Keenan Flynn & Txomin Chivite
Group 3

Economic Factors Affecting the Quantity of Sleep

1. Introduction

Humans spend a significant part of their life sleeping. This single activity accounts for roughly one-third of daily life. It is widely understood that quality and quantity of sleep are key indicators to personal health and task performance (Pilcher and Morris, 2020). Understanding how socioeconomic factors impact sleep can help scientists and researchers draw conclusions about human fitness and labor force productivity.

We propose that economic factors (age, income, hours worked, and full-time status) have some impact on the quantity of sleep reported. In this project, we will examine a broad-sweeping sample from the United States to be able to draw conclusions about population. We hope to provide awareness of the disparities of sleep on a socioeconomic level, and contribute to the collective understanding of broad factors that impact sleep.

Several previous studies have examined the relationship between income and sleep and determined that some correlation exists. Since these studies were published, several social and technological developments have occurred that could feasibly impact this relationship. Some examples of these factors include the widespread use of cellphones during all stages of daily activity (Rafique et al. 2020), a surging prevalence of telecommuting, (Costa et al. 2022), or questions on job security during a volatile post-pandemic market (Baird et al. 2022). It is not our goal to find the specific causation of the sleep-socioeconomic relationship changing over time, but rather to identify if the relationship has changed at all when compared to previous studies. The confirmation in variance between past versus modern sleep data could be a starting point for future research to examine more granular causal factors.

2. Literature Review

Sleep is widely assumed to hold a positive relationship with health, happiness, and socioeconomic status (SleepScore Labs 2022). A select number of studies have examined these relationships in closer detail. In this project we are chiefly concerned about the relationship between sleep and labor variables such as income and time spent working. We have chosen several papers to discuss below so as to: 1) Familiarize the reader with the relationship of sleep and health. 2) Examine how sleep and health impact productivity. 3) Discuss the mechanism by which productivity increases wages. 4) Assert the need for a study on sleep and income to be re-visited with modern data.

Sleep is well regarded as a key part of the daily cycle of humans. This cycle, called the circadian cycle, has been studied and found to be a significant factor in the overall health of an individual. (Gibson and Schrader 2014) Several of the papers examined in this project claim that long term sleep loss shortens lifespan by raising the mortality rate. (Gibson and Schrader 2014, Moore et al. 2002) Sleep also impacts the ability of memory to recall particular events, cardiovascular disease, and fatigue. (Gibson and Schrader 2014, Moore et al. 2002) One study examined how quantity of sleep impacts brain function performance. (Van Dongen et al. 2003) This study divided patients into groups and restricted their sleep to 4, 6, or 8 hours. Patients were given psycho-motor vigilance tests and groups who were getting 4 or 6 hours of sleep performed markedly worse than the group getting 8 hours even though their self-assessment on performance did not decrease. (Van Dongen et al. 2003) This gives us strong reason to believe that a lack of sleep negatively impacts cognitive performance.

Moore et al uncover additional information regarding the effects of sleep, income, and education on health. They find that sleep can mediate the relationship between income and physical health, although to a modest degree. The modest relationship “is not particularly surprising, given the myriad of factors influencing SES (socioeconomic status), sleep, and health.” (Moore et al 2002). They also go on to say that higher income is associated with better sleep quality and that sleep quality and sleep quantity share a positive relationship.

Socioeconomic status is shown to impact children of workers as well. A study examined sleep patterns of middle schoolers and showed that parent demographics (income, education, home environment) had some effect on the child's sleep (Marco et al. 2011). In particular, students with lower SES status tended to have less regular sleep schedules. Timing and consistency of school night sleep were associated with the parent demographics (Marco et al. 2011). As we have previously stated, sleep is correlated to cognitive function. If students are not able to function during school, the total level of education decreases. Moore et al (2011) demonstrate that education is correlated with income, and so a possibility for a negative feedback loop arises.

