1. Classify a model from a journal

(i) https://www.aeaweb.org/articles?id=10.1257/aer.20131193

(ii) Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. 2015. "Measuring Uncertainty." American Economic Review, 105 (3): 1177-1216.

(iii) Write down the mathematical or statistical model

(a) Uncertainty: common variation in the unforecastable components of macroeconomic indicators.

$$U_{jt}^{y}(h) = \sqrt{E[(y_{jt+h} - E[y_{jt+h}|I_t])^2|I_t]},$$

where y_{jt+h} is h-period ahead uncertainty, $U_{jt}^{y}(h)$ is the conditional volatility of the purely unforecastable component of the future value of the sereis, and the expectation is taken with respect to information I_t available to economic agents at time t.

(b) Macroeconomic uncertainty: aggregating individual uncertainty.

$$U_{t}^{y}(h) = E_w[U_{jt}^{y}(h)].$$

(c) Set of factor augmented regression.
This will replace the conditional expectation in (a) by a forecast.

$$y_{jt+1} = \phi_j^y(L)y_{jt} + \gamma_j^F(L)\hat{F}_t + \gamma_j^W(L)W_t + v_{jt+1}^y,$$

where $v_{jt+1}^y = \sigma_{jt+1}^y \epsilon_{jt+1}^y$ with $\epsilon_{jt+1}^y \sim^{iid} N(0,1)$ and

$$log(\sigma_{jt+1}^y)^2 = \alpha_j^y + \beta_j^y log(\sigma_{jt}^y)^2 + \gamma_j^y \eta_{jt+1}, \quad \eta_{jt+1} \sim^{iid} N(0,1)$$

$\hat{F}_t$ is introduced in the following factor structure and W_t consists of squares of the first component of $\hat{F}_t$ and factors in X_{it}^2 collected into the $N_G \times 1$ vectors $\hat{G}_t$ and X_{it}will be introduced in the following factor structure.

(d) Dynamic Factor Structure
$\mathbb{X}_t = (X_{1t},...,X_{Nt})'$ denote the predictor available for analysis and assume X_{it} has a following approximate factor structure

$$X_{it} = \Lambda_i^{F'} F_t + e_{it}^X$$

,where 12 forecasting factors F_t is chosen for the combined datasets $X^m$and X^f where each are 132 moslty macro economic time series variables and 147 financial time series variables respectively. The factors in the forecasting equation are estiamted by the method of static principal components (PCA).

(e) Vector Autoregression (VAR) model

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + e_t, \quad E(e_t e_t') = \Omega$$

where $p = 12$ and 11 macro variables

$$\text{X}=\begin{cases} log(Industrial\ production) \\ log(employment) \\ log(real\ consumption) \\ log(PCE\ deflator) \\ log(\ real\ value\ of\ new\ orders) \\ log(real\ wage) \\ log(hours) \\ federal\ funds\ rate \\ log(S\&P500) \\ growth\ rate\ of\ M2 \\ uncertainty. \end{cases}$$

(iv) List which variables are exogenous and endogenous
In this article, since it focuses on predictor and dynamics relying on its past activity, majority of the variables are endogenous. Factors in regression model that are extracted through factor model can be considered as exogenous. Other than that, variables in VAR are all endogenous.

(v) Classify the model as static vs dynamic, linear vs nonlinear, deterministic vs stochastic

(a) Uncertainty: Deterministic static nonlinear model

(b) Macroeconomic uncertainty: Deterministic static linear model

(c) Set of factor augmented regression: Stochastic Dyamic nonlinear model

(d) Factor Structure: Stochastic Dynmaic linear model

(e) VAR model: Stochastic Dynamic Linear model

(vi) List a vriable or feature that you think the model is missing that might be valuable
We can think of making macro variables and financial variables separate in doing an analysis. By relying on financial variable that has more frequent time window, I think researchers can try to calculate real time uncertainty in daily basis or in realtime.

2. Make your own model

(i) Write down a model of whether someone decides to get married
Here, we will use logistic regression model. Probability of getting married,

$$p = pr[y = 1|x] = F(x'\beta)$$
$$(By\ logit\ model) = \frac{exp(x'\beta)}{1 + exp(x'\beta)}$$

where $F(x'\beta)$ is the cdf of the logistic distribution. The predicted probabilities are limited between 0 and 1.

(ii) Explanatory variables are as follows (explanation on the right side):

$$x' = \begin{pmatrix} \text{age.husband} \\ \text{age.wife} \\ \text{pregnancy} \\ \text{dating.duration.in.month} \\ \text{age.first.dating} \\ \text{age.first.dating*dating.duration} \\ \text{income.husband} \\ \text{income.wife} \\ \text{wealth.parent.husband} \\ \text{wealth.parent.wife} \\ \text{education.husband} \\ \text{education.wife} \end{pmatrix}' = \begin{pmatrix} I[age_{husband} \geq 30] \\ I[age_{wife} \geq 30] \\ I[pregnancy] \\ \text{The longer the higher} \\ \text{The older the higher} \\ \text{The bigger the higher} \\ \text{The larger the higher} \\ \text{The larger the higher} \\ \text{The larger the higher} \\ \text{The larger the higher} \\ \text{middle/high/college/graduate} \\ \text{middle/high/college/graduate} \end{pmatrix}$$

(iii) Unlike the OLS regression model with error term, Logistic model doesn't have a error term however we can simulate it through a binary response outcome from Bernoulli distribution where each binary response is randomly generated according to the specified probabilities θ as a threeshold as follows

$$f(y) = \begin{cases} 0 & if\ y \leq \theta \\ 1 & if\ y > \theta \end{cases}$$

where $\theta = \frac{exp(x'\beta)}{1+exp(x'\beta)}$

(iv) The key factors are income of husband, age.first.dating*dating.duration, pregnancy, and age of wife.

(v) First and foremost, income of husband is considered as the most important factor for both wife and husband now that realistically it is very costly to get married. Second, if age of first dating start very late, it is more likely to decide to get married. Morever, even though couple start to date from the young age, long time relationship might increase the probability of deciding to get married. Age of wife is espectially important in Korean culture in that people usually believe younger the better for both wife and husband. Also if the wife gets pregnant, it is very likely to get married.

(vi) We can try both Wald Test and Likelihood Ratio test for the test of significance. The wald test is to use check whether the individual coefficients are significant or not as we do t-test in linear regression. The test statistics

$$t = \frac{\beta}{\sqrt{var(\beta)}}$$

can be used to test $H_0 : \beta = 0$

When testing the joint significance of several coefficients, we can use likleihood ratio test. Here null hypothesis is about any subset of several β's that equals to 0. It compares the likelihood of the data under the full model against the likelihood of the data under a model with fewer predictors.