Comparing total hours worked from the Norwegian LFS and QNA

– the major change to continuous data collection in Norway revisited, through seasonal – and break adjustment[1]

By Jørn Ivar Hamre[2]

Utilizing new preadjustment functionality in seasonal adjustment program X-13ARIMA-SEATS, we can now also more easily adjust for seasonal breaks in time series, a couple of years in the aftermath of major change. Prior to 1996, there were only one reference week a month in the Norwegian LFS, normally avoiding holidays. Preliminary unofficial monthly figures for total hours worked from the Norwegian LFS are seasonal- and break adjusted, aggregated to quarters, and compared to the total hours worked figures from the Labor account of our Quarterly National Accounts (QNA).

Introduction about the Norwegian LFS

The sampling design in the Norwegian Labour Force Survey (NLFS) is a one-stage county-stratified systematic sampling (with assumed random sorting) of family clusters from our Central Population Register(CPR). Lesser populated counties are overrepresented. The sample consists of about 12 000 family units (24 000 persons) each quarter. After 1995, the quarterly sample is distributed evenly among 13 reference weeks in the quarter. In other words, the NLFS is a continuous survey. A monthly sample in our LFS consists of either 4 or 5 whole reference weeks, so the monthly gross sample consists of about 24000*4/13 or 24000*5/13 persons. From 1996 onwards each sampled family participates in the survey 8 times *every 13 week* during a two year period. This means that 7/8 of the total sample overlap between two adjacent quarters, and that half of the total sample overlap between a quarter and the same quarter a year before. All resident persons aged 16 – 74 years in the selected families are interviewed by telephone (computer assisted). Proxy interviewing is to some extent allowed if direct interview is not possible. Nonresponse is now about 20 per cent.

We compute *monthly* weights based on monthly data sets. The *quarterly* weights in the *quarterly* data files are the *monthly* weights divided by 3, so monthly and quarterly estimates are consistent. The monthly weights (inflation factors) are adjusted in 3 steps, in the end with calibration on counties, register employment, 12 age groups and gender, after national post-stratification on register employment divided in 3 industry[3] groups, 12 age groups and gender. The initial weights take account of different sampling and non-response proportions in each county.

From 2006 the age limit was lowered to 15 years in the NLFS, and the questions about settled and actual working hours, overtime and temporary absence was rearranged to better follow Eurostat recommendations.

Preliminary estimation of total hours worked

Total number of hours worked during month m may be estimated by

$\hat{T}_m = F_m \times (\Sigma_i W_i \times t_i)$, where

t_i = number of hours worked during the reference week for interviewee *i*,

W_i = monthly weighting factor for interviewee *i* in month m,

$(\Sigma_i W_i \times t_i)$ is the estimate for the total number of hours worked ***per reference week*** on average during the reference month

F_m is a factor to transform the reference week average to a "monthly" sum in month m,

The easiest way is to set F_m=13/3=4.333 for all months, like a normalized standard months, but that would not work very well in quarters with 14 reference weeks (in years with 53 week).

Therefor we use:

F_m=13/3=4.333 for all months in quarters with 13 reference weeks, and then use

F_m=14/3=4.667 for the months in quarter with 14 reference weeks (in 53 week years).

---

[1] This Working Paper should not be reported as representing the views of Statistics Norway. The views expressed are those of the authors and do not necessarily reflect those of Statistics Norway.

[2] Senior Adviser, Department of social statistics, Statistics Norway. Email: ham@ssb.no

[3] The industry groups are 1) Primary industry, 2) Manufacturing and construction, and 3) Service industry. For the age group with the youngest and the oldest, the industry divided post-strata are collapsed to register employment (yes/no) in order to avoid empty or small post-strata.

