IVANA GANEVA & RANA MOHIE

Forecasting Currency Crises

Comparison of Conventional Methods and of a Neural Networks Model

Thesis Advisor: Prof. Luis Rojas

Motivation

To contribute to the literature on forecasting currency crises as such crashes continue to pose threats on developing economies and have a significant welfare impact.

To explore a new approach that leverages similar countries experience with crisis that can improve the prediction under the low data availability conditions in most developing countries

Some Theoretical Background

The Four-Generation Currency Crisis Model

FIRST GENERATION

'A crisis is predictable from Macroeconomic fundamentals' deterioration.'

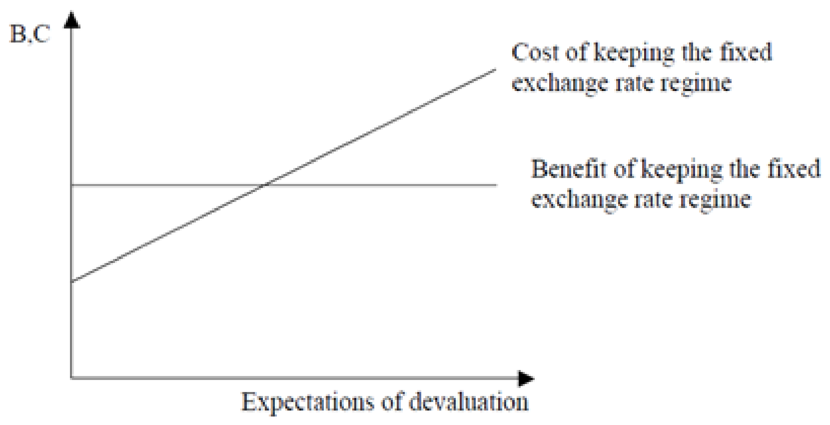
Fixed Exchange Regime
\+ Persistent Deficit

A Currency Crisis

SECOND GENERATION

Introducing the role of Expectations.

'The occurrence of a crisis is somewhat random.'



THIRD GENERATION

Introducing the impact from banking and the financial sector.

'A model for currency crises should include financial indicators.'

FOURTH GENERATION

Introducing the contagion effect and the political factors effect.

'Political indicators can be included in prediction.'

The Empirics - Selecting Explanatory Variables

Following the Approach of Kaminsky & Reinhart, 1998

Performing a survey of 8 Papers with 105 different indicators.

Creating a ranking of all indicators based on their

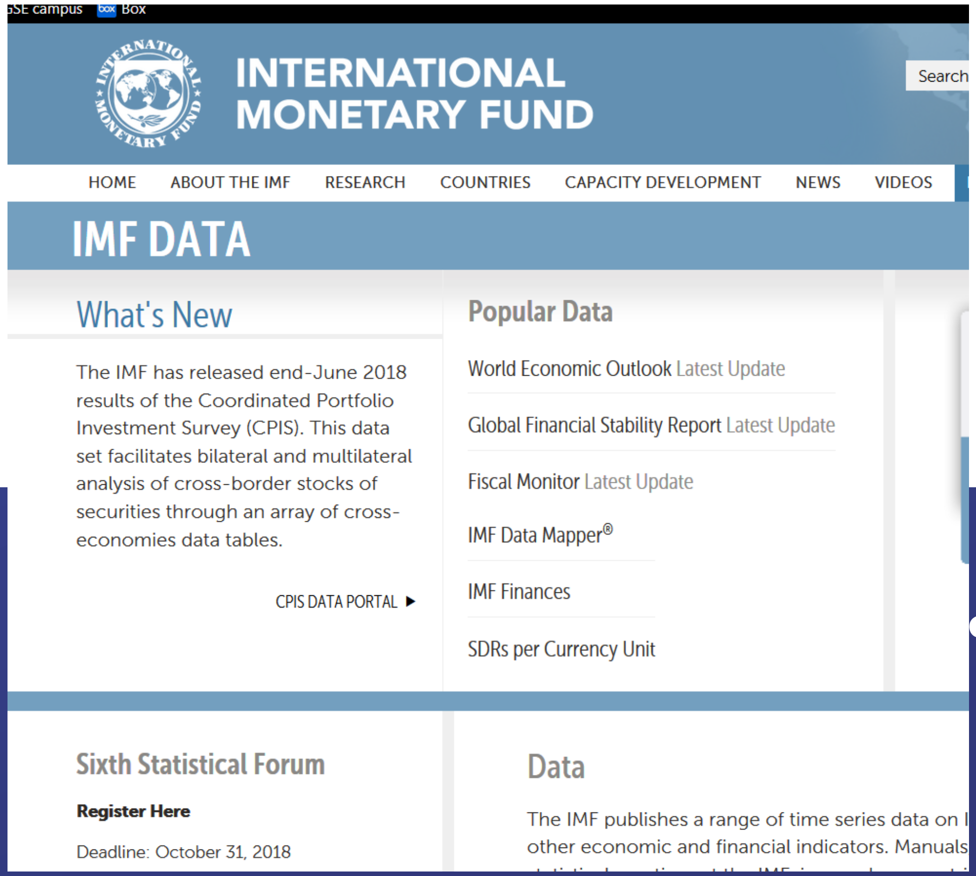
Predictive power: they were used to predict probability of crisis and their out of sample ability was formally tested.

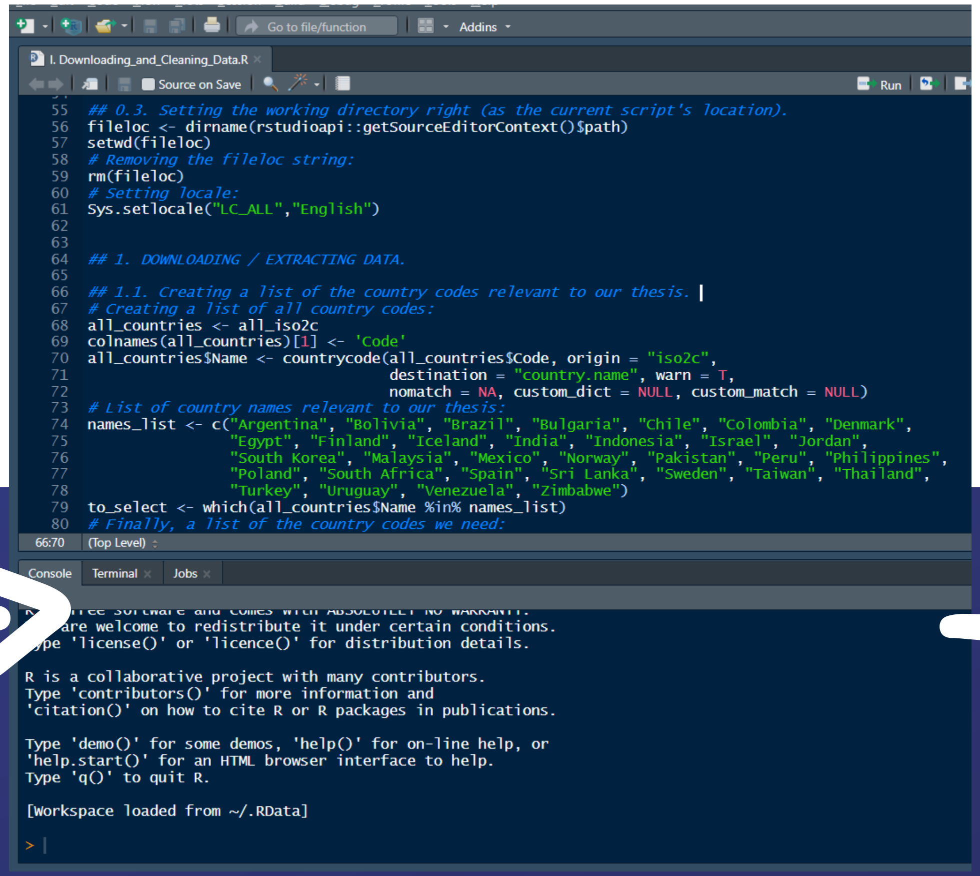
False Signals: their behaviour in crisis period was systematically compared to their behaviour in the 'tranquil' time.

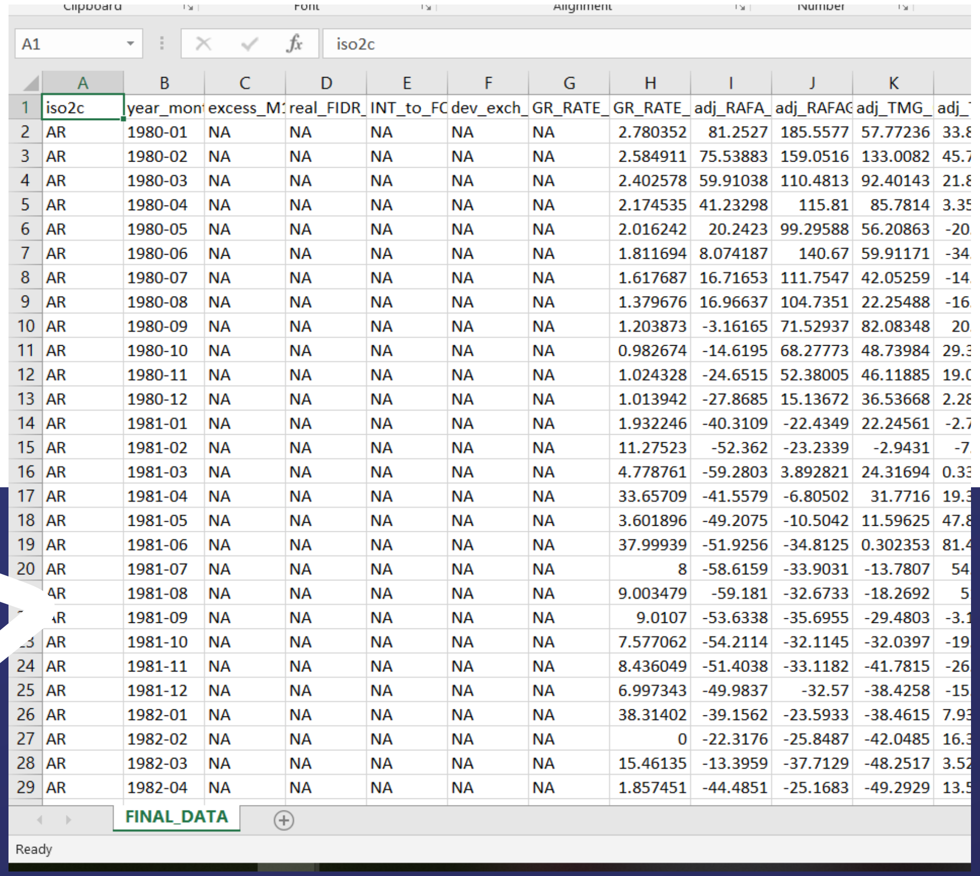
Indicator
International Reserves (measured in USD)
Imports (measured in USD)
Exports (measured in USD)
The Terms of Trade (defined as the unit value of exports over the
Deviations of the Real Exchange Rate from Trend (in percen
The Differential Between Foreign (U.S./German) and Domestic Real Interest Rates on Deposits (monthly rates, deflated using consumer prices and measured in perc
"Excess" Real M1 Balances
The Money Multiplier (of M2)
The Ratio of Domestic Credit to GDP
The Real Interest Rate on Deposits (monthly rates, deflated using consumer prices and measured in perc
The Ratio of (Nominal) Lending to Deposit Interest Rates
The Stock of Commercial Banks Deposits
The Ratio of Broad Money to Gross International Reserves
An Index of Output
An Index of Equity Prices (measured in USD)

