Multivariate Statistics Project: Multivariate Models for Predicting Country Fiscal Stress

S S K Pavan,[*] Nitin Gautam[†]
Submitted to Prof. Niladri Chatterjee[‡]

November 30th, 2021

Abstract

We explore Multivariate methods to determine factors of fiscal stress. We use a unbalanced panel data specification of 43 countries over the period 1992-2016. We use a LASSO logistic model to choose the variables which significantly influence fiscal stress periods in a country. We then use a dynamic factor model as a dimensionality reduction technique. Using the estimated factors we run a logistic regression further explore the relationship between stress periods and macroeconomic variables. We find that US Interest rate, change in national consumption, public debt, and current account balance significantly influence stress period incidence in a country.

Keywords: Logistic regression, LASSO, Wilcoxon test, Dynamic Factor analysis, Fiscal Stress

[*]hes207045@hss.iitd.ac.in
[†]hes207035@hss.iitd.ac.in
[‡]as part of the course requirement in MTL 766 Mutivariate Statistcal Methods

Methods Employed

- Non-parametric test: Wilcoxon
- LASSO Logit
  - All factor Specification
  - Restricted Specification
- Exploratory factor Analysis
  - Bartlett's test
  - Scree plot
- Dynamic Factor model with panel data logit

Stress Period

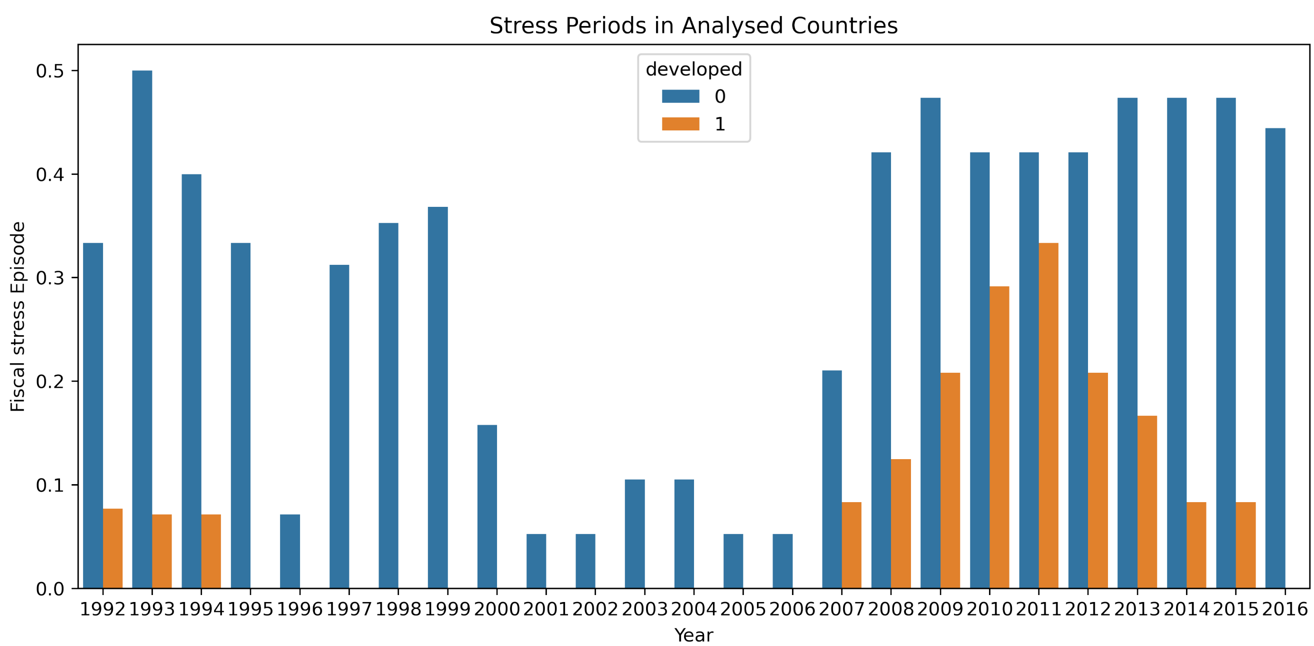

Figure 1: Stress Periods Incidence. 0 - developing, 1- developed country

We define that a country is in a fiscal stress period if at least one of the following conditions is met:

- Debt default or restructuring is present
- When a country avails aid from a large international institution

- Hyperinflation, defined as rapid, excessive, and out-of-control general price increases in an economy
- Deteriorating investor sentiment, defined as a period in which the spreads of a country's bond yield to a benchmark bond yield increase considerably.

The figure 1 plots the incidence of stress periods across time for developing and developed countries

- Among observations for developing countries 29% were stress periods.
- Among observations for advanced countries only 7% were stress periods. In particular, most of the stress periods in advanced countries are related to the fiscal stress in the euro area which started around 2010-2011.

Data and Variables

We use the variables listed in table 1, collected from different sources for determining the salient factors affecting fiscal stress periods of a country. We have collected these variables for 43 different developing and developed countries countries.

	Variable	Description	Source
0	cpi	Consumer Price index - inflation	OECD
1	dyn_gdp	GDP dynamics	OECD
2	dyn_gdp_china	China GDP dynamics	World Bank
3	dyn_GDP_US	US GDP dynamics	BP P.l.c
4	interest_rate_US	US interest rates	CBOE
5	oil_yoy	Oil price dynamics	World Bank
6	dyn_consum	Consumption dynamics	World Bank
7	dyn_fx_rate	FX rate dynamics	IMF WEO
8	diff_priv_credit_gdp	Credit to GDP change	IMF WEO
9	net_lending	Net lending	World Bank
10	public_debt	Public debt	World Bank
11	interest_on_debt	Interest on debt	IMF IFS
12	overvaluation	Currency overvaluation	IMF IFS
13	ca_balance	Current account balance	IMF WEO
14	dyn_fix_cap_form	Fixed capital formation dynamics	IMF WEO
15	dyn_export_share	Export share dynamics	IMF WEO
16	diff_unempl	Unemployment change	IMF WEO
17	dyn_prod_dol	Labor productivity dynamics	IMF WEO
18	VIX	Chicago Board Options Exchange Volatility Index	ILO
19	GDP_per_cap	GDP per capita	IMF WEO

Table 1: Variable Description

Multicollinearity is an issue with the logit model that we are estimating. We look at the var-covariance matrix of our variables in figure 2. We find High pairwise correlation for some variables which are reported below:

- dynamics of labor productivity and of GDP (76%)
- dynamics of fixed capital formation and of GDP (69%)
- currency overvaluation and GDP per-capita (67%)

