

Abstract

We consider the effect of economy on the amount of time Americans reported watching television on the ATUS between 2003 and 2012. Measures of economic performance include GDP, unemployment rate, and stock indices; these measures are summarized using their first two principal components. A GLM is used to test the association between economy and TV time use, with time being included smoothly. We report that watching television is associated with both PCs, and that the association changes with gender (but not with region, race or income). Then, using penalized regression, we determine that the strongest sociodemographic factors influencing television viewing are age, gender, presence of household children, and average number of hours worked in a week. The stratified sampling scheme is accounted for through regression techniques, and two separate models are used to deal with the zero-inflated nature of the data.

Main questions

1. What effect does the economy have on the amount of time spent watching TV? Does this vary by gender, race, income?
2. What are the strongest socio-demographic predictors of time spent watching TV?

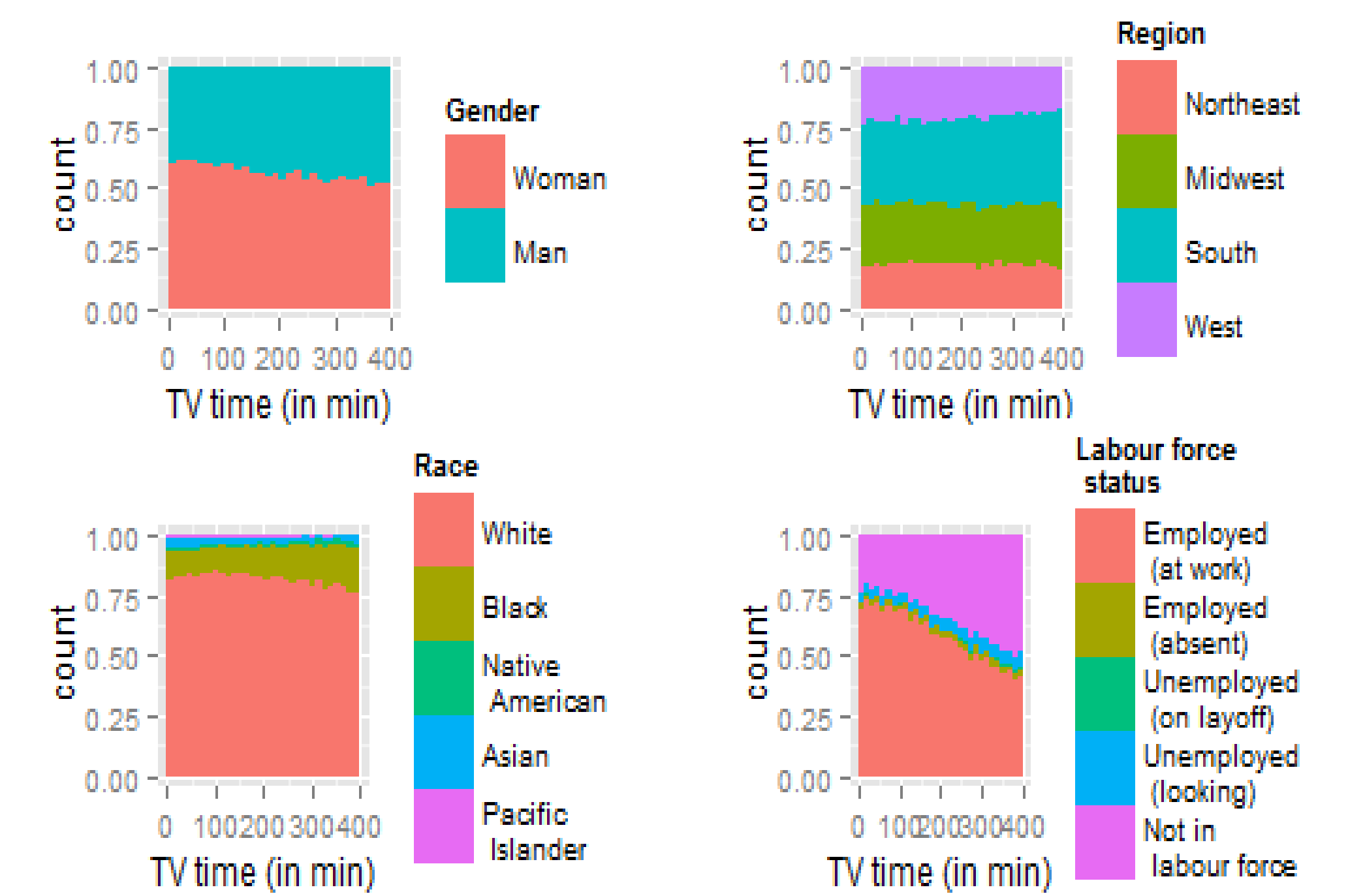
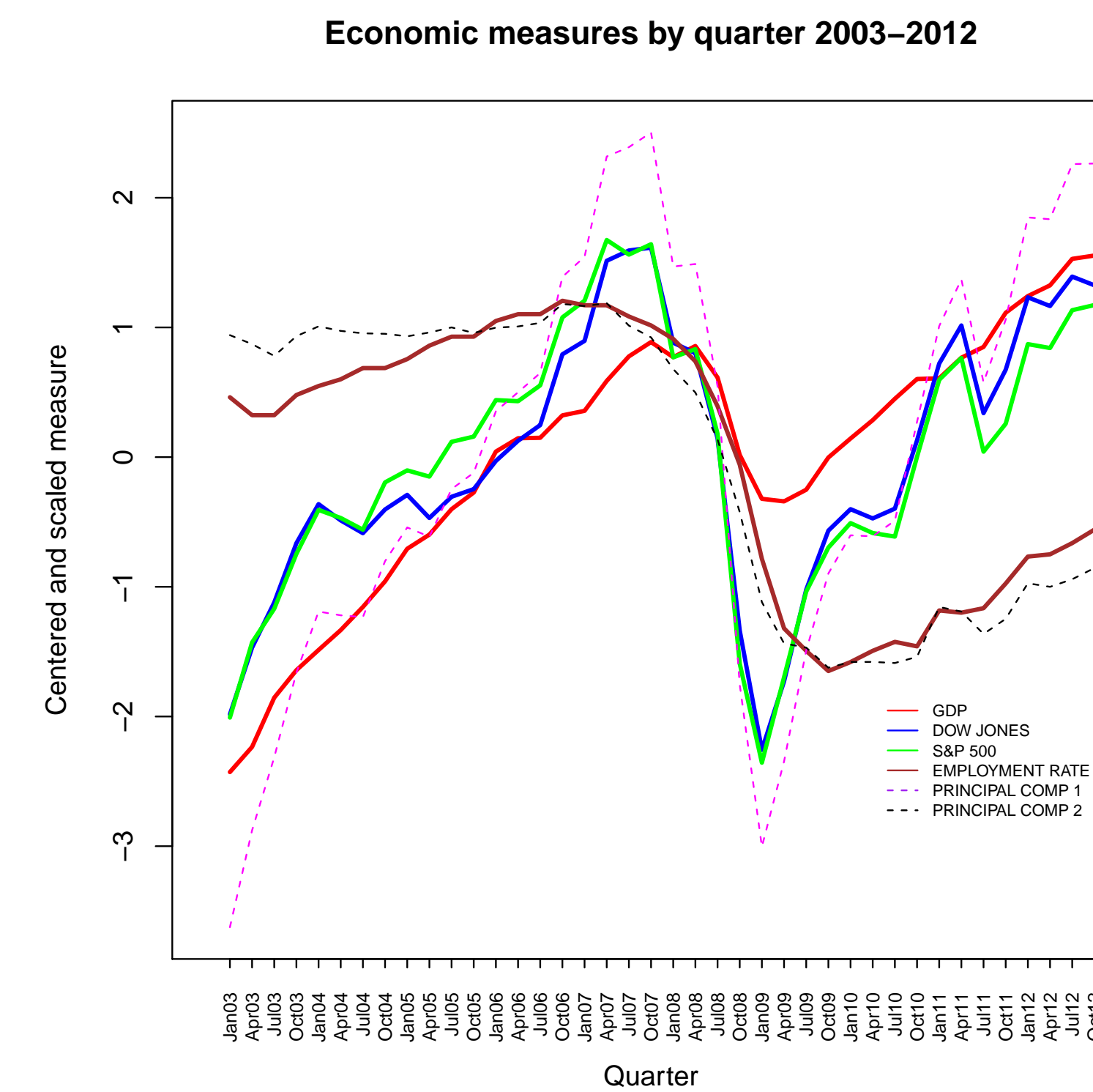
Description of the data