The Final Set of 15 Indicators.

Downloading & Cleaning Our Data







IMF DATA

BOP & IFS DATABASES

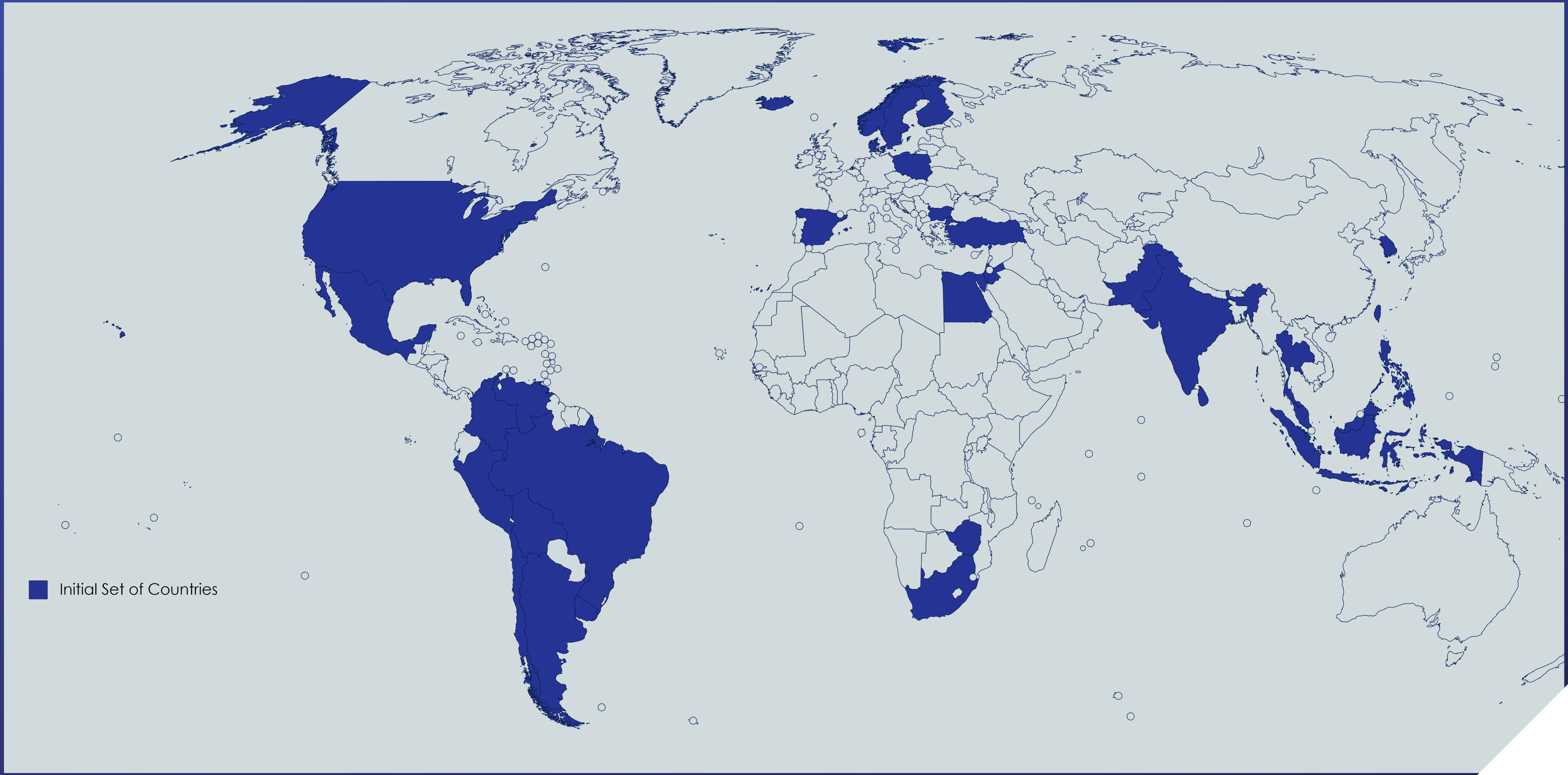
OUR R TOOL

FOR EXTRACTING & CLEANING DATA

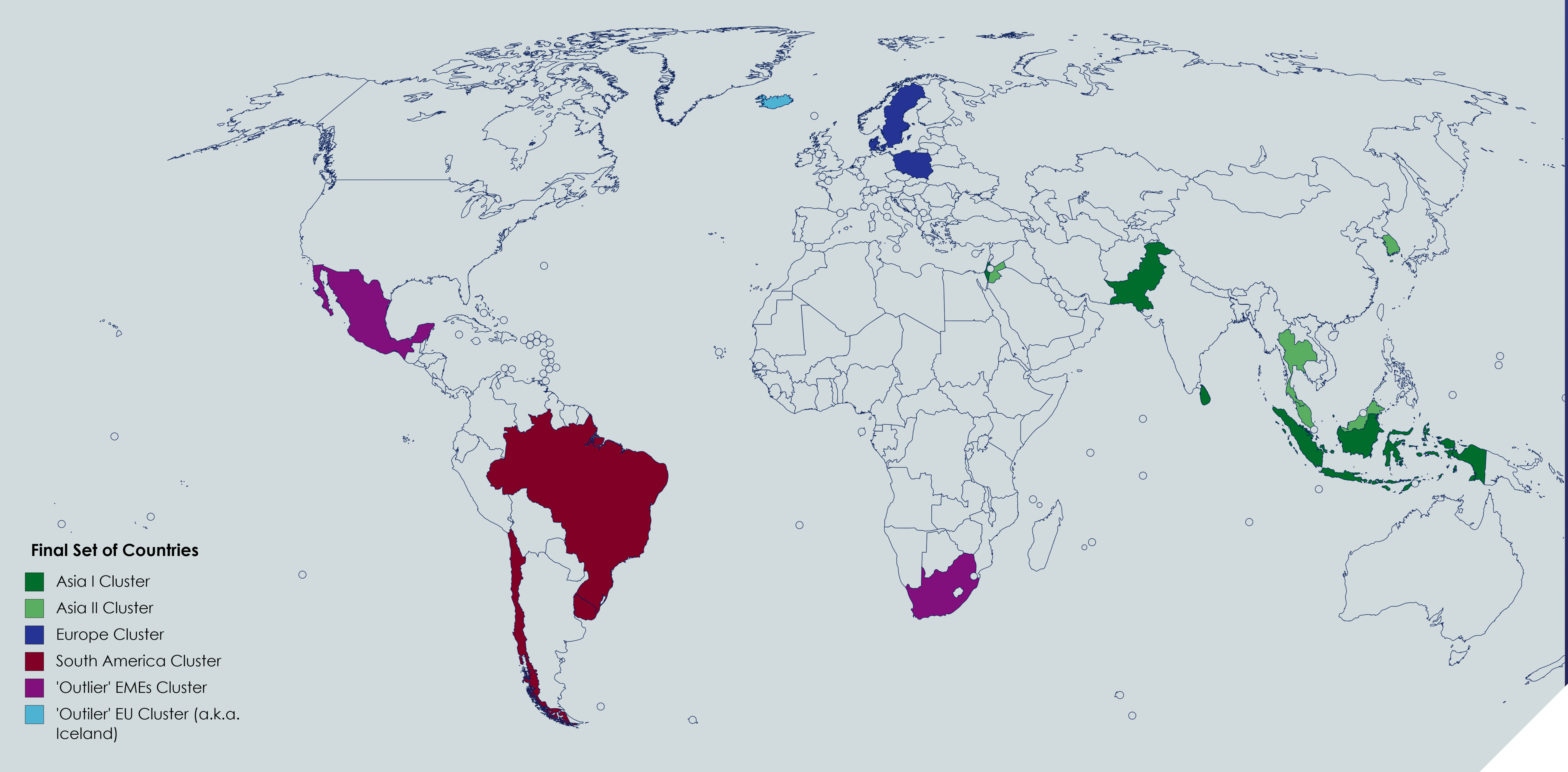
CSV FILES

WITH ALL DATA

Initial Set of 33 Countries Considered



Final Set of 17 Countries Considered



Defining What a Currency Crisis Is

DEFINITIONS FROM THE LITERATURE

DEFINITION I

Based on Exchange Rate - any depreciation above a pre-specified threshold considered a crisis.

DEFINITION II

Based on a Composite Index, which is the weighted average of exchange rate reserves and interest rates, e.g.