If we want to obtain $\hat{T}_m$ for ***actual LFS reference month,*** we may use

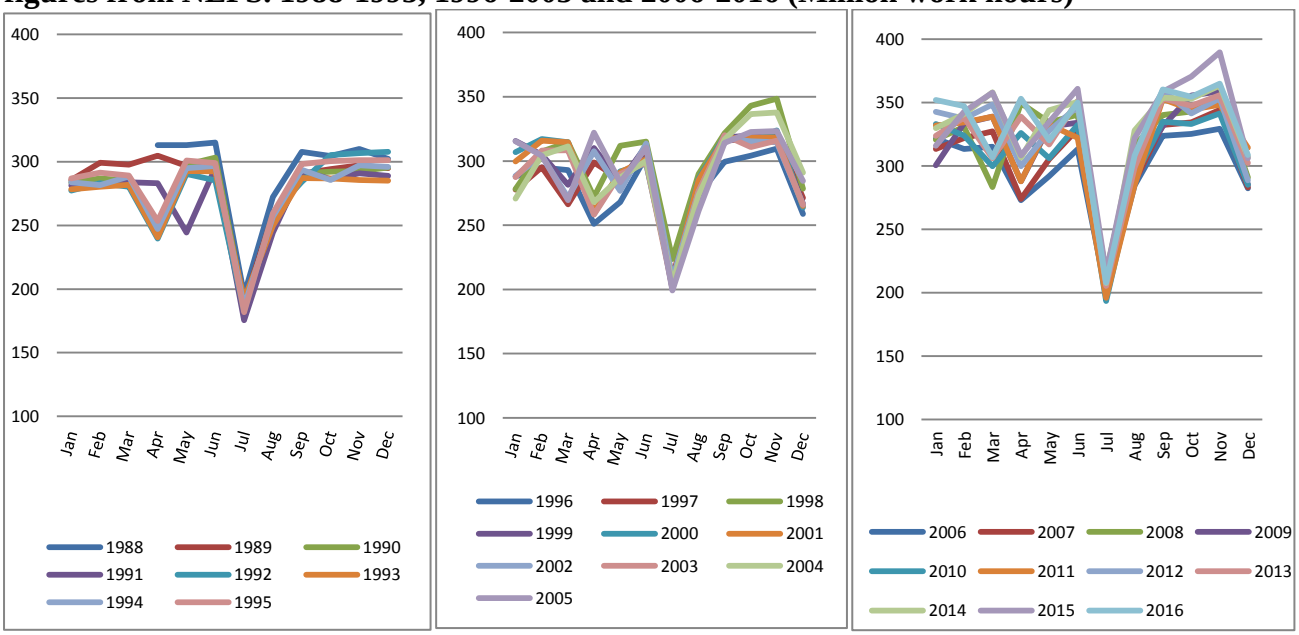
$$F_m = \begin{cases} 4 \ \ if\ a\ 4\ reference\ week\ month \\ 5 \ \ if\ a\ 5\ reference\ week\ month \end{cases}$$

Those also take account of the extra 5 week months (in stead of a 4 week month) in the quarter with 14 reference weeks. The drawback here is that a more or less random month in the quarter is about 25% larger than the other two[4]. For an illustration of this, please see the attached figure in the end. As far as I know, this makes seasonal adjustment difficult. The exception is if one use the same method for managing months every quarter, for example a permanent 4–4–5 calendar every quarter[5]. On the other hand, such deviation of reference months from the ISO-calendar is "banned" in the upcoming EU regulation. When the goal is to get quarterly total hours worked figures, I believe it's not so important to use the actual LFS reference months. The quarterly aggregates in the two suggested methods are about the same.

The reference week prior to 1996 and occasional breaks

From 2nd quarter 1988 to 4th quarter 1995, the NLFS was based on only one reference week each month. In this period the reference week was never on the week with Maundy Thursday and Good Friday, which are public holiday in Norway. Many take vacation the whole week. On the other hand the one and only surveyed reference week in April included Easter Monday[6] in the period 1988 – 1996, except in 1988, 1989 and 1991. But in 1991, the survey reference week in May included Ascension Day. All these factors effect the figures on man-hours worked, and needs to be handled specially in the seasonal adjustment process. These occasional breaks in April and May are quit visible in the left panel of the figure bellow. The lack of Christmas holidays in NLFS before 1996 is also rather visible in the left panel bellow compared to later years in the middle and right panel.

Figure 1. Total hours worked per normalized months (no. of ref.weeks in Q/3). Preliminary test figures from NLFS. 1988-1995, 1996-2005 and 2006-2016 (Million work hours)



User defined preadjustment variables[7]

A ***March Easter*** variable is constructed based on the "LFS-calender", with the value 1 for March and the value -1 for April if Good Friday is included in a reference week in March, otherwise zero, also zero for other months than March and April, and for all months prior to 1996. The March Easter variable is then deseasonalized based on monthly averages after 1996.

