Evaluating Hierarchical Logit and Probit Regression Models on Economic Recessions of 10 Organization for Economic Cooperation and Development (OECD) Countries

A Bayesian Computing Project by Yin Kwong John Lee, Yuanbin Sun, and Chenyi Yu

Abstract:

Economic cycles for businesses are inevitable, yet identifiable. While riding through the wave of an economic expansion is pleasant, stumbling through an unforeseeable economic contraction may be damaging. The economic contraction is also known as an economic recession. Recession is often associated with a downturn in economic performance. During an economic downturn, people sequentially sell off their assets and so do you, but before you notice, your assets have been sold at a price lower than their intrinsic values. Through the late 1920s, investors repeatedly sold off their shares due to bleak news of the economy. The public panicked and followed what the investors did- selling off shares like crazy. Consequently, shares were oversold at prices far below their true values. A recession commenced. The bearish outlook caused people to sell off shares at even bigger losses, which later accelerated the Great Depression in 1929. What do we learn? Identifying moments before a recession hits may protect us from unanticipated losses. In a statistical sense, many inferences may be drawn between the economic condition and the performance of financial instruments- stocks, bonds, options, futures, and exchange rate swaps to name a few. This paper will take a Bayesian approach to evaluate hierarchical models that fully leverage the top economic indicators as precursors toward recession events for 10 OECD countries. Identifying such economic conditions prior allows for time to strategize our holdings before the actual recession arrives.

1. Introduction:

1.1 Data collection and description:

A recession is defined by the National Bureau of Economic Research (NBER) as a significant decline in real Gross Domestic Products (GDP), real income, employment, industrial production, and wholesale-retail sales for more than a few months. We stood by this definition and collected primary data for these 5 economic indicators along with their economic outcomes. The data are collected via the U.S. Bureau of Economic Analysis (BEA) and OECD's Composite Leading Indicators (CLIs) repositories. The U.S. BEA stores macro and micro-economic measures on a multinational level as the OECD CLIs screens the conditions that dictates the economic booms and busts of OECD countries. Monthly data are collected for 10 OECD countries over the past 240 months (20 years). These countries include the United States, United Kingdom, Netherlands, Luxembourg, Japan, Korea, Germany, France, Denmark, and Canada. The observation period ranges from October 1999 to September 2019. Since data for real income were only available in quarterly terms, we converted them into monthly terms through cubic spline interpolation, a polynomial basis technique widely used by economists to estimate gap periods within each time interval. There are a total of 2400 observations, with 5 numeric covariates and 1 binary response (0= no recession and 1=recession) for each observation. Each observation is indexed its time and country of occurence.

1.2 Exploratory Data Analysis:

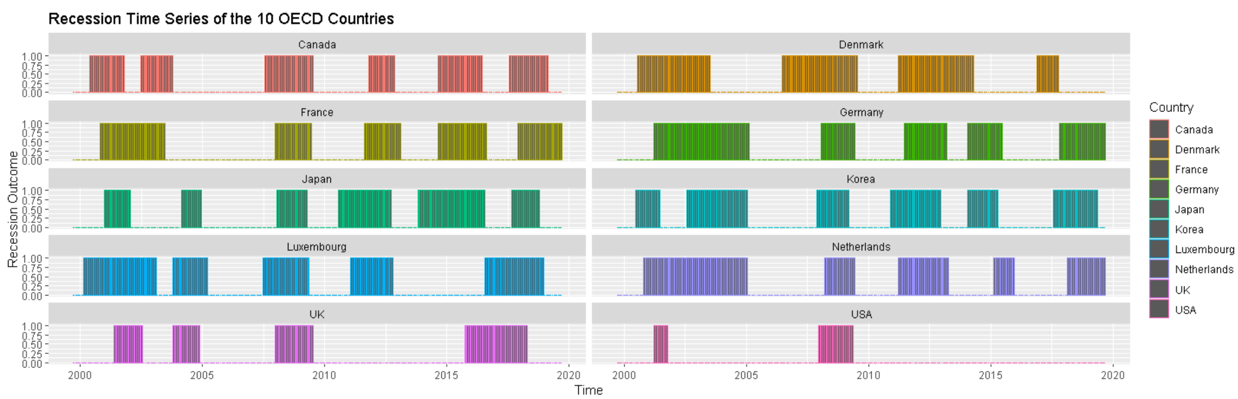

Figure 1 Recession time series of the 10 OECD countries from 1999-2019.