OUR DEFINITION

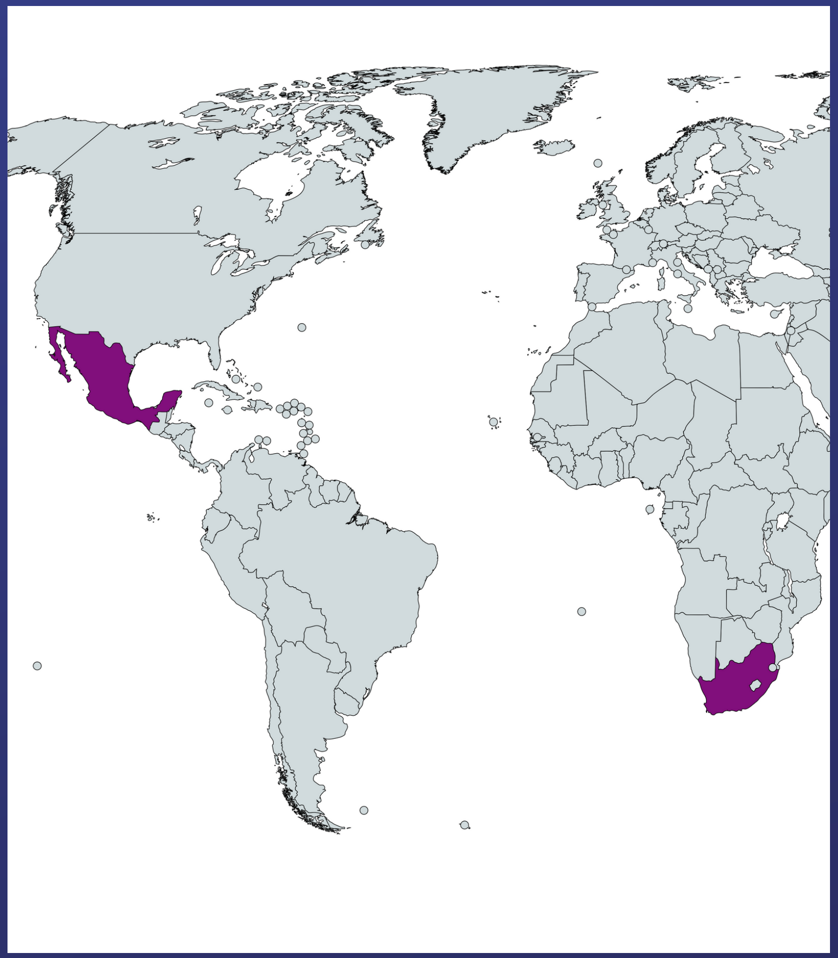
THE MIDDLE-GROUND DEFINITION

Suggesting a threshold that takes into account countries' specifics, without imposing the same frequency of crises among all of them.

The Clustering Exercise - Idea

IDEA: Countries that have similar economies, or/and are geographically close to each other, should go in the same cluster.

LOCATION



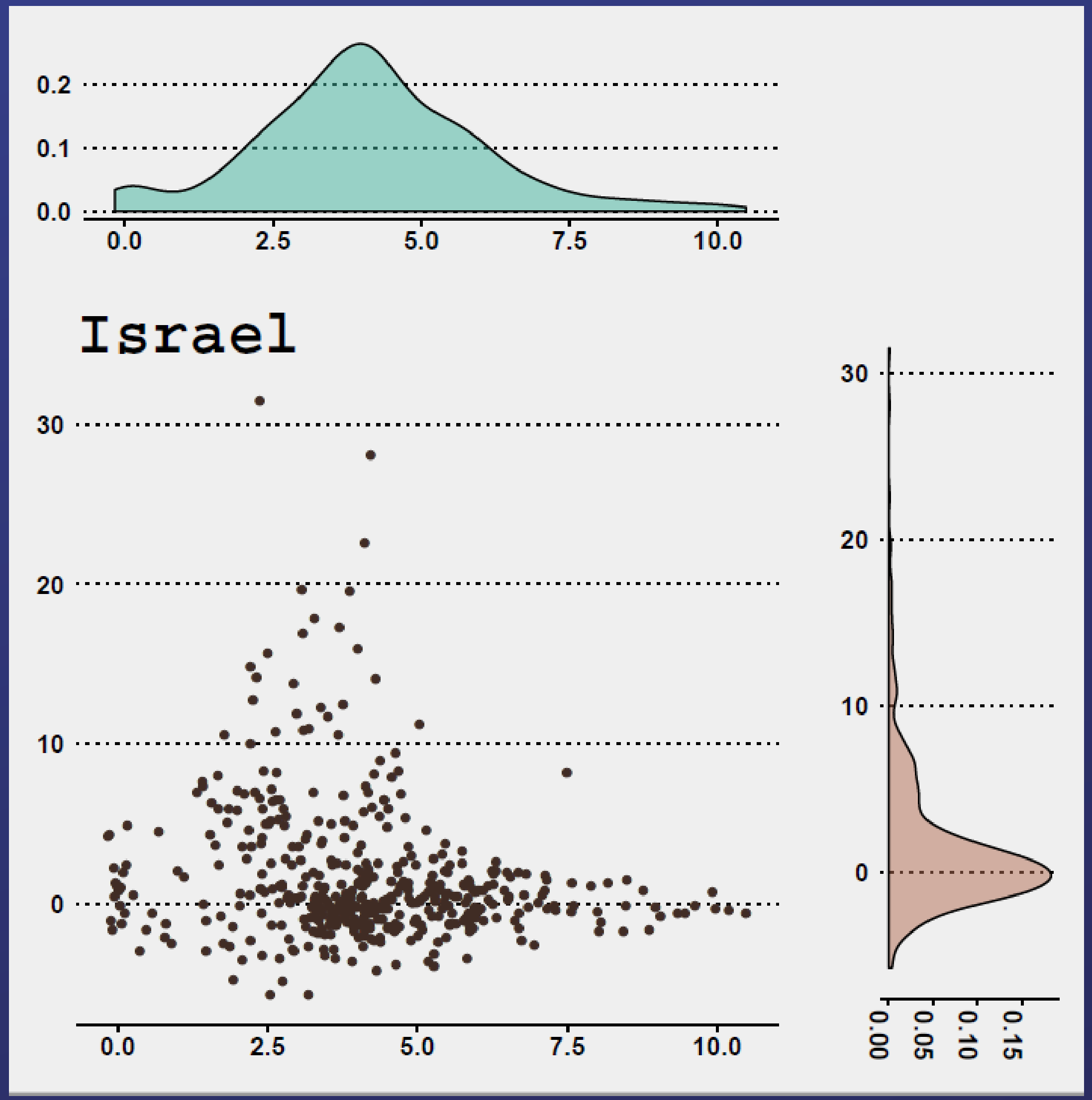
CHARACTERISTICS

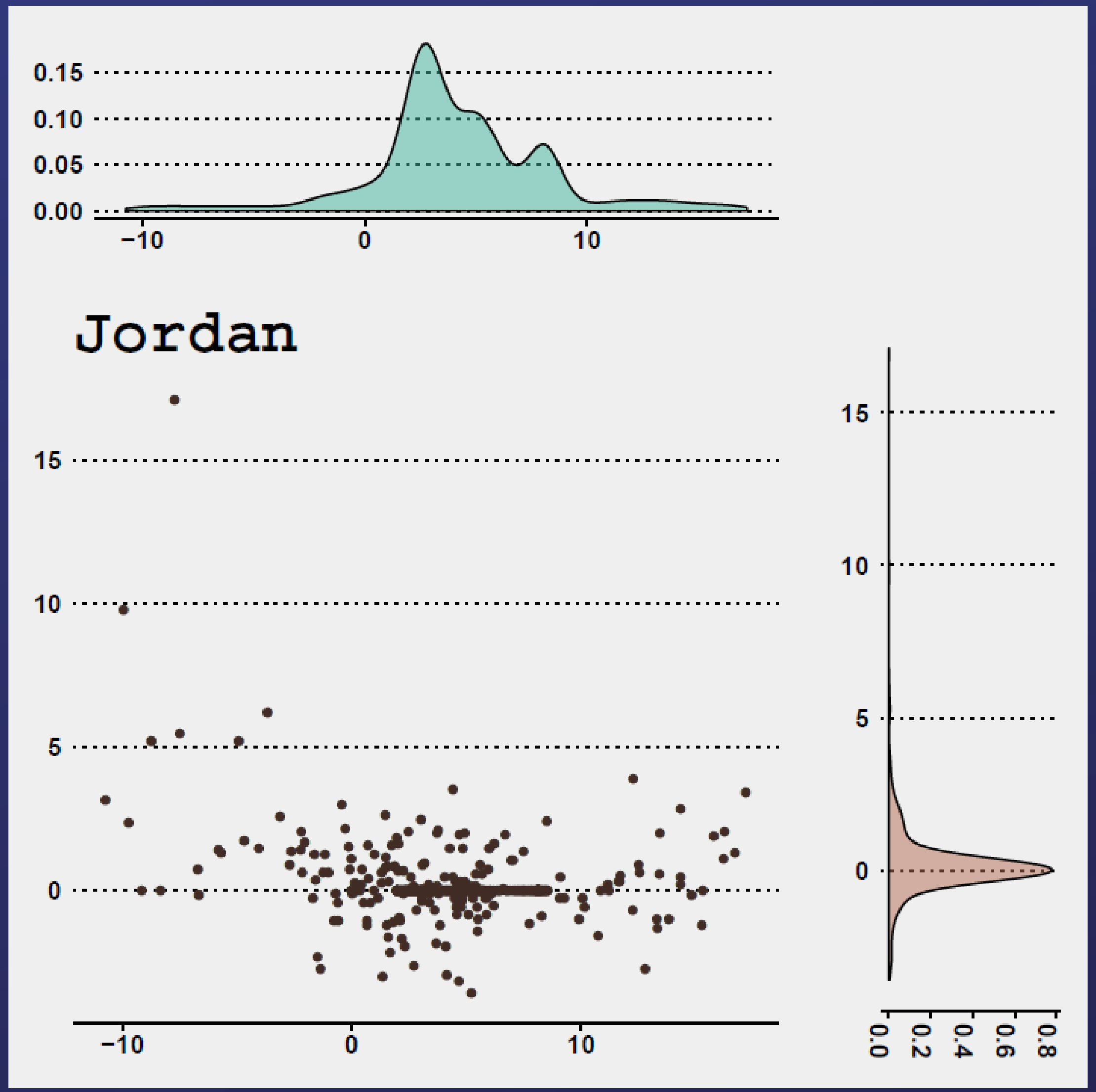
BOTH

The Clustering Exercise - Method

METHOD: Visual exploration of the countries' joint and marginal distributions on their GDP growth rate and on their exchange rate fluctuations. Formal testing for normality using Kolmogorov-Smirnov.