For most humans, the largest use of time is sleep. It would be intuitive to think if one spent less time sleeping, they could spend more time being productive. Biddle and Hammerash's (1990) conclusions seem to agree with this assumption to a degree, stating that an individual is given a choice on the allotment of time he spends sleeping. If his wages are high, then his opportunity cost of sleep is also high so he will choose to sleep less (although because sleep is related to productivity, there is a point where the cost diminishes) (Biddle and Hammerash 1990). Gibson and Schrader consider the inverse function, holding income as the predicted variable. They go on to state that sleep may be correlated to unmeasurable activities that produce higher wages, such as mental or physical ability. Their initial findings agree with Biddle and Hammerash (negative relationship between income and sleep), but they go a step farther by considering sunset time. Sunset time is important because the time at which humans go to sleep is correlated with the sunset. The time they wake up is not correlated with sunrise, but rather the fixed time that they are required to go to work. By accounting for sunset time, they find that an extra hour of sleep accounts for a positive relationship and roughly 16% increase of wages. We have chosen to not consider sunset time in this project.

Prior to 1990, scientists widely held sleep to be a fixed variable under an umbrella of 'personal need and care'. The study by Biddle and Hammerash was novel in that it posed that sleep was dependent on income. In our project we aim to complete a modern variation of this study. We believe that recent developments of technology (cell phones, TVs, etc) could have some adverse effect on sleep that is not accounted for by Biddle and Hammerash. We do not seek to identify the specific cause for a deviance in the relationship between sleep and income, but rather test whether a new relationship exists. The causal factors of this relationship changing over time could be motivation for future studies as previously stated.

3. Methodology

The purpose of this study is to identify the important income influences or factors that alter different time-use activities. This will help us observe if equity issues are related to activity time-use. In particular, we attempt to elucidate the impact of economic factors on the quantity of sleep. Minutes of sleep reported per week is a continuous variable, and therefore is modeled using multiple regression models. In this characteristic type of model, the relationship between a response variable, in this case sleep, and a series of explanatory variables is estimated (Tranmer et al., 2020). There are five key assumptions that must be met for this type of modeling to be valid.

The first assumption requires that the response variable the continuous, which applies in this case as sleep, it is a continuous variable. The second point is that the relationship between the response variable and the predictor variable must be linear. The third is that the residual of the regression follows a normal distribution. The fourth is that the error terms must have constant variance and not depend on any other variable in the model, a trait referred to as homoscedasticity. Finally, there must not be multicollinearity in the model where predictor variables are strongly correlated with each other, which would skew the reported beta values in the model (Tranmer et al., 2020).

Two multiple regression models are implemented in this specific study to identify the relationship between independent variables and a unique dependent variable.

<u>Model #1:</u>

Independent Variables: Age, Full Time, Weekly Earnings, Weekly Hours Worked
Dependent variables: Sleep

<u>Model #2:</u>

Independent Variables: Age, Part Time, Weekly Earnings, Weekly Hours Worked
Dependent variables: Sleep

By comparing two multiple regression models, we can evaluate the relative performance of the model and dictate if a stronger relationship exists. The logical representation of a multiple regression model takes the form as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i;\ i = 1,2,\ldots,k$$

where,

Y_i = is the i^{th} observation in the Dependent Variable

X_{ki} =Number of Observations (Independent Variables)

k = Regressors

u_i = Error term

The average relationship between Y and the regressors is given by the population regression line:

$$E(Y_i | X_{1i} = x_1, X_{2i} = x_2, X_{3i} = x_3, \ldots, X_{ki} = x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_i$$

where,

β_0 = Intercept (the predicted value of Y when Xs equal to 0)

$\beta_j,\ j = 1, 2, \ldots, k$ are the coefficients on $X_j,\ j = 1, 2, \ldots, k$.

- β_1 calculates the supposed change in Y_i caused by one unit change in X_{1i} while holding all other regressors constant.