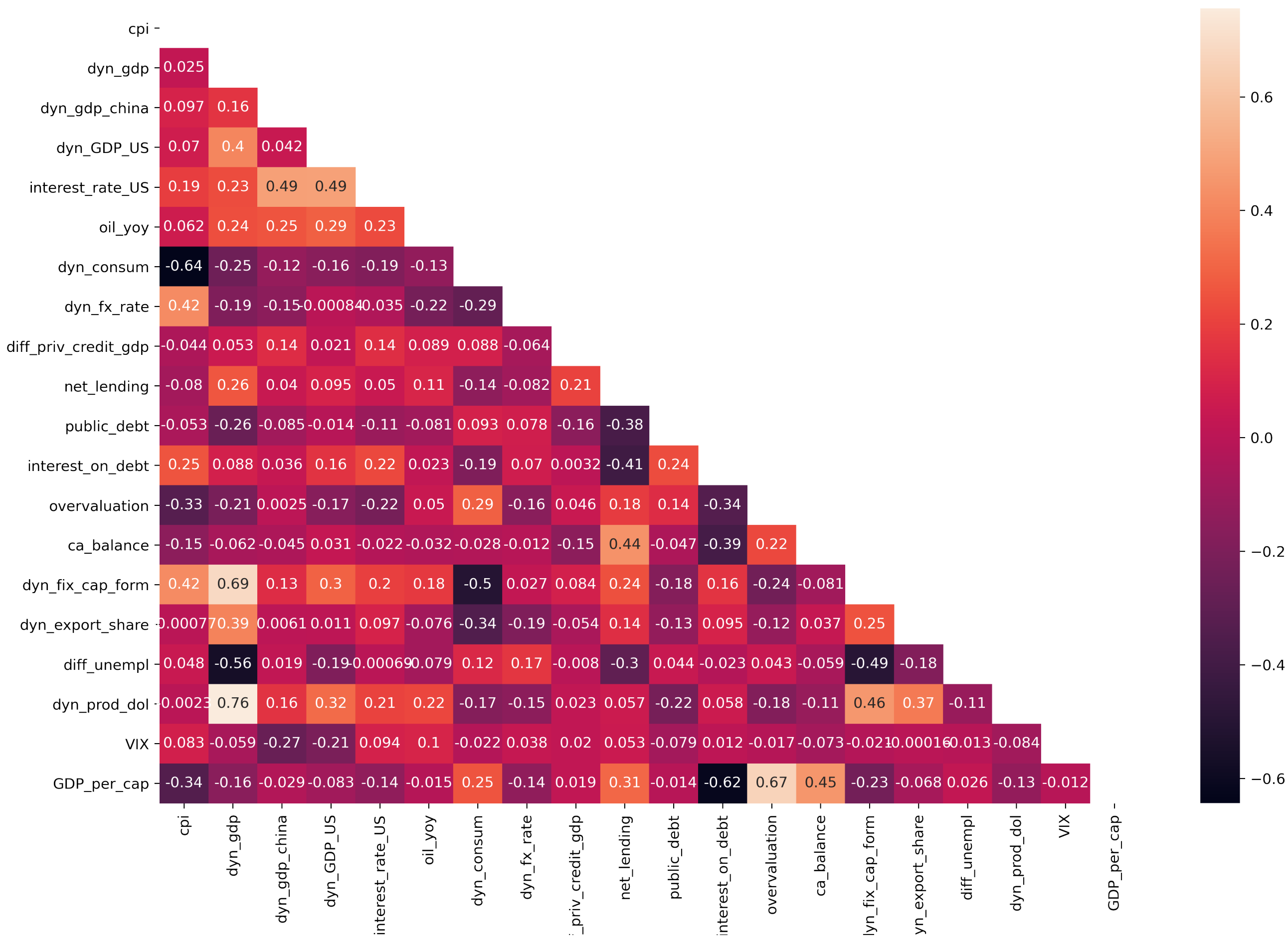

Figure 2: Cov-varince matrix

Leading up to a stress period, our hypothesis is that the macroeconomic variables of the country exhibit some signs of distress. We aim to exploit these signals of distress in the variables that we have chosen. We test whether there is a difference in the variables leading upto the stress period and the variables leading up to a tranquil (non-stress) period in the following section.

Wilcoxon Test

The Wilcoxon test is a non-parametric test that can be used to compare two unpaired groups with non-normal distributions. In general, the differences in means before the stress period and before the tranquil (non-stress) period by intuition should be statistically different. We test this hypothesis using the Wilcoxon test, which is considered an alternative to the t-test. We apply the test to all the explanatory variables in the table above.

$$H_0 : S_{t-k,t} - T_{t-k,t} = 0$$
$$H_a : S_{t-k,t} - T_{t-k,t} \neq 0$$

- Where $S_{t-k,t}$ is the mean of k years before the stress period
- $T_{t-k,t}$ is the mean of k years before the tranquil period (No-stress period)

The p-values for the test have been displayed below:

	Variable 1	p-value for k = 1	p values for k = 2
0	cpi	0.00	0.02
1	dyn_gdp	0.57	0.42
2	dyn_gdp_china	0.00	0.00
3	dyn_GDP_US	0.84	0.50
4	interest_rate_US	0.00	0.00
5	oil_yoy	0.30	0.00
6	dyn_consum	0.11	0.48
7	dyn_fx_rate	0.91	0.52
8	diff_priv_credit_gdp	0.22	0.03
9	net_lending	0.45	0.26
10	public_debt	0.03	0.01
11	interest_on_debt	0.67	0.18
12	overvaluation	0.14	0.59
13	ca_balance	0.00	0.00
14	dyn_fix_cap_form	0.56	0.32
15	dyn_export_share	0.19	0.21
16	diff_unempl	0.30	0.15
17	dyn_prod_dol	0.19	0.29
18	VIX	0.56	0.66
19	GDP_per_cap	0.04	0.03

Striking result: For 'US interest rates' we can reject the null hypothesis at 5% level of significance. Before stress period the interest rates are different than before tranquil period. This gives us evidence that US interest rates have significant influence on creating fiscal stress in other countries

Logistic Regression

Our main method of identifying stress periods is the logistic regression. We ran two types of logistic regressions to model the probability of occurrence of stress period using the macroeconomics variables. We use a standard logit model and a logit model with a least absolute shrinkage and selection operator (LASSO) penalisation.

LASSO penalisation performs simultaneous variable selection and regularisation in order to enhance the prediction accuracy and interpret-ability of the model. The difference between the logit with the LASSO penalisation and regular logit is the restriction imposed on estimated coefficients. If the LASSO penalisation is included, the sum of absolute values of estimated coefficients cannot exceed the pre-specified free parameter, which effectively determines the amount of penalisation imposed. We estimate a penalisation parameter using cross-validation technique.

The results of the cross-validation techniques are presented in the table 2. The column corresponding to the variable gives the number of times its coefficient went to 0. We run 100 loops to determine which variable goes to zero in majority of the loops. We discard any variable which goes to zero in more than 60 loops (highlighted red) in the model with LASSO penalization .

Variable	times coefficient = 0
cpi	0
dyn_gdp	3
dyn_gdp_china	62
dyn_GDP_US	86
interest_rate_US	0
oil_yoy	73
dyn_consum	0
dyn_fx_rate	4
diff_priv_credit_gdp	5
net_lending	27
public_debt	0
interest_on_debt	55
overvaluation	96
ca_balance	0
dyn_fix_cap_form	92
dyn_export_share	77
diff_unempl	0
dyn_prod_dol	15
VIX	71
GDP_per_cap	0
total_loops	100

Table 2: Cross-Validation

Results of the two regressions with fixed effects: model with the LASSO penalization and the model with all the variables is presented in 3 for easy comparison.

Table 3: Logistc Regression results: Dependent Variable Stress Period Next Year

	LASSO model	full specification model
Intercept	-3.722**	-7.170***
	(1.464)	(2.149)
cpi	0.019	0.022
	(0.031)	(0.050)
dyn_gdp	-0.112	-0.151
	(0.101)	(0.117)
interest_rate_US	-0.304**	-0.696***
	(0.132)	(0.220)
dyn_consum	-0.088***	-0.082**
	(0.030)	(0.037)
dyn_fx_rate	0.005	0.014
	(0.012)	(0.013)
diff_priv_credit_gdp	-0.014	-0.021
	(0.021)	(0.031)
net_lending	-0.092*	-0.103
	(0.054)	(0.063)
public_debt	0.039***	0.041***
	(0.009)	(0.012)
interest_on_debt	-0.105	-0.102
	(0.079)	(0.098)
ca_balance	-0.104***	-0.110**
	(0.037)	(0.047)
diff_unempl	0.316**	0.307*
	(0.145)	(0.167)
dyn_prod_dol	-0.013	-0.008
	(0.088)	(0.082)
overvaluation		-0.002
		(0.011)
dyn_gdp_china		0.320***
		(0.102)
dyn_export_share		0.034*
		(0.020)
oil_yoy		-0.007
		(0.007)
dyn_fix_cap_form		-0.016
		(0.017)
VIX		0.057**
		(0.026)
Log-likelihood Ratio	-219.054	-211.581
Log-likelihood Null	-445.508	-445.508
Pseudo-R sq	0.508	0.525
No. observations	987	987

