Rethinking Early Warning Systems: Using the Radial Based Support Vector Machine to Forecast Currency Crises

by
Matthew John McMahon

Claremont Graduate University
2019

APPROVAL OF THE DISSERTATION COMMITTEE

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Matthew John McMahon as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy.

Thomas D. Willett, Chair
Claremont Graduate University
Horton Professor of Economics

Paul J. Zak
Claremont Graduate University
Professor of Economics, Psychology, and Management

Graham Bird
Claremont Graduate University
Clinical Professor of Economic Sciences

Abstract

Rethinking Early Warning Systems: Using the Radial Based Support Vector Machine to Forecast Currency Crises

by Matthew John McMahon

Claremont Graduate University: 2019

This paper proposes a new Early Warning System model for currency crises, based on the support vector machine. The support vector machine introduces the concept of structural risk minimization, an inductive process that uses resampling methods as a way to seek an upper-bound for the generalization error. Traditional methods such as limited dependent regressions use empirical risk minimization, which seeks to minimize training error. The support vector machine stems from the field of statistical machine learning and helps to overcome the shortcomings of traditional approaches such as non-linear relationships, the variance-bias tradeoff, forecasting accuracy, and the ability to gage the impact of individual variables within large datasets. The paper compares the support vector machine to the multinomial logit model of Busierre and Fratzscher (2006) to establish the SVM's usefulness in forecasting currency crises compared to the performance of conventional modelling techniques. The results show that the support vector machine would have correctly predicted a majority of crises between 1996 and 2014 for 17 emerging market economies and tends to perform more accurately than existing methods. Thus, it is likely to be a useful tool for economists that wish to forecast varying types of economic phenomena.

Table of Contents

I. Introduction:

After the Mexican "Tequila" Crisis of the mid 1990's the frequency of currency crises increased across the globe, particularly in emerging market economies. As such, investigating early warning systems (henceforth, EWS) came to the forefront of importance for international institutions private and public alike. Afterwards the interest in EWS declined but has increased again in the aftermath of the 2008-09 global financial crisis. The pioneering work in this field was done by Kaminsky, Lizondo, and Reinhart (1999) which was heavily influenced by the empirical work of Frankel and Rose (1996). From this, several other models have emerged. Namely, Berg and Pattillo (1999b), and the work of Bussiere and Fratzscher (2006) amongst other models developed by the US Federal Reserve (Kamin and Babson, 1999, and Kamin et al., 2007) and Bundesbank (Schnatz, 1998, and Schnatz 2007) to allow policy makers to perform policy interventions. In addition, several private sector institutions including Credit Suisse First Boston (Amlan and Tudela, 2000), Morgan Stanley (Kimbrough and Li 2001), Goldman Sachs (Ades et al. 1998), JP Morgan (Persaud, 1998), the International Monetary Fund (Abiad, 2003), and The Deutsch Bank (Garber et. al, 2001) have also undertaken the development of EWSs focused on profit opportunities.

The mere existence of so many models raises the question: which model does the best job of predicting currency crises for which institutions? Also, considering that these models failed to deliver accurate predictions often providing no more insight than a random guess; can we formulate better predictions using a different methodology? Answering these questions will require the examining of the factors that influence the evaluation and comparison of models. Model performance will depend among other things on the thresholds that minimize the signal to noise ratio. These may be model specific. As such, the models were compared using three different thresholds: 25%, 50%, and 75%. This was done so that other researchers may analyze the results across varying thresholds, and to allow for a fair comparison between models. An important takeaway from this is that the modeler may be able to increase accuracy through threshold selection.

In essence, by choosing a lower threshold the occurrence of type 2 errors (signals crisis but no crisis occurs) or number of false negatives increase. Inversely, if a larger threshold is chosen type 1 errors (does not signal crisis but crisis occurs) or number of false positives increase. Thus, threshold selection can be chosen to meet the needs of the modeler in terms of what errors they choose to weight more heavily. When applied to the problem of predicting currency crises it is clear to see that different modelers may have different preferences based on their objectives. For example, policy makers and private actors may not value type 1 errors in the same way that they value type 2 errors. That is because for policy makers the real effects of large depreciations are more important to their constituents. Therefore, they view type 1 errors as costly, since missing a crisis can have very serious implications for the economy as a whole. However, private businesses may have different needs for predicting large currency devaluations related to arbitrage opportunities, hedging, or diversification to mitigate foreign market risk. Thus, they may tend to view type 2 errors as costly since conducting overseas business operations or making investment decisions around future depreciations that do not occur, would be more important than simply missing a depreciation.

This research applies the methodology employed in Busierre and Fratzscher's (2006) EWS to a data set using a common definition of currency crises, analyzes the results across a one-year time horizon, and compares the forecasting accuracy of this model to a new approach for EWS. The intuition behind studying across a one-year time horizon is that academia and governments tend to use longer time horizons to be able to intervene in markets and the prevent currency crises while private institutions are focused on shorter time horizons for the purpose of profiting from exchange rate volatility and knowing when to hedge more aggressively in order to avoid losses. Also, different types of businesses may have specific time horizons of analysis inherent to the type of business. Therefore, analyzing over a one- year window makes sense as opposed to analyzing the model at different time periods. However, graphic analysis will also aid in illustrating the performance across different time periods.

The paper will be organized in the following way: Section I will introduce the support vector machine and why it may be useful as a forecasting tool, Section II will provide an overview of the theoretical literature concerning the causes of currency crises, Section III will review current EWS methods, Section IV will discuss the rationale behind currency crisis measures and establish the measure for use in the analysis, Section V will cover the EWS models to be analyzed for comparison, Section VI will provide a review of the particular data and methodology applied in this research, Section VII will provide results and robustness testing, and Section VIII will provide conclusions and future areas for research.

I. The Support Vector Machine

The support vector machine is a powerful modelling technique from the field of statistical machine learning that was introduced by Vapnik (1998) and has not yet made it to the mainstream techniques of the social sciences. The approach is particularly fitting for EWSs because it focuses on the predictive power of the model rather than the explanatory power. It is based on the principle of structural risk minimization and uses structural correlations to forecast, allowing the model to capture non-linear relationships underlying the data-set and incorporate large amounts of data within the analysis, often with high predictive accuracy. It is important to note that the SVM can be used for both classification and regression. Also, we must note that at this point in time the SVM is econometrically speaking a black-box of inputs and outputs and may not be very useful for policy analysis. However, minimizing the generalization error or out of sample error via structural risk minimization makes it particularly useful for forecasting as it allows for some tolerance to random or unforeseen events.

The basic concept employed within this empirical analysis is a typical binary classification problem. In other words, the SVM will be trained on a randomly -selected sample then used to predict a crisis or non-crisis episode. The SVM does this by finding the optimal separating hyperplane, then derives the decision boundary for the classification using the upper-bound of the generalization error while still allowing for some slack in the classification. The algorithm then uses this structure to divide the data into crisis and non-crisis episodes, given the underlying dependent variables[1]. When given new data the support vector machine will predict by assigning the outcome to one category or another based on the structure created in training.

Essentially, the SVM is a machine learning algorithm that can identify patterns and forecast by minimizing structural risk. Where structural risk refers to the portion of the variance that remains after the upper bound of the generalization error has been found by the algorithm. Thus, the concept of structural risk minimization is an inductive process aimed at placing an upper-bound on the generalization or out of sample error thereby

[1] Please see Appendix A for a review of hyperplanes and the SVM adapted from Witten et al. (2013)

minimizing the structural risk. It uses repeated sampling methods to cross-validate the underlying structure of the dependent variable to the independent variables and allows for some slack in formulating the upper-bound of the separating decision boundary. In order to better understand this, we will review how the support vector classifier achieves this and what constitutes a support vector machine as presented by Witten et al. (2013) below.

The support vector classifier or soft margin classifier allows for this type misclassification to occur. It separates most of the training data into their respective classes but allows some leeway by solving the following optimization problem:

$$\max_{\beta_0, \beta_1, \ldots, \beta_p, \varepsilon_1, \ldots, \varepsilon_n} M \quad (11)$$

subject to,

$$\sum_{j=1}^{p} \beta_j^2 = 1 \quad (12)$$

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \geq M(1 - \varepsilon_i) \quad (13)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C (14)$$

Here, C is a non-negative tuning parameter that allows some slack in the classification of training observations up to a certain margin of error. As with the example of the optimal separating hyperplane margin M is still being maximized, but we allow some slack to individual observations based on the slack variables $\varepsilon_1, \ldots, \varepsilon_n$.

The classification is then still determined by solving (11) – (14) then by deciding on which side of the hyperplane the observation lies based on the sign of the output $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$. The slack variable ε_i now allows the location of the i-th variable to be known. Consider if $\varepsilon_i = 0$, then the test observation would lie on the correct side of the hyperplane. Likewise, if $\varepsilon_i \geq 0$ then the test observation will lie on the wrong side of the margin, or if $\varepsilon_i > 1$ then it will be on the wrong side of the hyperplane. Note now that sum of ε_i's is bounded by C, or the tuning parameter. This dictates the number and severity of violations that the system will tolerate. In practice, it is chosen via cross-validation more commonly referred to as repeated sampling and is used to control the bias to variance trade-off pertaining to this methodology. When C is large the system of equations allows for more violations leading to a lower fit, increase in bias, and lower variance. However, when C is small the system allows less violations thereby leading to a higher fit, decreasing in bias, and an increased variance. One advantage of this is that although optimal parameters will be found through training, and cross validation the modeler can choose parameters that allow for more or less slack given the likelihood of random events impacting the forecast.

It is important to notice a special property of the optimization problem characterized by equations (11) – (14), and that is that only observations that are lying on the margin or violate the margin will affect the formation of the separating hyperplane or decision boundary. The fact that the support vector classifier's decision rule is based only on a small subset of the training observations (the support vectors), implies that it is robust to the behavior of observations that are far away from the hyperplane. This is a property that distinguishes the support vector classifier from limited dependent regressions and may add to its usefulness as a forecasting method. Also, note that the support vector classifier still only works in the case of a linear decision boundary. In the next section a general mechanism for producing non-linear decision boundaries will be discussed, and the support vector machine introduced.

When the data is linearly separable the support vector classifier is the obvious choice. However, when there is a non-linear relationship between the predictors and the outcome, we can enlarge the feature space using functions of the predictors to address this problem of non-linearity. The support vector machine is an extension of the support vector classifier that expands the feature space in a unique way by using the correlation of functions between individual observations known as kernels. When solving equations (11) – (14) only the inner product of the observations, not the observations themselves, matter. The inner product of two observations x_i, x_i' is given by:

$$\langle x_i, x_i' \rangle = \sum_{j=1}^{p} x_{ij} x_{ij}'$$

Thus, it can be shown that,

- The support vector classifier can be written:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle$$

or,

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i K\langle x, x_i \rangle$$

where there are n parameters α_i, or one per training observation, K is a kernel. Note that to estimate parameters $\alpha_1, \ldots, \alpha_m$ *and* β_0 all that is needed are the inner products and the number of pairs among the n observations (given by n(n-1)/2). The solution for the support vector classifier can now be rewritten in terms of the inner product which allows the use of functions of the inner product that relate the similarity in the position of two variables to extend the feature space. Below are some examples of commonly used Kernels:

1. Linear Kernel (e.g. Standard Pearson Correlation):

$$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j},$$

Using the linear kernel gives the support vector classifier and is still only for linear analysis.

2. Polynomial Kernel of Degree d:

$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d.$$

Here, d is a positive integer greater than one. This is the same as fitting the support vector classifier in a higher dimensional space of polynomials of degree d, rather than the original feature space. This leads to a much more flexible decision boundary. Usually, degrees 3 and 4 are the most robust.

3. Radial Kernel:

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2).$$

This is an example of another positive non-linear kernel where gamma is a positive constant. Note that anytime a non-linear kernel is used in a support vector classifier we call it a support vector machine.

The support vector machine offers many valuable advantages in modelling crisis episodes as documented by Ahn et al. (2011), and general advantages as reported by Witten et al. (2013). One advantage is that the output of the SVM can be interpreted as a simple probability similar to limited dependent regressions. This allows the analysis to compare models across various metrics, including but not limited to probability thresholds and mean square errors. The theory used to construct the SVM output has also been shown to add robustness to individual variables when compared to logit, probit, neural networks, and linear discriminant analysis (Nikam, 2015). This stems from the SVM's ability to incorporate slack or support vectors which allow for the control over the variance-bias trade-off in fitting the model.

Also, by allowing some tolerance or misclassification for data points to violate the underlying decision boundary creates more robustness in out of sample forecasting in light of random or unforeseen events: a key advantage of this methodology over traditional techniques. Another important advantage to note is that the SVM uses kernels or correlation functions to form different types of decision boundaries allowing the SVM to model underlying structural non-linearities and incorporate large amounts of data in terms of the number of variables and the observations per variable. This also makes it particularly fitting for currency crisis prediction. Current literature (Eichengreen et al. 1996, Reinhart and Rogoff (2008,2009), Amri and Willett (2017)) suggest that currency crises are usually caused by several factors coming together to form a "perfect storm" or zone of vulnerability and then a trigger or shock that leads to the actual crisis. If we consider the perfect storm to be a cluster of linearly non-separable factors, then applying a support vector machine with a non-linear kernel should be able to separate these triggers more easily. In turn, allowing for some slack in the classification may allow us to provide some robustness to those triggers within the cluster.

Allowing for some misclassification also gives the SVM an advantage over traditional limited dependent regressions by being able to gage the impact of individual variables when a large number of variables are considered in the analysis. This shortcoming of limited dependent regressions is rooted in the fact that logistic and probabilistic functions are bounded between 0 and 1. The support vector machine produces decision values and probability estimates and allows for the decision values to exceed these bounds of the measure with some slack it then attributes the probability of the outcome by using Bayesian maximum likelihood estimation. These are feats that other techniques have yet to fully address and may be crucial to forecasting the occurrence of currency crises.

