# Methods of Combining Time-use Diary Data with Economic Survey Data: How the COVID-19 Pandemic Affected Gendered Work Patterns in India in 2019-2020

By Wendy Olsen*, Manasi Bera**, Amaresh Dubey+, Jihye Kim*, Ioana Macoveciuc*, Samantha Watson++, Arkadiusz Wiśniowski*
*University of Manchester **India Institute of Dalit Studies +Jawaharlal Nehru University ++ILO and University of Southampton

Appendix

Table A1: Correlation Matrix From ITUS 2019 and CMIE Waves 18-21, ie Late 2019-and All of 2020

## ITUS

| | fractional H&V time | fractional work time | binary for unemployment | age* | age*-squared | female | hhsize* | ever married | depndt_child* | lockdown | wave | caseweight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| fhtime | 1.00 | -0.83 | -0.17 | -0.09 | -0.10 | 0.94 | -0.07 | -0.13 | 0.26 | -0.04 | -0.08 | 1.00 |
| fetime | -0.83 | 1.00 | -0.16 | 0.24 | 0.21 | -0.86 | -0.07 | 0.02 | 0.11 | -0.01 | -0.02 | -0.83 |
| unemployed | -0.17 | -0.16 | 1.00 | -0.35 | -0.32 | -0.15 | 0.01 | 0.00 | -0.52 | -0.16 | -0.03 | -0.17 |
| age_new | -0.09 | 0.24 | -0.35 | 1.00 | 1.00 | -0.13 | -0.42 | 0.12 | 0.78 | -0.35 | -0.17 | -0.09 |
| age2_new | -0.10 | 0.21 | -0.32 | 1.00 | 1.00 | -0.13 | -0.42 | 0.13 | 0.74 | -0.37 | -0.15 | -0.10 |
| female | 0.94 | -0.86 | -0.15 | -0.13 | -0.13 | 1.00 | -0.05 | -0.09 | 0.13 | -0.05 | -0.09 | 0.94 |
| hhsize_new | -0.07 | -0.07 | 0.01 | -0.42 | -0.42 | -0.05 | 1.00 | -0.60 | -0.29 | 0.42 | -0.09 | -0.07 |
| evermarried | -0.13 | 0.02 | 0.00 | 0.12 | 0.13 | -0.09 | -0.60 | 1.00 | -0.10 | -0.42 | -0.12 | -0.13 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| depndt_child_new | 0.26 | 0.11 | -0.52 | 0.78 | 0.74 | 0.13 | -0.29 | -0.10 | 1.00 | -0.14 | -0.20 | 0.26 |
| lockdown | -0.04 | -0.01 | -0.16 | -0.35 | -0.37 | -0.05 | 0.42 | -0.42 | -0.14 | 1.00 | -0.12 | -0.04 |
| wave | -0.08 | -0.02 | -0.03 | -0.17 | -0.15 | -0.09 | -0.09 | -0.12 | -0.20 | -0.12 | 1.00 | -0.08 |
| fwt | 1.00 | -0.83 | -0.17 | -0.09 | -0.10 | 0.94 | -0.07 | -0.13 | 0.26 | -0.04 | -0.08 | 1.00 |
| | -0.83 | 1.00 | -0.16 | 0.24 | 0.21 | -0.86 | -0.07 | 0.02 | 0.11 | -0.01 | -0.02 | -0.83 |

Note: * the variable is centred and standardised using the weighted mean and weighted standard deviation
Key: H&V = household and voluntary services
WOHVS = work other than household and voluntary services
Depndt children are those age 12 and under
The dependent child ratio is DepndtChildN/Hhsize, standardised
Lockdown is a constant term in each Wave, defined in the paper
Caseweight is a post-hoc combined weight variable

## CMIE

| | fractional H&V time | fractional work time | binary for unemployment | age* | age*-squared | female | hhsize* | ever married | depndt_child* | lockdown | wave | caseweight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| fH&Vtime | 1.00 | -0.75 | -0.14 | 0.09 | 0.09 | 0.87 | -0.15 | 0.19 | -0.05 | -0.18 | -0.19 | 0.04 |
| fWODVStime | -0.75 | 1.00 | -0.31 | 0.32 | 0.30 | -0.83 | -0.09 | 0.29 | 0.04 | -0.12 | -0.07 | -0.11 |
| Unemployed | -0.14 | -0.31 | 1.00 | -0.49 | -0.46 | -0.15 | 0.04 | -0.62 | -0.18 | 0.06 | 0.02 | 0.00 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| age_new | 0.09 | 0.32 | -0.49 | 1.00 | 1.00 | 0.02 | -0.57 | 0.90 | -0.35 | -0.21 | -0.16 | -0.21 |
| age2_new | 0.09 | 0.30 | -0.46 | 1.00 | 1.00 | 0.02 | -0.58 | 0.87 | -0.39 | -0.19 | -0.14 | -0.19 |
| Female | 0.87 | -0.83 | -0.15 | 0.02 | 0.02 | 1.00 | -0.08 | 0.14 | -0.02 | -0.12 | -0.12 | -0.03 |
| hhsize_new | -0.15 | -0.09 | 0.04 | -0.57 | -0.58 | -0.08 | 1.00 | -0.39 | 0.62 | -0.17 | -0.19 | -0.17 |
| Evermarried | 0.19 | 0.29 | -0.62 | 0.90 | 0.87 | 0.14 | -0.39 | 1.00 | 0.00 | -0.33 | -0.29 | -0.26 |
| depndt_child_new | -0.05 | 0.04 | -0.18 | -0.35 | -0.39 | -0.02 | 0.62 | 0.00 | 1.00 | -0.26 | -0.27 | -0.15 |
| lockdown | -0.18 | -0.12 | 0.06 | -0.21 | -0.19 | -0.12 | -0.17 | -0.33 | -0.26 | 1.00 | 0.79 | 0.37 |
| Wave | -0.19 | -0.07 | 0.02 | -0.16 | -0.14 | -0.12 | -0.19 | -0.29 | -0.27 | 0.79 | 1.00 | 0.20 |
| Fwt | 0.04 | -0.11 | 0.00 | -0.21 | -0.19 | -0.03 | -0.17 | -0.26 | -0.15 | 0.37 | 0.20 | 1.00 |
| | 1.00 | -0.75 | -0.14 | 0.09 | 0.09 | 0.87 | -0.15 | 0.19 | -0.05 | -0.18 | -0.19 | 0.04 |