The barplots (Figure 1) highlight recession events for each country over the past 20 years. On one end, the U.S stands out to be where the least number of recession events occured, and also where the shortest recession took place, for instance, the 2001 recession. On the other end, Canada, Japan, and Korea seem to experience economic recession most frequently; there were a total of 6 recessions over the past 20 year period. Moreover, it seems that the recession time and duration of one country depends on those of the other countries. For example, similar recession patterns are observed for Canada, France, Germany, Japan, Korea, and Netherlands. This signifies evidence for correlations in the economic outcomes between the countries throughout time.

1.3 Data Pre-processing:

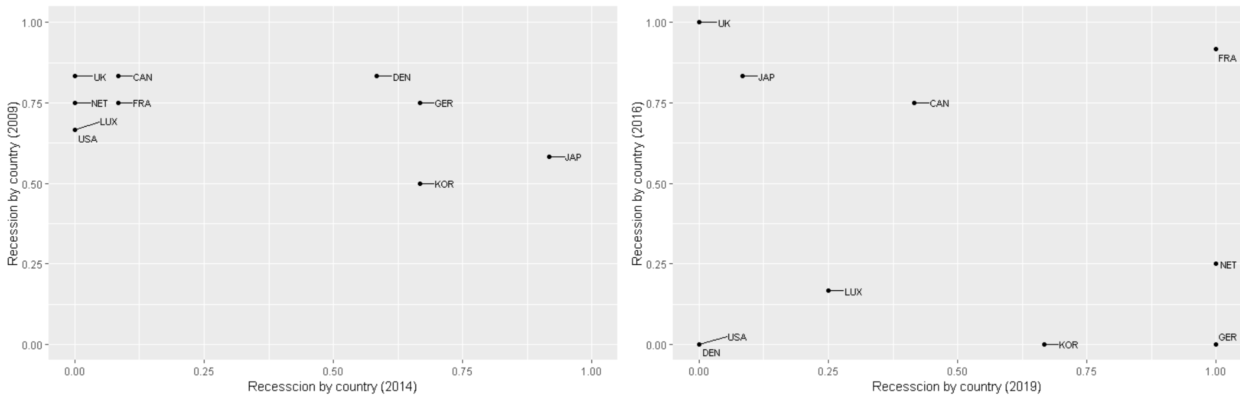

Figure 2 (a) Percentage of time throughout the year that recession persisted, in 2009 and 2014. (b) Percentage of time throughout the year that recession persisted, in 2016 and 2019.

The left scatterplot (Figure 2) substantiates our claim about the existence of correlations. The duration of the 2004 recession for one country seems to be linearly dependent on that of the 2009 recession for the other countries. The right scatterplot (Figure 2) shows the reverse case. The duration of the 2019 recession for one country does not correlate with that of the 2016 recession for the other countries. We may deduce this from the dispersed pattern of the plot. The scatterplots thus suggest some collinearity between the variables time and country. We eliminate the country variable to preserve the identifiability of our estimated predictors under the logit and probit regression models. The rest of the covariates are standardized for modeling fitting on the same scale.

The rest of our report is organized as follows. The settings of our models are presented in Section.2. The sampling methods and results are detailed in Section.3. Model selection results are reported in Section.4. We then conclude our findings and raise some thoughts in Section.5.

2. Fitting Time-Variant Logistic and Probit Regression Models:

2.1 Weakly Informative Priors:

$(1)\ \alpha \sim N(0,10)$ and $\beta_k \sim N(0,2.5), k=1,...,5$

We impose minimal prior information on predictor coefficients using a $Normal$ distribution centered at 0, with scale at 10 for the intercept term and 2.5 for other coefficients (shown in (1)). Assuming weak prior knowledge can stabilize the estimated coefficients by constraining them to fall within a certain range on a logit scale. Thus, larger scale factors guarantee weaker prior information, and mitigates the problem of separation that arises from a linear combination of predictors being perfectly predictive of the outcome. $Cauchy$ and t_0

distributions with the same parameter choices may also be prior candidates for the logistic and probit regression coefficients due to their ability to capture coefficient variations while remaining conservative. $Uniform$ priors may not be appropriate for our models because they are overly conservative to bring about the identifiable features in the regression.