Exploratory factor Analysis

Factor analysis is a statistical method used to describe the variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. In the next section we employ factor analysis to reduce the 20 variables considered in the logit model to 7 factors. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. We perform two statistical tests before we move on to factor analysis:

Bartlett's Test of Sphericity

Bartlett's Test of Sphericity compares an observed correlation matrix to the identity matrix. Essentially it checks to see if there is a certain redundancy between the variables that we can summarize with a few number of factors.

This test is often performed before we use a data reduction technique such as principal component analysis or factor analysis to verify that a data reduction technique can actually compress the data in a meaningful way.

The null hypothesis of the test is that the variables are orthogonal, i.e. not correlated. The alternative hypothesis is that the variables are not orthogonal, i.e. they are correlated enough to where the correlation matrix diverges significantly from the identity matrix. We present the results of the test here:

$$\chi^2 = 8606.20 \qquad \text{p-value} = 0.0$$

The test is statistically significant. We accept the alternate hypothesis that the observed correlation matrix is not an identity matrix.

Kaiser–Meyer–Olkin test

KMO is a test conducted to examine the strength of the partial correlation (how the factors explain each other) between the variables. KMO values closer to 1.0 are consider ideal while values less than 0.5 are unacceptable. We estimate an overall KMO of data = 0.6019, which indicates high strength of the partial correlation giving evidence that factor analysis can be appropriate.

Factor Analysis

We estimate the eigenvalues in order to choose the number of factors required. The scree plot is shown in figure 3. A Scree Plot is a simple line segment plot that shows the eigenvalues for each individual factor in model. The eigenvalues for 7 factors are greater than 1. Therefor we choose 7 factors in the dynamic factor model estimated in the next section.

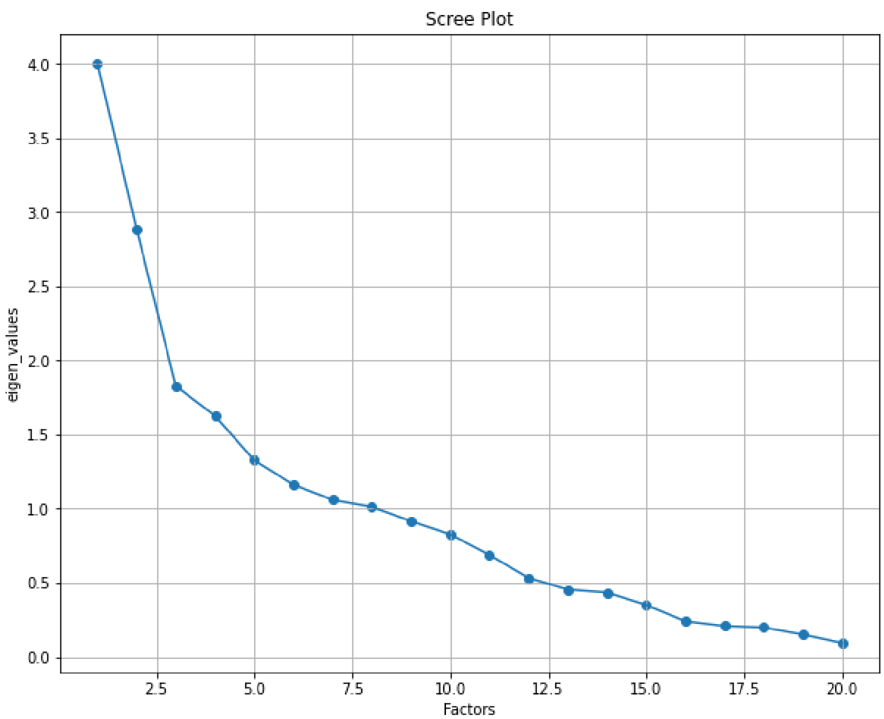

Figure 3: Scree Plot

Dynamic Factor Model with panel data logit

In this section we use the time series version of factor analysis i.e Dynamic Factor analysis as a dimensionality reduction technique. We estimate 7 dynamic factors and use these factors in determining the the probability that a country is in a stress period. The steps in the model are described below:

- Step 1: We estimate a different dynamic factor model for each country
- Step 2: We use the factors obtained to estimate a panel logit regression with fixed effects and heteroskedasticity and autocorrelation corrected standard errors.