Why didn't Israel and Jordan end up in the same cluster?





Model I - The Probit & Its Methodology

As seen in Berg & Patillo (1999)

STEP 1

Dependent variable considered to be the probability of observing a crisis in the 18-months window to follow.

STEP 2

Estimating a loop of Probit models on the dependent variable, considering one indicator from our set at a time.

$$\mathbb{P}(y_t|x_t) = (\psi((x_t)'\beta)$$

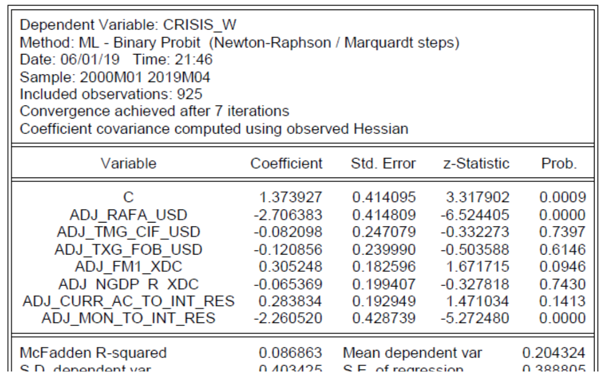
STEP 3

Estimating the General Multivariate Probit model using only the indicators that showed significance in the previous step.

$$\mathbb{P}(y_t|X_t) = (\psi((x_{it})'\beta)$$

STEP 4

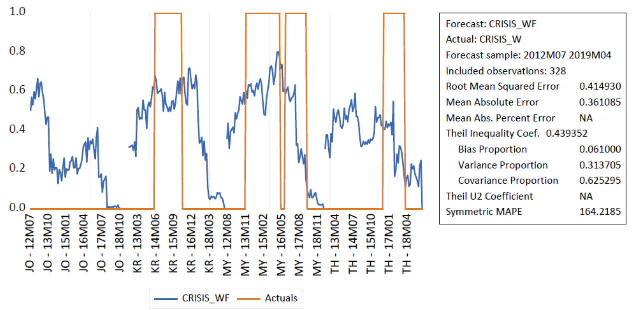
Interpretation.



Dependent Variable: CRISIS_W
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)
Date: 06/01/19 Time: 21:46
Sample: 2000M01 2019M04
Included observations: 925
Convergence achieved after 7 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.373927	0.414095	3.317902	0.0009
ADJ_RAFA_USD	-2.706383	0.414809	-6.524405	0.0000
ADJ_TMG_CIF_USD	-0.082098	0.247079	-0.332273	0.7397
ADJ_TXG_FOB_USD	-0.120856	0.239990	-0.503588	0.6146
ADJ_FM1_XDC	0.305248	0.182596	1.671715	0.0946
ADJ_NGDP_R_XDC	-0.065369	0.199407	-0.327818	0.7430
ADJ_CURR_AC_TO_INT_RES	0.283834	0.192949	1.471034	0.1413
ADJ_MON_TO_INT_RES	-2.260520	0.428739	-5.272480	0.0000

McFadden R-squared	0.086863	Mean dependent var	0.204324
S.D. dependent var	0.403425	S.E. of regression	0.388805



Note: We perform these steps twice - once for our country-by-country data, and once as a panel regression on our clusters data.

Model I - The Probit & Our Results

Comparing Our Out-of-Sample Results to the Literature Ones

% OF CORRECTLY CALLED OBSERVATIONS			
Threshold	Berg & Patillo, 1999	Our Probit	
	Panel on all the countries	Country-by-country	Cluster panels
0.25	79%	-	-
0.50	78%	71%	77%
0.80	-	71%	76%

% OF FALSE ALARMS (OUT OF TOTAL ALARMS)			
Threshold	Berg & Patillo, 1999	Our Probit	
	Panel on all the countries	Country-by-country	Cluster panels
0.25	49%	-	-
0.50	-	**62%**	**67%**
0.80	-	61%	66%

% OF CRISES CORRECTLY CALLED			
Threshold	Berg & Patillo, 1999	Our Probit	
	Panel on all the countries	Country-by-country	Cluster panels
0.25	80%	-	-
0.50	0%	**80%**	**54%**
0.80	-	63%	9%

Quick Summary:

Berg & Patillo's clustering of all countries together generally works worse than our way of clustering the data.

Our choice of a 'middle-ground' crisis definition.

Our attempts not to miss important country- or cluster-specific traits.

Model II - The Markov Switching Algorithm

As seen in Abiad (2007)

STEP 1

Dependent variable considered to be the depreciation of the nominal exchange rate itself.
(Not binary & no need of introducing a window to the crises like in the Probit.)

STEP 2

Giving the estimation an initial value for the probability of being at a crisis at period 1. This choice depends on the stationarity of our indicators.

$$y_t|S_t \quad \text{i.i.d.}, \sim \mathcal{N}(\mu_{S_t}, \sigma^2_{S_t})$$

STEP 3

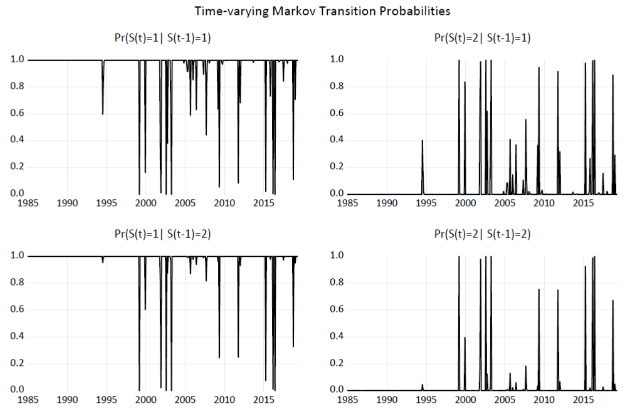
Forming forecasts by iterating a set of equations that give us the conditional probability of observing a crisis at t+1 given the information until t.

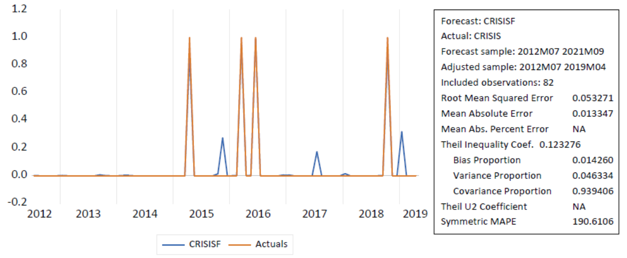
$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \cdot \phi_t}{\mathbb{I}'(\hat{\xi}_{t|t-1} \cdot \phi_t)}$$

$$\hat{\xi}_{t+1|t} = \mathbb{P}'_{t+1}(\hat{\xi}_{t|t})$$

STEP 4

Letting EViews do the whole job here and simply interpreting.





The Models I & II - In-Sample Performance

Comparing our In-Sample Results Across Models and to the Literature Results

% OF CORRECTLY CALLED OBSERVATIONS, IN-SAMPLE					
Threshold	Berg & Patillo, 1999	Our Probit		Abiad, 2007	Our MS
	Panel on all the countries	Country-by-country	Cluster panels	Country-by-country	Country-by-country
0.25	78%	-	-	-	-
0.50	84%	88.0%	79.0%	71%	70%
0.80	-	83.7%	79.3%	-	-

In Summary: We get **comparable results** to the ones in the literature when using the conventional methods.

A Very Mild Introduction to Neural Networks

1

The universal approximation property (Hornik, Stinchcombe, White, 1989)

Any function uniformly continuous on compact sets can be approximated by a multilayer feedforward network by (arbitrarily) increasing the number of nodes in the hidden layers.[a]

[a] Hornik, Stinchcombe, White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359-366

2

A kernel k is a function

$$\begin{aligned} k : X \times X &\to \mathbb{R} \\ (\boldsymbol{x}, \boldsymbol{y}) &\to k(\boldsymbol{x}, \boldsymbol{y}) \end{aligned}$$

define as

$$k(\boldsymbol{x}, \boldsymbol{y}) = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{y}) \rangle$$

where Φ maps into some dot product space F, called the feature space

3

Thm.: A function $k : X \times X \to \mathbb{R}$ is a kernel iff it induces a positive definite marix

positive-(semi)definite: for any *finite* family of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of X, the matrix

$$K = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & k(\boldsymbol{x}_1, \boldsymbol{x}_2) & \ldots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ k(\boldsymbol{x}_2, \boldsymbol{x}_1) & k(\boldsymbol{x}_2, \boldsymbol{x}_2) & \ldots & k(\boldsymbol{x}_2, \boldsymbol{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1) & k(\boldsymbol{x}_n, \boldsymbol{x}_2) & \ldots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix}$$

is positive semidefinite iff $\forall \boldsymbol{z}, \boldsymbol{z}^t K \boldsymbol{z} \geq 0$

K is called the Gram (or kernel) matrix of $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$

4

The problem with slack variables is

minimize $\frac{1}{2}\|\boldsymbol{w}\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$

subject to $\begin{cases} y_i - \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$

- The constant $C \geq 0$ determines the trade off between the flatness of f and the accuracy of the model
- This formulation corresponds to dealing with the so called ε-insensitive loss function $|\xi|_\varepsilon$

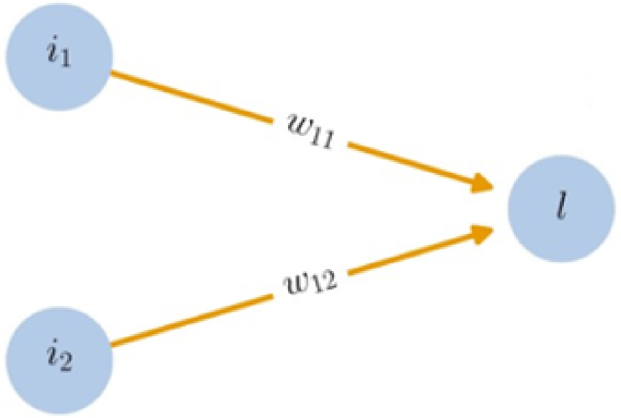
$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \end{cases}$$

No worries if you forgot the theorem already.

