129	0.164	0.231	0.398
Age	-0.028*	-0.028*	-0.028*	-0.028*	-0.028*
Age^2	0.001***	0.001***	0.001***	0.001***	0.001***
Female*age	0.015	0.027	0.030	0.035	0.048*
Female*age^2	0.0001	0.0002	-0.0002	-0.0003	-0.0004*
In work	-0.695***	-0.742***	-0.755***	-0.781***	-0.846***
Female*in work	0.474***	0.421***	0.409***	0.386**	0.328**
Spouse in work	0.167*	0.178*	0.181**	0.186**	0.200**
Female*spouse in work	-0.264*	-0.238*	-0.230*	-0.215*	-0.176
Log(household income)	0.047	0.057	0.059	0.064	0.076
Female*log(household income)	0.032	0.012	0.007	-0.004	-0.031
Log(wage)	0.009	0.002	0.000	-0.004	-0.013
Female*log(wage)	-0.090*	-0.092*	-0.093*	-0.093*	-0.094*
Cohabitating [b]	0.059	0.061	0.062	0.063	0.067
Female*cohabitating	-0.035	-0.046	-0.052	-0.061	-0.086
Intercept	4.510***	4.556***	4.575***	4.611***	4.701***
R^2	0.140	0.142	0.143	0.144	0.147

Note: see Table 10.

7 Conclusions

In this study, we have investigated the nature of measurement error in time use data. We have assumed diary estimates are systematically unbiased but subject to random measurement errors and stylised estimates are potentially prone to both systematic and random errors. Given these assumptions, we have found that there is systematic error in the stylised estimates of time use. Nonetheless the main source of error is random because the variance of measurement error is much larger for the stylised estimate than for the diary estimate.

We have also examined the implications of these results for three common kinds of statistical analyses in which: (*i*) the time use estimate is a dependent variable, (*ii*) the estimate is an explanatory variable, and (*iii*) time use estimates are used in cross-tabulations. Our findings show that the coefficient biases are small in the first case, because most of the errors are random rather than systematic.

In the second case, when the time use estimate is a right-hand side variable, we have found that its coefficient is strongly biased towards zero. Nevertheless this bias can be corrected if we have an external data source like the Home On-line Survey to estimate the variances of the mean and the error. The mean error bias is larger in this case compared to the first case. Still there is a positive correlation between the error on D_i and that on S_i.

In the third case, we have found that the estimates of average time use of given population groups are influenced by systematic error. The biases are particularly significant for some groups, such as women with dependent children.

In conclusion, our findings suggest that measurement error in time use data will bias the results of statistical models. We have proposed methods to correct these biases. We recommend that future research should give more efforts to develop data sources that contain both stylised and diary time use data, because such data sources are essential for the estimation of measurement error in time use data.

References

Baxter, J. and Michael Bittman. 1995. "Measuring time spent on housework: A comparison of two approaches." *Australian Journal of Social Research* 1:21-46.

Bianchi, Suzanne M., M. A. Milkie, L. C. Sayer, and John P. Robinson. 2000. "Is anyone doing housework? Trends in the gender division of household labor." *Social Forces* 79:191-222.

Bonke, Jens. 2005. "Paid work and unpaid work: Diary information versus questionnaire information." *Social Indicators Research* 70:349-368.

Gershuny, Jonathan. 2000. *Changing times: work and leisure in postindustrial society*. Oxford: Oxford University Press.

Juster, F. Thomas, Hiromi Ono, and Frank P. Stafford. 2003. "An assessment of alternative measures of time use." *Sociological Methodology* 33:19-54.

Juster, F. Thomas and Frank P. Stafford. 1985. *Time, goods, and well-being*. Ann Arbor, MI: The University of Michigan.

Kan, Man Yee. 2006. "Measuring housework participation: the gap between "stylised" questionnaire estimates and diary-based estimates." in *Institute for Social and Economic Research Working Paper 2006-11*. Colchester, UK: University of Essex.

Kitterød, Ragni Hege and Torkild Hovde Lyngstad. 2005. "Diary versus questionnaire information on time spent on housework - The case of Norway." *Electronic International Journal of Time Use Research* 2:13-32.

Marini, Margaret Mooney and Beth Anne Shelton. 1993. "Measuring household work: Recent experience in the United States." *Social Science Research* 22:361-382.

Niemi, Iiris. 1993. "Systematic error in behavioural measurement: Comparing results from interview and time budget studies." *Social Indicators Research* 30:229-244.

Press, Julie E. and Eleanor Townsley. 1998. "Wives' and husbands' housework reporting: Gender, class, and social desirability." *Gender and Society* 12:188-218.

Robinson, John P. 1985. "The validity and reliability of diaries versus alternative time use measures." Pp. 33-62 in *Time, goods, and well-being*, edited by F. T. Juster and F. P. Stafford. Ann Arbor, MI: The University of Michigan.