2.2 Likelihoods/Models:

(2) likelihood for logistic regression: $\prod_{t=1}^{240}[(logit^{-1}(X_t\beta))^{y_t}(1-logit^{-1}(X_t\beta))^{1-y_t}]$

(3) likelihood for probit regression: $\prod_{t=1}^{240}[(\Phi^{-1}(X_t\beta))^{y_t}(1-\Phi^{-1}(X_t\beta))^{1-y_t}]$

where $y_t = \{0, 1\}$, and $X_t, \beta \in \mathbb{R}^5$, for $t = 1, ..., 240$

We specify $X_t\beta$ as the linear predictor and pass it through the logit and probit link functions, $logit^{-1}$and Φ^{-1}, to construct linear dependency between y_t and X_t. Here, y_t tells whether recession exists with the weights $(\beta_1, ..., \beta_5)$ on their respective recession measures $(X_1, ..., X_5)$ at time t. The logit link function provides an estimate for the odds ratios as the probit link function transforms latent variables to follow a standard normal for better approximations. The transformed parameters are then used as probabilities for the Bernoulli trials. The product of the likelihoods at all 240 time points (shown in (2)&(3)) forms the likelihoods of the logistic and probit models.

2.3 Posterior Distributions:

(4) $P_{logit}(y \mid X_t, \alpha, \beta) \propto \prod_{t=1}^{240}[(\frac{e^{X_t\beta}}{1+e^{X_t\beta}})^{y_t}(1-\frac{e^{X_t\beta}}{1+e^{X_t\beta}})^{1-y_t}] \times \prod_{k=1}^{5} e^{-\frac{(\beta_k^2)}{2\times 2.5^2}} \times e^{-\frac{(\alpha)}{2\times 10^2}}$

(5) $P_{probit}(y \mid X_t, \alpha, \beta) \propto \prod_{t=1}^{240}[(\Phi^{-1}(X_t\beta))^{y_t}(1-\Phi^{-1}(X_t\beta))^{1-y_t}] \times \prod_{k=1}^{5} e^{-\frac{(\beta_k^2)}{2\times 2.5^2}} \times e^{-\frac{(\alpha)}{2\times 10^2}}$

Multiplying the likelihoods and prior distributions returns the posterior density (shown in (4)&(5)) for our models. Note that these are posterior distributions of one country. In principal, we need to construct different posterior densities if the economic features of other countries follow different distributions. We assume these features to be exchangeable across countries at all times. This means coefficients and economic responses for all 10 countries are sampled under the same posterior framework.

2.4 Posterior Predictive Check on Independent Bernoulli Trials:

To examine the assumption of independent observations, α_m and $(\beta_{1,m}, \ldots, \beta_{5,m})$ are sampled from their posterior distributions to simulate 10,000 Bernoulli trials, with $m$denoting the m-th trial. The number of switches between 0 and 1 from each trial is used to construct a test quantity, $T(recBinary^{rep})$. Consequently, m such quantities are compared with the observed number of switches, $T(recBinary) = 96$ (Figure 3), where 58.94% of the prior are found to be larger than the later. The posterior predictive p-value, $\hat{p}$, carries the same interpretation, and suggests that the independence assumption is not horribly violated. This is because its value lies around the median of the posterior predictive distribution of $T(recBinary^{rep})$, where $\hat{p} = 0.5$, meaning that independent observations are decently likely to be drawn in posterior replications under the fitted models if they were true.

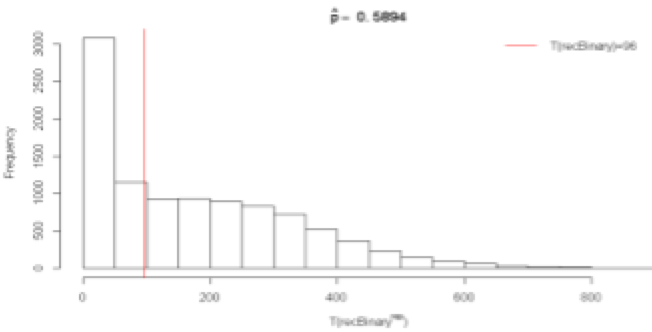

Figure 3 Observed number of switches (red vertical line at T(recBinary)=96), compared to 10,000 simulations from the posterior predictive distribution of the number of switches.

3. **Sampling Methods:**

3.1 No-U-Turn Sampling (NUTS)- An extension to the Hamiltonian Monte Carlo (HMC) Sampling:

<u>3.1.1 Weighted-time (WT) Approach:</u> We add a time variable to the pool of coefficients such that all time points are associated with their estimated weights. These estimated weights are predictor coefficients,$(\beta_1, \ldots, \beta_6)$, sampled from their priors and updated via their posterior distributions at each draw. This approach is risky because other covariates may be highly correlated with time. For instance, a -0.86 correlation exists between real personal income and time. Collinearity the models tends to inflate the variance on parameter estimates and throw off our inferences.