---

[4] A more complicated and probably better method is used in the Swedish LFS. For more information, please see Background Facts for Labour and Education Statistics 2012:1 from SCB. But also these monthly estimates are difficult to seasonally adjust, I believe, under the ISO calendar distribution of weeks into months.

[5] This also is challenged in 53-week years.

[6] Also public holiday in Norway.

[7] All these (user defined) regression variables may be ***fine tuned by a week-multiplier*** to take account of that some months represent 4 reference weeks and other 5 reference weeks. It makes the March Easter variable even more significant, because the estimated Easter effect in the population on total hours worked are "watered down" by being on out of 5 weeks that are estimated together some years. The week-multiplier should be 1.087031 (=4.348125/4) if it's a 4 week months, and 0.869625 (=4.348125/5), since the long run average number of weeks in a months is 4.348125 (17,75% of years have 53 weeks).

Due to sensitivity to single holidays or common days off, the total hours worked series in addition are pre-adjusted for Easter Monday, May 1st (Labour day), May 17th (our constitution day), Whit Monday and Ascension Day[8]. Assuming the effect of these common days off have the same and constant effect on total hours worked, one numeric variable called "***springfree***" is constructed, counting the number of these common days off when they fall on weekdays (Monday – Friday) in the reference months according to our "LFS-calendar".
Prior to 1996, the values on this variable is sett to zero for all months, except the occasional breaks mentioned earlier. Due to that, the value of ***springfree*** is set to ***4.333*** for 1991M5 (because of Ascension Day in the one and only reference week that month), and for 1990M4, 1992M4, 1993M4, 1994M4 and 1995M4 (because of Easter Monday in the one and only reference week those months).

In addition we pre-adjust total hours worked series for the ***number of common days off that falls on week days in December***, and for the ***number of normal days in the week between Christmas and New Year's Eve that falls in December/January*** in the LFS. These variables are set to zero for all months prior to 1996. All regression variables are deseasonalized by subtracting the long term monthly average after 1996. The aim of these two variables is to stabilize the irregularities in regular working days and common days off inn the last week or two of December.

Seasonal- and break adjustment

Seasonal – and break adjustment are done through X-13ARIMA-SEATS. Level shifts and different seasonal outliers are tested both for the major change to continuous data collection from 1996 in Norway and for major change in 2006. We also specified temporary level shifts for the months in quarter with 14 reference weeks[9]. Regression Model in table 1 seemed to work best, where I kept only regressors with a t-value>1.96.

Table 1. Regression Model with ARIMA: (0 1 1)(0 1 1) and no log transformation. Variables, parameter estimates, standard errors and t-values

Variables		Parameter Estimate	Standard Error	t-value
SO1996.4	(Seasonal Outlier)	-35,4	3,5	-10,0
SO1996.12	(Seasonal Outlier)	-28,4	4,2	-6,8
SO2006.7	(Seasonal Outlier)	-18,2	3,3	-5,6
TL1992.10-1992.12	(14-week quarter)	18,2	3,4	5,4
TL1998.10-1998.12	(14-week quarter)	24,5	3,4	7,2
TL2004.10-2004.12	(14-week quarter)	21,3	3,4	6,3
TL2009.10-2009.12	(14-week quarter)	18,3	3,4	5,4
TL2015.10-2015.12	(14-week quarter)	21,4	3,5	6,1
User-defined				
March_easter (after '96)*		-41,1	1,4	-29,7
Spring_free (after '96 and occasional breaks)*		-9,4	0,5	-20,0
Red_days_i_Dec (after '96)*		-11,6	0,8	-14,0
Christmas_week_jobdays (after '96)*		-4,1	0,6	-6,8

* Deseasonalized with monthly means after '96

Utilizing the model in table 1 for preadjustment of the time series gives a clear and completely regular monthly pattern without any visible breaks, as seen in Figure 2. Therefore the seasonal adjustment also works well.

[8] We also take account of the effect when two of these days occur on the same date, for instance Ascension Day on May 17th in 2007 and 2012 and on May 1st in 2008.