Now that we have established the principles on which the support vector machine functions, let us turn to a conceptual discussion of the underlying causes of currency crises. This will help create a stronger argument for why the SVM may be particularly useful to currency crisis prediction, as theory has yet to provide a parsimonious answer.

II. Theoretical Reasons for Why Currency Crises Occur:

The literature has Identified several reasons for the occurrence of currency crises. The first-generation models invented by Paul Krugman (1979) focus on incongruences between exchange rate commitments and the underlying economic fundamentals. Krugman argues that currency crises stem from balance of payment imbalances that lead the government to deplete reserves in order to protect the exchange rate regime. This behavior is also noted to cause sudden growth in the domestic credit market causing further fiscal issues. It is also important to note that these models only apply to fixed exchange regimes. These models give a specific timing, but since the underlying assumptions are unrealistic and give little insight into actual timing.

The second generation of Currency Crisis models, such as, Obstfeld (1986, 1994), Flood and Garber (1994), and Eichengreen et al. (1995) view currency crises as shifts between multiple monetary equilibria in a zone of vulnerability that arise in response to self-fulfilling speculative attacks. In this case, the timing of crises is no longer possible because the prospect of depreciation depends on the government's objective loss function among other factors, for example, the government's credibility, reputation, and the political costs of associated with tight monetary policy versus allowing the currency to depreciate.

Third generation Models provide in depth analysis of other possible factors usually related to problems in the financial sector. For example, McKinnon and Pill (1995) argue that financial liberalization and insurance schema can lead to institutional lending booms that expand both foreign and domestic credit leading to banking and currency crises. Radelet and Sachs (1998) show how the increase of short-term debt denominated in foreign currency caused bank runs resulting in solvency and liquidity problems in the market for international assets during the Asian crisis of 1997. Others, including Burnside et al. (2004) point to moral hazard as the culprit, contending that government guarantees give banks an incentive to take on foreign debt making them vulnerable via contagion channels. Chang and Velasco (2002) argue that under regulated banks in open economies that issue short-term deposit in both foreign and domestic currency, while still holding long-term assets that lack liquidity, this in turns leads to international liquidity crises. Over all most third-generation models emphasize the causes that lead to both banking and currency crises. The simultaneity of these phenomena was coined the "twin crises" (Kaminsky and Reinhart, 1999). While the theoretical considerations provide us with insight into the nature and causes of currency crises, it is the empirical literature in the next section which will give us an understanding of the current modelling techniques that have been used for EWS thus far.

III. Review of Methods for EWS:

a. The Signal Approach

The first method employed for formal early warning systems was developed by Kaminsky, Lizondo, and Reinhart (1998) and is commonly referred to as the signal approach. It is a bivariate method in which market behavior during tranquil periods is compared to behavior right before and during the crisis to monitor behavior that signals a crisis. In this approach an arbitrary threshold is chosen for the independent variables or determine whether or not they are giving a reliable signal. The threshold is usually related to the probability distribution. For example, if the data point falls within the top twenty percent of the variable's distribution, then that value is coded as a signal of the crisis.

As mentioned earlier the threshold can also be chosen to strike a balance between Type 1 Errors (failure to predict a crisis that occurs), and Type 2 Errors (predicting a crisis that does not occur). In other words, if the threshold is too low then the model will fail less in predicting crises that occur but will also predict more crises that in fact do not occur. If the threshold is set too high, then the model will send less false signals but will miss crises that occur. Therefore, it is important to employ a mechanism that optimizes this trade-off when choosing a threshold. The mechanism used is known as the noise-to-signal ratio. In general, the threshold is chosen to maximize the good signal to noise ratio. However, in some cases the modeler may not simply maximize the noise to signal ratio, but rather maximize the ratio after specific costs have already been weighted.

It is important to note that the Signal Approach has various shortcomings. The most obvious lies in the fact that the model issues binary signals based on arbitrary thresholds, as such it does not say anything about the strength of a signal and how it relates to the extent of a crisis, thereby sacrificing valuable information. It also evaluates variables individually failing to pick up on how possible interactions between variables may affect the on-set of a crisis. This is especially important since currency crises generally result from the inconsistency among a set of variables and this can occur in many different combinations (See Amri and Willett (2017) on the political economy of currency crises). Finally, it ignores the correlations between certain variables. The next family of methods helps to overcome some of the shortcomings of the signal approach.

b. Limited Dependent Regressions (Logit/Probit)

The term limited dependent regressions refers to the family of probit/logit regressions used to analyze the problem of predicting currency crises. They capture the probability relationship between the discrete dependent variable and a set of explanatory variables. These regressions take into account correlations between all variables and allow for testing the significance of independent variables. In addition, the output can be read as a simple probability. Models that use this type of methodology include the work of Frankel and Rose (1996), Berg and Pattillo (1998) in addition to all the private sector models. This approach was extended by Busierre and Fratzscher (2006) to a multinomial logit model, in which they control for pre-crisis, crisis, and post-crisis behavior by analyzing the probabilities with respect to the behavior witnessed in each of the crisis episodes. This is considered the most accurate EWS to date.

The logit/probit regression is not without fault. Its shortcomings include the inability to accurately demonstrate the impact of individual variables. The non-linear and bounded nature of the logistic and probabilistic functions imply that the impact of a certain variable will be dependent on the magnitude of other variables. This means that the change in the variable might not accurately reflect that individual variable's contribution to predicting a crisis. This is problematic, as the causes of currency crises are many, requiring a methodology that can incorporate a large number of variables.

c. Methods from Statistical Machine Learning/ Computational Finance

Due to the inability of current methodologies to incorporate some of the specific characteristics of economic problems, like non-linear relationships or random behavior of the variables, significant growth has been experienced in the applications of financial engineering to applied economic problems (Chen and Wang, 2004;). Since the mid 2000's economists and computer scientists alike began to explore the usefulness of statistical machine learning techniques to help forecast economic phenomena such as currency crises (Ahn et al., 2011; Kim et al., 2004, Yu et al., 2006). Although the literature is scarce there are a few significant contributions to bridging the gap between statistical machine learning and currency crisis forecasting.

It is important to note that all of these use some form of the exchange market pressure (EMP) measure of currency crises with different weighting schema as will be discussed in the following section. Yu et al. (2006) construct a forecasting model using general regression neural networks and find them to outperform traditional methods. Ramli et al. (2013, 2015) create models using the k-nearest neighbor method, and statistical ensemble. A few papers have also tried preliminary construction of EWSs using support vector machines. A working paper by Chaduri (2014) constructs an early warning system using a support vector machine, however he only examines one country (Argentina), and while the accuracy of his model was reported to have a mean square error of less than 10%, he provides no comparisons or robustness testing over the sample. Zhang (2014) uses a wavelet-based support vector machine on a small sample of 2 East Asian economies, using daily exchange rate fluctuations for a period of roughly one year. This paper also, compares the approach to the logit and signals approach. However, the small sample size and use of different data and time schema limits the robustness of the analysis with respect to other more cutting-edge techniques. In proposing the support vector machine to analyze this particular type of phenomena these papers help to bridge the gap between economics and financial engineering but fall short in their overarching analysis by choosing data that is not comparable to other models and using small sample sizes of usually less than two years of data and examining one or two countries. This research plans to address these pit-falls by using a sample of 17 countries, with monthly data over a period of about 18 years and using a common definition for comparison. That being said, the next section will outline some of the theoretical considerations behind constructing measures of currency crisis, provide some background to measures employed in EWS, and establish a measure for this analysis.

IV. Currency Crisis Definitions and Measures:

In order to create, recreate, and examine the models, it is important to decide on a working definition or measure of a currency crisis. Unfortunately, a universal definition does not exist making it necessary to examine various concepts from the literature and decide on an appropriate measure for the task at hand. Early papers define a currency crisis as the necessity for a government to abandon a fixed exchange rate regime due to fiscal pressures (Krugman 1979, Flood and Garber 1984, Obstfeld 1984). However, this definition may not only be too restrictive for our purposes because they exclude speculative crises that do not lead to large depreciations [4] and also exclude

[4] Almahmood et. al. (2018) show that a substantial majority of speculative attacks do not end in large depreciations.

floating regime countries that are documented as having experienced heavy pressure caused by speculative attacks.

To address the issue of floating regime countries many of the early studies took to the idea the currency crisis should be measured as a large depreciation in the exchange rate itself. For example, Frankel and Rose (1996) define a currency crisis as an annual depreciation of 25%, preceded by a depreciation of 10% in the previous year. Other studies such as Kaminsky and Reinhart (1999) and Laeven and Valencia (2008) have also used varying thresholds of percentage changes to define crisis periods.

Others measure the severity of the depreciation by using fluctuation in the nominal exchange rate and defining crises where the fluctuation exceeds a threshold defined by a country specific mean and standard deviation. On example is Glick and Moreno (1999) who define crisis when the monthly change in the exchange rate diverges from the country specific mean by two standard deviations. Other studies that employ this type of measure use thresholds between 1 and 3 standard deviations.

In order to address the nature of speculative attacks and market pressures, much of the literature has taken to the idea that conceptually measures of currency crisis should capture various behaviors that could place pressure on the currency. Typically, national authorities have several policy options to respond to this pressure, such as allowing the currency to depreciate, raising interest rates, and selling reserves to bolster the currency. It is very common for national authorities to respond with combinations of policy. Therefore, the EWS literature tends to employ exchange market pressure indices composed of two to three factors, and various weighting schemes. For example, Kaminsky and Reinhart, (1999), Berg and Pattillo (1999a, 1999b), Bussiere and Mulder (2000); Kamin et al. (2001), and all use a two-factor index comprised of the nominal or real exchange rate, and reserves.

Others consider three factors by adding in nominal or real interest rates (Eichengreen et al., 1996; Bussiere and Fratzscher, 2006) when data is available. Often, they will use equal weights or precision weights for the factors. Where the precision weight is defined as the inverse variance of the variable. As described below this paper will use the latter concept in choosing a working measure[5].

For our purposes the exchange market pressure mechanism suggested by Bussiere and Fratzscher (2006) is used. There are three reasons for this. The first being that their model is of particular interest for comparison due to its high forecasting accuracy and incorporation of the post-crisis bias. The second being that in order to truly compare the models it is important to use the same working definition. Finally, they employ a three-factor EMP, which helps capture the concept of speculative pressure from all three theoretical perspectives. The EMP proposed by Busierre and Fratzscher (2006), is given by:

$$EMP_{i,t} = w_{REER}\left(\frac{REER_{i,t} - REER_{i,t-1}}{REER_{i,t-1}}\right) + w_r(r_{i,t} - r_{i,t-1}) + w_{Res}\left(\frac{Res_{i,t} - Res_{i,t-1}}{Res_{i,t-1}}\right)$$

Where the EMP of country (i) at time (t) is the weighted average of the change in the respective Real Effective Exchange Rate (REER), interest rate (r), and reserves (Res). The weight assigned to a variable is traditionally the relative precision of that variable. Where precision is defined as the inverse of the variance, therefore the less volatility the higher the assigned weight.

Next, we define a currency crisis episode (CCi, t) as the event when the exchange market pressure (EMPi, t) variable is two standard deviations (SD) or more above its country average EMPi, such that:

[5] See Efrimidze(2011), Abiad(2003), and Schnatz(1998), for more comprehensive review of currency crisis measures. Also see Appendix B for list of definitions commonly used in the EWS literature.

$$CC_{i,t} = \begin{cases} 1 & if \;\; EMP_{i,t} > EMP_i + 2SD(EMP_i) \\ 0 & if \;\; Otherwise \end{cases}$$

For all intents and purposes this will be the definition employed in our empirical review and in the proposition of the Support Vector Machine. It is important to note that other studies such as Alhamood (2018) have tested measures between 1.5 and 3 standard deviations and shown only small variances in the accuracy of the measures. Therefore, we assume that this is a fairly accurate measure for our analysis.

It should be noted that this is the definition commonly applied in the EWS literature and has been modified for the SVM by using values of 1 and -1 as opposed to the traditional 0 or 1 binary metric. This was done so that the SVM could classify the phenomena, as the algorithm sets the optimal separating hyperplane to a value of 0 by default[6]. On another note, the definition is altered for the multinomial logit model as well. This is done to capture the post crisis bias presented by Busierre and Fratzscher (2006). Therefore, the crisis measure is altered to take a value of 0,1, or 2. Where 0 represents a tranquil period, 1 the pre-crisis period leading up to the actual event reported by the binary metric, and 2 represents the post-crisis period. In this case the post crisis or recovery period was determined by crisis measures that use inflationary pressures to establish the duration of the crisis.

V. Models for Comparison:

While many EWS studies have been conducted many of them have been shown to have poor performance. In this study we tried to choose the models that have shown to perform best on their respective criteria and data. These include state of the art techniques that were reported to predict more than 50% of crises (Berg and Pattillo, 2005). To do this

[6] Please see Appendix A for further details relating to structural risk and constructing the optimal hyperplane

several models from the literature were reviewed, and those that had the highest predictive power were selected[7]. The list of models included the multinomial logit model of Bussiere and Fratzscher (2006), the markov-switching process proposed by Abiad (2003), and the Artificial Neural Networks proposed recently by Yu et. al 2016.

However, the markov-switching process, and neural networks did not converge in the maximum likelihood estimation of the probability output for this particular dataset. Therefore, reliable estimates could not be gathered, and these models were omitted from the final analysis. The research also tried to apply a linear support vector machine, as well as SVMs using 3rd and 4th degree kernels to the dataset in order to capture the effect of incorporating non-linearities. Because these methods did not converge, the final analysis will only include the comparison the of the multinomial logit and the SVM that employs a radial kernel. It is of the upmost importance to note that although it could not be proven, the inability of the aforementioned models to converge may indicate that capturing structural non-linearities may be critical to predicting future currency crises.