Note: * the variable is centred and standardised using the weighted mean and weighted standard deviation
Key: H&V = household and voluntary services
WOHVS = work other than household and voluntary services
Depndt children are those age 12 and under
The dependent child ratio is DepndtChildN/Hhsize, standardised
Lockdown is a constant term in each Wave, defined in the paper
Caseweight is a post-hoc combined weight variable.

Combining Time-use Diary Data…, V1 Appendix. Cite as: Wendy Olsen, Manasi Bera, Greg Dropkin, Amaresh Dubey, Jihye Kim, Ioana Macoveciuc, Samantha Watson, and Arkadiusz Wiśniowski (2021), Combining CMIE-ITUS Data With Bayesian Estimation, https://github.com/WendyOlsen/combining-cmie-itus-data/, Creative commons attribution.

Figure A1:

Table A2: Variables Used from ITUS 2019 and CMIE Waves 18-21, ie Late 2019-and All of 2020

GITHUB CONTENTS
About:
This repository holds code and figures as well as appendix material for papers combining-cmie-itus-data. The papers look at time-use of women and men in India during a period 2019-2020 when a tertile (4-month interval) survey, CMIE was used to augment the Indian Time-Use Survey of 2019. Both used large national samples, which had a multi-stage clustered and stratified selection procedure. The CMIE however did not have lists of village or town residents except insofar as they created such lists. Therefore, their samp ling method was different at the final stage within the final sampled geographic units; see Centre for Monitoring Indian Economic (CMIE) Consumer Pyramids survey (no date).

**CMIE data: Introduction to the Survey Methods Used**

CMIE data is usually collected from the field on hand-held GPS enabled phone devices in a face-to-face interview. However, during the 2020 lockdown a telephonic interview and later both were used as possible. As a result of both lockdown and switch to telephone interview, acceptance per day dropped from an average of 1,219 in January and February to 425 in April. While response rate was 84.5% in Wave 18, it dropped to 64.4% in Wave 19. Also, during the lockdown weeks the response accepted shifted towards rural households and farmers and daily wage earners. The table below presents the sample size and response rate for each wave considered.

**Table: Sample size and response rate for CMIE CPS Wave 18-21**

| Wave No. | Wave | Sample Households | Household Members | Accepted Households | Response Rate |
|---|---|---|---|---|---|
| Wave 21 | Sep-Dec 2020 | 1,74,405 | 8,59,608 | 1,23,188 | 70.63 |
| Wave 20 | May-Aug 2020 | 1,74,405 | 8,55,808 | 76,386 | 43.8 |
| Wave 19 | Jan-Apr 2020 | 1,74,405 | 8,54,355 | 1,12,289 | 64.38 |
| Wave 18 | Sep-Dec 2019 | 1,74,405 | 8,51,786 | 1,47,319 | 84.47 |

Source:
https://peopleofindiadx.cmie.com/kommon/bin/sr.php?kall=wshowadv&code=poiwv_cpdx&sectid=21

**Table: Sample for analysis, CMIE CPS Wave 18-21**

| Wave No. | Total household members | Accepted household members | Living household members |
|---|---|---|---|
| Wave 21 | 8,59,608 | 6,16,485 | 4,71,635 |
| Wave 20 | 8,55,808 | 3,83,592 | 2,97,668 |
| Wave 19 | 8,54,355 | 5,57,729 | 4,31,683 |
| Wave 18 | 8,51,786 | 7,24,904 | 5,70,296 |

**Further Details of CMIE Sampling Methods**

The initial sample for the survey was created on the basis of Census 2001 districts, revamped in 2013 to accommodate new districts in Census 2011. The survey omits these states/UTs: Arunachal Pradesh, Nagaland, Manipur, Mizoram, Andaman & Nicobar Islands, Lakshadweep, Dadra & Nagar Haveli, Daman & Diu.