[9] The preliminary specification here of variables for 14-week quarter and 53 week years should probably be done in another way in order to lift the level in 53 week years, in a smooth way, without disorganizing the seasonal pattern. Maybe one use-defined variable that capture all the months effected by 14-week quarter (assuming equal constant effect), and then a deseasonalizing would do the job. Then we wouldn't have to specify new 14-week quarter periods as they come. Or maybe subtracting annual averages instead could work as well? The problem is that an entire extra week in a quarter in 53 week years deviates quite a bit from the Gregorian calendar that the QNA and everybody else use.

Figure 2: Prior adjusted LFS-figures for total hours worked per months in Norway. 1988-2016 (million workhours)

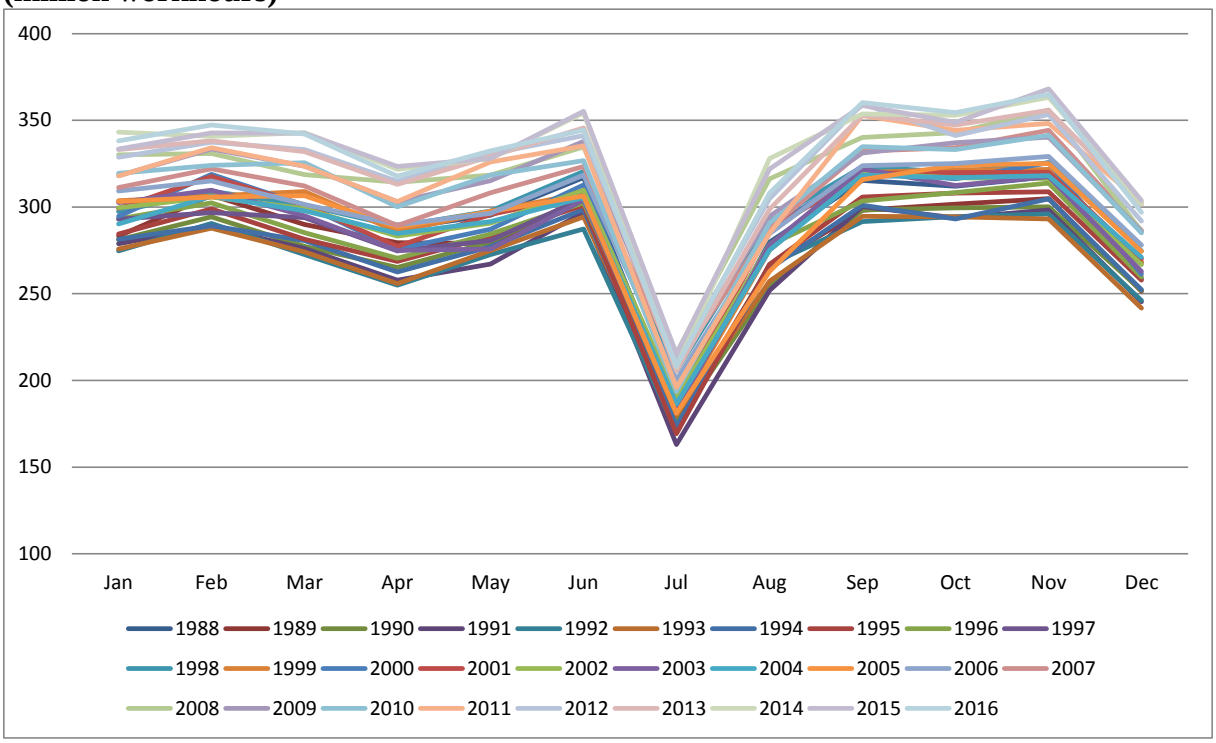


Table 2 shows that the most important seasonal adjustment diagnostics looks quite well, even though M3 and M5 have a value above 1.

Table 2 Monitoring and Quality Assessment Statistics

The relative contribution of the irregular component to the stationary portion of the variance (from Table F 2.F)	**M2 = 0.07**
The amount of moving seasonality present relative to the amount of stable seasonality (from Table F 2.I)	**M7 = 0.06**
Same as 8, calculated for recent years only.	**M10 = 0.16**
Same as 9, calculated for recent years only.	**M11 = 0.15**
A weighted average of M1 - M11	**Q = 0.63**

NOTE: The measures above has a range between 0 – 3, where values less than 1 is satisfactory.

Quarterly LFS figures are then made by taking the sum for the 3 months in the quarter for the seasonal adjusted and trend-cycle figures.