DFM Specification

For each country we estimate the following set of equations

$$Y_t = \Lambda f_t + \epsilon_t \tag{1}$$
$$f_t = \beta f_{t-1} + u_t \tag{2}$$

- Y_t contains all the explanatory variables
- Λ is the matrix of factor loadings
- f_t is the matrix of the common factors which has a VAR(1) specification

- u_t is the error term which is modelled as iid Multivariate Normal RV

We use an ad-hoc specification for the characteristics of the model. The following decisions are not based on scientific reasons. Rather they are based on intuition. A through analysis and scientific logic is required for the correct model specification. The decision for the number of factors is taken:

- Number of Factors: We divide the variables into four groups (shown in table 4) based on the macroeconomic category they fall under. For each of these four groups we estimate one factor which only loads on the variables present in the group. Additionally we estimate two factors which load on all the variables. (Instead of doing var-max after the factor estimation we are specifying beforehand which factor loads on which variable)

	Variables	Macro	Domestic	Ficial	Fiscal	Labor
0	US interest rates	Yes				
1	US GDP dynamics	Yes				
2	China GDP dynamics	Yes				
3	Oil price dynamics	Yes				
4	VIX	Yes				
5	GDP dynamics	Yes				
6	GDP per capita	Yes				
7	Currency overvaluation		Yes			
8	Current account balance		Yes			
9	Export share		Yes			
10	Fixed capital formation		Yes			
11	Consumer Price index		Yes			
12	Consumption		Yes			
13	FX rate			Yes		
14	Credit to GDP change			Yes		
15	Net lending				Yes	
16	Public debt				Yes	
17	Interest on debt				Yes	
18	Unemployment change					Yes
19	Labor productivity dynamics					Yes

Table 4: grouping of variables

Obtaining the 7 factors for each country, we use these as exogenous variables to run a logit model. Due to the inclusion of these factors, we cannot clearly interpret the results of the regression coefficients. To exploit the panel nature of the data, we use a fixed effects specification (ϕ_i). In the regression we also include the lagged term of the dependent variable i.e stress during current period to predict the stress during next period.

$$stress_{it} = \gamma f_{it} + stress_{it-1} + \alpha_i + \phi_{it} \tag{3}$$

We have to correct for autocorrelation due to the inclusion of lagged dependent variables. This is done using heteroskedasticity-autocorrelation robust standard error estimation. The

Table 5: Logit model using Dynamic factors

	stress factors
crisis_next_year.shift(1)	2.910***
	(0.277)
global_1	0.286***
	(0.063)
global_2	0.008
	(0.071)
Macro	-0.011
	(0.092)
domestic	-0.162
	(0.122)
fiscal	0.038
	(0.107)
labor	0.239**
	(0.118)
Log-likelihood Ratio	-197.096
Log-likelihood Null	-443.716
Pseudo-R sq	0.556
No. observations	986

results of the regression are presented in table 5. Variables "global_1" and "global_2" are the factors which load on all the variables. Rest of the factors load on their corresponding variables

Conclusion

Aim of the study was to determine which macroeconomic variable was instrumental in determining fiscal stress period in a country. We find that US Interest rate, change in national consumption, public debt, and current account balance significantly influence stress period incidence in a country. Effective management and monitoring of public debt and current account balance can help policy makers manage fiscal risks.

We use extensions of different multivariate methods taught in class. The personal aim of this analysis was to expand our toolkit of statistical models so that we can effectively apply them to economic problems.