- ▶ Since 2003, the American Time Use Survey (ATUS) has been collecting information on how Americans spend a day, what kind of activities they perform, with whom, and for how long.
- ▶ Over a decade, some 135,000 respondents were asked to walk an interviewer through a selected day.
- ▶ The dataset itself contains socio-demographic information about the respondents (e.g. sex, race, income), as well as time-use variables (e.g. sleeping, working, watching TV)

Main challenges

- ▶ About 20% of respondents reported no TV watching; two separate models are used to address this issue
- ▶ How can economy be measured? **Note:** economy cannot have a direct, causal effect on TV watching
- ▶ Since economy varies with time, how can we separate the effect of time from the effect of economy?
- ▶ How to take into account the stratified sampling scheme?
 - ▶ Through regression (Gelman, 2007)
- ▶ How to handle missing data

Economic measures and Distribution of TV use



Reported TV time, stratified according to gender, race, region, and labour force status

1. Effect of economy

Methods

To test for the effect of economy, we use two GLMs: gamma regression and logistic regression.

$$g(\mu_i) = \beta_1 ECON1_i + \beta_2 ECON2_i + \alpha^t X_i + \lambda(t_i),$$

where X_i are the covariates accounting for the stratified sampling scheme (gender, day of the week, region, hispanicity, race) and $\lambda(t)$ is a smooth function of time (measured in months). The smoothing was done via linear extrapolation:

$$E(\lambda(t)|\lambda(t-1), \lambda(t-2), \dots, \lambda(1)) = 2\lambda(t-1) - \lambda(t-2).$$

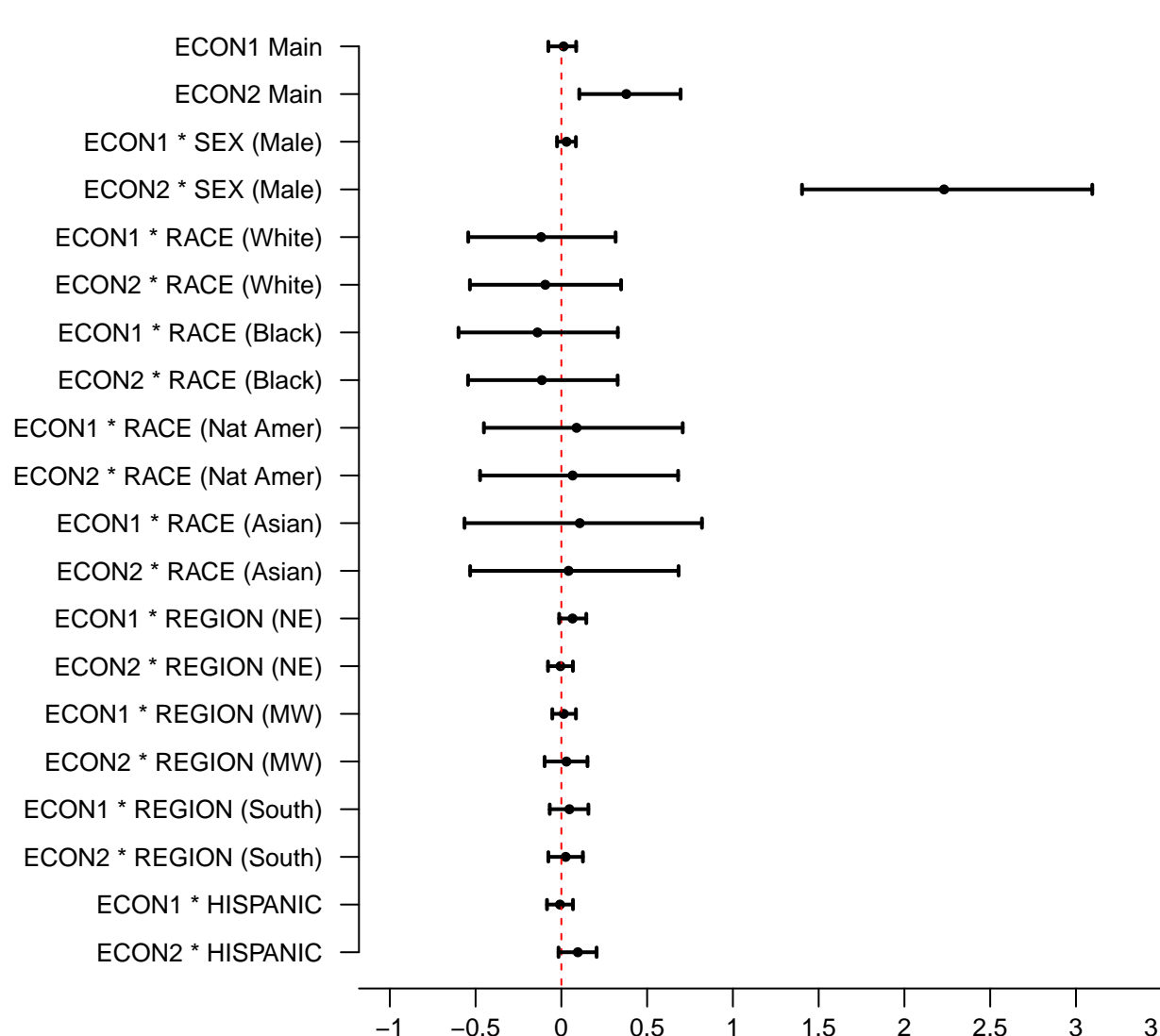
The stochastic part of the models are respectively