In order to evaluate the predictive accuracy of these models, one should compare the predicted probability of the crisis with the actual probability. However, the latter is not directly observable. Therefore, it follows that one must compare the predicted probabilities with the actual occurrence of crises given by the aforementioned exchange market pressure mechanism. Since the predicted probabilities are given as a continuous series it is of the upmost importance to choose threshold or cut-off level at which a given probability signals a pending crisis. The key issue here lies in finding the "optimal" level for this threshold.

[7] Please see Appendix B for a summary of the models reviewed.

In general, the lower the threshold is set the larger the number of signals will be. However, the lower the cut-off the larger the number of erroneous signals (Type 2 errors). Choosing a threshold requires making a judgment on the relative importance of Type 1 and Type 2 errors. This may be dependent on the aim of the EWS. As mentioned earlier, policy makers may put more emphasis on type 1 errors, as type 2 errors tend to be less worrisome from a welfare perspective. Also, Type 2 errors may not be due to predictive failures of the model, but rather reflect vulnerability which was addressed via timely policy intervention and a crisis was averted. Since "optimal" thresholds tend to be model specific, we will examine the classification rates at varying thresholds for our comparative analysis. This will allow us to compare the models across similar thresholds, see the effect of threshold selection on erroneous signals (type 1 vs. type 2), and compare the best results from each model. The last point is important since the best thresholds will likely be model specific.

VI. Data and Methodology:

Data was gathered from the IMF's International financial statistics, and balance of payments statistics at https://www.imf.org/en/Data, The World Bank Data base, and Joint External Debt Hub. In this study we use a sample of 17 emerging market countries, for which monthly data was collected ranging from 1996M03 to 2014M12. The sample includes, 6 Latin American countries: Argentina, Brazil, Chile, Colombia, Mexico, and Venezuela; 9 Asian economies: China, Hong Kong, Indonesia, South Korea, Malaysia, Philippines, Singapore, Taiwan, and Thailand; and 4 Eastern European economies: Czech Republic, Hungary, Poland, Russia. Turkey was also included in the analysis.

Generally speaking, while there are many documented variables of interest that may explain the occurrence of currency crises most studies tend to rely on a few areas of interest.

The first is external competitiveness, which tends to rely on fundamental trade imbalances, real exchange rate overvaluation, terms of trade, current account balances, and import or export growth. These variables tend to be founded in the principles of first-generation crisis models. Secondly, many models point to factors relating to external exposure of the country's debt composition such as external short-term debt to reserves, foreign direct investment, foreign exchange reserves, and capital inflows and outflows. It is important to note that these are derived from models in the second-generation but have been primarily extracted from third generation crisis literature. Other variables represent vulnerabilities from the real, public, and financial sectors, and are derived from second and third generation literature and include variables such as: real GDP growth, inflation, domestic investment ratios, domestic credit to the private and public sectors, money multipliers, equity market indices, and bank deposits. The other two areas of interest, which were omitted in this analysis for parsimony, are contagion measures such as trade channels or financial interdependence. The final area consists of the vast array of political variables provided by the World Bank and like institutions.

In this research we investigated 5 variables of interest: the current account to GDP ratio (CAGDP), real GDP growth (RGDP), exchange rate overvaluation (OREER), a lending boom index (LBI), and the external short-term debt to reserves ratio (STDR). The sample was reduced from a larger set of variables; however, these variables were finally chosen for two important reasons. The first being to allow for parsimony in the analysis by using variables commonly employed in EWS, and that have been shown by Edison (2003) and others such as Frankel and Saravelos (2012) to be the best leading indicators of more recent crises. The second is a combination of availability and for a fair comparison to Bussiere and Fratzscher's

2006 EWS. It is important to note that the contagion measure used by Busierre and Fratzscher was omitted in this analysis due to availability of data for the country sample. As such, we compare our Multinomial Logit Regression's performance to the performance of 4 other models documented by Berg et al. (2004) including the IMF-DSCD, KLR, Goldman Sachs, and CSFB early warning system. Although these models are not explicitly recreated, this comparison will give us a good idea of how well the multinomial logit model suggested here performs compared to other more traditional methods. Note that the models from Berg et al. (2004) were not recreated on this dataset.

Most of the data was given at the monthly and quarterly frequencies. Data that was recorded at or less often than the quarterly rate were converted to monthly frequencies using a cubic spline. The cubic spline was chosen as it uses all the data for the interpolation and tends to produce smoother curves. Here we note that frequency also impacts the recorded level. For example, data that is given in yearly frequencies will be the cumulative total for the whole year whereas data given in quarterly frequencies will be the cumulative total of that quarter. This must be considered when constructing ratios of variables given in different frequencies. Cubic splines do not address this issue in the same way as moving average interpolation. Therefore, variables given by ratios were calculated by first finding the ratio then applying the cubic spline.[8] For our measure of exchange rate valuation we used the deviation from the trend created using a hp-filter. In this case we recreated the measure used by Bussiere and Fratzscher (2006).

[8] Please see Appendix C for specifics concerning the data transformation and compilation. Cubic spline was chosen because it tends to produce smoother curves with smaller error than other methods such as moving average interpolation.

The multinomial logit model proposed by Bussiere and Fratzscher (2006) was recreated using the updated dataset and compared to the radial SVM, and evaluated using thresholds of 25%, 50%, and 75% to gage performance at different levels. It is important to note that the Multinomial Logit approach used by Busierre and Fratzscher (2006) uses pooled data, thereby negating time-series, and country specific effects. The SVM is applied on a country by country basis, thereby incorporating these effects and producing a tighter fit to the data. We also included estimates of Mean Square Error, and Monte Carlo simulations to establish the robustness of the SVM. In and out of sample estimates were also made.

VII. Results

1. Comparison of Models In Sample:

Table 1 shows the results produced using STATA 15, on the pooled data for the multinomial logistic regression that was performed on the dataset described above. It shows rather poor results, as some of the signs are incorrect and some of the variables have weak or no statistical significance. However, this dissertation will not put too much emphasis on the results of the multinomial logit on this particular dataset but rather how the relative probability estimate compares to the SVM. Tables 2,3, and 4 show the classification for the in-sample results of the multinomial logistic regression at signaling thresholds of 25%, 50%, and 75%. Tables 5,6, and 7 provide the performance of the support vector machine for the respective thresholds. Table 8 provides the performance of the four other models tested in the paper by Berg et al. (2004). Note that the models in table 8 were not recreated in this exercise. The performance is given by 5 measures. The first measure is the percent of observations correctly called or overall the classification rate. The second is the percent of crises correctly called or the percentage of crisis signals that actually resulted in crises over the total number of crisis signals. The third is the percent of false alarms or the number of crisis signals that did not result in crises over the total number of actual crisis episodes. The fourth is the probability of crisis given an alarm or the number of correct crisis signals over the total number of actual crisis episodes. Finally, the fifth measure is probability of crisis given no alarm or the number crisis signals when no crisis occurred over the total number of actual non-crisis periods.

Table 1. Multinomial Logit Regression Results

Variable	Coefficient	Standard error	Z	$P>\|z\|$
Pre-crisis period $Y_{i,t}=1$				
CAGDP	-0.211	0.516	-0.41	0.683
STDR	-0.608	0.300	-2.02	0.043
OREER	0.157	0.010	15.44	0.000
RGDP	0.013	0.005	2.41	0.016
LBI	-0.129	0.070	-1.85	0.065
Constant	-1.329	0.061	- 21.81	0.000
Post-crisis period $Y_{i,t}=2$				
CAGDP	3.230	0.807	4.00	0.000
STDR	-0.548	0.399	-1.37	0.170
OREER	0.007	0.011	0.68	0.498
RGDP	-0.118	0.011	-10.34	0.000
LBI	0.044	0.052	- 0.85	0.396
Constant	-2.308	0.095	-24.25	0.000
No. of observations	3,631			
Pseudo R^2	0.0905			

Table 2. Performance of Multinomial Logit Model at 25% threshold

	$S_{i,t}=0$	$S_{i,t}=1$	Total
$Y_{i,t}=0$	2,516	30	2,816
$Y_{i,t}=1$	543	330	873
$Y_{i,t}=2$	211	12	216
Total	3,270	362	3,632
% of observations correctly called			78.4
% of crises correctly called			37.8
% of false alarms of total alarms			8.3
% probabilities of crisis given an alarm			91.6
% probabilities of crisis given no alarm			17.6

Table 3. Performance of Multinomial Logit Model at 50 % threshold

	$S_{i,t} = 0$	$S_{i,t} = 1$	Total
$Y_{i,t} = 0$	3406	64	3,470
$Y_{i,t} = 1$	48	110	158
$Y_{i,t} = 2$	3	1	4
Total	3,457	165	3,632
% of observations correctly called			96.9
% of crises correctly called			69.6
% of false alarms of total alarms			36.8
% probabilities of crisis given an alarm			63.2
% probabilities of crisis given no alarm			1.4

Table 4. Performance of Multinomial Logit Model at 75% threshold

	$S_{i,t} = 0$	$S_{i,t} = 1$	Total
$Y_{i,t} = 0$	3,457	72	3,476
$Y_{i,t} = 1$	15	26	152
$Y_{i,t} = 2$	50	12	4
Total	3,522	110	3,632
% of observations correctly called			96.5
% of crises correctly called			17.1
% of false alarms of total alarms			73.5
% probabilities of crisis given an alarm			26.5
% probabilities of crisis given no alarm			0.4

The results from the multinomial logit show that it is more accurate at lower thresholds. We see the best overall performance at the 50% threshold, and the poorest at the 75% threshold. Where at a 50% threshold of analysis, 96.9% of observations were correctly classified by the model, along with 69.6% of crises being correctly predicted. However, we notice that at the 50% threshold that the number of false alarms increases from 8.3% to 36.8% when compared to the results at the 25% threshold. We also see that the conditional probability of a crisis given an alarm drops from 91.6% to 63.2%. Therefore, by choosing the model analyzed at the 50% threshold we sacrifice some of the accuracy as more signals will in fact be false alarms. When considering this from a policy makers perspective the model analyzed at the 50% threshold would be the best of the three documented results.

However, it is also important to note that private investors may prefer the lower threshold as it will give fewer false alarms, which is preferable for profit-maximizing or loss avoiding behavior. We see also that the model performs relatively well compared to other models described in Table 8 on the one-year time horizon, making suitable for private investors, and policy makers alike. Now we turn to the results from the SVM in order to compare the performance.

Table 5. Performance of Support Vector Machine at 25%

	$S_{i,t} = 0$	$S_{i,t} = 1$	Total
$Y_{i,t} = 0$	2,404	14	2,418
$Y_{i,t} = 1$	1097	117	1,214
Total	3501	131	3,632
% of observations correctly called			69.4
% of crises correctly called			9.6
% of false alarms of total alarms			10.7
% probabilities of crisis given an alarm			89.3
% probabilities of crisis given no alarm			31.3

Table 6. Performance of Support Vector Machine at 50% threshold

	$S_{i,t} = 0$	$S_{i,t} = 1$	Total
$Y_{i,t} = 0$	2,732	15	2,747
$Y_{i,t} = 1$	757	128	885
Total	3,489	143	3,632
% of observations correctly called			78.7
% of crises correctly called			14.5
% of false alarms of total alarms			10.5
% probabilities of crisis given an alarm			89.5
% probabilities of crisis given no alarm			21.7

Table 7. Performance of Support Vector Machine at 75% threshold

	$S_{i,t} = 0$	$S_{i,t} = 1$	Total
$Y_{i,t} = 0$	3,454	32	3,486
$Y_{i,t} = 1$	44	102	146
Total	3,498	136	3,632
% of observations correctly called			97.9
% of crises correctly called			69.8
% of false alarms of total alarms			23.5
% probabilities of crisis given an alarm			75.0
% probabilities of crisis given no alarm			1.3

The results from the support vector machine show the best overall performance at a 75% threshold of analysis. Here we see that 97.9% of observations were called with 69.8% of crises correctly called. It also provides relatively fewer false alarms when compared to the multinomial logit at 50% but does not provide fewer when compared to the multinomial logit model at a 25% threshold of analysis. Although the SVM analyzed at 75% performs better than any of the models listed in any of the tables, it does not seem to perform well below that 75% threshold. This is primarily a result of using the support vector machine analysis on individual country data, and in using only 5 variables. It seems that the algorithm has produced conditional probabilities that never fall below 50%. Hong Kong, Korea, and Poland are prime examples of this. We also see that with a one a one-year time horizon this model would be suitable for policy makers or profit-maximizers. However, it must be noted that the support vector machine tends to categorize less observations as crisis signals within the one year prior to a crisis, and the signals issued tended to be much closer to the actual crisis date. Therefore, it follows that the SVM may be more suited for profiteers as the time horizons between signals and the occurrence of the crisis will be shorter. As policy maker may require more time to prepare for policy interventions.

The table below documents the results provided by Berg (2004) as to the relative performance of the models. Note that these models were not reproduced over this dataset as the multinomial logit has shown a consistently better track record of performance. The table is only included to provide the reader an idea of how the recreated results stand up to the reported performance of other EWS models in the literature.