A stratified multi-stage survey design was used to draw a sample of households. First, the 640 districts of the 2011 Census were organised into 110 Homogeneous Regions (HR). HRs bring together neighbouring districts with similar agro-climatic conditions, urbanisation levels, and female literacy. Exceptionally, the set of included north-eastern states is considered as one HR. Next, each HR is divided into rural and urban sub-strata. The Primary Sampling Units (PSUs) were the villages and towns of the 2011 Census. The sampling frame for rural PSUs was the list of villages, and that for urban PSUs was the list of towns. To represent different town sizes, all towns of each HR were stratified: very large towns with more than 200,000 households in 2011; large towns between 60,000 and 200,000 households; medium sized towns with between 20,000-60,000 households; and small towns with less than 20,000 households. In rural regions, all the villages within an HR form one rural stratum, giving 412 strata in all.

Due to inaccessibility, a few strata were omitted at the field stage. All towns within a stratum formed the sampling frame from which sampled towns were selected. At least one town was selected from each such stratum. 328 towns were selected through this process. From each selected town, 21 Census Enumeration Blocks (CEB) – each one a cluster of 100-125 neighbouring households- were randomly selected. Households were selected through a process of geographically localised sampling from each of the CEBs (this also helped to avoid the difficulty of creating household roster). In doing so, every *nth* household from a village street was selected, *n* being a randomly selected integer between 5 and 15. First, households were selected from the main street, every alternative selection being made from opposite sides of the street. Afterwards, as required, households were selected from the inner streets. Similarly, for the rural sample, first, villages were selected through a simple random sampling process. Then households were selected from each selected village through a process of systematic convenience sampling, partly planned through the geography of each village (Mohanon, et al., 2021). As published in summary format, the representativeness of these samples in Karnataka state (after using the post-hoc adjustment weights) was acceptable on age-groups and sex (Mohanan, et al., 2021). The post-hoc case weights available to us are the same ones used by Manohan, et al.. From two weights options, we used the one aimed at studies that omit children ages 0-14 years.

## CODE SAMPLES

We enable replication of our work by providing several snippets of code showing key elements of the data-cleaning and modelling strategy. Note, this code is all under Creative Commons license, see Github https://github.com/WendyOlsen/combining-cmie-itus-data/
and https://github.com/WendyOlsen/combining-cmie-itus-data/blob/main/LICENSE

The citation for the code should be:

Code from Wendy Olsen, Manasi Bera, Greg Dropkin, Amaresh Dubey, Jihye Kim, Ioana Macoveciuc, Samantha Watson, and Arkadiusz Wiśniowski (2021), Combining CMIE-ITUS Data With Bayesian Estimation, https://github.com/WendyOlsen/combining-cmie-itus-data/, Creative commons attribution.

## SAMPLE 1

Pseudo-code for ITUS Data Cleaning

Authors Ioana Macoveciuc, Jihye Kim
Obtain the ITUS data from its government source and clean it as follows in four steps.

1 convert the stints to time in minutes/day
2 obtain the weights of the cases as a post-hoc adjustment weight, fwt
3 create additional variables such as standardised expenditure per capita of the household
4 save results including a summary of minutes/day

