Determining Income Dynamics and Optimizing Tax Rate Over Permanent Income

Hassan Rahim Kamil

Department of Economics
University of Illinois at Urbana-Champaign

April 25, 2019

(Ongoing)

Contents

Motivation

Research Question

What determines income? Would it make sense to tax based on future/permanent income?

- Given the objective function, what would be the optimal tax rate?
- How would one accurately identify and predict future income?

Decision Theory

Payoff Function

$$\underset{D \in \mathcal{D}}{\operatorname{argmax}} \pi(Y, D, X) = B(Y(X, D), D) + C(X, D) \tag{1}$$

- Where $B(.)$ and $C(.)$ are the benefit and cost functions, and $Y = f(X)$.
- For a concave payoff function and differentiable $\mathcal{D}$, the FOC of (1) would yield:
  $B_Y(Y(X, D), D) \times Y_D(X, D) + B_D(Y(X, D), D) = C_D(X, D)$.

Choosing our payoff function as:

Permanent Income Hypothesis (Friedman)

$$c_t = \frac{r}{(1+r)+(1+r)^{-(T-t)}} \left\{ A_t + \sum_{k=0}^{T-t} \left(\frac{1}{1+r}\right)^k E_t[y_{t+k}(1-\tau_t)] \right\} \quad (2)$$

- Where r is the interest rate, A asset, y income, τ tax rate and $E_t[.]$ is the expected value at time t.

Why?

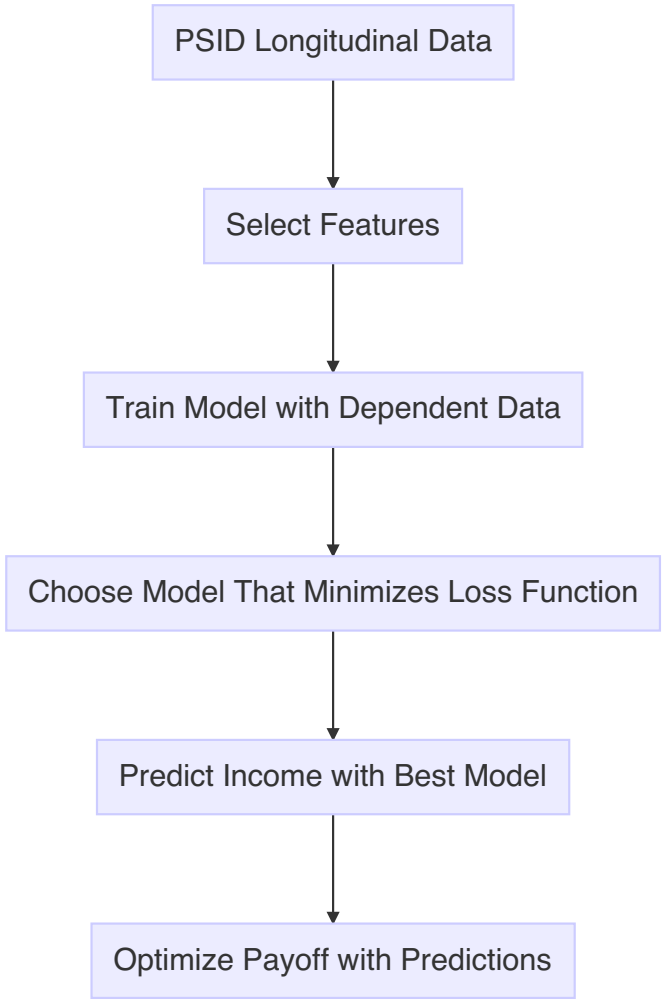
1. Nonlinear
2. $\frac{\partial^2 C}{\partial Y_d^2} < 0$, decreasing MPC.
3. Defined temporally.