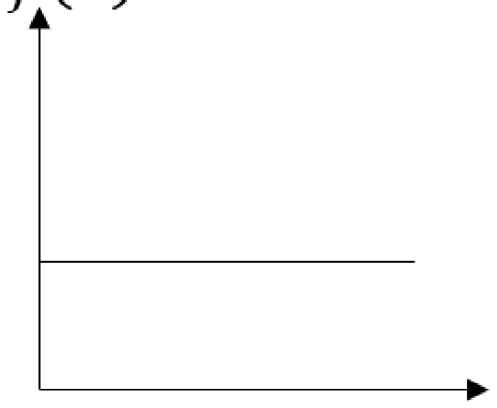
IT WAS A JOKE.

From a Linear Regression to an ANN

$$y = \beta_1 x_1 + \beta_2 x_2$$

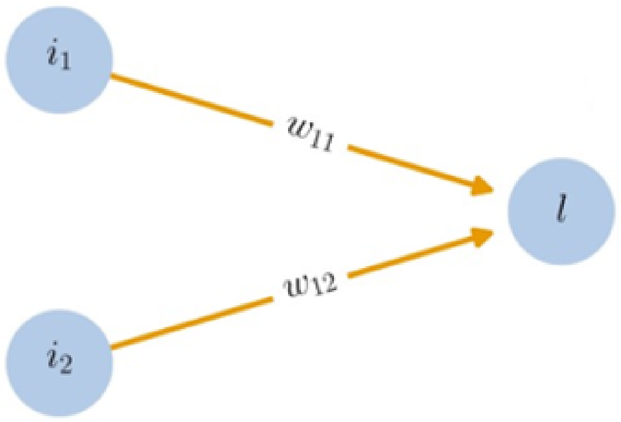


$f(x) = 1$

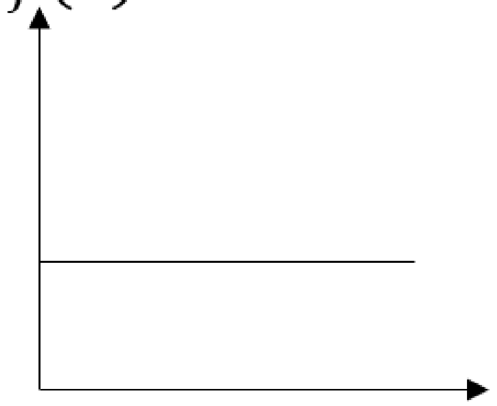


From a Linear Regression to an ANN

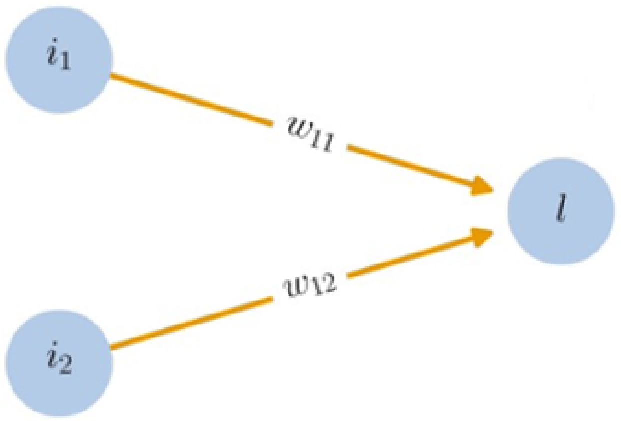
$$y = \beta_1 x_1 + \beta_2 x_2$$



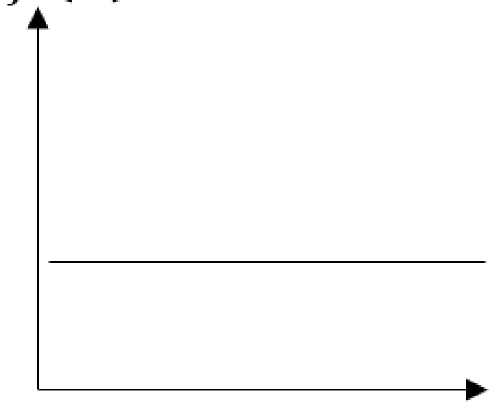
$f(x) = 1$



$$P(y|x_1, x_2) = \psi(\beta_1 x_1 + \beta_2 x_2)$$

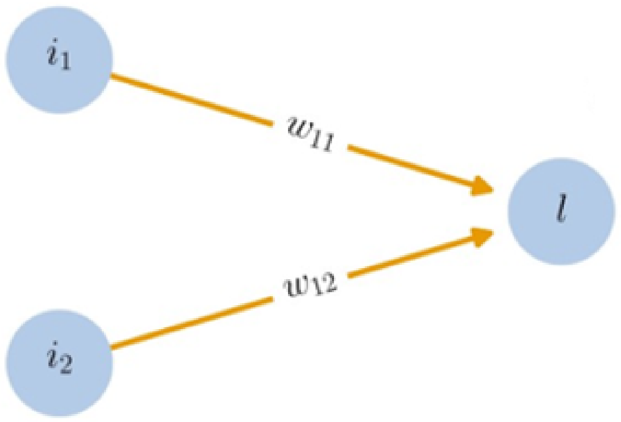


$f(x) = 1$

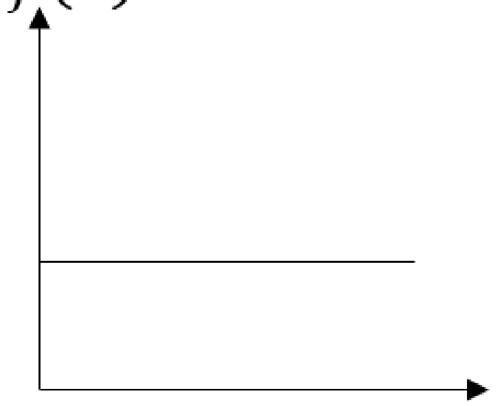


From a Linear Regression to an ANN

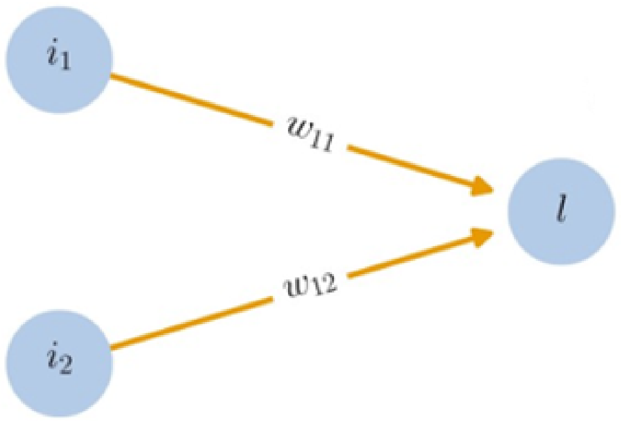