Examples are provided here for each of these steps.

```
#Convert time from hours to minutes
data$time <- data$time *60
#Indicate dominance of the activity for each stint
#Make uniquestint_id (psid+stintid)
data$stint_id <- as.character(data %>% group_indices(data$time_from))
data$psid_string <- as.character(data$psid)
data <- data %>% mutate(uniquestint = paste0(psid_string, stint_id))
data <- data[order(data$uniquestint),]

#Flag the WOHVS* activities denoted as major
#* WOHVS refers to work other than household services & voluntary services.
data <- data %>% mutate(product0=as.numeric(majoractivity==1 & (
                              activity_code >= 110 & activity_code <=142|
                              activity_code >= 211 & activity_code <=242
                                    )))

#Thus a WOHVS_stint=a stint that has WOHVS work as major activity

data$product_stint <- ave(data$product0, data$uniquestint, FUN =
function(x) max(as.numeric(x)))
#Measuring HVS:  domestic_stint=a stint that has domestic work or voluntary
services
data <- data %>% mutate(domestic0=as.numeric(activity_code >=311 &
activity_code <=372 |
                               activity_code >=390 & activity_code <=439 |
                               activity_code >=490 & activity_code <= 519 |
                               activity_code == 521 |
                               activity_code == 522 |
```

```
                               activity_code == 529))
data$domestic_stint <- ave(data$domestic0, data$uniquestint, FUN =
function(x) max(as.numeric(x)))

#Generate aggregated data
time_productive_0 <- data %>% filter( majoractivity==1 & (
          activity_code >= 110 & activity_code <=142|
                              activity_code >= 211 & activity_code <=242
                                    ))
temp_0 <- data.frame(with(time_productive_0, cbind(time, psid)))
etime <- aggregate.data.frame(temp_0, by=list(temp_0$psid), sum)
etime <- etime[, c(1, 2)]
etime <- rename(etime, c("psid" = "Group.1", "etime" ="time"))

#Generate aggregated data with HVS work-time variable (htime)
time_domestic_0 <- data %>% filter (data$product_stint!=1 &
domestic_stint==1 &
                               (activity_code >=311 & activity_code <=372 |
                               activity_code >=390 & activity_code <=439 |
                               activity_code >=490 & activity_code <= 519 |
                               activity_code == 521 |
                               activity_code == 522 |
                               activity_code == 529
                                    ))

#To avoid duplicated counts, divide aggregate by the number of domestic
activities per stint
time_domestic_0$no_domestic_activity <- ave(time_domestic_0$domestic_stint,
time_domestic_0$uniquestint, FUN = function(x) sum(as.numeric(x)))

temp_1 <- data.frame(with(time_domestic_0,
cbind(time/time_domestic_0$no_domestic_activity, psid)))
htime <- aggregate.data.frame(temp_1, by=list(temp_1$psid), sum)
htime <- htime[, c(1, 2)]
htime <- rename(htime, c("psid" = "Group.1", "htime" ="V1"))

#Merge aggregates(htime, etime) with the original data and select one row
per person
df <- data %>% filter(activity_id==1)
df <- merge(df, htime, by= "psid", all = TRUE)
df <- merge(df, etime, by= "psid", all = TRUE)
df$htime[is.na(df$htime)] <- 0
df$etime[is.na(df$etime)] <- 0

#Generate the number of dependent children
df$children <- as.numeric(df$age<15)
df$depndt_child_under15 <- ave(df$children, df$hhid, FUN = function(x)
sum(as.numeric(x)))
df$children_new <- as.numeric(df$age<13)
df$depndt_child_under13 <- ave(df$children_new, df$hhid, FUN = function(x)
sum(as.numeric(x)))
df$havingchildunder13 <- as.numeric(df$depndt_child_under13>=1)
df$havingchildunder15 <- as.numeric(df$depndt_child_under15>=1)

#Create a new dataset for modelling
 … [some steps are omitted]  …
data_itus <- df %>%  mutate (
        wave = 1,
        lockdown = 0,
        wtime  = htime+etime,
```

```
        fhtime = htime/1440,
        fetime = etime/1440,
        fwtime = (htime+etime)/1440,
        age2 = age*age,
        urban=as.numeric(sector==2),
        expenditure=as.numeric(expenditure_monthly/hhsize),

evermarried=as.numeric(maritalstatus==2|maritalstatus==3|maritalstatus==4),
        unemployed=as.numeric(usualstatus==81),
        st=as.numeric(socialgroup==1),
        sc=as.numeric(socialgroup==2),
        )

data_itus <- data_itus %>% filter(age >=15 & age <=65 & !is.na(expenditure)
& !is.na(belowprimary) ) %>% select(psid, hhid, wave, fwt, wtime, htime,
etime, fwtime, fhtime, fetime, age, age2, female, urban,  hhsize,
expenditure, evermarried, depndt_child_under15, state, unemployed, st, sc,
lockdown, depndt_child_under13, weight, havingchildunder13,
havingchildunder15)

# Standardise variables
…
data_itus$age_c  <-  data_itus$age - weighted.mean(data_itus$age,
wt=data_itus$fwt)
data_itus$age2_z <-  (data_itus$age*data_itus$age -
weighted.mean(data_itus$age,  wt=data_itus$fwt))/ weighted.sd
(data_itus$age,  wt=data_itus$fwt)
#data_itus$expenditure_z <- (data_itus$expenditure-
weighted.mean(data_itus$expenditure,
wt=data_itus$fwt))/weighted.sd(data_itus$expenditure,wt= data_itus$fwt)

#Save .csv files
write.csv(data_itus, "filename1.csv")
write.csv(summary(data_itus), "filename1_summary.csv")
data_itus <- read_csv("filename1.csv")
```

## S A M P L E 2

Pseudo-code for CMIE Data Cleaning

Authors Greg Campbell, Jihye Kim, Manasi Bera.
Obtain the CMIE data from its proprietary source and clean it as follows.

1. Import data
2. Rearrange dataframe to make it easier to read
3. Filter dataset to balance the panel and exclude any individuals younger than 15 and older than 65
4. Generate new variables including age-groups, worktime, and hworktime, etc. For flogit, create the new fractional variables, and create an unemployment binary variable. Note: personal labour statuses such as 'unemployed' in the ITUS do not utilise the time-use stint information. Refer to the documentation.
5. Creating a new dataset for modelling

Detailed Code Examples
First, creating case-weights for CMIE. The variable weight_15yrsplus is provided by CMIE.

```
#### GENERATING FWT VARIABLE ####
```

```
df$fwt <- as.integer(df$weight_15yrsplus/mean(df$weight_15yrsplus))
df$fwt[df$fwt==0] <- 1
```

Second, adapting time to fractional units (minutes per day).

```
## GENERATING NEW FRACTIONAL VARIABLE  ####
df$frac_othertime<- (df$othertime)/1440
df$frac_worktime<- (df$worktime)/1440
df$frac_hworktime<- (df$hworktime)/1440
```

Third, create an unemployment binary.

```
cbind(table(df$employment_status, useNA = "always"))
df$unem<-as.numeric(as.factor(df$employment_status))
table(df$employment_status,df$unem)
df$unemployed<- 0
df$unemployed[ df$unem %in% c(3,4)] <- 1
table(df$unem,df$unemployed)

df$unem<-NULL
```

When CMIE variables are standardised for regression, a weighted mean and weighted standard deviation is used to create the Z score.