$$TVTIME_i \sim \text{Gamma}(\nu, \lambda_i), \quad g(x) = \log(x), \quad \text{and} \quad TVIND_i \sim \text{Bernoulli}(\pi_i), \quad g(x) = \text{logit}(x).$$

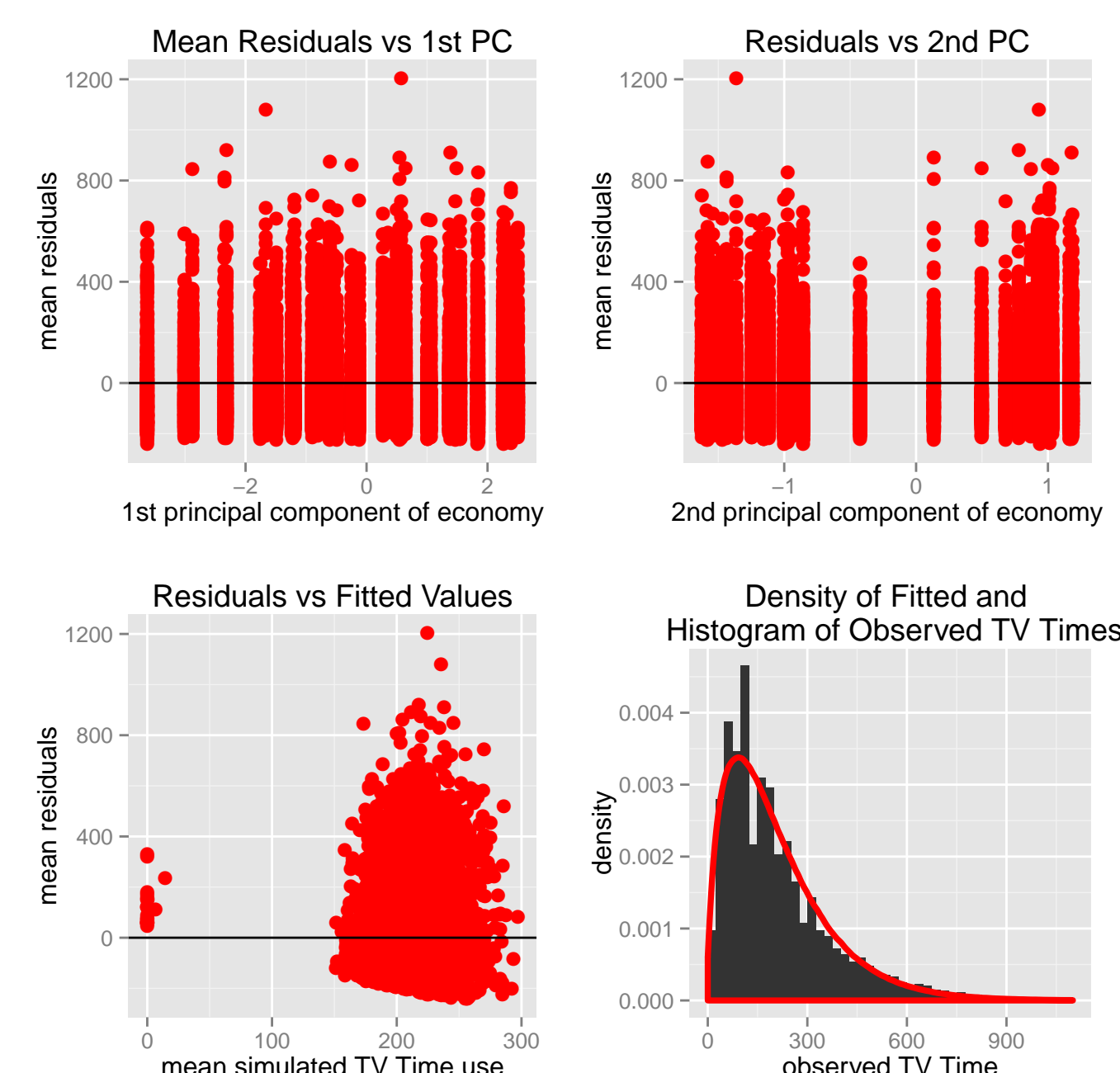
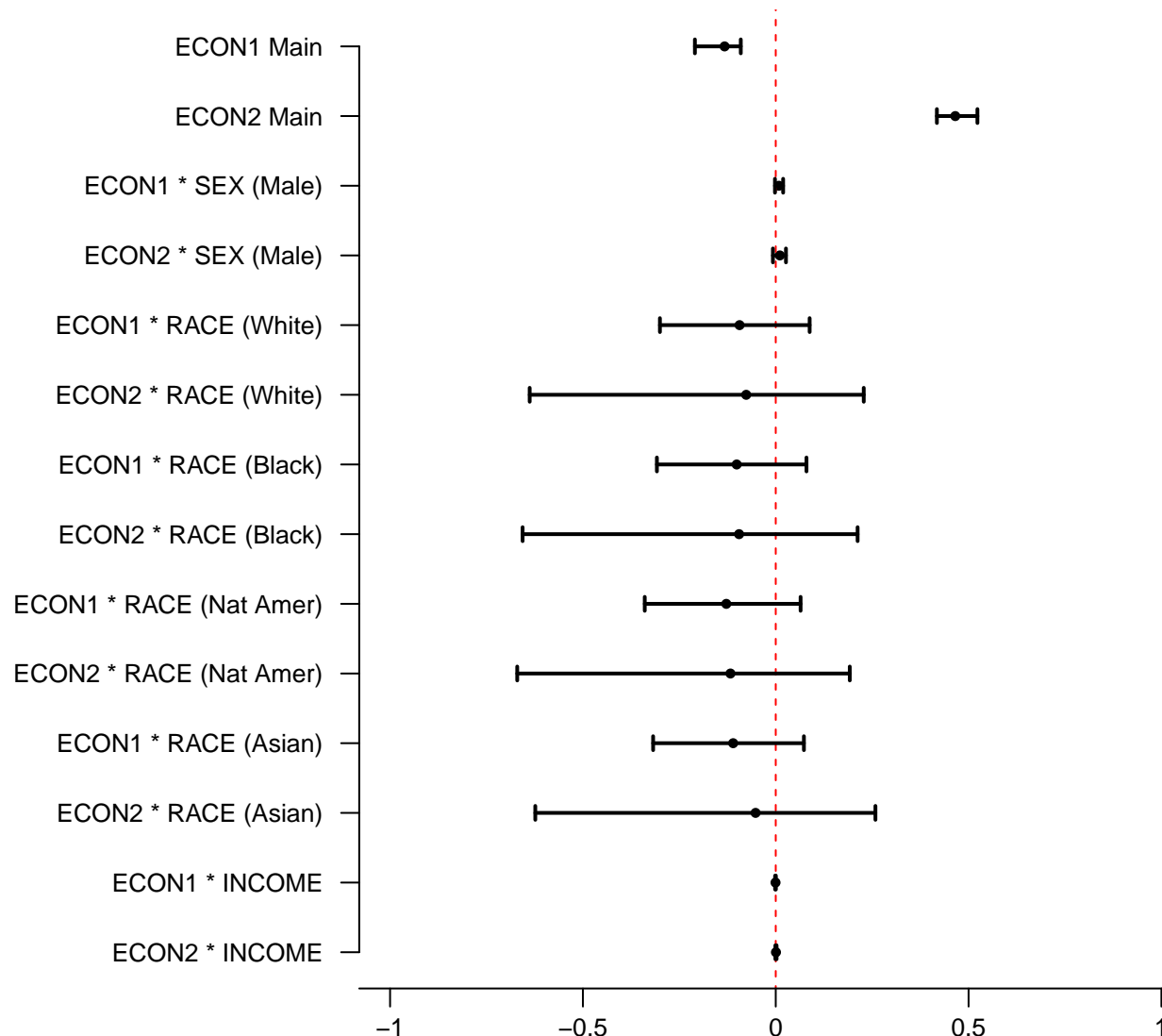
The model was fitted using Gibbs sampling, via JAGS (Plummer, 2003) and R, using a 10% random sample.