$$y = \beta_1 x_1 + \beta_2 x_2$$



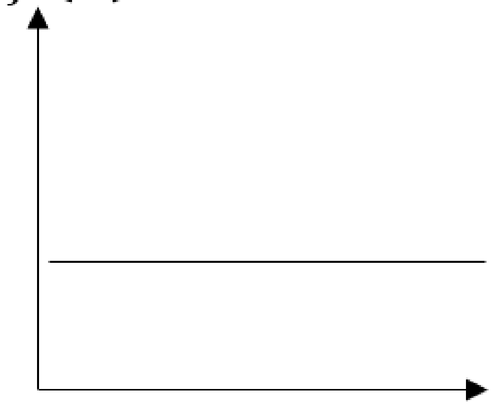
$f(x) = 1$



$$P(y|x_1, x_2) = \psi(\beta_1 x_1 + \beta_2 x_2)$$

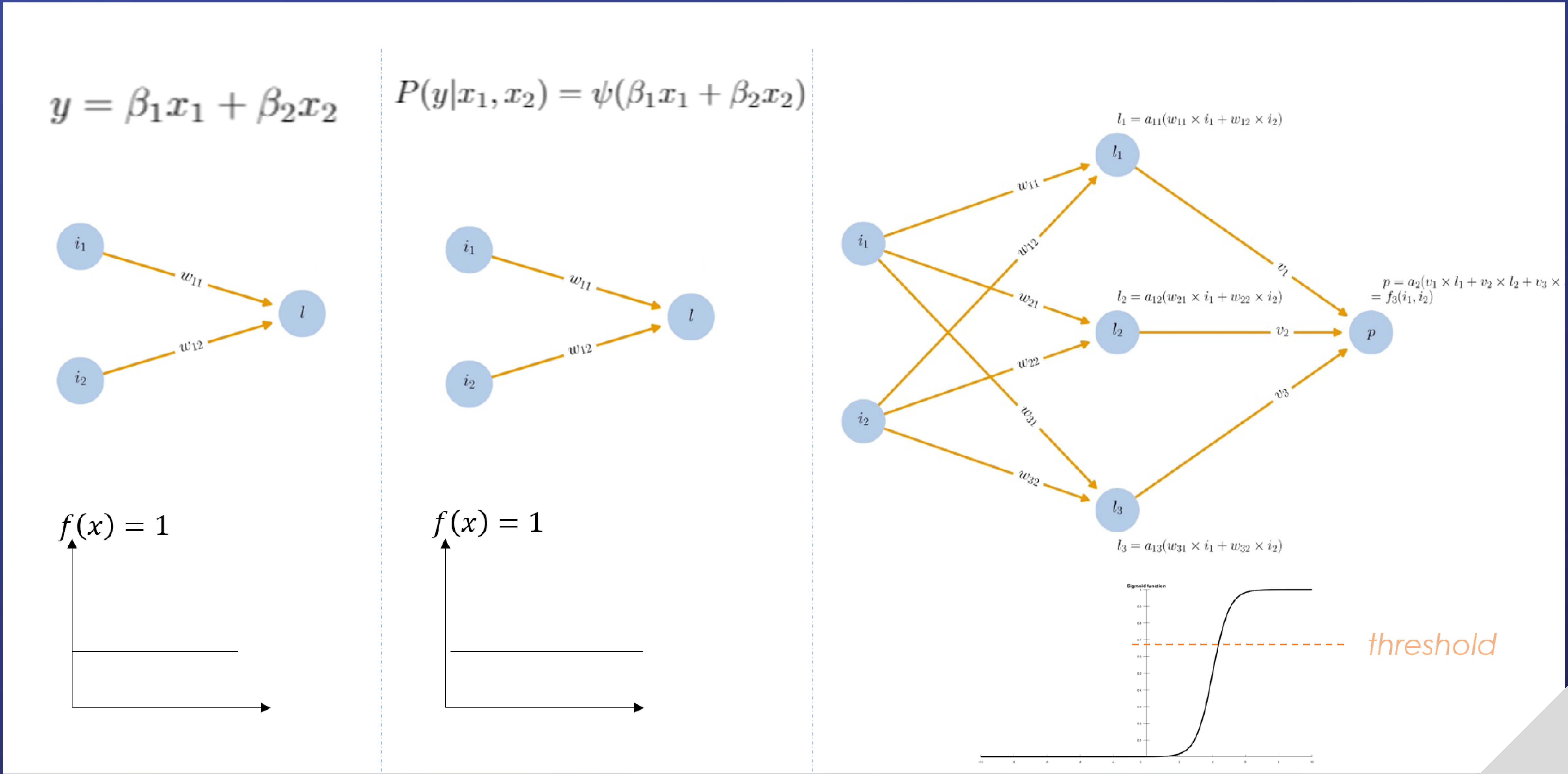


$f(x) = 1$



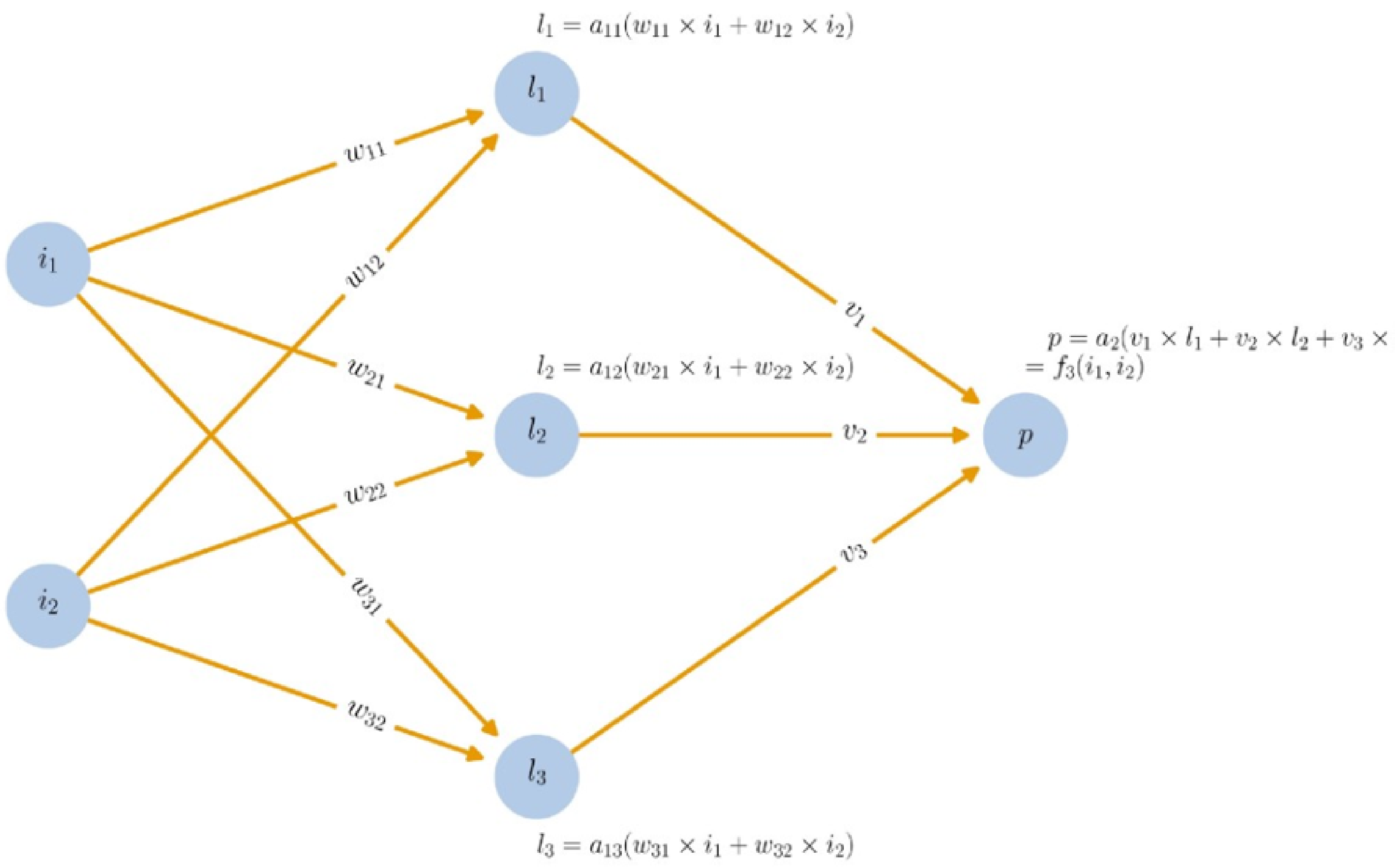
But…What if we didn't want to make assumptions about the model?

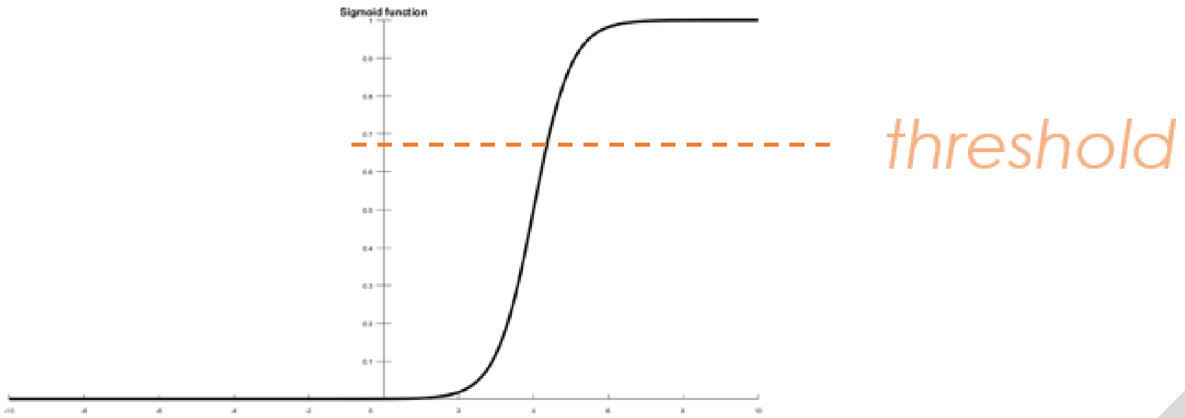
From a Linear Regression to an ANN



From a Linear Regression to an ANN

- Layers (Input, Output, Hidden)
- Nodes
- Activation Function.
- Weights -> β in Regression
- Threshold.
- Decay





Our Model III - An Artificial Neural Network

- Layers (Input, Output, Hidden)
- Activation Function.