SAMPLE 3

```
############################################################
# Flogit Model - Data combining model
# Implementation in STAN
# Code for the manuscript:
#
```

**#Methods of Combining Time-use Diary Data with Economic Survey Data: How the**

**#COVID-19 Pandemic Affected Gendered Work Patterns in India in 2019-2020**

#By Wendy Olsen*, Manasi Bera**, Amaresh Dubey+, Jihye Kim*, Ioana Macoveciuc*, Samantha #Watson++, Arkadiusz Wiśniowski*

#*University of Manchester **India Institute of Dalit Studies +Jawaharlal Nehru University ++ILO and #University of Southampton

#Acknowledgements:

#We thank the International Labour Office, whose statistics department has helped to fund this #research. We thank Greg Campbell, who helped with data cleaning at an early stage in the project. #We thank our Delhi hosting institution: Indian Institute for Dalit Studies (IIDS), which hosted one #of the large datasets, and the academics Dr. Bera and Prof. Dubey who were critical to success of #this project.

#October 2021

```
############################################################
# Code Authors:
# Jihye Kim and Wendy Olsen
# Social Statistics Department, UOM
############################################################
# Run the regression with the explicit likelihood function
# and priors using STAN
# Requires: rstan, tidyverse, tictoc
```

```
…
library(tidyverse)
library(rstan)
library(tictoc)
…

#########################################################
######### download data and sampling for model  #########
#########################################################

setwd("D:/")
sample_itus <- read.csv[filename]
sample_cmie <- read.csv[filename]
#sample_combined <- bind_rows(sample_cmie, sample_itus)
#write.csv(sample_combined, [filename in target location on drive])
#sample_combined <- read.csv([filename] )
########################################################
##########DATA cleaning for models #####################
########################################################

line.data <-list( "fetime" = c(sample_itus$fetime,sample_cmie$fetime),
                 "fhtime" = c(sample_itus$fetime,sample_cmie$fhtime),
                 "fwtime" = c(sample_itus$fwtime,sample_cmie$fwtime),
                 "umemployed" =
c(sample_itus$unemployed,sample_cmie$unemployed),
                  "female" = c(sample_itus$female,sample_cmie$female),
                  "age" = c(sample_itus$age_c,sample_cmie$age_c),


… etc., etc.,
… "fwt" = c(sample_itus$fwt,sample_cmie$fwt),
                  "lockdown" =
c(sample_itus$lockdown,sample_cmie$lockdown),
                  "wave" = c(sample_itus$wave,sample_cmie$wave),
                  "urban" = c(sample_itus$urban,sample_cmie$urban),
                  "unemployed" =
c(sample_itus$unemployed,sample_cmie$unemployed),
                  "N"=80000)

####################################################################
################ ITUS & CMIE - Fetime With INTERACTION ############
####################  Weak Priors Model  ##########################
####################################################################


Model3.e.int <- "
 data {

   // Covariates
   int  N;
   int <lower=0, upper=1> female[N];
   int <lower=0, upper=1> evermarried[N];
   int <lower=0, upper=1> urban[N];
   int <lower=0, upper=1> unemployed[N];
   int <lower=1, upper=26> fwt[N];
   int <lower=1, upper=4> wave[N];
   real age[N];
```

```
   real age2[N];
   real depndt_child[N];
   real hhsize[N];
   real lockdown[N];
   real expenditure[N];
   // outcome
   real <lower=0, upper=1> fetime[N];
}

parameters {
real beta[9];
real kappa[4];
real kappa2[4];
}


model {

 //Priors
  beta[1] ~ normal(0.0, 1);
  beta[2] ~ normal(0.0, 1);
  beta[3] ~ normal(0.0, 1);
  beta[4] ~ normal(0.0, 1);
  beta[5] ~ normal(0.0, 1);
  beta[6] ~ normal(0.0, 1);
  beta[7] ~ normal(0.0, 1);
  beta[8] ~ normal(0.0, 1);
  beta[9] ~ normal(0.0, 1);

  kappa ~ normal(0.0, 1);
  kappa2 ~ normal(0.0, 1);

//ITUS
  for (i in 1:17723) {
  1 ~ bernoulli(pow(inv_logit(beta[1]*female[i] + beta[2]*age[i] +
beta[3]*age2[i]  + beta[4]*evermarried[i] + beta[5]*hhsize[i] +
beta[6]*depndt_child[i] + beta[7]*urban[i] + beta[8]*expenditure[i] +
beta[9]*lockdown[i] + kappa[wave[i]] + kappa2[wave[i]]*female[i]),
fetime[i]*fwt[i])*
                 pow(1-inv_logit(beta[1]*female[i] + beta[2]*age[i] +
beta[3]*age2[i]  + beta[4]*evermarried[i] + beta[5]*hhsize[i] +
beta[6]*depndt_child[i] + beta[7]*urban[i] + beta[8]*expenditure[i] +
beta[9]*lockdown[i] + kappa[wave[i]]+ kappa2[wave[i]]*female[i]), (1-
fetime[i])*fwt[i]));
}

 //CMIE
  for (i in 17724:N) {
  1 ~ bernoulli(pow(inv_logit(beta[1]*female[i] + beta[2]*age[i] +
beta[3]*age2[i]  + beta[4]*evermarried[i] + beta[5]*hhsize[i] +
beta[6]*depndt_child[i] + beta[7]*urban[i] + beta[8]*expenditure[i]+
beta[9]*lockdown[i] + kappa[wave[i]]+ kappa2[wave[i]]*female[i]),
fetime[i]*fwt[i])*
                 pow(1-inv_logit(beta[1]*female[i] + beta[2]*age[i] +
beta[3]*age2[i]  + beta[4]*evermarried[i] + beta[5]*hhsize[i] +
beta[6]*depndt_child[i] + beta[7]*urban[i] + beta[8]*expenditure[i]+
```

```
beta[9]*lockdown[i] + kappa[wave[i]]+ kappa2[wave[i]]*female[i]), (1-
fetime[i])*fwt[i]));
}
}
"
init <- function() list(beta=rep(0.1, 9), kappa=rep(0.1, 4),
kappa2=rep(0.1, 4))

tic()
set.seed(4838282)
fit3.e.int <- stan(model_code = Model3.e.int, data = line.data,
init=init, iter = 800, warmup=300, chains = 2, cores=2, seed=4838282)
toc()

saveRDS(fit3.e.int, "fit3.e.int.rds")
print(fit3.e.int, pars = c("beta", "kappa",  "kappa2"))