Results

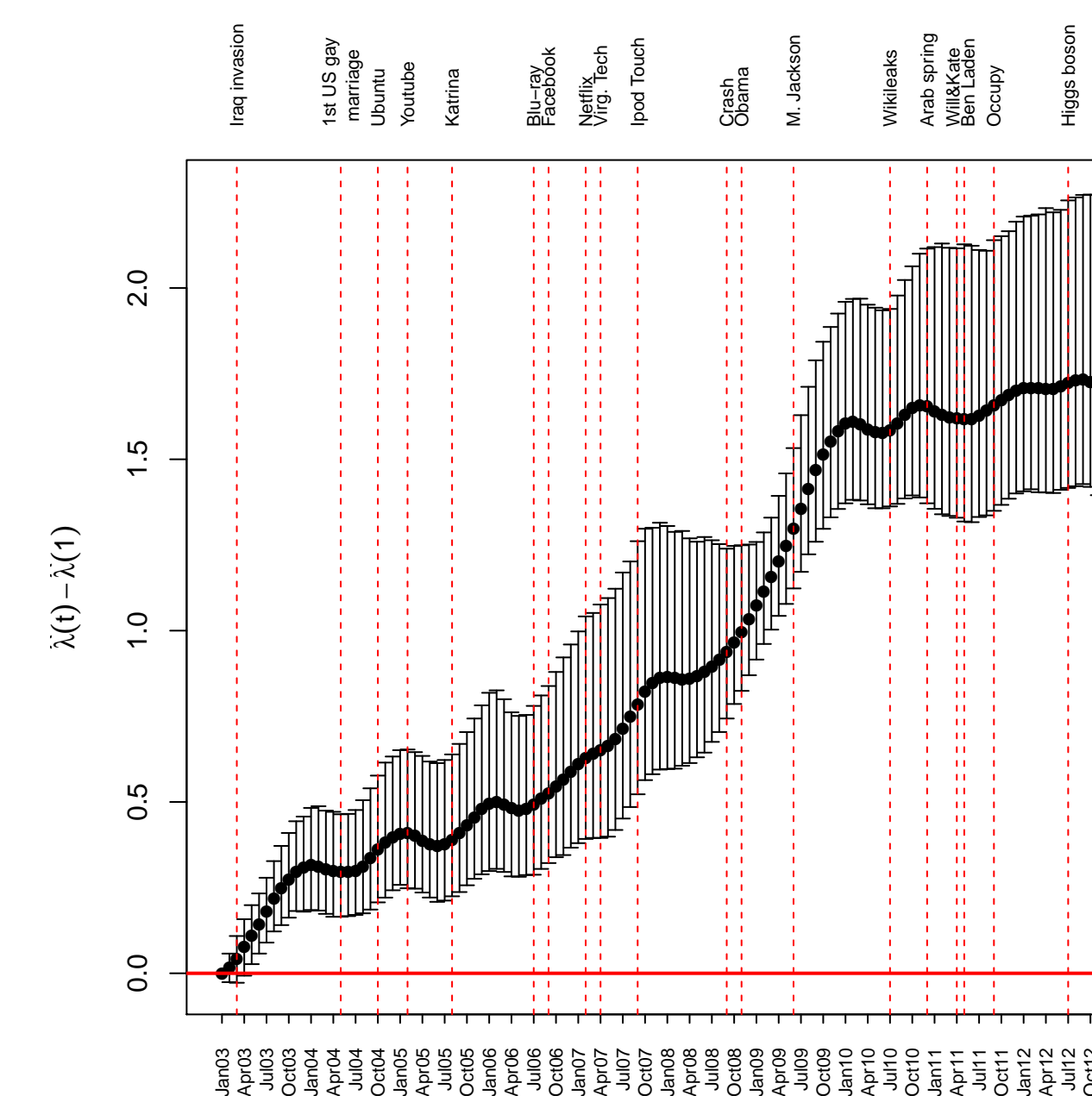
Estimates and credible intervals for parameter estimates in logistic model