Methodology

Functional Form

$$Y = f(X) + \epsilon, \;\; Y|X \sim N(0, \sigma^2) \tag{3}$$

- Typically, if Y is income, then $\log(Y)|X \sim N(0, \sigma^2)$.
- Y is often undeterministic.
- Want to estimate $\hat{Y} = \hat{f}(x)$, parametrically or nonparametrically.
- Employ feature selection method (like LASSO, Random Projection, Spike-and-Slab) on sparse matrix and prediction method (like Random Forest, Boosting, SVM).

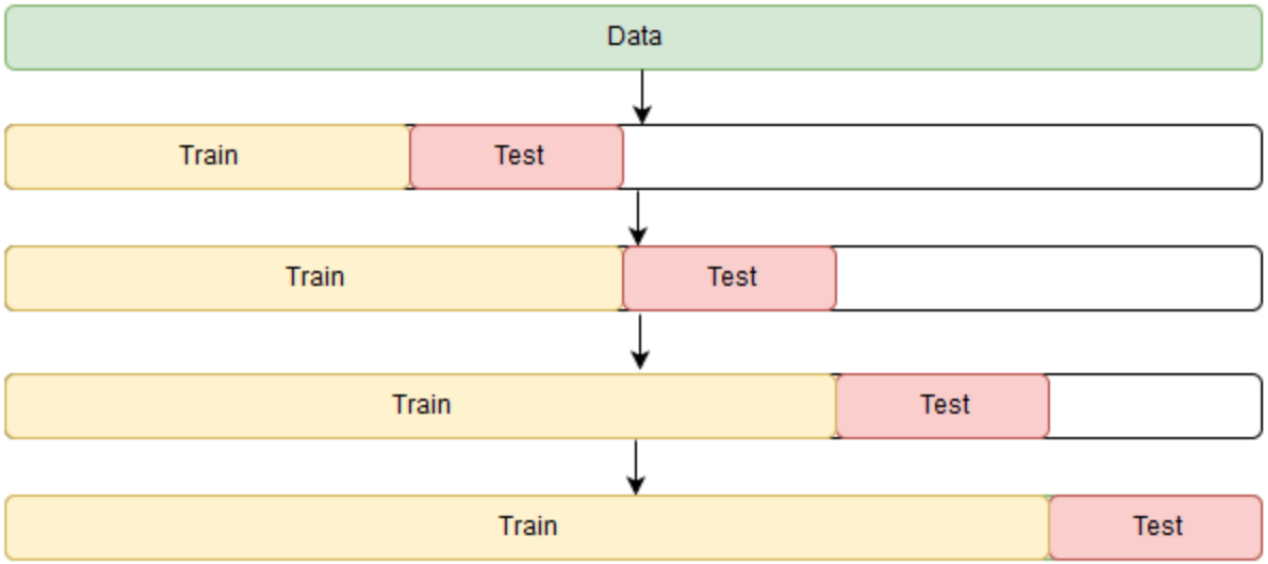


1

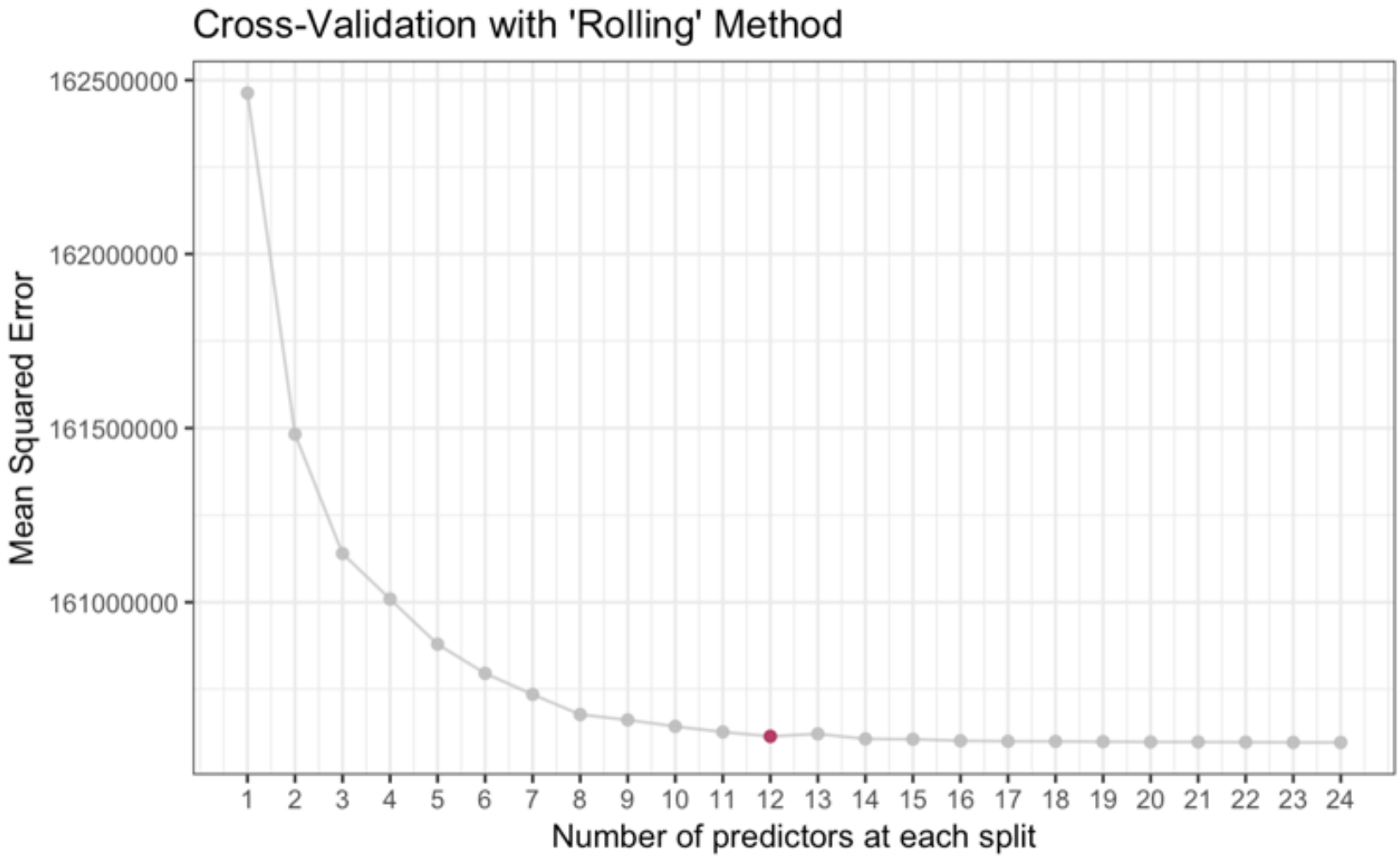
[1]The Panel Study of Income Dynamics (PSID) is the longest running longitudinal household survey in the world with sample of over 18, 000 individuals in 5, 000 families in the United States.

Analysis

- Generally, we want to predict on a dataset that is independent. Or, $Y|X \perp Y'|X'$.
- Due to the nature of our longitudinal data, we would like train our model using a cross-validation method employed with time series data. (e.g. "Rolling" Cross-Validation).
- Roll the train and test sets forward with respect to time t.