<u>3.1.2 Simple Random Intercept (SRI) Approach:</u> As the name states, we add a random term,w_t, as an intercept into our original linear predictor $(\alpha + X_t\beta \rightarrow \alpha + w_t + X_t\beta)$, and hope that it accounts for the uncertainty caused by time. The random term adds some stochasticity to our regression models and allows for heteroskedasticity at each time step as we sample from the

posterior distributions of the coefficients. This approach regulates the noise in the time-series sequence, and intends for "fairness" in posterior sampling.

3.1.3 Random Walk (RW) Approach:

$$(6)\ \alpha_t \sim N(\alpha_{t-1}, 10) \text{ and } \beta_{t,k} \sim N(\beta_{t-1,k}, 2.5),\ t = 2, \ldots, 240,\ k = 1, \ldots, 5$$

At the first time step, predictor coefficients are drawn from their weakly informative priors proposed above. These samples are used to update the centers for the draw at the next time step. The next draw then updates the center for the draw at the time step after. This procedure renews the posterior knowledge of the coefficients at each time step rather than fixing minimal knowledge at all times. Most likely, informative priors are used from time points 2 and on. This makes sense because as time passes the population gains richer knowledge about the behaviors of the economic measures, e.g., real GDP and unemployment rate. However, as priors grow stronger throughout time, the models may be overconfident in estimating the coefficients, and vice versa, thus offsetting the quality of our inferences.

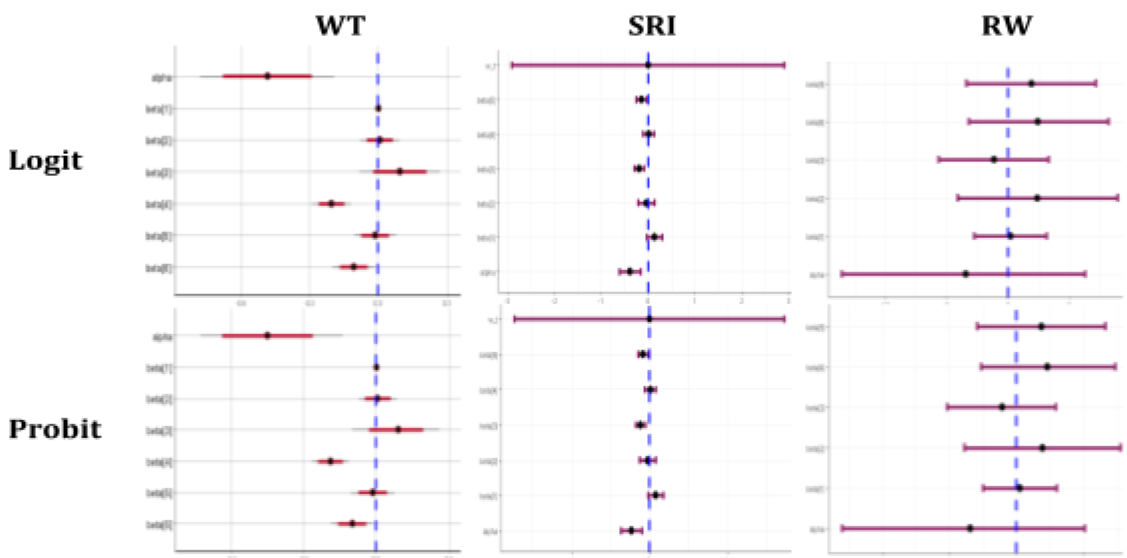

Figure 4 Averaged estimates and errors for relevant coefficients of the WT, SRI, and RW models under the logit and probit frameworks. A N(0,10) prior for intercept and N(0,2.5) for other coefficients are used for all models.
Note: The orders of coefficients on the y-axis differ for each graph

The bar-and-error plots (Figure(4)) illustrate the average sampling quality for the predictor coefficients under the weighted-time (WT), simple random intercept (SRI), and random walk (RW) logit and probit regression models. Results for the logit models seem no different than that of the probit model. What's said from now on applies for both the logit and probit models. The black dots represent the averaged centers of the estimated coefficients while the red lines indicate their mean standard errors. Centers for the WT model, in particular, the intercept term, α, of the regression seems to stray from 0 (away from the blue dashed line) and has a wide average standard error, meaning that this model allows for some additive variations, and these variations may be coming from predictors such as time. We bear with this hypothesis to examine the average quality of the coefficients drawn under the SRI model. All centers are averaged to hover closely around the zero dashed line as their mean standard errors are sand-like, except for that of the additional noise term, w_t. This result aligns with the hypothesis we are bearing because the wide range of variations added to our model accounts for the additional noise at each time step, which grants consistent non-informative prior information (centers close to the dashed line) on average. We inspect error plots of the RW model and infer a risky performance in sampling model coefficients. This is due to the relaxing of the assumptions of independent observation and weakly informative prior at each time step, as time volatility at one time step is incorporated based on information of the previous time step. To sum up, it is more sensible to implement the WT and SRI model since it handles time-variability fairly better than the RW model. For this reason, we eliminate the RW model and move on to compare the two models, WT and SRI regression models.