Estimates and credible intervals for parameter estimates in gamma model



Model diagnostics



Smoothed time trend and major events

Socio-demographic variables

- ▶ Gender, age, region, race, hispanicity, marital status, education level, presence of household children or spouse, US citizenship status
- ▶ Weekly earnings, labour force status, number of jobs, average number of hours worked in a week
- ▶ Month and year the interview was conducted, the day of the week, whether it was a holiday

Methods

- ▶ We use a penalized GLM to select the most relevant socio-demographic factors. The group-Lasso penalty (Yuan & Lin, 2007) is:

$$\tau \sum_{\ell=1}^L \sqrt{p_\ell} \|\beta_\ell\|_2.$$

This penalty is used to account for the fact that some predictors are categorical. Here, p_ℓ is the number of features in group $\ell = 1, \dots, L$.

- ▶ The tuning parameter τ is selected using 10-fold cross-validation and the one-standard error rule (Hastie *et al.*, *Elements of Statistical Learning*)

Results

- ▶ For the **logistic model**, the selected factors are: presence of household children, age, gender, average number of hours worked in a week, and the number of jobs.
- ▶ For the **gamma model**, the selected factors are: presence of household children, age, gender, weekly earnings, presence of spouse in household, average number of hours worked in a week, diary day, education level, race, and the number of jobs.

Discussion

- ▶ **Why Bayesian:** convenient computational tools, smooth time effect, and exact inference
- ▶ **Survey weighting is overrated:** regression methods are sufficient
- ▶ **Zero inflation:** two phenomena, two models
- ▶ **Computational efficiency:** Group-Lasso fast, Gibbs sampling slow

Strengths

- ▶ The model is effectively separating the economic trend from the time trend
- ▶ The measure of economy being used is quite flexible
- ▶ The model was able to differentiate between the effect of unemployment and the effect of market behaviour

Limitations

- ▶ The gamma GLM assumes a constant dispersion parameter; model diagnostics seem to indicate otherwise
- ▶ The model does not include the natural auto-regressive structure of the economic variables
- ▶ The measure of association with economy may be too coarse
- ▶ To measure economy, we used PCA, which ignores the relationship with the response

Acknowledgements

We would like to thank Dr. Celia Greenwood, Dr. Olli Saarela, Dr. Abbas Khalili, and Dr. James Hanley for their help and input on this project. The training and advice they have provided and continue to provide has been influential on this project and beyond.