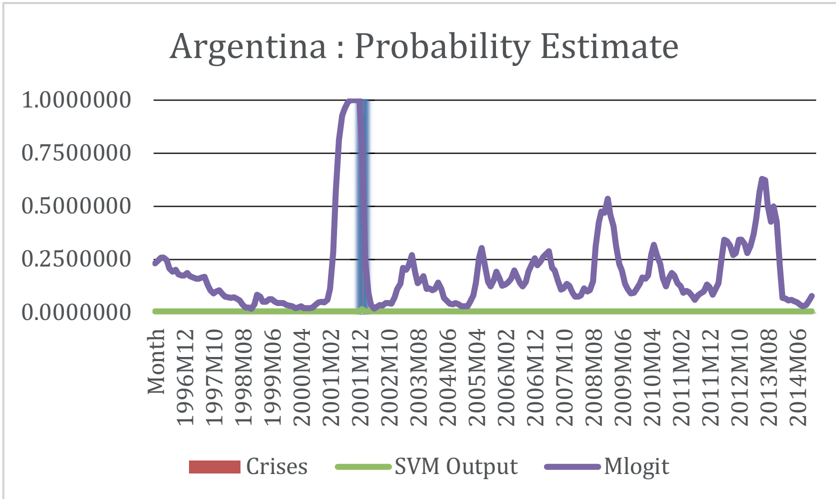
Table 8. Performance of Alternate Models (Berg. 2004)[9]

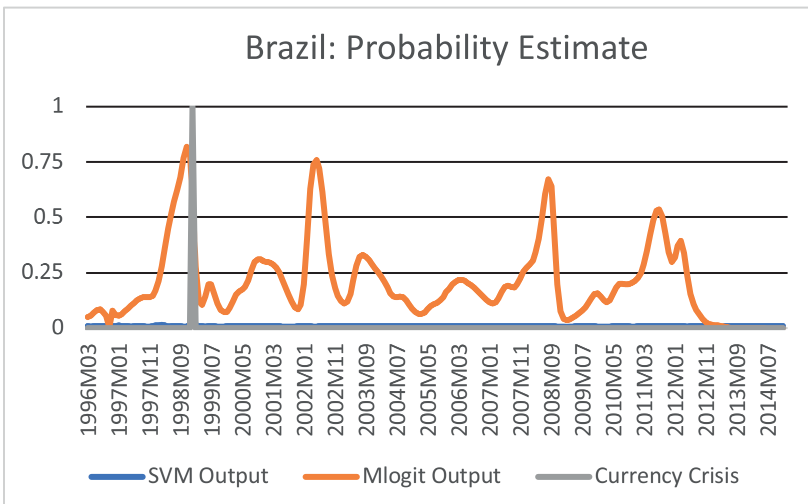
Models	IMF-DCSD	IMF-KLR	Goldman-Sachs	CSFB
% of observations correctly called	76.7	70.2	66.1	75.3
% of crises correctly called	65.1	59.8	66.2	61.1
% of false alarms of total alarms	62.8	70.3	74.0	93.5
% probabilities of crisis given an alarm	37.2	29.7	26.0	6.5
% probabilities of crisis given no alarm	7.8	9.8	8.4	1.4

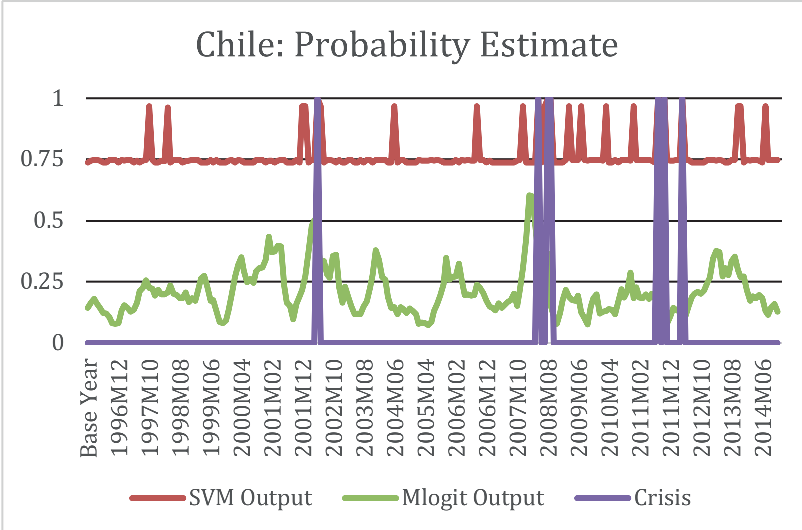
When compared to the results of table 8 we see that both the SVM and Multinomial logit, at their respective thresholds, tend to outperform conventional methods. This supports the idea that the SVM will be useful for forecasting currency crises as it tends to produce more accurate, and better fitted estimates. Below we present the graphic comparison to better illustrate the aforementioned statement of the results.

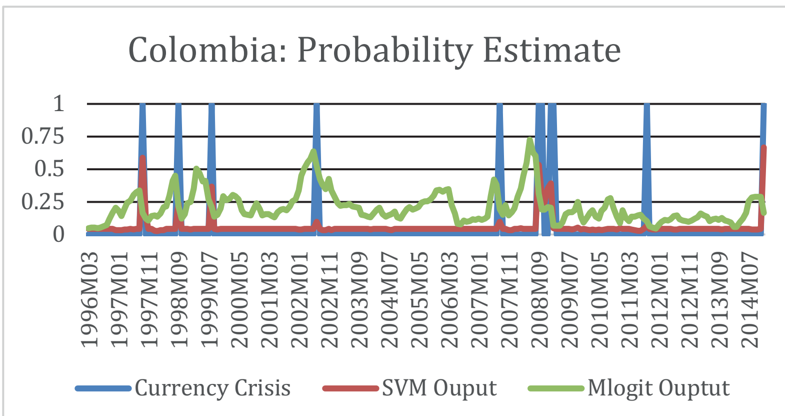
[9] Please see Appendix B for definitions and parameters of individual models.

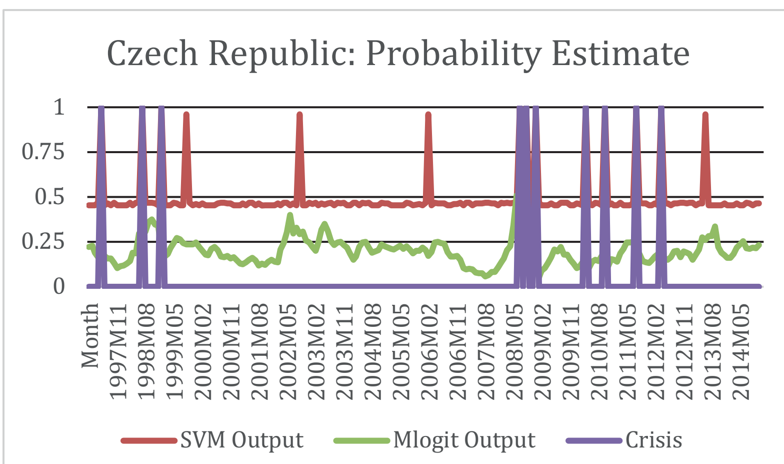
Graphs: Multinomial Logit and SVM Estimates

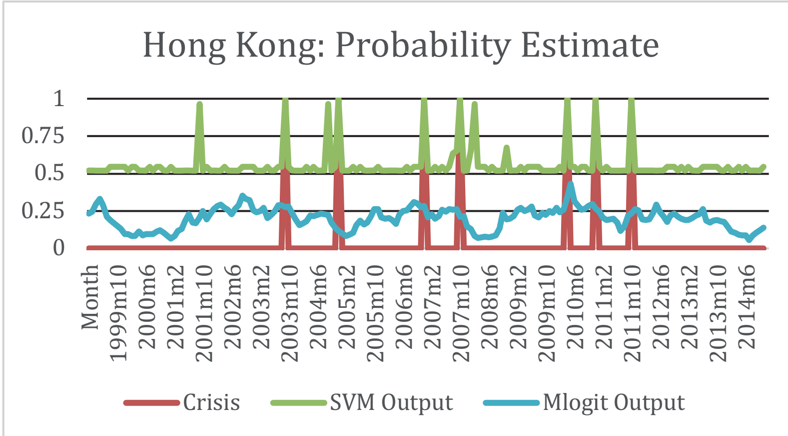


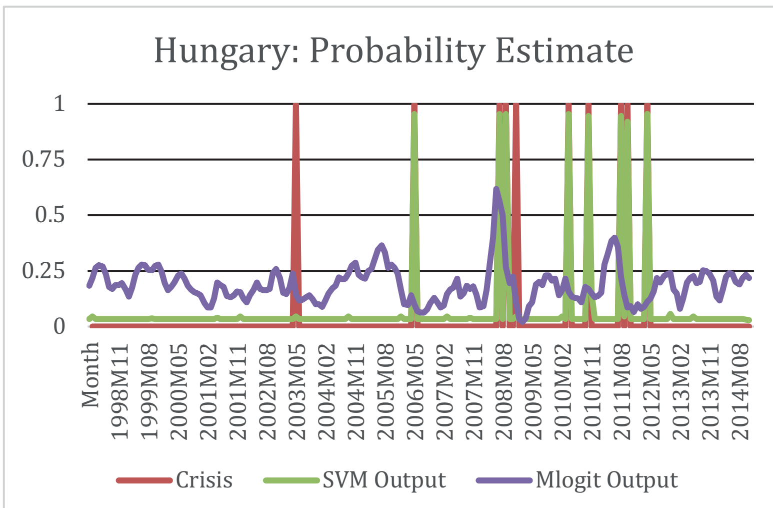


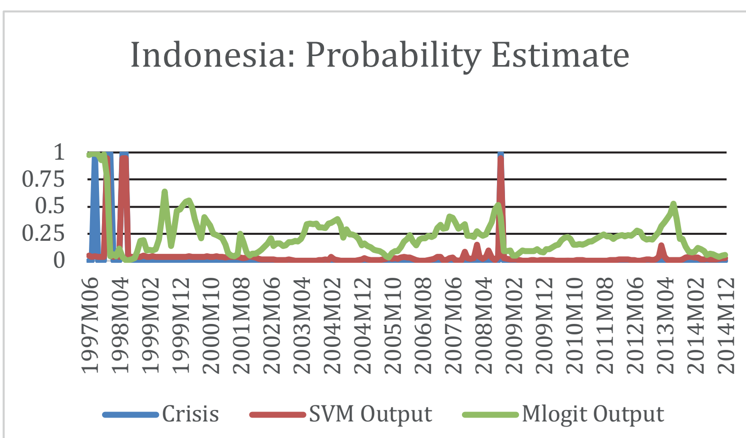

Chile: Probability Estimate
1
0.75
0.5
0.25
0
Base Year
1996M12
1997M10
1998M08
1999M06
2000M04
2001M02
2001M12
2002M10
2003M08
2004M06
2005M04
2006M02
2006M12
2007M10
2008M08
2009M06
2010M04
2011M02
2011M12
2012M10
2013M08
2014M06
SVM Output
Mlogit Output
Crisis

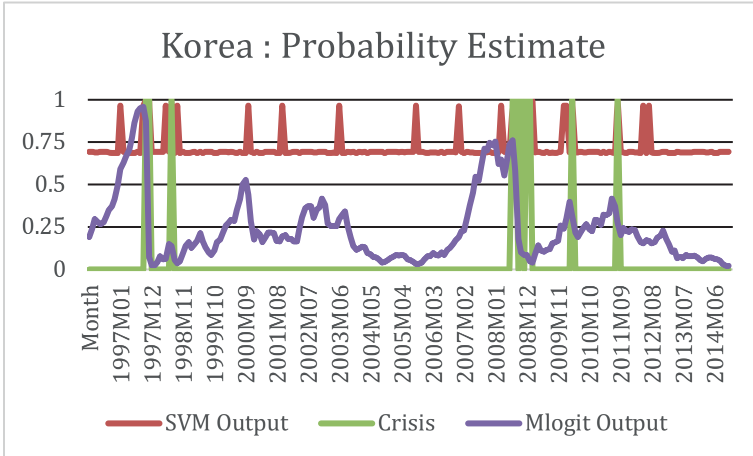

Colombia: Probability Estimate
1
0.75
0.5
0.25
0
1996M03
1997M01
1997M11
1998M09
1999M07
2000M05
2001M03
2002M01
2002M11
2003M09
2004M07
2005M05
2006M03
2007M01
2007M11
2008M09
2009M07
2010M05
2011M03
2012M01
2012M11
2013M09
2014M07
Currency Crisis
SVM Ouput
Mlogit Ouptut

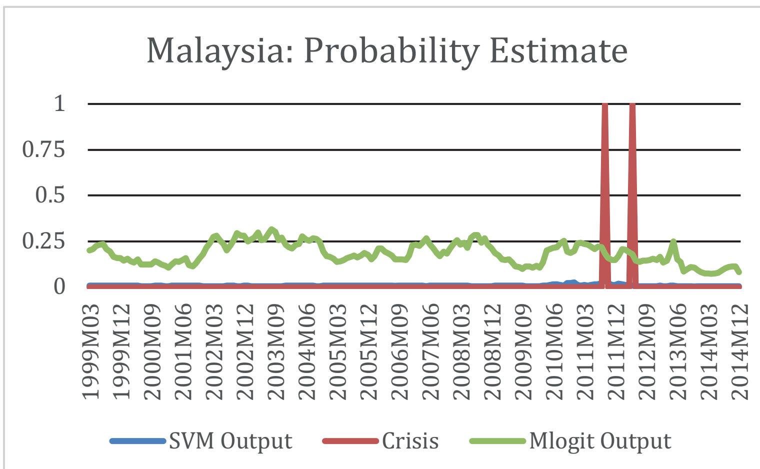

Czech Republic: Probability Estimate
1
0.75
0.5
0.25
0
Month
1997M11
1998M08
1999M05
2000M02
2000M11
2001M08
2002M05
2003M02
2003M11
2004M08
2005M05
2006M02
2006M11
2007M08
2008M05
2009M02
2009M11
2010M08
2011M05
2012M02
2012M11
2013M08
2014M05
SVM Output
Mlogit Output
Crisis