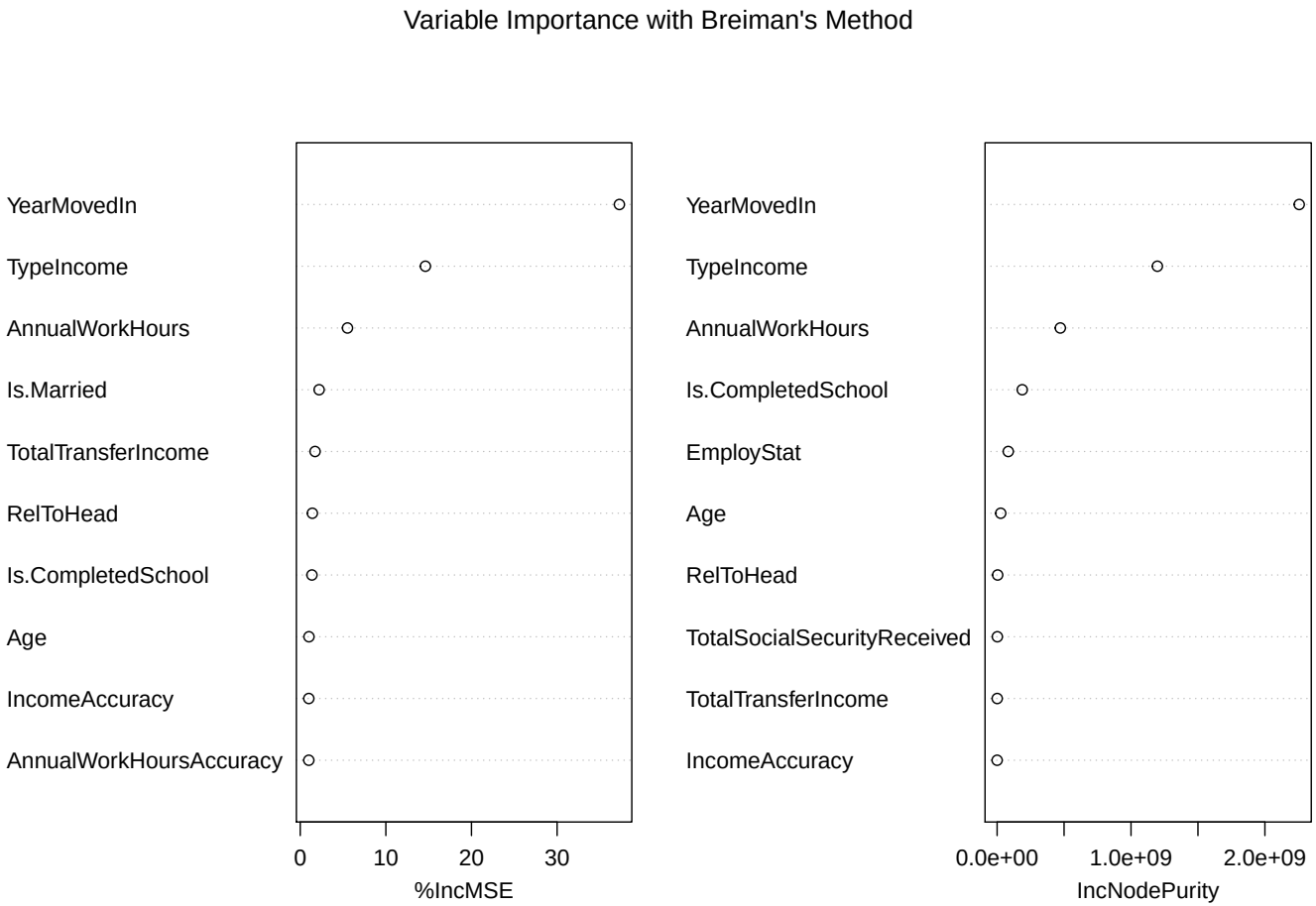


- Suppose our best model is the Random Forest:

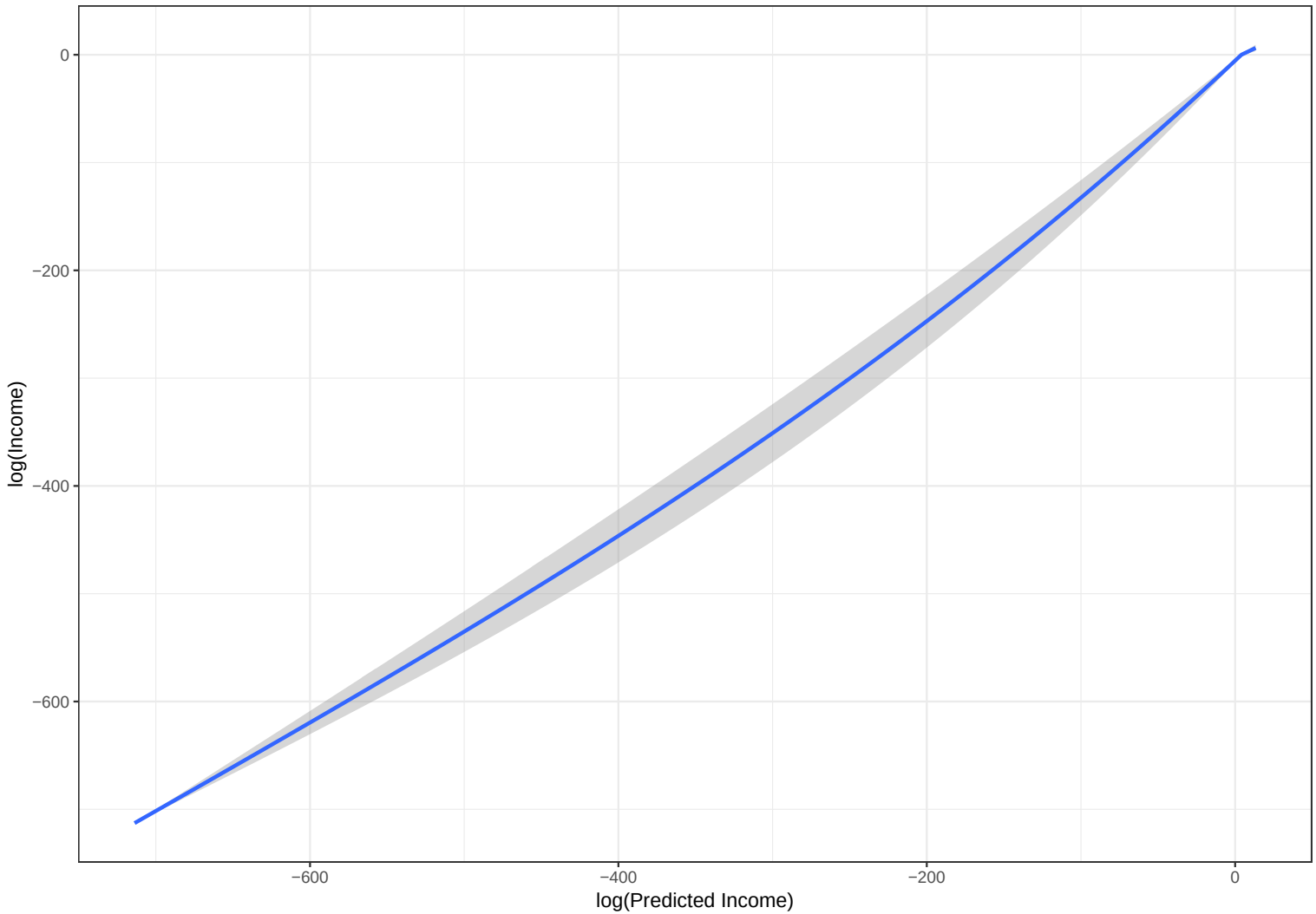


Variable Selection

- Top 10 variables selected by Random Forest:



- Comparison between actual income and predicted income.



Optimization

- Given the predicted incomese want to optimize the following:

$$h(\hat{y}_{t=1,\ldots,T}) = \begin{cases} c_t & c_t \geq 0 \\ \tau & 0 \leq \tau \leq 1 \end{cases}$$

Limitations

1. Lacking causal interpretation.
   - Honest Causal Trees (Athey, 2018).
2. Lacking consistency and asymptotic normality.
   - Inconsistent variance, unable to construct confidence interval. (Athey, 2018)
3. Hard to understand.
   - Often a black box.