Robinson, John P. and Geoffrey Godbey. 1997. *Time for life: The surprising ways Americans use their time*. Pennsylvania: Pennsylvania States University Press.

Zellner, A. 1962. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association* 57:348-368.

Appendix 1: Survey questions and pre-coded activities

A1.1 HoL and BHPS Stylised survey question

The HoL survey and the BHPS ask respondents about their housework time using the same wording:

> "*About how many hours do you spend on housework in an average week, such as time spent on cooking, cleaning and doing the laundry?*"

Information about the distribution of responses can be found in Table 1.

A1.2 The HoL time-use diary

The HoL asks respondents to maintain a 7-day diary based on 15-minute time slots. They were asked to fill the following pre-coded activities in the slots of the diary. The responses are summarised in Table A1.

Table A1 Pre-coded activities used in the HoL time-use diaries

	Activity	Mean weekly hours as primary activity (*SD*)	Mean % of time as primary activity
1	Sleep, rest	59.39 (7.72)	35.5
2	Washing, dressing	5.50 (2.26)	3.3
3	Eating at home	8.00 (4.03)	4.8
4	Cooking	4.29 (3.87)	2.6
5	Child-adult care	4.66 (11.50)	2.8
6	Housework	6.21 (6.32)	3.7
7	DIY, odd jobs, gardening	4.19 (6.40)	2.5
8	Travel	5.78 (4.68)	3.5
9	At the workplace	20.64 (20.36)	12.3
10	Paid work at home	0.83 (3.45)	0.5
11	Study at home	0.36 (1.71)	0.2
12	Study out of home	0.60 (3.02)	0.4
13	Voluntary work	1.05 (2.91)	0.6
14	Shopping etc	4.12 (3.49)	2.5
15	Concerts, cinema, theatre, etc	1.36 (2.79)	0.8
16	Walking, outings etc	1.77 (3.13)	1.1
17	Eating/drinking out	2.67 (3.75)	1.6
18	Visiting friends	3.80 (5.24)	2.3
19	Sports activity	1.02 (2.34)	0.6
20	Hobbies, games	1.36 (3.17)	0.8
21	TV	17.36 (9.67)	10.4
22	Videos	0.91 (2.10)	0.5
23	Radio, CD etc	1.04 (2.24)	0.6
24	Books/papers/magazines	3.73 (4.51)	2.2
25	Visited by friends	1.76 (3.32)	1.1
26	Getting phone calls	0.41 (0.71)	0.2
27	Making phone calls	0.51 (0.86)	0.3
28	PC games or consoles	0.34 (1.61)	0.2
29	Emailing	0.23 (1.37)	0.1
30	Using the internet	0.22 (1.06)	0.1
31	Using a PC for education	0.19 (1.27)	0.1
32	Using a PC for work at home	0.28 (1.93)	0.2
33	Using a PC for other purposes	0.40 (1.82)	0.2
34	Doing nothing, may include illness	0.75 (3.19)	0.4
35	Doing other things, not elsewhere specified	1.66 (7.06)	1.0

Notes: Data from Home On-line Survey (1999-2001). Weighted *N* = 1, 248.

Appendix 2: Correlation between measurement errors

When Corr(U_i, V_i) is non-zero, equations (11)-(12) must be replaced by the following:

$$\mathrm{Var}(R_{1i}) = \sigma_\varepsilon^2 + \sigma_u^2 \tag{A2.1}$$

$$\mathrm{Var}(R_{2i}) = \sigma_\varepsilon^2 + \sigma_v^2 \tag{A2.2}$$

$$\mathrm{Cov}(R_{1i}, R_{2i}) = \sigma_\varepsilon^2 + \rho\sigma_u\sigma_v \tag{A2.3}$$

If ρ is known, we can substitute (A2.1) and (A2.2) into (A2.3) to eliminate σ_u^2 and σ_v^2 and give a quadratic in σ_ε^2. This has a pair of roots:

$$\sigma_\varepsilon^2 = \frac{2V_{12} - \rho^2(V_{11} + V_{22}) \pm \sqrt{\left[2V_{12} - \rho^2(V_{11} + V_{22})\right]^2 - 4\left(1-\rho^2\right)\left(V_{12}{}^2 - \rho^2 V_{11}V_{22}\right)}}{2\left(1-\rho^2\right)} \tag{A2.4}$$

where V_{rs} = cov(R_r, R_s). If ρ is set at a positive value, there is at most one root lying below Cov(R_{1i}, R_{2i}) and thus we select the smaller root as the estimate of σ_ε^2. For some values of ρ there may exist no admissible root. After σ_ε^2 is calculated using (A2.4), σ_u^2 and σ_v^2 can be calculated directly from (A2.1)-(A2.2).