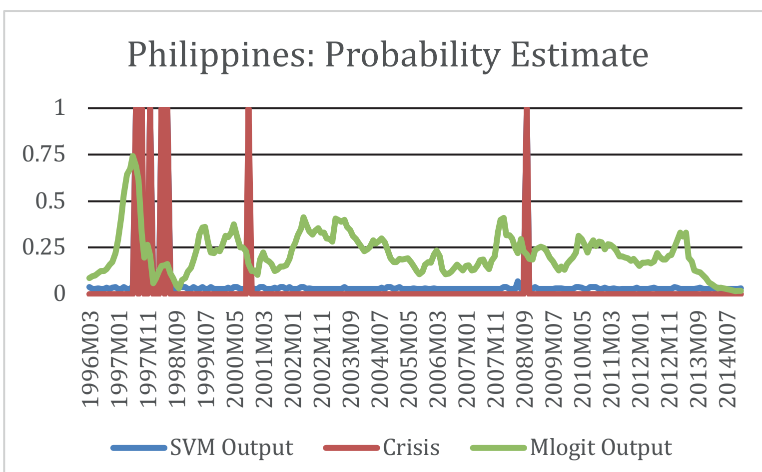

Hong Kong: Probability Estimate
1
0.75
0.5
0.25
0
Month
1999m10
2000m6
2001m2
2001m10
2002m6
2003m2
2003m10
2004m6
2005m2
2005m10
2006m6
2007m2
2007m10
2008m6
2009m2
2009m10
2010m6
2011m2
2011m10
2012m6
2013m2
2013m10
2014m6
Crisis
SVM Output
Mlogit Output

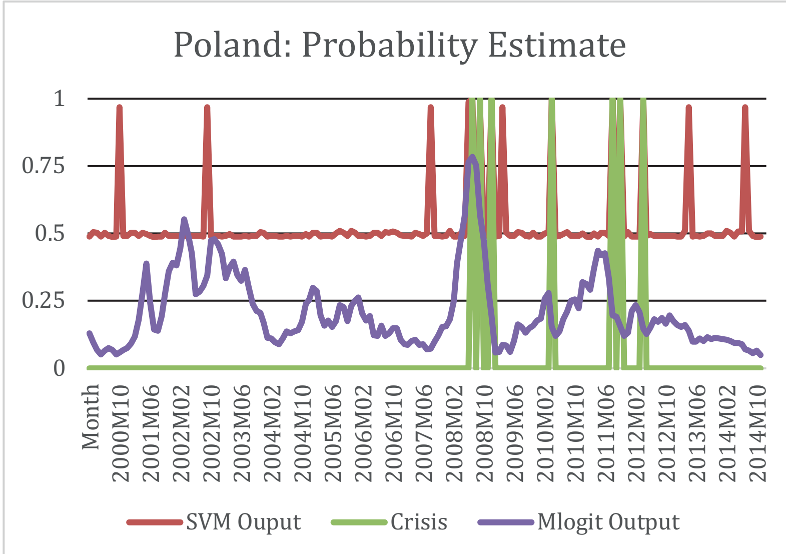

Hungary: Probability Estimate
1
0.75
0.5
0.25
0
Month
1998M11
1999M08
2000M05
2001M02
2001M11
2002M08
2003M05
2004M02
2004M11
2005M08
2006M05
2007M02
2007M11
2008M08
2009M05
2010M02
2010M11
2011M08
2012M05
2013M02
2013M11
2014M08
Crisis
SVM Output
Mlogit Output

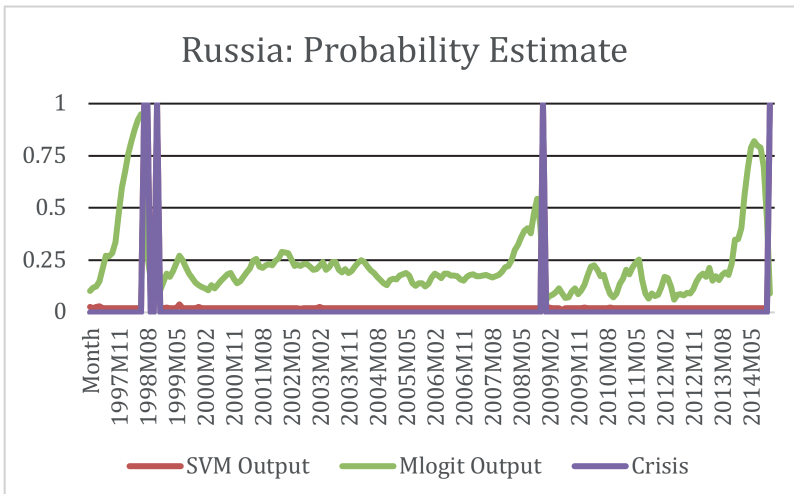

Indonesia: Probability Estimate
1
0.75
0.5
0.25
0
1997M06
1998M04
1999M02
1999M12
2000M10
2001M08
2002M06
2003M04
2004M02
2004M12
2005M10
2006M08
2007M06
2008M04
2009M02
2009M12
2010M10
2011M08
2012M06
2013M04
2014M02
2014M12
Crisis
SVM Output
Mlogit Output

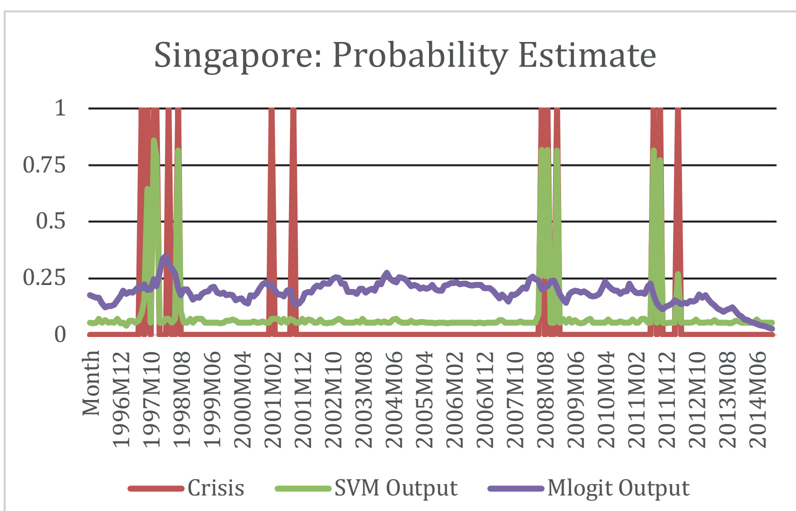

Korea : Probability Estimate
1
0.75
0.5
0.25
0
Month
1997M01
1997M12
1998M11
1999M10
2000M09
2001M08
2002M07
2003M06
2004M05
2005M04
2006M03
2007M02
2008M01
2008M12
2009M11
2010M10
2011M09
2012M08
2013M07
2014M06
SVM Output
Crisis
Mlogit Output

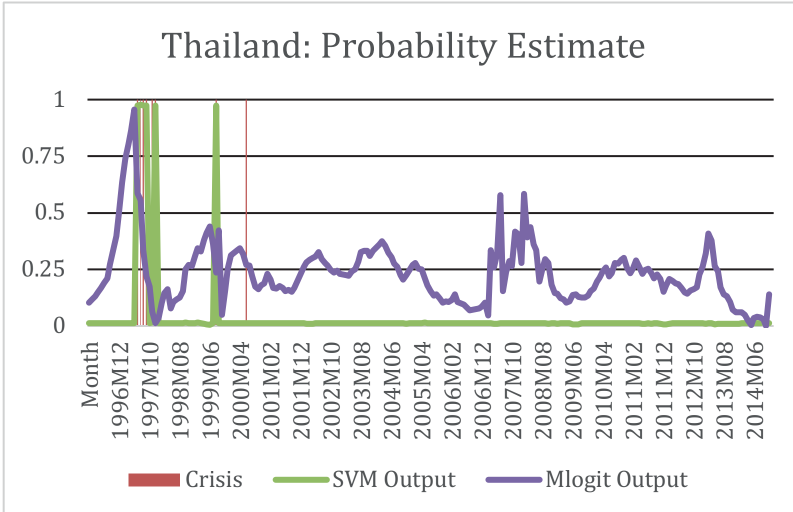

Malaysia: Probability Estimate
1
0.75
0.5
0.25
0
1999M03
1999M12
2000M09
2001M06
2002M03
2002M12
2003M09
2004M06
2005M03
2005M12
2006M09
2007M06
2008M03
2008M12
2009M09
2010M06
2011M03
2011M12
2012M09
2013M06
2014M03
2014M12
SVM Output
Crisis
Mlogit Output

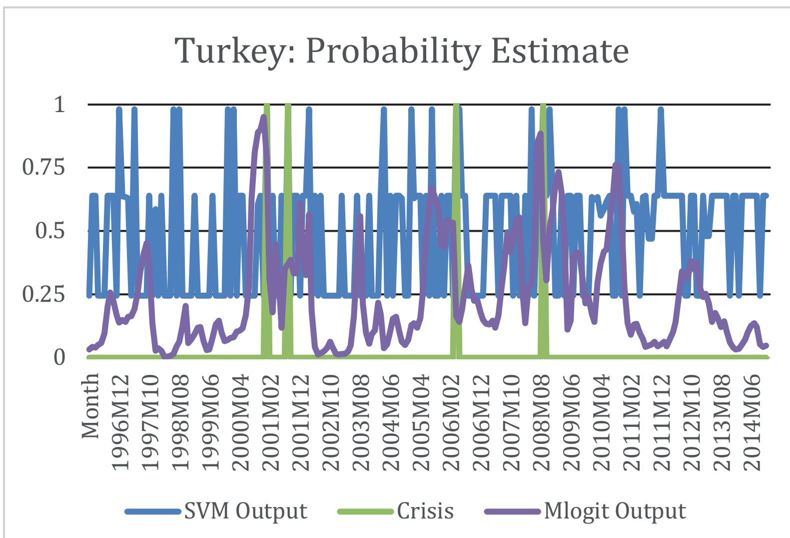

Philippines: Probability Estimate
1
0.75
0.5
0.25
0
1996M03
1997M01
1997M11
1998M09
1999M07
2000M05
2001M03
2002M01
2002M11
2003M09
2004M07
2005M05
2006M03
2007M01
2007M11
2008M09
2009M07
2010M05
2011M03
2012M01
2012M11
2013M09
2014M07
SVM Output
Crisis
Mlogit Output

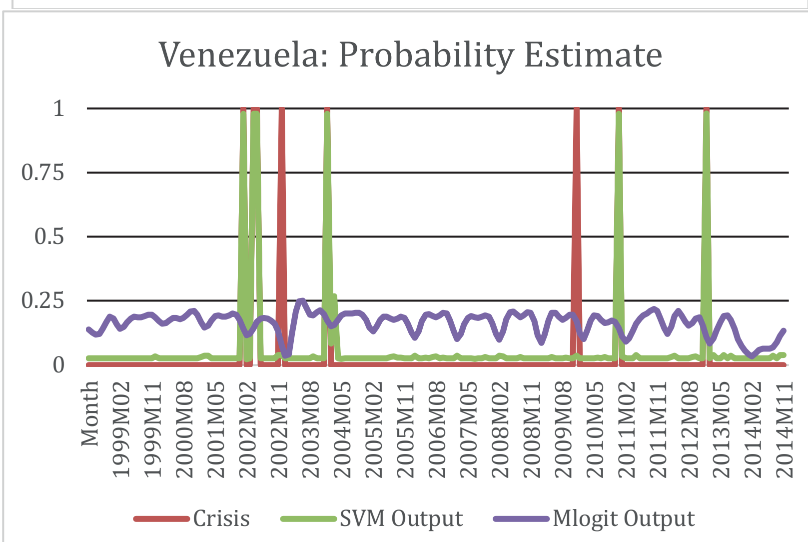

Poland: Probability Estimate
1
0.75
0.5
0.25
0
Month
2000M10
2001M06
2002M02
2002M10
2003M06
2004M02
2004M10
2005M06
2006M02
2006M10
2007M06
2008M02
2008M10
2009M06
2010M02
2010M10
2011M06
2012M02
2012M10
2013M06
2014M02
2014M10
SVM Ouput
Crisis
Mlogit Output


Russia: Probability Estimate
1
0.75
0.5
0.25
0
Month
1997M11
1998M08
1999M05
2000M02
2000M11
2001M08
2002M05
2003M02
2003M11
2004M08
2005M05
2006M02
2006M11
2007M08
2008M05
2009M02
2009M11
2010M08
2011M05
2012M02
2012M11
2013M08
2014M05
SVM Output
Mlogit Output
Crisis


Singapore: Probability Estimate
1
0.75
0.5
0.25
0
Month
1996M12
1997M10
1998M08
1999M06
2000M04
2001M02
2001M12
2002M10
2003M08
2004M06
2005M04
2006M02
2006M12
2007M10
2008M08
2009M06
2010M04
2011M02
2011M12
2012M10
2013M08
2014M06
Crisis
SVM Output
Mlogit Output


Thailand: Probability Estimate
1
0.75
0.5
0.25
0
Month
1996M12
1997M10
1998M08
1999M06
2000M04
2001M02
2001M12
2002M10
2003M08
2004M06
2005M04
2006M02
2006M12
2007M10
2008M08
2009M06
2010M04
2011M02
2011M12
2012M10
2013M08
2014M06
Crisis
SVM Output
Mlogit Output


Turkey: Probability Estimate
1
0.75
0.5
0.25
0
Month
1996M12
1997M10
1998M08
1999M06
2000M04
2001M02
2001M12
2002M10
2003M08
2004M06
2005M04
2006M02
2006M12
2007M10
2008M08
2009M06
2010M04
2011M02
2011M12
2012M10
2013M08
2014M06
SVM Output
Crisis
Mlogit Output