```
```

```
####################################################################
################ ITUS & CMIE - Fhtime INTERACTION #################
####################################################################

…  …  …  Etc. through the four equations.  We also ran through five
samples:  ITUS only, CMIE only, CMIE with sex-wave interactions,
combined data, and lastly combined data with sex-wave interactions.
```

Pseudo-Code of the Combined-Data Regression Estimation with MCMC (Shown above)

For running an MCMC estimate, four steps are followed, and the model code follows Stan norms.

1 Bring the data into R objects and prepare a list of the data as a series of standardised variables, each being a named vector.
2 Create a model object, which is a command consisting of a series of chunks in a list. The line separator is ;.
3 State the desired parameters, taking the hyperpriors for granted as they will be set by Stan using its defaults, along with the model structure and its likelihood function. Express this function twice, once for dataset A and once for dataset B, with N running first through the cases of A and then through the cases of B.
4 Run r package Stan using the desired number of iterations, with the appropriate chains and/or machine cores to optimise speed whilst retaining convergence. The number of iterations shown here is 1,000 after dropping the warmup of 300 and noting there are two chains, each with 500 cases (which is 800-300).

S A M P L E   4: Predictions

Pseudo-Code of the Predictions Routine
For predicting the now-cast from the combined data, nine steps can be followed. The pseudocode must be adapted to your environment.

1 Clear the computer's memory a bit and run RStudio as administrator, making sure to install and load GMCM which can carry out inverse logit calculations for predicting out of a logit model.
```

(If you have a memory shortage, e.g. 8 Gb not 16 Gb of memory, then using Task Manager, turn off competing memory-resident programmes.)

2 Read the saved results from the 'Strong Priors' or 'Weak Priors' logit regression.

3 Create empty vectors with the results column names as headings, one set for Beta-hat, one set for Kappa-hat, and one set for Kappa2-hat, then link them in a list so they make an empty data frame. (cbind) Our predictions for 'strong priors' had 9 elements in Beta-hat and 4 each in the other two.

4 Simulate 1,000 resamples of the 17,723 rows of combined data X under the above headings, shaped as vectors.

5 Create additional variables in this data frame to reflect wave and the sex-wave interaction term.

The dimensions of this data frame are 17,723,000 x 17 columns.

6 Open the raw data, and create another data frame of dimension (N*1,000 * 17 ) by resampling the raw data 1,000 times.

7 We aim for every case to have 1,000 predictions, an arbitrarily chosen number sufficient to display its uncertainty. The calculation has these dimensions:

$$\begin{bmatrix} N*1{,}000 * K \end{bmatrix} \begin{bmatrix} K \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \text{ N*1,000 predictions}$$

The corresponding calculation can also be seen as: $\hat{B}X = \hat{Y}$ for 1,000 simulated samples of **X** and **Y**.

Where N is sample size and K is the number of slope parameters, 17 in this case.

For flogits, take the inverse logit of (XB) where (XB) = xBetahat + Kappa*t + Kappa2*sex*t ,

where t = wave. We used wave = 1, 2, 3, 4.

8 Aggregate by sex within each wave. Future research can aggregate by social groups once these variables are included in the regression equation. If a variable is supplementary (ie not in the regression equation), alter the equations sufficiently to create predictions for the factor, based on its mean **X** values for sub-groups, where **X** is the variables in the equation.

9 Graph the 90%, 80%, 50%, 20% and 10% or other fan diagram levels of uncertainty.

In conclusion, these prediction diagrams will tend to show more uncertainty for sub-groups where there were more zeroes in the data. However the fan will not exceed 1 or become negative because the best methods were used in both the estimation and prediction stages.