The numbered beta x terms are repeated for the number of predictor variables that are included in the model. Beta 0 is the regression constant, otherwise interpreted as the intercept, and in this context is the number of hours of sleep if all predictor variables are zero.

One predictor used in this model is whether the participant is employed on a full time or a part time basis. This is a binary variable and can be included in the multiple regression model. However, there may be a different impact of other variables on the amount of sleep based on employment status. Therefore, two different multiple linear regression models will be developed, one for full time employees and one for part time employees. These will be compared to see if there is a significant difference between the two models.

The multiple linear regression model permits investigation into the relationship between the dependent variable, defined by a linear combination of the regressors, and the numbers of observations or independent variables. In this research we considered variables such as (1) age, (2) level of education, (3) labor force status, (4) weekly earnings, (5) number of household children, (6) total hours worked in a week, (7) time spent in different time-use activities. Ultimately several of these variables were cut, our final models only consider the regressors of age, earnings, hours worked, and full time status.

4. Data and Samples

Our dataset consists of information gathered by the United States Census Bureau and is publicly available through the United States Bureau of Labor Statistics. The specific survey that we are using is called the American Time Use Survey (ATUS). The Census Bureau defines ATUS as a survey that “measures the amount of time people spend doing various activities, such as paid work, childcare, volunteering, and socializing” (US Census Bureau 2022). The survey is conducted on a yearly basis with the goal being to discover how people allocate their time. The Census Bureau describes that measuring time as a resource helps economists and social scientists measure the total economic output of United States citizens. Each record in the survey corresponds to a unique respondent, which is identified by a Case ID.

The particular iteration of the dataset we will be using is the 2021 ATUS Summary File. This file contains data for the total number of minutes that each respondent spent doing a particular activity. Examples of activities include time spent sleeping, eating, washing, grooming, watching TV, reading, and more. Each activity field is indicated by a 6 digit numerical code. For simplicity, we have translated the codes into more common nomenclature in the table below.

In addition to activity fields, the Summary File contains selected variables from the Current Population Survey (CPS). This survey is defined as “a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics and provides a comprehensive body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics” (U.S. Bureau of Labor Statistics). Selected variables from the CPS that are included in the ATUS include demographic and employment information such as gender, age, sex, number of children, income, hours worked, and others. Each of the selected variables from the CPS is given an encoded name but has also been translated into common nomenclature below.

4.1 Data Preparation

To be able to apply the dataset for our research purpose, the data was first prepared in a preprocessing manner. The raw ATUS dataset contains 9087 records and 400 columns. Many of these columns are registered as optional. Additionally, many of the time use fields contain zeros because there are some activities which people never do. These factors create sparsity in the dataset. To be able to have enough data for a significant econometric model, we hand selected fields based on whether they had a significant number (> 1000) of non-null and non-null records. This eliminated a large number of time use fields but retained most of the demographics fields. After eliminating columns that did not meet our required record count, we were left with 1838 records. This is a significant amount of records and will give us enough data to create a logistic model.

4.2 Summary Statistics

The variables retained in the dataset that will be used for modeling are sleep, age, weekly earnings, weekly hours worked, and employment status (full time or part time). These variables are summarized in Table 1.

In this study, only approximately 84.65% of the employed participants are full time and half are part time. These groups will be separated in the model creation process to see if there is a difference in how the other variables predict sleep based on employment status because of the different sample sizes between the groups. The variable of interest, sleep, has a large range from 400 minutes to 1420 minutes, so there is space to create meaningful predictions within this range.

Table 1: Summary Statistics of Participant Attributes

	Mean/Percent	Median	Standard Deviation	Maximum	Minimum
Sleep	526.0	510.0	123.0	1420.0	400.0
Age	39.6	40.0	9.78	80	15
Weekly Earnings	$1,280	$1,061	$821	$2,885	0
Weekly Hours Worked	40.7	40.0	11.2	110	1
Full-Time	84.65%				

5. Results and Discussions

Multiple regression model is a statistical technique that is helpful in a variety of circumstances, such as when comprehending how several causal factors are connected to one another or how they influence a certain conclusion. In this specific scenario, we introduced 2 multiple regression models, where the models contain 4 independent variables (differing by one variable) and one dependent variable.