Venezuela: Probability Estimate
1
0.75
0.5
0.25
0
Month
1999M02
1999M11
2000M08
2001M05
2002M02
2002M11
2003M08
2004M05
2005M02
2005M11
2006M08
2007M05
2008M02
2008M11
2009M08
2010M05
2011M02
2011M11
2012M08
2013M05
2014M02
2014M11
Crisis
SVM Output
Mlogit Output

Robustness of SVM: Monte Carlo Simulation

Table 9. Monte Carlo and Error Statistics for SVM

Country	Mean Square Error		Monte Carlo Mean Error Rate SVM	
	Mlogit	SVM	MSE	SD
Argentina	0.03500	0.00002	0.00000	0.00000
Brazil	0.00392	0.00002	0.00000	0.00000
Chile	0.03160	0.00953	0.00096	0.01638
Colombia	0.02642	0.00196	0.00266	0.01966
Czech Republic	0.02261	0.00218	0.00219	0.02818
Hong Kong	0.02713	0.64000	0.00131	0.02584
Hungary	0.01895	0.00246	0.00246	0.02522
Indonesia	0.95329	0.89845	0.00134	0.01827
Korea	0.05788	0.00196	0.01825	0.01886
Malaysia	0.03034	0.00011	0.00005	0.06081
Philippines	0.03941	0.00008	0.00093	0.02519
Poland	0.02890	0.00015	0.00000	0.00000
Russia	0.03888	0.00009	0.00479	0.02741
Singapore	0.01944	0.00049	0.00390	0.03454
Thailand	0.04151	0.00124	0.00101	0.02737
Turkey	0.04894	0.00031	0.00031	0.01907
Venezuela	0.014972	0.00010	0.00176	0.05755

The table above provides mean square error statistics for both the multinomial logit and SVM, and the results of Monte Carlo Simulations for the SVM. We see that the SVM always provides a lower mean square error statistic reflecting an overall tighter fit to the data. To test the Robustness of the SVM. The Monte Carlo process was chosen primarily because the SVM is based on structural risk minimization, so we do not actually produce residuals, nor do we make assumptions over their distribution. Instead we incorporate tolerance for residual error through the support vectors or "slack variables" described in Appendix A. Therefore, traditional techniques such as adding or omitting variables from our equation or using alternative definitions and analyzing the impact on the residuals becomes meaningless in the context of the SVM. This Monte Carlo simulation recreated the results of the radial SVM 1000 times, using a different sample for training each time, it then calculated the respective error rates and standard deviations. This type of repeated simulation is ideal for establish the reliability and robustness of the SVM as a modelling technique. Above the results show that the error estimate of the Monte Carlo process is very close to the original results and all of them are well within the standard deviations of the simulation meaning that it is very unlikely that the SVM will produce highly varied results and is in fact robust. However, we also see an exception in the case of Argentina, and Brazil. Where thresholds were low and these countries only experienced one crisis each within the dataset. This implies that the SVM will produce relatively poor results when the degrees of freedom are limited within the sample, making it unreliable when little information is provided about the crisis within the model. As a final gage of performance, estimates for out of sample probabilities are provided in the next section.

Out of Sample Estimation:

As a final test for the usefulness of the Support Vector Machine and its applications. We conducted a comparison of out of sample forecasts. This was done by training the models on data up to a certain period then forecasting the probability of crisis in the next year. It was of upmost importance to choose dates that would include crisis periods to train the models, and where they subsequently experienced crises so that we could establish some meaning to the results. One question of particular interest would be whether or not the models could accurately predict the aftermath of the 2008-2009 global financial crisis? Or maybe even the currency implications of the crisis itself? As such, the periods chosen on which to train the data were 2004M12, 2007M12, and 2010M12. The probabilities calculated referred to years 2005, 2008, and 2011. This allowed for the smallest training set to be established over 6-8 years. Below Table 10 provides the results for the multinomial logit regression performed of the respective training periods. Table 11 shows a comparison of the out of sample probabilities for these time periods.

Table 10. Results of Multinomial Logit : Out of Sample

Variable (Yi,t=1)	End Point : 2004M12	End point: 2007M12	End point: 2010M12
CAGDP	-1.119 (0.858)	-2.216*** (0.702)	-1.696*** (0.584)
STDR	-0.840* (0.435)	-0.991*** (0.435)	-0.899*** (0.383)
OREER	0.159*** (0.015)	0.155***(0.0131)	0.156*** (0.011)
RGDP	-0.046*** (-0.014)	-0.285***(0.012)	-0.027***(0.010)
LBI	-0.324***(.109)	-0.197***(0.090)	-0.072(0.075)
Constant	-1.277***(.103)	-2.327***(0.090)	-1.083***(0.073)
No. Of Observations	1,592	2,204	2,816
Pseudo R^2	0.133	0.115	0.100

Note: Standard errors in parentheses. ***, **, * indicate statistical significance at the 99%, 95% and 90% levels,

Table 11. Predicted Probabilities Out of Sample

Country	Out-of-sample end date: 2004M12		Crisis in 2005	Out-of-sample end date: 2007M12		Crisis in 2008	Out-of-sample end date: 2010M12		Crisis in 2011
	M-logit	SVM		M-logit	SVM		M-logit	SVM	
Argentina	0.019	0.016	No	0.119	0.012	No	0.224	0.008	No
Brazil	0.091	0.015	No	0.256	0.011	No	0.251	0.006	No
Chile	0.094	0.009	No	0.162	0.975	08M6	0.271	0.957	11M9
Colombia	0.136	0.059	No	0.160	0.963	08M9	0.082	0.961	11M9
Czech Republic	0.206	0.032	No	0.070	0.982	08M8	0.123	0.953	11M9
Hong Kong	0.099	0.976	05M5	0.161	0.964	No	0.286	0.968	11M1
Hungary	0.234	0.015	No	0.162	0.985	08M8	0.142	0.953	11M9
Indonesia	0.148	0.045	No	0.237	0.972	08M10	0.195	0.014	No
Korea	0.046	0.980	No	0.688	0.975	08M9	0.215	0.952	11M9
Malaysia	0.120	NEDF	No	0.127	NEDF	No	0.170	NEDF	11M9
Philippines	0.138	0.951	No	0.302	0.957	08M10	0.295	0.966	No
Poland	0.204	NEDF	No	0.132	NEDF	08M8	0.243	0.974	11M9
Russia	0.083	0.042	No	0.126	0.034	No	0.086	0.064	No
Singapore	0.141	0.072	No	0.105	0.032	08M8	0.010	0.934	11M9
Thailand	0.170	0.114	No	0.321	0.187	No	0.311	0.034	No
Turkey	0.064	0.982	No	0.348	0.942	08M10	0.462	0.018	No
Venezuela	0.131	0.051	No	0.123	0.029	No	0.140	0.0978	11M1

NEDf Lack info.*

The out of sample results show us several things. One being that we see the multinomial logit and SVM become more accurate given more information. We also see that the Lending Boom Index Variable was more important in earlier periods and subsequently becomes insignificant. This may be due to the fact that latter crisis episodes had less to do with the growth of credit to the private sector. As opposed to crises, such as the East Asian crisis, that occurred during 1996-2004 where credit to the private sector was documented as a cause of many crises during that period. It could also be that governments demonstrated austerity in lending policy in the aftermath of those crises; leading to crisis episodes that simply had less to do with credit expansion.

We also see that in the cases of Malaysia and Poland, where the crisis episodes happened outside the sample period, that the support vector machine was unable to converge because it had no information on which to train. We see that for countries like Brazil and Argentina the models converged and were accurate even though they failed the Monte Carlo test for robustness. This is likely an artificial construct in that if we look at the in-sample results for these countries we see that they missed the crises in the sample completely. Both of these facts reinforce the idea that the SVM is not a good choice when trying to predict crises in countries with little available information on which to train the algorithm. However, when the SVM is fed a fair amount of information we see that it out performs the conventional multinomial approach. Providing higher signal levels and more accuracy in those signals. It is also interesting to see that it correctly predicts much of the currency turmoil surrounding the 2008-2009 financial crisis. We also see that the multinomial logit must be evaluated on a much lower threshold out of sample to provide good results. If we look at the out of sample prediction for 2008, we see that a 20% threshold would produce 6 alarms: 2 false alarm, 4 correct alarms. The model would also miss 6 crises for that sample period. Meaning it only called roughly 33% of crises correctly out of sample. Increasing the threshold only decreases the accuracy for the multinomial logit. However, the SVM performs well at low or high threshold, and tends to call over 50% of crises correctly out of sample. This implies that the SVM may allow for more flexibility than traditional approaches in respect to the out of sample analysis for private practitioners and policy makers alike.

VIII. Conclusions and Further Research:

This exercise established the usefulness of the support vector machine in helping to predict the occurrence of currency crisis. The literature was used to provide some background and demonstrate the shortcomings of the existing models and how introducing the concept of structural risk minimization and the support vector machine to overcome some of these issues. The literature was then used to establish the methodologies and variables of interest for our comparison. The results from the support vector machine were then compared to the multinomial logit model of Busierre and Fratzscher (2006). The accuracy of the model was analyzed and compared to the empirical results of other models on different datasets from Berg et al. (2004) in order to gage its usefulness for predicting currency crises. It was shown that the support vector machine is indeed useful. As it predicts the majority of observations correctly in and out of sample within our dataset of 17 countries between 1996M03 and 2014M12. The support vector machine also tends to perform more accurately than traditional methodologies, particularly when it comes to producing out of sample forecasts. However,, it was also noted that the support vector machine performs poorly when there is little information provided on which to train the algorithm. This limits its usefulness to countries that exhibit a large number of currency crises and will be more useful for forward looking out of sample predictions. Note that the latter is actually an advantage of the support vector machine.

This study suggests a number of avenues for further research. One direct extension would be to look at the performance of the support vector machine for predicting currency crashes as opposed to crises used in this study. This may be particularly useful to actors such as international investors. Who may be more interested in capturing large exchange rate changes over shorter time horizons. This study also opens the door to many future areas of research regarding the overlap of statistical machine learning and economic phenomena.

One area of interest may be to develop a formal methodology for the application of the support vector machine. This would include approaches to verify the best type of SVM to apply. In this case, issues related to the degrees of freedom were dealt with by altering the tuning function. For example, if the country had only one crisis episode then fixed sampling validation was used instead of k-fold cross validation. Since k-fold would not converge. Brazil and Argentina were examples of this. Also, since many approaches were tried, but many did not converge they were excluded by default (i.e. markov-switiching model, and the linear SVM). However, in other cases with other data sets the sampling methods and testing of various models may produce varying results that are open to interpretation in the absence of a methodological rubric. As such, undertaking research to study the relationships between traditional methods and the support vector machine may be fruitful in establish such a methodology.

Another idea for future research may be to take advantage of its ability to incorporate large amounts of information and apply it to areas in which economics has failed to find parsimonious of universal solutions. In fact, one idea might be to look at currency crises using 20-30 lead indicators. Typically, this tends to reduce the accuracy of limited dependent regressions, thereby limiting their usefulness in predicting phenomena that can occur through varying channels of transmission. Creating new EWS for these types of phenomena will more than likely be very fruitful as an area of research. Potentially using other measures of currency crises based on the logic of large depreciations in the exchange rate and incorporate research on the cost of Type 1 and Type 2 errors to better suit the needs of agents. Ultimately this research would be a labor of data science, as the coding and data structuring would be key.

On a final note, another area of potential research may lie in using other facets of the support vector machine. It can also conduct regression analysis and provides decision values for its output rather than probability values. Although for this exercise the probabilities were of better use for our comparison analysis the decision values may be used for other purposes. For example, may be one could take the values and run a regression against of crises index to produce a vulnerability index, where decision values would not be arbitrary in the same sense that threshold selection is in nature generally arbitrary. The values could be used to create scores that relate to a country's vulnerability to speculative pressure comparable to the "z-scores" designed by Altman for manufacturing firms. In many ways using support vector machines to create these types of scores or indices maybe useful as well.

Bibliography:

1. Abiad, Abdul, 2003, "Early Warning Systems: A Survey and a Regime-Switching Approach," Working Paper 03/32 (Washington: International Monetary Fund).
2. Ades, A., Masih, R., Tenengauzer, D., December 1998. GS-Watch: a new framework for predicting financial crises in emerging markets. Goldman Sachs.
3. Ahn, J. J., Oh, K. J., Kim, T. Y., & Kim, D. H. (2011). Usefulness of support vector machine to develop an early warning system for financial crisis. *Expert Systems with Applications*, *38*(4), 2966-2973.
4. Amlan R., and Tudela, M., 2000, "Emerging Market Risk Indicator (EMRI): Re-Estimated Sept 2000" (New York: Credit Suisse First-Boston).
5. Amri, P. D., & Willett, T. D. (2017). Policy Inconsistencies and the Political Economy of Currency Crises. *Journal of International Commerce, Economics and Policy*, *8*(01), 1750004.
6. Almahmood, H., Munyif, M. A., & Willett, T. D. (2018). Most Speculative Attacks Do Not Succeed: Currency Crises and Currency Crashes. *Journal of International Commerce, Economics and Policy*, *9*(01n02), 1850001.
7. Berg, A., & Pattillo, C. (1999a). Are currency crises predictable? A test. *IMF Staff papers*, *46*(2), 107-138.
8. Berg, A., & Pattillo, C. (1999b). Predicting currency crises: The indicators approach and an alternative. *Journal of international Money and Finance*,*18*(4), 561-586.
9. Berg, Andrew, Eduardo Borensztein, Gian Maria Milesi-Ferretti, and Catherine Pattillo, 2000, *Anticipating Balance of Payments Crises: The Role of Early Warning Systems,* IMF Occasional Paper 186 (Washington: International Monetary Fund).
10. Berg, A., Borensztein, E. and Pattillo, C. (2005). "Assessing early warning systems: How have they worked in practice?" IMF Staff Papers 52(3).
11. Burnside, C., Eichenbaum, M. and Rebelo, S. (2004). "Government guarantees and self-fulfilling speculative attacks." Journal of Economic Theory 119, 31-63
12. Bussiere, M., & Fratzscher, M. (2006). Towards a new early warning system of financial crises. *journal of International Money and Finance*, *25*(6), 953-973.
13. Calvo, G. (1997). "Varieties of capital-market crises" in Calvo, G. & King, M. (eds.) *The debt burden and its consequences for monetary policy.* London: MacMillan Press.