Combining Time-use Diary Data…, V1 Appendix. Cite as: Wendy Olsen, Manasi Bera, Greg Dropkin, Amaresh Dubey, Jihye Kim, Ioana Macoveciuc, Samantha Watson, and Arkadiusz Wiśniowski (2021), Combining CMIE-ITUS Data With Bayesian Estimation, https://github.com/WendyOlsen/combining-cmie-itus-data/, Creative commons attribution.

Full Model Results [Embedded Excel File]
Column 1 is ITUS data only. Column 2 is CMIE data only, without sex-wave interactions. Column 3 is CMIE with sex-wave interactions. Column 4 is combined data without sex-wave interactions. Column 5 is the best model (shown here as Model 1), with combined data and sex-wave interactions. In the rows, the first row is work time WOHVS, second row is HVS only, third row is these two summed together, it total work time WOHVS+HVS as the dependent variable. The last row is the logit model for unemployment.

| ITUS [N=17723] | | | | | | |
|---|---|---|---|---|---|---|
| Model 1.e ( | mean | se_mean | sd | 2.50% | 25% | 50% |
| **female** | **-1.46** | **0.00** | **0.05** | **-1.55** | **-1.49** | **-1.4** |
| **age** | **0.17** | **0.00** | **0.01** | **0.14** | **0.16** | **0.1** |
| **age2** | **-0.03** | **0.00** | **0.00** | **-0.03** | **-0.03** | **-0.0** |
| **evermarrie** | **0.17** | **0.00** | **0.08** | **0.02** | **0.12** | **0.1** |
| hhsize | -0.01 | 0.00 | 0.01 | -0.03 | -0.01 | -0.0 |
| depndt_chi | -0.20 | 0.01 | 0.14 | -0.46 | -0.30 | -0.2 |
| urban | -0.02 | 0.00 | 0.05 | -0.11 | -0.05 | -0.0 |
| expenditur | -0.02 | 0.00 | 0.02 | -0.07 | -0.03 | -0.0 |
| **wave1** | **1.52** | **0.01** | **0.25** | **1.05** | **1.35** | **1.5** |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

## Predicted Results by Sex (Model 1 at Top; Model 2 at Bottom)

Predicted Fetime [N=41123]

| female | wave | mean | sd | n | lower2.5 | lower25 | mid | upper75 | upper97.5 |
|---|---|---|---|---|---|---|---|---|---|
| Male | 1 | 0.227 | 0.093 | 12098 | 0.065 | 0.143 | 0.245 | 0.306 | 0.359 |
| Male | 2 | 0.203 | 0.089 | 3149 | 0.060 | 0.113 | 0.222 | 0.288 | 0.320 |
| Male | 3 | 0.185 | 0.082 | 3149 | 0.053 | 0.101 | 0.201 | 0.263 | 0.294 |
| Male | 4 | 0.224 | 0.096 | 3149 | 0.068 | 0.128 | 0.246 | 0.316 | 0.349 |
| Female | 1 | 0.056 | 0.024 | 11475 | 0.013 | 0.037 | 0.060 | 0.076 | 0.093 |
| Female | 2 | 0.021 | 0.009 | 2701 | 0.004 | 0.014 | 0.023 | 0.029 | 0.036 |
| Female | 3 | 0.018 | 0.008 | 2701 | 0.004 | 0.012 | 0.019 | 0.024 | 0.030 |
| Female | 4 | 0.021 | 0.009 | 2701 | 0.004 | 0.014 | 0.023 | 0.028 | 0.035 |

Observed Fetime [N=41123]

| female | wave | weighted.mean | sd | n | lower2.5 | lower25 | mid | upper75 | upper97.5 |
|---|---|---|---|---|---|---|---|---|---|
| Male | 1 | 0.227 | 0.159 | 12098 | 0.000 | 0.000 | 0.292 | 0.333 | 0.458 |
| Male | 2 | 0.204 | 0.169 | 3149 | 0.000 | 0.000 | 0.333 | 0.333 | 0.417 |
| Male | 3 | 0.186 | 0.164 | 3149 | 0.000 | 0.000 | 0.250 | 0.333 | 0.417 |
| Male | 4 | 0.226 | 0.165 | 3149 | 0.000 | 0.000 | 0.333 | 0.333 | 0.417 |
| Female | 1 | 0.056 | 0.104 | 11475 | 0.000 | 0.000 | 0.000 | 0.042 | 0.333 |
| Female | 2 | 0.021 | 0.073 | 2701 | 0.000 | 0.000 | 0.000 | 0.000 | 0.333 |
| Female | 3 | 0.018 | 0.066 | 2701 | 0.000 | 0.000 | 0.000 | 0.000 | 0.292 |
| Female | 4 | 0.021 | 0.071 | 2701 | 0.000 | 0.000 | 0.000 | 0.000 | 0.333 |

Predicted Fhtime [N=41123]