The model was developed to elucidate the predictive power of various work and economic related variables on the amount of sleep reported per night. Variables were selected because of their potential impact on sleep. After dividing the dataset up into two groups for employment status, the final model contains 3 predictor variables: age, weekly earnings, and weekly hours worked. In the part time regression model, only weekly earnings had a statistically significant coefficient at the 0.05 level of significance. However, in the full time regression model, all predictor variables meet the level of significance of 0.05. This suggests a different impact of economic variables on sleep between the two employment status groups. Table 2 summarizes the results of the multiple regression models. The R-squared of the part time employed model was 1.59% and the R-squared of the full time employed model was 3.56%.

Table 2: Parameter Estimation Results of Multiple Logistic Regression

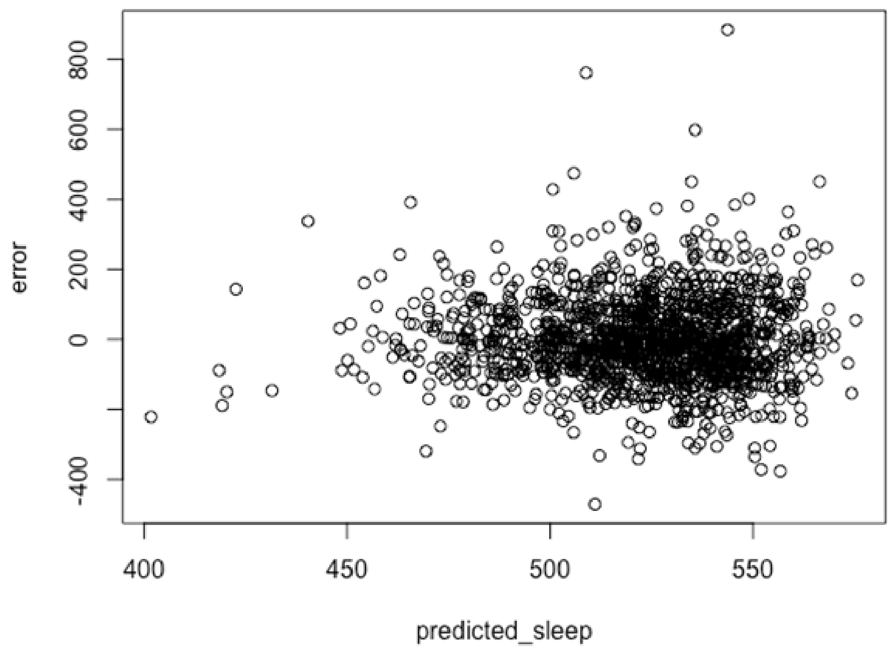
	Part Time Employees			Full Time Employees		
	B	**Std. Error**	**P-Value**	**B**	**Std. Error**	**P-Value**
Intercept	570	25.63	0.000	668	21.29	0.000
Age	-0.369	0.551	0.504	-1.111	0.360	0.002
Weekly Earnings	-0.039	0.017	0.018	-0.013	0.004	0.001
Weekly Hours Worked	0.121	0.889	0.892	-1.862	0.395	< 0.001

For part time employees, only weekly earnings was statistically significantly predictive of sleep among these variables. As weekly earnings increased, the amount of sleep decreased. For full time employees, there was a similar trend with decreasing sleep as earnings increased. This group also experienced decreasing sleep with increasing age and decreasing sleep with increasing weekly hours worked. This observation could provide evidence for a hypothesis that states higher work leads to higher stress, which is correlated to decreased sleep. This aligns with several of the studies that we previously reviewed.