14. Chang, R. and Velasco, A. (2001). “A model of currency crises in emerging markets.” *Quarterly Journal of Economics* 116(2), 489-517.

15. Chen, S. H., & Wang, P. P. (2004). Computational intelligence in economics and finance. In *Computational Intelligence in Economics and Finance* (pp. 3-55). Springer Berlin Heidelberg.

16. Chaudhuri, A. (2014). Support Vector Machine Model for Currency Crisis Discrimination. Working Paper. *arXiv preprint arXiv:1403.0481*.

17. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

18. Dooley, M. (2000). "A model of crises in emerging markets." *Economic Journal* 110(460), 256-272.

19. Edison, H. (2003). “Do indicators of financial crises work? An evaluation of an early warning system.” *International Journal of Finance and Economics* 8(1), 11-53.

20. Efremidze, L, SM Schreyer and O Sula (2011). Sudden stops and currency crises. *Journal of Financial Economic Policy*, 3(4), 304–321.

21. Eichengreen, B., Rose, A. K., & Wyplosz, C. (1995). Exchange market mayhem: the antecedents and aftermath of speculative attacks. *Economic policy*, *10*(21), 249-312.

22. Flood, R. P., & Garber, P. M. (1994). *Speculative bubbles, speculative attacks, and policy switching*. MIT Press.

23. Frankel, J. A., & Rose, A. K. (1996). Currency crashes in emerging markets: An empirical treatment. *Journal of international Economics*, *41*(3), 351-366.

24. Frankel, J. A., & Saravelos, G. (2010). *Are leading indicators of financial crises useful for assessing country vulnerability? Evidence from the 2008-09 global crisis* (No. w16047). National bureau of economic research.

25. Garber, Peter M., Robin L. Lumsdaine, and Longato, Paolo. 2001, "Deutsche Bank Alarm Clock: Forecasting Exchange Rate and Interest Rate Events in Emerging Markets" (New York: Deutsche Bank).

26. Goldstein, M., Kaminsky, G. L., & Reinhart, C. M. (2000). *Assessing financial vulnerability: an early warning system for emerging markets*. Peterson Institute

27. Hutchison, M. and Noy, I. (2006). "Sudden stops and the Mexican wave: Currency crises, capital flow reversals and output loss in emerging markets." *Journal of Development Economics* 79,225–248.

28. International Monetary Fund (2002a). "Early warning system models: the next steps forward." *Global Financial Stability Report*, Ch 3, March.

29. Kamin, S. B., Schindler, J. W., & Samuel, S. L. (2001). The contribution of domestic and external factors to emerging market devaluation crises: An early warning systems approach. *FRB International Finance Discussion Paper*, (711).

30. Kamin, S. B., Schindler, J., & Samuel, S. (2007). The contribution of domestic and external factors to emerging market currency crises: an early warning systems approach. *International Journal of Finance & Economics*, *12*(3), 317-336.

31. Kaminsky, G. L., Lizondo, S., & Reinhart, C. M. (1998). Leading indicators of currency crises. *Staff Papers*, *45*(1), 1-48.

32. Kaminsky, G. L., & Reinhart, C. M. (1999). The twin crises: the causes of banking and balance-of-payments problems. *American economic review*, 473-500.

33. Kim, T. Y., Oh, K. J., Sohn, I., & Hwang, C. (2004). Usefulness of artificial neural networks for early warning system of economic crisis. *Expert Systems with Applications*, *26*(4), 583-590.

34. Kimbrough, K., Li Jen, S., July 2001. An early warning system for currency crises. Morgan Stanley Dean Witter.

35. Krugman, P. (1979). A model of balance-of-payments crises. *Journal of money, credit and banking*, *11*(3), 311-325.

36. Laeven, L. and Valencia, F. (2008). "Systemic banking crises: a new database." *IMF Working Paper* WP 08/224.

37. Mariano, R. S., Gultekin, B. N., Ozmucur*, S., Shabbir, T., & Alper, C. E. (2004). Prediction of currency crises: case of Turkey. *Review of Middle East Economics and Finance*, *2*(2), 87-107.

38. McKinnon, R. and Pill, H. (1995). “Credible liberalizations and international capital flows: The ‘overborrowing syndrome’.” In *Financial deregulation and integration in East Asia*, Ito, T.& Krueger, A.O. (eds.). Chicago: University of Chicago Press.

39. Mulder, Christian, Roberto Perrelli, and Manuel Rocha, 2002, "The Role of Corporate, Legal, and Macroeconomic Balance Sheet Indicators in Crisis Detection and Prevention," IMF, Working Paper 02/59 (Washington: International Monetary Fund).

40. Obstfeld, M (1986): “Rational and Self-Fulfilling Balance-of-Payments Crises”. American Economic Review, 76, pp. 72-80.

41. Obstfeld, M. (1994). “The logic of currency crises.” *Cahiers Economiques et Monétaires* 43,189–213.

42. Persaud. A, JP Morgan (1998): “Event Risk Indicator Handbook”. JP Morgan, London (can be downloaded from: (www.jpmorgan.com)

43. Radelet and Sachs, 1998: “The East Asian Financial Crisis: Diagnosis, Remedies, Prospects.”, Brookings Papers on Economic Activity, Vol. 1 (1998), pp. 1–90.

44. Ramli, N. A., Ismail, M. T., & Wooi, H. C. (2013). *Designing Early Warning System: Prediction Accuracy of Currency Crisis by Using k-Nearest Neighbour Method*. LAP LAMBERT Academic Publishing.

45. Ramli, N. A., Ismail, M. T., & Wooi, H. C. (2015). Measuring the accuracy of currency crisis prediction with combined classifiers in designing early warning system. *Machine Learning*, *101*(1-3), 85-103.

46. Reinhart, C. M., & Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press.

47. Reinhart, C. M., & Rogoff, K. S. (2008). This Time Is Different: A Panoramic View of Eight Centuries of Financial Crises. NBER Working Paper 13882. In *National Bureau of Economic Research*.

48. Reinhart, C.M. (2002). Default, Currency Crises, and Sovereign Credit Ratings. *World Bank Economic Review,* 16 (2), 151-170.

49. Schnatz, B., September 1998. Macroeconomic determinants of currency turbulences in emerging markets. Discussion Paper 3/98. Economic Research Group of the Deutsche Bundesbank.

50. Sevim, C., Oztekin, A., Bali, O., Gumus, S., & Guresen, E. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, *237*(3), 1095-1104.

51. Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.

52. Velasco, A. (1987). Financial crises and balance of payments crises: a simple model of the Southern Cone experience. *Journal of development Economics*, *27*(1-2), 263-283.

53. Veropoulos, K., C. Campbell and N. Cristianini (1999), "Controlling the Sensitivity of Support Vector Machines", Proceedings of the International Joint Conference on Artificial Intelligence, Sweden, Workshop ML3, pp. 55 – 60.

54. Vlaar, P. J. (2000, March). Early warning systems for currency crises. In *BIS conference papers* (Vol. 8, pp. 253-274).

55. Witten J.G, Hastie, T.D., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 6). New York: springer.

56. Yu, L., Lai, K. K., & Wang, S. Y. (2006). Currency crisis forecasting with general regression neural networks. *International Journal of Information Technology & Decision Making*, *5*(03), 437-454.

57. Q. L. Zhang, "Currency Crisis Prediction Using Wavelet-Based Support Vector Machine", Advanced Materials Research, Vols. 989-994, pp. 2560-2564, 2014

Appendix A: A Review of Hyperplanes and the SVM

The Support Vector Machine (henceforth, SVM) was developed in 1995 by Vapnik and is a novel classification schema from the field of statistical machine learning. It is a learning algorithm, in that it uses past data to train and formulate the underlying structural relationship between the dependent variables and the probability of crisis and assigns new observations to one category or the other. The SVM allows for non-linear decision boundaries with non-separable data, which has been a major challenge to limited dependent regressions. The support vector machine is known to supersede the performance of these types of models by providing more robustness to the individual variables, and generally better classification (Witten et al., 2013). The following sections introduce the theory behind the support vector machine and have been adapted from Witten et al. (2013).

A Review of Hyperplanes: How do they relate to the SVM?

The support vector machine is an extension of the simplest linear classifier known as the maximal margin classifier. In essence, the maximal margin classifier is a simple one-dimensional hyperplane which works well on linearly separable data in Cartesian space. In general, if the data can be perfectly separated using a linear hyperplane then there will exist an infinite number of such hyperplanes that separate the data. The intuition is that marginal shifts in a perfectly separating hyperplane will produce different hyperplane that can still separate the data in continuous space. It is important to note that a linear hyperplane is defined as a flat *affine* (meaning it does not have to pass through the origin) subspace of dimension p-1, where p is the number of separable classes. For example, in three-dimensional space (three variables of interest) a hyperplane would be a flat two-dimensional subspace. In two dimensions the definition of a hyperplane is quite simple, and is given as:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \quad (1)$$

In this case, the hyperplane is simply the equation for a line, since in two-dimensional space the flat affine subspace is indeed a line. It is important to note that equation (1) means that for any vector $X = (X_1, X_2)^T$ for which (1) holds is a point on the hyperplane and determined by parameters β_0, β_1, and β_2. Also, one may consider the fact that hyperplanes, while difficult to visualize, can be extended to dimensions beyond three-dimensional space. Below the equation in a p-dimensional setting is given:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0 \;\; (2)$$

Here this equation defines a p-dimensional hyperplane in that if point $X = (X_1, X_2, \dots X_p)^T$ or a vector of length p satisfies equation (2) then it lies on the hyperplane. If a vector X is above the hyperplane then the point will be defined by the equation:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p > 0 \;\; (3)$$

If the vector X is below the hyperplane it will be defined by an equation of form:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p < 0 \;\; (4)$$

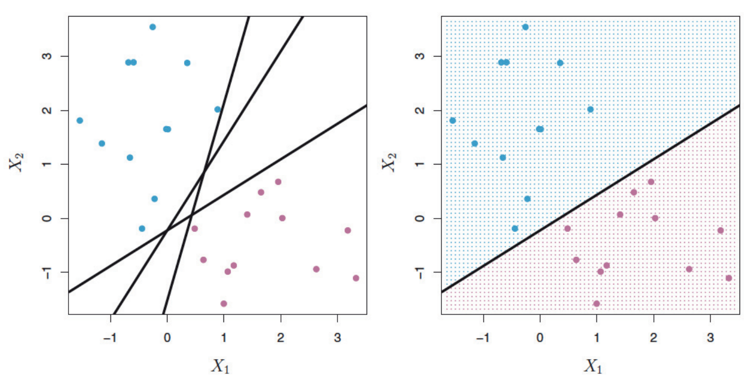
where the same definitions and parameters are used. We will be using these definitions to establish a mathematical foundation for classification techniques.

Classification using Hyperplanes: Basic Concepts

Suppose you are given an n x p matrix X that consists of n-training observations in a p-dimensional space such that,

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \dots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

and that the outcomes fall into one of two classes they are either 1 or -1 (as in our crisis definition), and in more technical terms this can be stated as $y_1, \dots, y_n \epsilon \{1, -1\}$. There also exists a test observation or a p-vector of observed features $x^* = (x_1^* \dots x_p^*)^T$. Now, suppose that it is possible to construct a hyperplane that perfectly separates the data according to the class labels. (See examples of three such hyperplanes from Witten et al. 2013 below),



On the left, there are two classes of observations, shown in blue and in purple, each of which has measurements on two variables. Three separating hyperplanes, out of many possible, are shown in black. On the right a separating HPERPLANE is shown in black. The blue and purple grid indicates the decision rule made by a classifier based on this separating hyperplane. The observations are classified by the respective portion of the grid on which they appear.

From this diagram, we see that we can label blue observations as $y_i = 1$ and purple observations as $y_i = -1$. This ensures that the separating hyperplane has the property that:

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} > 0 \ if \, y_i = 1 \quad (5)$$

and,

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} < 0 \ if \, y_i = -1 \quad (6)$$

or equivalently,

$$y_i \left(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \right) > 0 \qquad (7)$$

$\forall \ i = 1, \ldots, n$

This leads to the foundational notion that if a separating hyperplane exists that we can use it to construct a very basic yet natural classifier. In that, by constructing hyperplanes we can assign test observations to either side as an event 1 or -1. That is to say, that we classify the test observation x^* based on the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$ if $f(x^*)$is positive then the observation is assigned to class 1, if negative then to class -1. We can also make probability decisions based on the magnitude of $f(x^*)$ in that the closer the magnitude is to zero the less confident is the classification, whereas when it is further away, we can be more confident. The next section will begin to bring these properties and ideas together in order to create a technically sound classifier based on underlying relationships.