Chosen based on the financial ANN literature

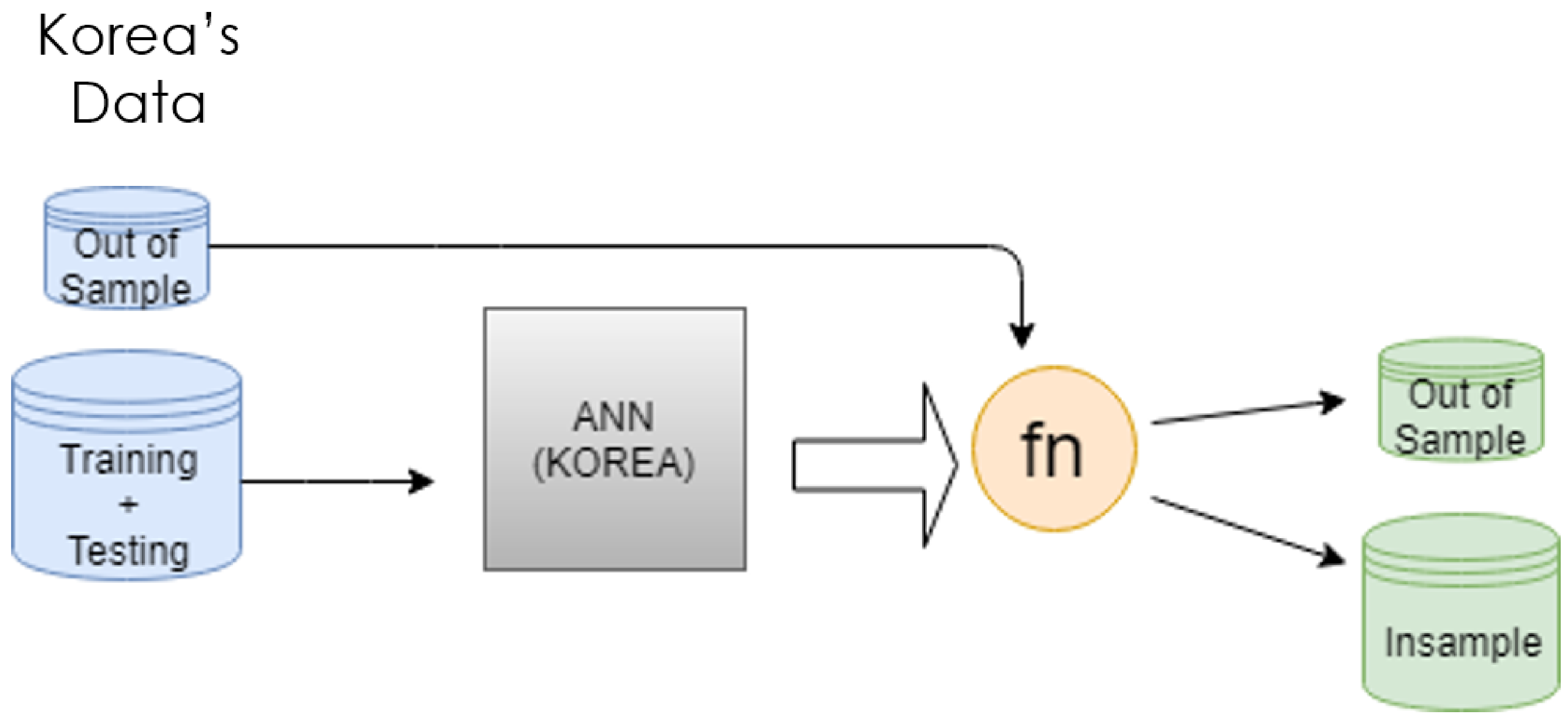
- Nodes
- Decay

Optimized by the Network to minimize the loss function in the Tuning. *(Known/initial value is given)*

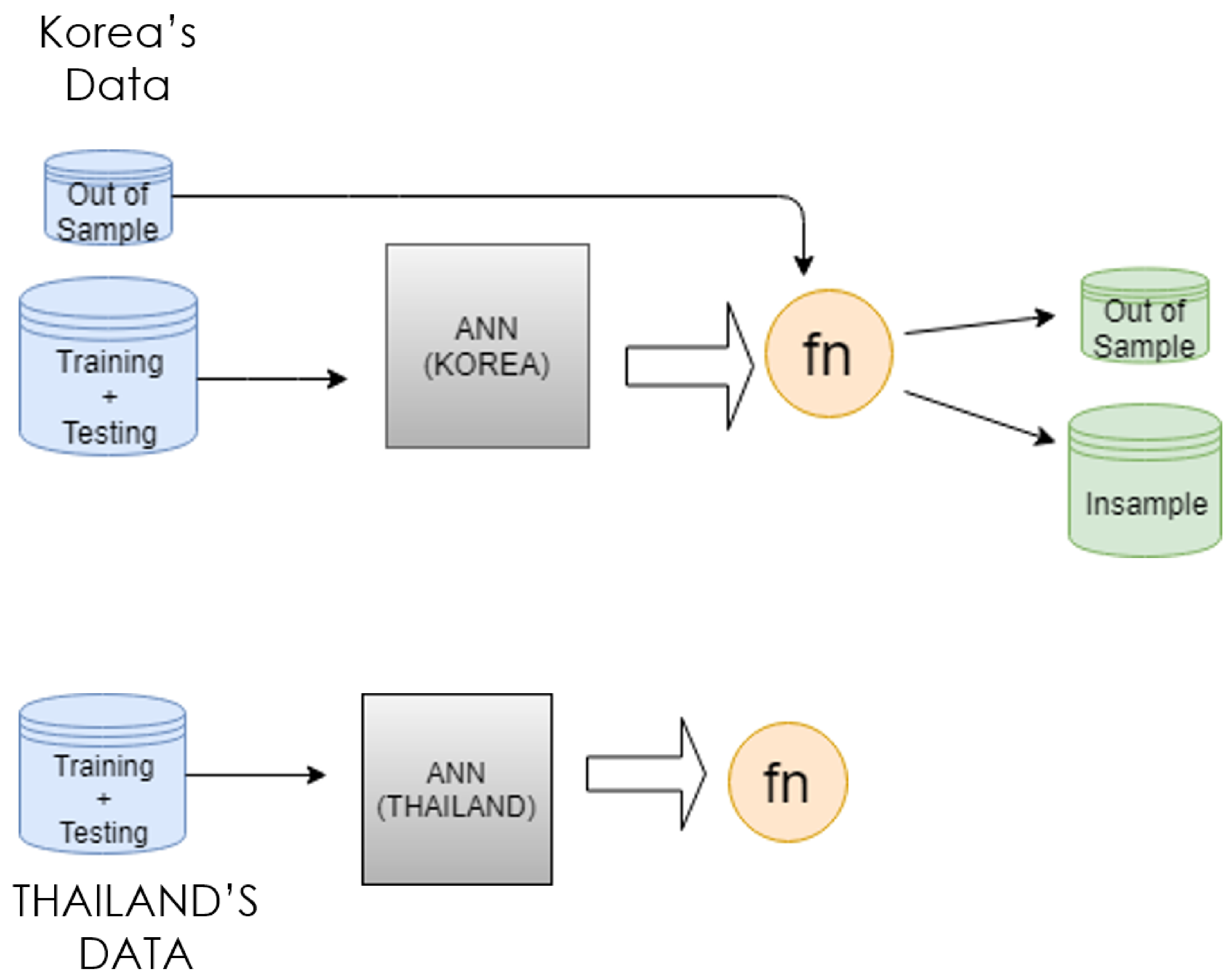
- Threshold.
- Weights -> β in Regression

Chosen by the Network in the training phase *(hidden)*

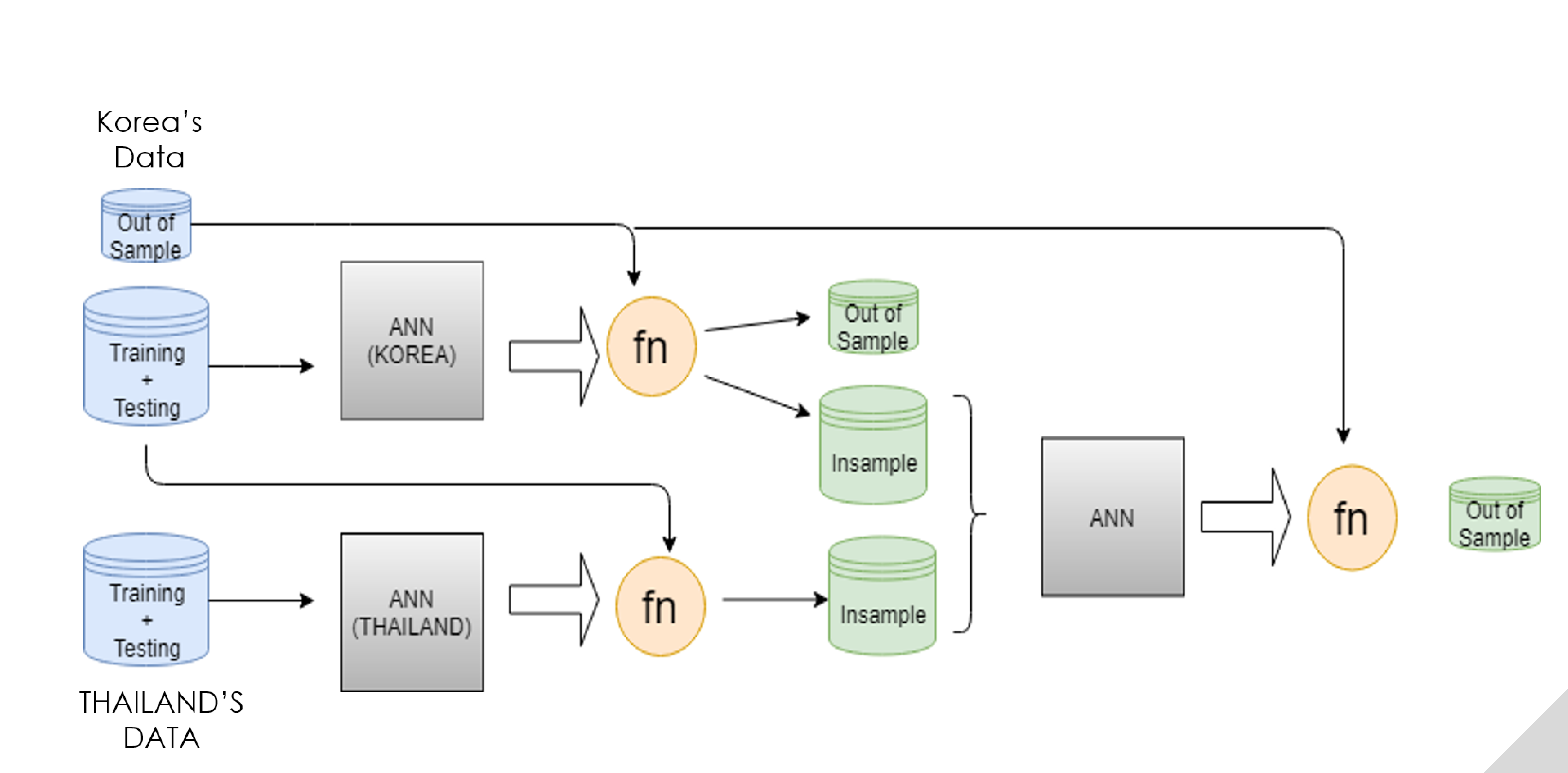
The ANN: One-Level and the Extended Model



The ANN: One-Level and the Extended Model



The ANN: One-Level and the Extended Model



The ANN: The One Thing to Remember

RESULTS: Clusters vs. Country-by-Country

Out-of-Sample Performance

	50% Threshold			
	Probit Cluster	Probit Country	ANN Cluster	ANN Country
% of observations correctly called	**77**	71	74	75
% of crisis called out of total crises	54	80	**80**	73
% of false alarms out of total alarms	67	62	**55**	58
Total Number of Crises	11			

	80% Threshold			
	Probit Cluster	Probit Country	ANN Cluster	ANN Country
% of observations correctly called	**76**	71	75	75
% of crisis called out of total crises	9	63	**63**	45
% of false alarms out of total alarms	66	61	**56**	61
Total Number of Crises	11			

Key Result: The the extension on our original ANN model is actually an improvement. That is, letting one country learn from other's experience is NOT a distortion here!