5.1 Linearity

The plot of error versus the predicted value for the full time model shows a linear trend (Figure 1). Most of the data is clustered around longer sleep times, but there is no clear curve or other trend line, indicating that the linearity requirement is met.

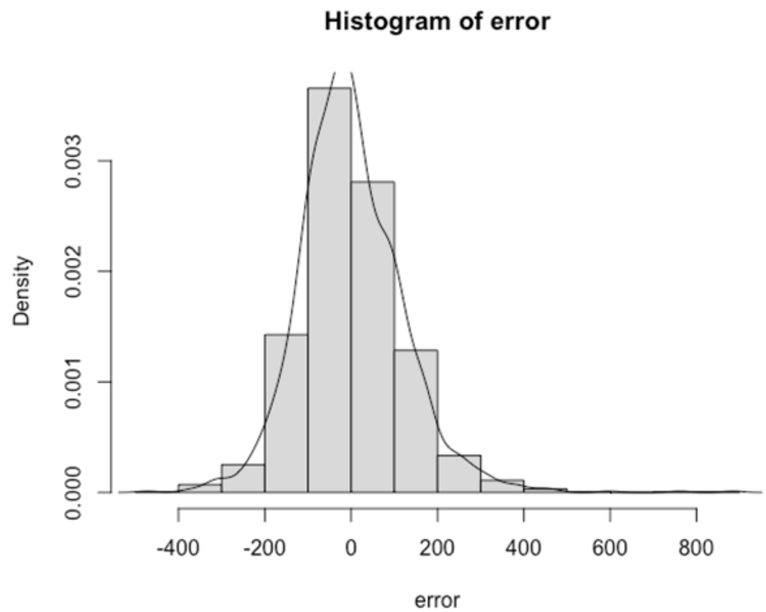
Figure 1: Predicted Sleep versus Error for the Full Time Model



5.2 Normal Residuals

The histogram of model error shows a general normal distribution (Figure 2). There is no clear skew to the data and the distribution about the center of zero is approximately symmetric.

Figure 2: Histogram of Full Time Model Error



5.3 Homoskedasticity

The null hypothesis of this test is that there is homoscedasticity The results of the studentized Breusch-Pagan test indicate that this assumption is not met, BP(3) = 16.292, $p < 0.001$. This indicates that in this model, the variance of the residual is not constant and that the error term changes as the predictor variable changes. The model's predictive strength, therefore, changes throughout the range of the predictor variables.

5.4 Multicollinearity

To determine multicollinearity, the variable inflation factor for each predictor was calculated. This metric determines the strength of the correlation between the independent variables and starts with a value of 1. Values above 10 are considered to represent high multicollinearity and therefore invalidates the model. The VIF for age (1.06), weekly earnings (1.10), and hours worked (1.05) are all well below this threshold, and therefore the model meets this assumption.

6. Conclusion

Sleep is an important factor of healthy living and one of the fundamental metrics for tracking sleep is hours of sleep obtained per week. Understanding this variable, as well as the factors that impact the amount of sleep can be impactful in improving health outcomes, overall happiness, and even socioeconomic status based on previous studies cited above. This study found that, particularly for full time workers, age, weekly earnings, and weekly hours worked had negative impacts on the amount of sleep reported. From a policy perspective, this can inform regulations on the maximum number of hours an employee can be asked to work each week, or drive future research into developing guidance on how to improve sleep for those who work large numbers of hours.

Ultimately, this study has demonstrated that economic factors are not the main driving force to the relationship between income, physical and mental health. However, they can mediate the relationship up to a modest degree given the myriad of factors influencing SES (socioeconomic status), sleep and health. However, This modest relationship does indicate the influence of the sleep variable, but ultimately, it does not show a strong correlation with age, (full–time, part-time), weekly earnings and weekly hours which is demonstrated with our results of R-squared (part time employed model was 1.59% and full time employed model was 3.56%).