Comparison of total hours worked, seasonal- and break adjusted, from the NLFS and QNA

QNA figures for total hours worked are only available from 1995Q1. NLFS figures for total hours worked are preliminary unofficial in Figure 3.

The development in the QNA and the LFS figures up to 2004 and after 2006 seems similar, and reasonable.

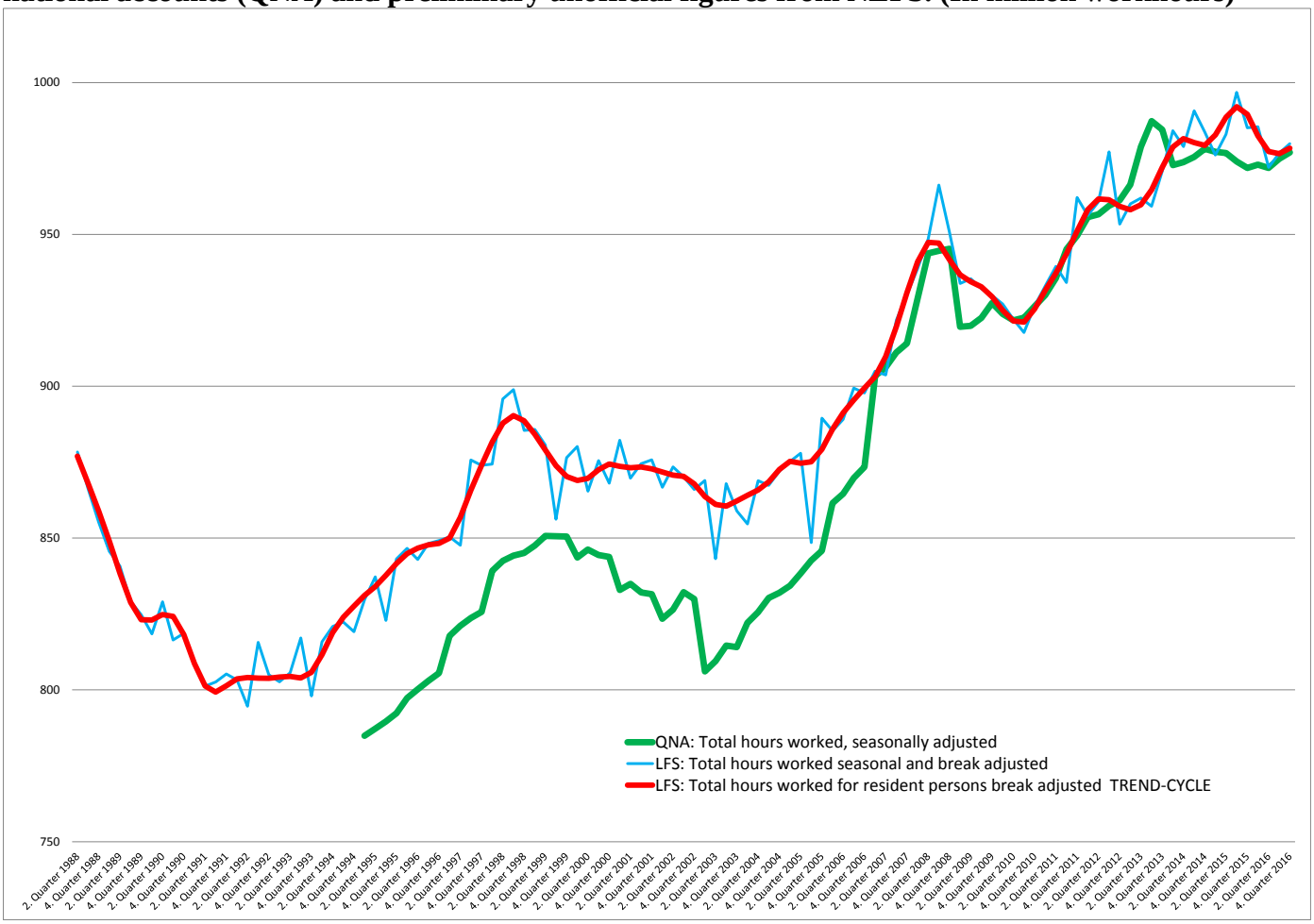
The levels for total hours worked in the QNA and NLFS are rather different in Figure 3. Due to different populations, the LFS is higher or the QNA is lower than expected.