4. Model Selection:

Table1: Information Criterions among models

Models	Logit-WT	Probit-WT	Logit-SRI	Probit-SRI
DIC	-771.015	-770.424	-2327.045	-2348.118
WAIC	3270.492	3270.849	2677.913	2667.240
LOO-CV	3270.502	3270.880	2682.576	2674.054

Deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC), and Leave-one-out cross-validation (LOOCV) are options for approximating expected predictive accuracy (elppd). These three methods are used to compare the WT and SRI regression models. Akaike information criterion (AIC) is not used here since it is not suitable for evaluating hierarchical models. DIC and WAIC are methods designed to correct the bias of the within-sample predictive accuracy (lppd). LOOCV, where the data is spitted to training and test data sets, estimates the leave-one-out sample predictive accuracy of the test data by fitting the models to the training data. For DIC, models are penalized by the value of the deviance, D, and the number of parameters. Hence, models with smaller DIC values are preferred. WAIC estimates the amount of information lost in a model. Therefore, lower values of WAIC are preferred as it indicates better prediction of the model. Similar to the previous, smaller values of LOOCV are preferred. In summary, for each method, the lower the value, the better the prediction. Therefore, the table above shows that the SRI probit regression model has the best predictive performance for all information criteria.

5. Conclusion and Thoughts:

To conclude, the weighted-time (WT), simple random intercept (SRI), and random walk (RW) logit and probit regression models are built for recession prediction among ten countries. The bar-and-error plots display the average sampling quality for the predictor coefficients, and suggest only the WT and SRI models to be further analyzed since they deal with time-variability better than the RW model. In addition, Deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC), and Leave-one-out cross-validation (LOOCV) are used to check the predictive performance among these models. The SRI probit regression model exhibits the lowest value for all information criteria, and therefore yields the best predictive performance.

For Bayesian hierarchical models, one of the concerns is the choice of the prior. We proposed a weak normal prior for our models and arrived at the above conclusion. Different choices of priors, such as $Cauchy$ and t_0 distributions, may urge different conclusions, and are worth future exploration. We may also define the uncertainties by countries rather than by time so that countries are able to strategize their decisions accordingly in the real economy.

Pooled and unpooled models could also be evaluated to make holistic comparisons between models. Moreover, the countries analyzed in our project are all OECD countries, who may have some special characteristics in common that other countries in the world do not have, such as high income level per capita. This exogeneity problem on one hand may improve our prediction accuracy of rich countries, but on the other hand makes it difficult to generalize our results to other countries in the world, i.e. medium income countries.

6. References:

[1] Bruce Western, “Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach”, American Journal of Political Science, Vol. 42, No. 4 (Oct., 1998), pp. 1233-1259

[2] George Y. Wong and William M. Mason, “The Hierarchical Logistic Regression Model for Multilevel Analysis”, Journal of the American Statistical Association, Vol. 80, No. 391 (Sep., 1985), pp.513-524

[3] Andrew Gelman, “Understanding Posterior P-values”, Electronic Journal of Statistics 2013

[4] OECD, “Interpreting OECD Composite Leading Indicators (CLIs)”, 2009

[5] Thomas B. King, Andrew T. Levin, and Roberto Perli, “Financial Market Perceptions of Recession Risk, ”(2007), Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.

[6] Constantinos Daskalakis,Nishanth Dikkal,Ioannis Panageas, “Regression from Dependent Observations”, (October 9, 2019)

[7] Lalonde, T. L., Nguyen, A. Q., Yin, J., Irimata, K. and Wilson, J. R. “Modeling Correlated Binary Outcomes with Time-Dependent Covariates”, Journal of Data Science 11(2013), 715-738

[8] By Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su, “A WEAKLY INFORMATIVE DEFAULT PRIOR DISTRIBUTION FOR LOGISTIC AND OTHER REGRESSION MODELS”, The Annals of Applied Statistics 2008, Vol. 2, No. 4, 1360–1383

[9] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis (Third Edition). Chapman & Hall/CRC Texts in Statistical Science, pg. 141-159, 165-187, 353-376, 381-398, 405-428.