Finally, this study analyzed the survey responses from ATUS (American Time Use Survey) which provided enough data to understand the premises of this study. Even though in western societies it is well known the idea that quality and quantity of sleep are key indicators to personal health and task performance, this equation does not translate the other way around. This study concludes that there is not a clear cause that indicates that socioeconomic factors impact the amount of sleep, however, there exists a small correlation but not strong enough to push a clean and establish hypothesis.

Contributions

Keenan Flynn: Manuscript preparation (Introduction, Literature Review, Data and Samples), Data pre-processing, Exploratory Data Analysis, Model and Study Conception, Proofread and File Edits.

Txomin Chivite: Manuscript preparation (Methodology, Summary Statistics, Results, Conclusion), Data processing, Model refinement, Analysis and Interpretations of Results

References

Baird, M. D., Dubowitz, T., Cantor, J., & Troxel, W. M. (2022). Examining the impact of employment status on sleep quality during the COVID-19 pandemic in two low-income neighborhoods in Pittsburgh, PA. Sleep, 45(3), zsab303. https://doi.org/10.1093/sleep/zsab303

Biddle, J. E., & Hamermesh, D. S. (1990). Sleep and the Allocation of Time. *Journal of Political Economy*, *98*(5), 922–943. http://www.jstor.org/stable/2937618

Costa, C., Teodoro, M., Mento, C., Giambò, F., Vitello, C., Italia, S., & Fenga, C. (2022). Work Performance, Mood and Sleep Alterations in Home Office Workers during the COVID-19 Pandemic. International Journal of Environmental Research and Public Health, 19(4), 1990. MDPI AG. Retrieved from http://dx.doi.org/10.3390/ijerph19041990

Gibson, M., & Shrader, J. (2014). Time Use and Productivity: The Wage Returns to Sleep. UC San Diego: Department of Economics, UCSD. Retrieved from https://escholarship.org/uc/item/8zp518hc

Marco, C. A., Wolfson, A. R., Sparling, M., & Azuaje, A. (2011). Family socioeconomic status and sleep patterns of young adolescents. Behavioral sleep medicine, 10(1), 70–80. https://doi.org/10.1080/15402002.2012.636298

Moore, P. J., Adler, N. E., Williams, D. R., & Jackson, J. S. (2002). Socioeconomic status and health: the role of sleep. Psychosomatic medicine, 64(2), 337–344. https://doi.org/10.1097/00006842-200203000-00018

Pilcher, J. J., & Morris, D. M. (2020). Sleep and Organizational Behavior: Implications for Workplace Productivity and Safety. Frontiers in psychology, 11, 45. https://doi.org/10.3389/fpsyg.2020.00045

Rafique, N., Al-Asoom, L. I., Alsunni, A. A., Saudagar, F. N., Almulhim, L., & Alkaltham, G. (2020). Effects of Mobile Use on Subjective Sleep Quality. Nature and science of sleep, 12, 357–364. https://doi.org/10.2147/NSS.S253375

SleepScore Labs. *How Income and Socioeconomic Status Played a Role in Sleep Health During COVID-19*, 12 Mar. 2022, https://www.sleepscore.com/blog/sleep-uncovered-part-3.

Tranmer, M., Murphy, J., Elliot, M., and Pampaka, M. (2020) Multiple Linear Regression (2nd Edition); Cathie Marsh Institute Working Paper 2020-01.
https://hummedia.manchester.ac.uk/institutes/cmist/a
rchive-publications/working-papers/2020/2020-1- multiple-linear-regression.pdf

U.S. Bureau of Labor Statistics. (n.d.). *Labor Force Statistics from the Current Population Survey*. U.S. Bureau of Labor Statistics. Retrieved December 14, 2022, from https://www.bls.gov/cps/

US Census Bureau. (2022, July 8). *American Time Use Survey (ATUS)*. Census.gov. Retrieved December 14, 2022, from https://www.census.gov/programs-surveys/atus.html

Van Dongen, H. P., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep, 26(2), 117–126. https://doi.org/10.1093/sleep/26.2.117