Resident persons aged 15-74 (16-74 before 2006) is the population in the LFS, while QNA don't have an age limit, and also include employed short-term immigrants, who work in Norway without being registered as residents in the CPR. They are expected to be in Norway for less than six months.

The number of ***short-term immigrants***[10] who are registered wage earners increased from around 15 thousand in 2004 to about 38 thousand persons in 2008. If we assume 37 hours of work pr. week for these short-term immigrants and multiplying by 13 weeks, we get a rough estimate of 7 million hours worked in 2004 and 18 million hours in 2008. This gives an estimated ***increase*** of about 11 million work hours in the period, so this probably can't explain the whole difference in development. For the difference in levels, this takes them further apart.

[10] Source: http://www.ssb.no/en/table/06135 and http://www.ssb.no/en/table/08435

Figure 3. Total hours worked seasonally adjusted and trend-cycle. Official figures from the Quarterly national accounts (QNA) and preliminary unofficial figures from NLFS. (In million workhours)



Discussion

The new preliminary unofficial figures for total hours worked from the NLFS have not been utilized in the Labor account of the Norwegian QNA. The uncertainty for the quarterly industry divided LFS-figures are too high, so only quarterly employment totals from the NLFS is normally used in the QNA. In stead detailed register based employment statistics by industry, and settled working hours are used in an indirect approach, starting with annual national accounts. Statistics on overtime, absence, wage and wage sums statistics, and many other sources are used together with calendar information and law and deals about holiday[11].

There may be some uncertainty about the reporting of settled working hours in the employee register used in the NA, and our employment register used to only cover jobs with minimum 4 hours per week, before 2015.

The hours worked variable in the LFS is also especially uncertain. There are some measurement problems in the LFS for the working time variables. They are sensitive to both proxy answers and memory issues, especially if the date of the interview and the end of the reference week is high. There may be quality issues regarding rounding, to nearest 5 or 10 hour that may have a systematic effect. Also too high working hours numbers may be answered by interviewees, because they felt that higher working hours give impression of more status.

In the new official statistics for overtime from the NLFS, proxy interviews about overtime and overtime hours are dismissed. A statistical method of imputation is used in stead for these variables. The actual hours worked per week are not yet consistent with settled working hours and the imputed over time hours for proxy interviews. The quality of the information on temporarily absence part of the week from proxy interviews has not yet been analyzed.

[11] The main principle in the QNA system is that the national accounts figures for the current quarter are computed based on the development of short-term indicators from the last base year of the annual national accounts.

The order of the questions about settled and actual working hours, overtime and absence may also be of some importance. As from 2006, the order of these questions in the NLFS was changed to the recommended way by EUROSTAT. Only one seasonal outlier (SO2006.7) was clearly significant, but no level shift from 2006.

According to Villund (2009)[12], LFS based contractual working hours are systematically longer, and vary more, than the register based, used in (Q)NA. Also, proxy interviews gives overestimation of contractual working hours in the NLFS.

Bias exploration in Zhang, Thomsen ang Kleven (2013)[13] suggests that the estimation procedure may not remove all nonresponse bias. Therefore we cannot rule out the possibility of an upward biased employment rate, also influencing total hours worked estimates.

These seasonally adjusted figures on total hours worked in the months and quarter are also preliminary. Due to very different seasonal pattern for youths and adults, and different long term trends in part time jobs for men and women, indirect seasonal adjustment of age and gender groups would be preferable. For this start up investigation and low-key presentation, only direct seasonal adjustment of the total was chosen.

More investigation about the difference in levels between LFS and QNA figures on total hours worked must be done; before we consider publishing such NLFS[14] figures.

Conclusion

Utilizing the new preadjustment functionality in seasonal adjustment program makes it easier to adjust for the break in the time series, a couple of years in the aftermath of a major change. But in the LFS case, user defined preadjustment variables must be supplied to take into account of the special LFS-calendar and national common holiday, especially for total actual hours worked or man-weeks worked.