RESULTS: Out-of-Sample Model Performance

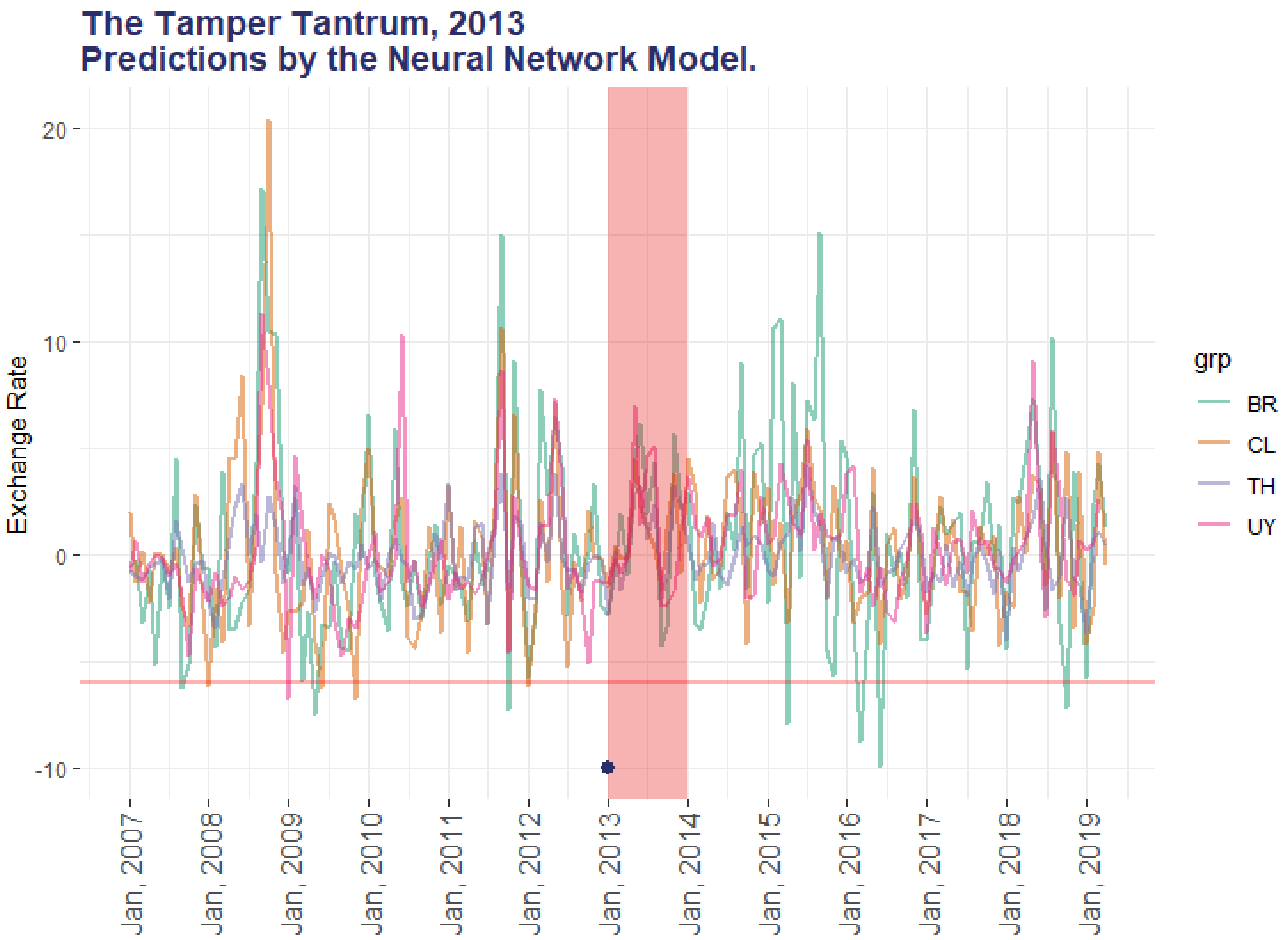
	50% Threshold			
	Probit Cluster	Probit Country	ANN Cluster	ANN Country
% of observations correctly called	77	71	74	75
% of crisis called out of total crises	54	80	**80**	73
% of false alarms out of total alarms	67	62	**55**	58
Total Number of Crises		11		

	80% Threshold			
	Probit Cluster	Probit Country	ANN Cluster	ANN Country
% of observations correctly called	76	71	75	75
% of crisis called out of total crises	9	63	63	45
% of false alarms out of total alarms	66	61	56	61
Total Number of Crises		11		

Key Result: All in all, the Neural Network models' performance is superior to the Probit's and should indeed be considered a better choice for forecasting currency crises.

RESULTS: 'Is it a false alarm indeed?'

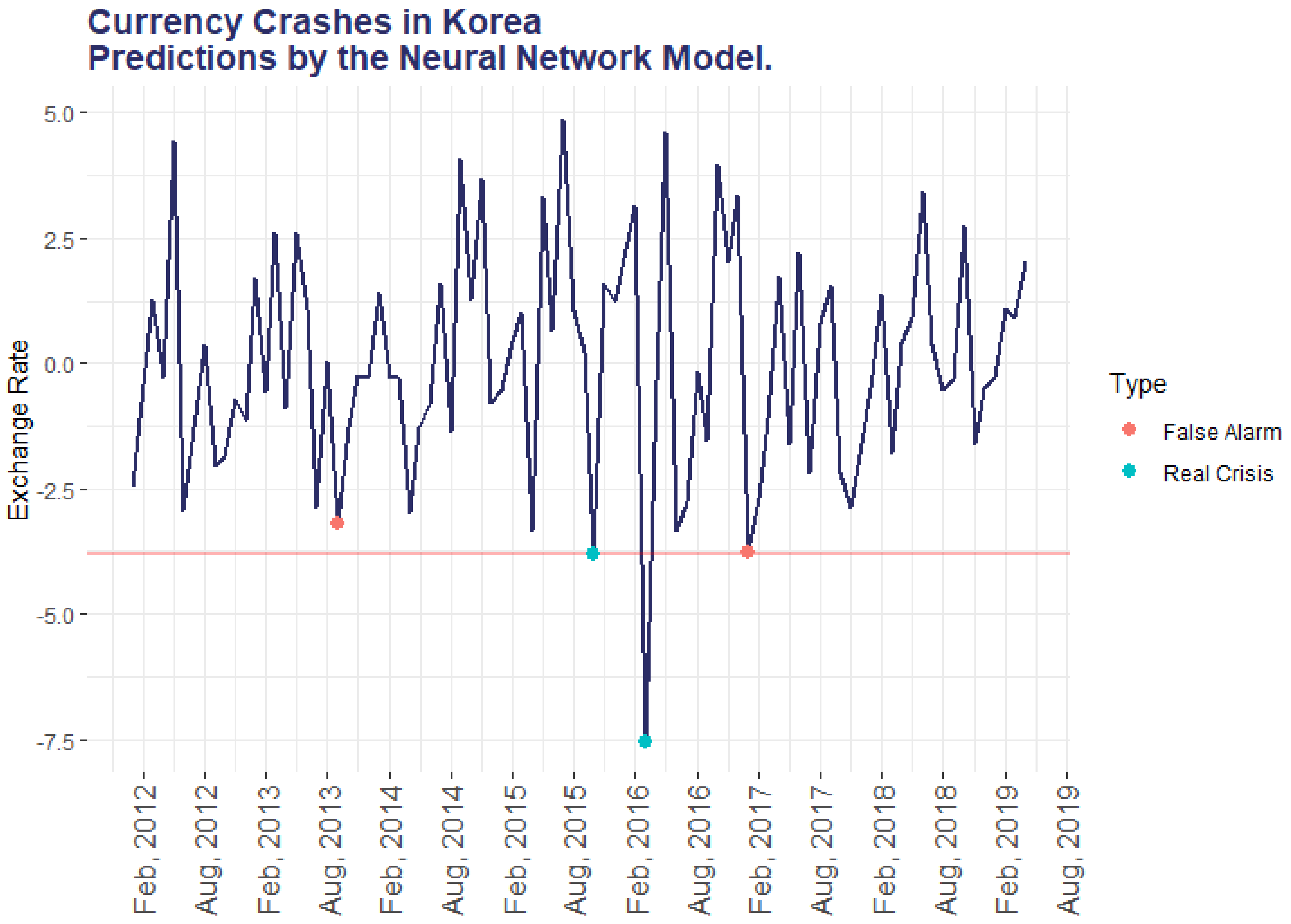
Answer: Yes, but there is some good reason behind it.



Our Suggestion: Might be wise to add an EME index, similar to the OECD composite growth index, that would help the network learn about any potential general EME effects.

RESULTS: 'Is it a false alarm indeed?'

Answer: It depends!



Our Suggestion: Create a measure that 'counts' how many of the currency crises are in the very neighbourhood of our pre-defined threshold. Also, run an ordered Probit instead of the standard one we have used, in order to capture crises' severity.

Our Humble Suggestions for Potential Further Research

THE THRESHOLD

Calculating it as the solution to an optimization problem.

PREDICTING INTENSITY

Estimating an Ordered Probit Model to allow for crises' intensity as well.

OPTIMIZING THE NN

Optimizing the number of variable lags which we have arbitrarily chosen so far.

MIXTURE OF EXPERTS

Developing our Neural Network further and including a Mixture of Experts element.

Conclusion

DATA SET

Comparing different models across a large data set, including observations for 17 countries.

CRISIS DEFINITION

Introducing the 'Middle-Ground' definition resulted in obtaining better results in Probit than the benchmark.

ML TECHNIQUES

Applying some image processing techniques to our Neural Network which has not been done before.

REPRODUCIBLE RESEARCH

Downloading & Cleaning Data has been automated, so that the results can easily be updated.

Q & A

Thank you for your attention!