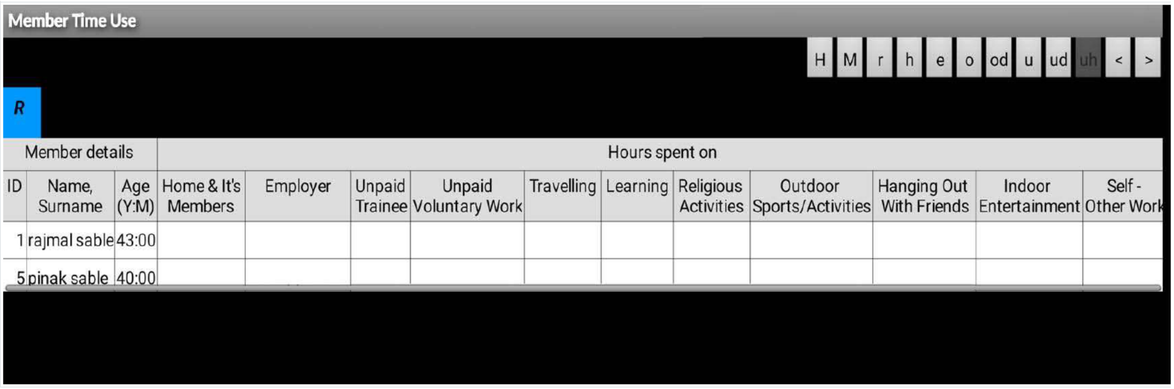
| female | wave | mean | sd | n | lower2.5 | lower25 | mid | upper75 | upper97.5 |
|---|---|---|---|---|---|---|---|---|---|
| Male | 1 | 0.187 | 0.040 | 12098 | 0.110 | 0.153 | 0.197 | 0.221 | 0.243 |
| Male | 2 | 0.091 | 0.022 | 3149 | 0.052 | 0.070 | 0.095 | 0.110 | 0.123 |
| Male | 3 | 0.094 | 0.023 | 3149 | 0.053 | 0.072 | 0.098 | 0.114 | 0.128 |
| Male | 4 | 0.086 | 0.021 | 3149 | 0.049 | 0.066 | 0.090 | 0.105 | 0.118 |
| Female | 1 | 0.101 | 0.022 | 11475 | 0.055 | 0.087 | 0.106 | 0.118 | 0.131 |
| Female | 2 | 0.220 | 0.043 | 2701 | 0.127 | 0.194 | 0.235 | 0.252 | 0.271 |
| Female | 3 | 0.211 | 0.041 | 2701 | 0.121 | 0.186 | 0.225 | 0.242 | 0.261 |
| Female | 4 | 0.209 | 0.041 | 2701 | 0.120 | 0.184 | 0.223 | 0.240 | 0.258 |

Observied Fhtime [N=41123]

| female | wave | weighted.mean | sd | n | lower2.5 | lower25 | mid | upper75 | upper97.5 |
|---|---|---|---|---|---|---|---|---|---|
| Male | 1 | 0.044 | 0.062 | 12098 | 0.000 | 0.000 | 0.021 | 0.083 | 0.208 |
| Male | 2 | 0.090 | 0.070 | 3149 | 0.000 | 0.042 | 0.083 | 0.104 | 0.292 |
| Male | 3 | 0.094 | 0.081 | 3149 | 0.000 | 0.042 | 0.083 | 0.125 | 0.333 |
| Male | 4 | 0.086 | 0.070 | 3149 | 0.000 | 0.042 | 0.083 | 0.104 | 0.250 |
| Female | 1 | 0.250 | 0.139 | 11475 | 0.000 | 0.146 | 0.250 | 0.333 | 0.521 |
| Female | 2 | 0.220 | 0.110 | 2701 | 0.021 | 0.125 | 0.229 | 0.292 | 0.417 |
| Female | 3 | 0.211 | 0.109 | 2701 | 0.021 | 0.125 | 0.208 | 0.292 | 0.417 |
| Female | 4 | 0.209 | 0.112 | 2701 | 0.021 | 0.125 | 0.208 | 0.271 | 0.458 |

## Time-Use Measurement in the Centre for Monitoring Indian Economy Data

### Figure A*: Time-Use Grid in CMIE Data 2019-2020


Member Time Use

H M r h e o od u ud uh < >

R

| Member details | | | Hours spent on | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | Name, Surname | Age (Y:M) | Home & It's Members | Employer | Unpaid Trainee | Unpaid Voluntary Work | Travelling | Learning | Religious Activities | Outdoor Sports/Activities | Hanging Out With Friends | Indoor Entertainment | Self - Other Work |
| 1 | rajmal sable | 43:00 | | | | | | | | | | | |
| 5 | pinak sable | 40:00 | | | | | | | | | | | |

# Description:

This screen captures the distribution of the 24 hours of a day into various activities by the members of the household. The activities accounted for here are - work for home and its members; work for employer; work as an unpaid trainee; work as an unpaid voluntary worker; time spent in travelling; time spent in learning; time spent in religious activities; time spent in outdoor sports activities; time spent hanging out with friends, time spent on indoor entertainment; and time spent on self.

The key "members'" time-use data-collection grid in CMIE has one row per person, and the household's key respondent tells what they do in a 24-hour period. No clear instructions are written about the recall period ('usual', 'yesterday' not stated). As shown, the 11 categories are filled in to sum to 24 hours. This grid is completed once per Wave, ie every four months. The grid allows no record of payments for each work stint, and makes no record of simultaneous activities.