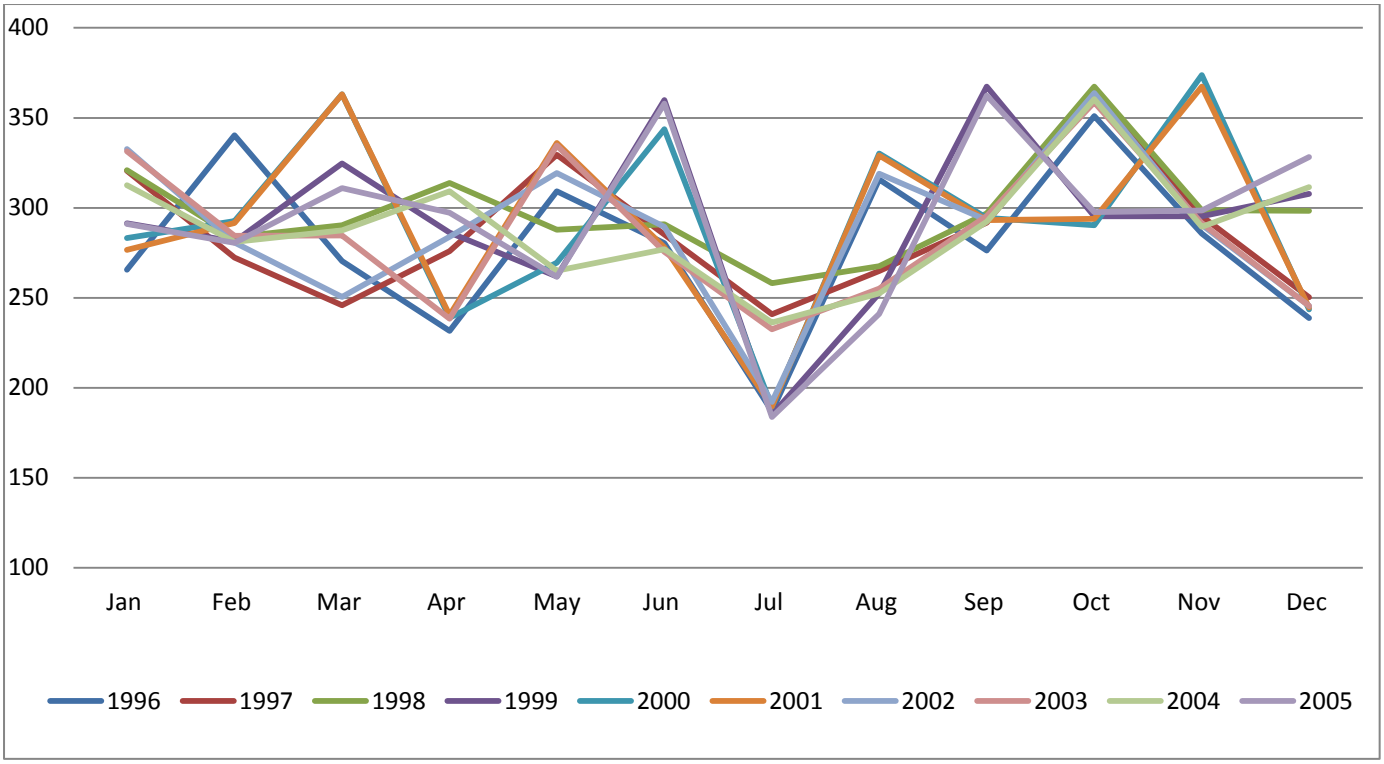
Attachment 1

Figure 4. Total hours worked per LFS-reference months for resident persons aged 16-74 Preliminary test figures from NLFS in Million work hours. Year 1996-2005

[12] Reports 2009/3 by Ole Villund: Measuring working hours in the Norwegian LFS - A pilot study of data quality using administrative Registers.

[13] Zhang, Li-Chun, Thomsen, Ib and Kleven, Øyvin (2013): On the Use of Auxiliary and Paradata for Dealing With Non-sampling Errors in Household Surveys. International Statistical Review.

[14] The estimation procedure for actual hours of work per week could have been better if we had calibrated on register employee ***by settled working hours (part/full time).*** That is proposed in the new estimation procedure we are working on. Also if total hours worked by industry was top priority for the NLFS, Regression **C**omposite estimation would probably also be very beneficial.



NOTE: Here average actual hours worked pr. week multiplied by 4 or 5 depending on the number of (reference) week in the LFS calendar months.