The Maximal Margin Classifier: Bringing some ideas together:

As mentioned above, if one separating hyperplane exists then an infinite number of separating hyperplanes exist. This is because in continuous space the hyperplane can be shifted or rotated by a small margin without touching the given observations; to construct an efficient classifier a method for choosing the optimal separating hyperplane must be established. In very simple terms the optimal separating hyperplane will be the one that is furthest from the test observation on both sides. To establish this classifier, we compute the perpendicular distance between each observation and a given hyperplane. The smallest distance between a hyperplane and the observations is known as margin. The optimal separating hyperplane (or maximal margin hyperplane) is the one that has the largest margin given the test observations. Once the optimal separating hyperplane has been established it is then possible to classify an event based on which side of the it the observation lies. This is known as the maximal margin classifier. In technical terms if $\beta_0, \beta_1, \ldots, \beta_p$ are the coefficients of the optimal separating hyperplane, then the maximal margin classifier classifies the test observation x^* based on the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$. It is important to note that while the maximal margin classifier is often useful it can lead to overfitting when p is large.

Next, the details that underlie task of constructing the optimal separating hyperplane will be discussed. Suppose a set of n training observations $x_1, \dots, x_n \in \mathbb{R}^p$associated with class labels $y_1, \dots, y_n \epsilon \{1, -1\}$ is given. Then the optimal separating hyperplane is the solution to:

$$\max_{\beta_0, \beta_1, \dots, \beta_p} M \quad (8)$$

subject to,

$$\sum_{j=1}^{p} \beta_j^2 = 1 \quad (9)$$

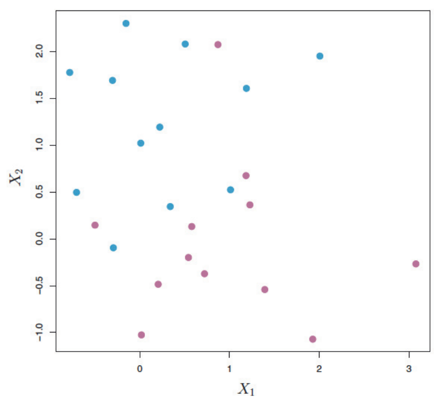
and,

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \geq M, \forall i + 1 \dots n \quad (10)$$

Note that equation (10) guarantees that the observations will be on the correct side given a bit of cushion (M), given that M is positive. Also note that equation (9) is not really a constraint on the hyperplane, rather a constraint that when paired with equation (10) ensures that you maximize the perpendicular distance from the i-th observation. In this case, the perpendicular distance is given by $y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})$. Therefore, these constraints ensure that the observations are always on the correct side of the hyperplane at a least distance M. Where M is the maximal margin and the optimization problem chooses $\beta_0, \beta_1, \dots, \beta_p$ to maximize M. Thus, the optimal separating hyperplane is established, and the maximal margin classifier can be used for classification analysis. However, this is only true in a perfectly separable and linear case. The next few sections will extend the classification schema to non-separable and non-linear cases, ultimately, leading to the construction of the support vector classifier and the support vector machine.

<u>The Non-Separable Case: Introducing the Support Vector Classifier</u>

Sometimes the data cannot be easily separated. In this case, there will be no efficient solutions to equations (8) - (10). Specifically, no solution with M > 0 will exist. This causes the optimal separating hyperplane (maximal margin classifier) to break down and become undesirable as a classifier. Below we see an example of non-separable data.



Above: There are two classes of observations, shown in blue and purple. In this case, the two classes are not separable by a hyperplane, and so the maximal margin classifier cannot be used.

A classifier based on the theory behind the maximal margin classifier will inherently classify the training observations correctly, leading to hypersensitivity to the addition of new information and lead to a margin M that is too small, hence over-fitting the data. In this case, it may be necessary to choose a classifier that does not separate the classes perfectly. Thereby misclassifying a few training observations for the sake of classifying more of them correctly. This will allow for more robustness to individual observations, and better classification of most of the data.

The support vector classifier or soft margin classifier allows for this type misclassification to occur. It separates most of the training data into their respective classes but allows some leeway by solving the following optimization problem:

$$\max_{\beta_0, \beta_1, \dots, \beta_p, \varepsilon_1, \dots, \varepsilon_n} M \quad (11)$$

subject to,

$$\sum_{j=1}^{p} \beta_j^2 = 1 \quad (12)$$

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \geq M(1 - \varepsilon_i) \quad (13)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C (14)$$

Here, C is a non-negative tuning parameter that allows some slack in the classification of training observations up to a certain margin of error. As with the example of the optimal separating hyperplane margin M is still being maximized, but we allow some slack to individual observations based on the slack variables $\varepsilon_1, \dots, \varepsilon_n$. The classification is then still determined by solving (11) – (14) then by deciding on which side of the hyperplane the observation lies based on the sign of the output $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$. The slack variable ε_i now allows the location of the i-th variable to be known. Consider if $\varepsilon_i = 0$, then the test observation would lie on the correct side of the hyperplane. Likewise, if $\varepsilon_i \geq 0$ then the test observation will lie on the wrong side of the margin, or if $\varepsilon_i > 1$ then it will be on the wrong side of the hyperplane. Note now that sum of ε_i's is bounded by C, or the tuning parameter. This dictates the number and severity of violations that the system will tolerate. In practice, it is chosen via cross-validation, and is used to control the bias to variance trade-off pertaining to this methodology. When C is large the system of equations allows for more violations leading to a lower fit, increase in bias, and lower variance. However, when C is small the system allows less violations thereby leading to a higher fit, decreasing in bias, and an increased variance.

It is important to notice a special property of the optimization problem characterized by equations (11) – (14). Note that only observations that lie on the margin or violate the margin will affect the hyperplane. The fact that the support vector classifier's decision rule is based only on a small subset of the training observations (the support vectors), implies that it is robust to the behavior of observations that are far away from the hyperplane. This is a property that distinguishes the support vector classifier from limited dependent regressions. Also, note that the support vector classifier still only works in the case of a linear decision boundary. In the next section a general mechanism for producing non-linear decision boundaries will be discussed, and the support vector machine introduced.

Classification with Non-Linear Decision Boundaries: The Support Vector Machine

When the data is linearly separable the support vector classifier is the obvious choice. However, when there is a non-linear relationship between the predictors and the outcome, we can enlarge the feature space using functions of the predictors to address this problem of non-linearity. The support vector machine is an extension of the support vector classifier that expands the feature space in a unique way, by using the correlation of functions between individual observations known as kernels. When solving equations (11) – (14) only the inner product of the of observations, not the observations themselves matter. The inner product of two observations $x_i, x_i^{'}$ is given by:

$$\langle x_i, x_i^{'} \rangle = \sum_{j=1}^{p} x_{ij} x_{ij}^{'}$$

Thus, it can be shown that,

- The support vector classifier can be written:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle$$

or,

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i K\langle x, x_i \rangle$$

where there are n parameters α_i, or one per training observation, K is a kernel

- To estimate parameters $\alpha_1, \ldots, \alpha_n, and\, \beta_0$ all that is needed are the inner products and the number of pairs among the n observations (given by n(n-1)/2).

The solution for the support vector classifier can now be rewritten in terms of the inner product which allows the use of functions of the inner product that relate the similarity in the position of two variables to extend the feature space. Below are some examples of commonly used Kernels and what the decision boundaries look like when computed:

4. Linear Kernel (e.g. Standard Pearson Correlation)

$$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j},$$

   Using the linear kernel gives the support vector classifier and is still only for linear analysis.

5. Polynomial Kernel of Degree d

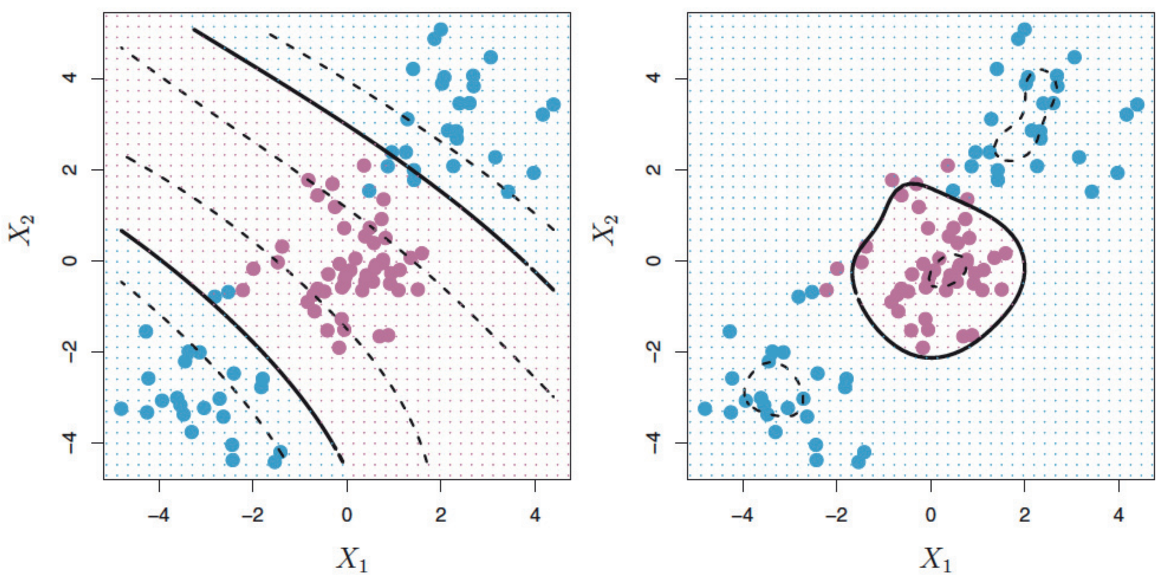
$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d.$$

   Here, d is a positive integer greater than one. This is the same as fitting the support vector classifier in a higher dimensional space of polynomials of degree d, rather than the original feature space. This leads to a much more flexible decision boundary (shown below with d=3). Usually, degrees 3 and 4 are the most robust.

6. Radial Kernel

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2).$$

This is another positive non-linear kernel where gamma is a positive constant. Anytime a non-linear kernel is used in a support vector classifier we call it a support vector machine. Below are examples of non-linear, and non-separable data that has been converted to a working feature space using polynomial and radial kernels.



On the left we see An SVM with polynomial kernel of degree 3 is applied to non-linear data resulting in a far more appropriate decision rule. On the right, we see an SVM with a radial kernel (n.b. in this example either kernel captures a fairly accurate decision boundary).

Appendix B : Summary of Models Reviewed

Model	Time Horizon	Crisis Definition	Method	Variables Used
KLR (1998)	2yrs	Weighted Average of one month changes in exchange rate and reserves more than three deviations above the country average	Signal: Weighted average of indicators, where variables are measured as binary indicators based on minimizing signal to noise ratio.	Overvaluation, Current Account, Reserve Loss, Export Growth, Reserves/M2 (level), Reserves/M2(growth), Domestic Credit Growth, Money Multiplier Change, Real Interest Rate, Excess M1 Balances
Berg and Pattillo (1999)	2yrs		Probit framework in which the independent variable takes a value of one if there is a crisis in the subsequent 24 months and zero otherwise	Real Exchange Rate Deviations, Current Account, Reserve growth, Export Growth, M2/ Reserves (level and growth)
Bussiere and Fratzscher (2006)	1yrs	EMP: Weighted average of indicators, where weights correspond to inverse variance of vars. 2SD from country specific mean	Multinomial Logit: Controlling for pre, during, and post crisis periods using standardization techniques of respective variance	REER overvaluation, Current account (%GDP), Short-term debt/reserves, Real GDP growth rate, Domestic credit to private and government sector (level and growth rate)

Goldman Sachs (Ades et. Al, 1998)	3mos	Weighted Average of three-month changes in exchange rate and reserves above country specific threshold	Logit regression with most of the RHS variables measured as binary based on thresholds found in autoregression with dummies (SETAR)	Overvaluation, Export Growth, Reserves/M2 (level), Financing Requirement, Stock Market, Political Event, Global Liquidity, Contagion
Deutsch Bank Alarm Clock (Garber et al., 2001)	1mos	Various "trigger points" typically depreciation > 10% and interest rate increase >25%	Logit two equation simultaneous systems on exchange rate and interest rate events of different magnitude	Overvaluation, Industrial Production, Domestic Credit Growth, Stock Market, Devaluation Contagion, Market Pressure Contagion, Regional Dummies, Interest Rate events
CSFB – EMRI (Amlan and Tudela, 2001)	1mos	Depreciation of > 5% and at least double the preceding month	Logit regression with RHS measured in logs, then deviation from mean and standardized	Overvaluation, Debt/Exports, Private Sector Credit Growth, Reserves/Imports (level), Oil Prices, Stock Price Growth, GDP Growth, Regional Contagion,
Federal Reserve (Kamin et al., 2001/2007)	2yrs	Weighted average of 2-month percentage changes in the real bilateral exchange rate against the dollar and in international reserves, with the weights being	Multivariate Probit	GDP growth, Fiscal deficits, Bank loans, M2/reserves, External debt, Terms of trade, US real short-term interest rates, Industrial country GDP growth,

		proportional to the inverse of the standard deviation of these series. Declines in these weighted averages more than 1.75 standard deviations indicate a crisis month		
Bundesbank (Schnatz, 1998/ 2007)	1yrs	2-factor EMP using percentage change in reserves and percentage change in real exchange rate. Where a crisis is when the EMP exceeds 1.5 SD from the country average for the first time.	Bivariate Logit	Change Domestic credit/GDP, Change in Claims of mon. auth. on cent. Gov/GDP, Change in Narrow money/GDP, Money market indicator (narrow money), Mon. reserves/narrow money, Change Mon. Reserves/narrow money, Inflation differential, Real exchange rate (Dev. from the trend), Change in real exchange rate, Exports/imports, Current account/GDP, Change in Exports, US money market rate, Change Output index,
DCSD (Berg et al., 2000)	2yrs	Weighted Average of one-month changes in exchange rate and reserves more than three deviations above the	Probit regression with RHS variables measured in country-specific percentage terms	Overvaluation, Current Account, Reserve Losses, Export Growth, ST Debt/Reserves

		country average		
IMF (Abiad, 2003)	3mos	Month to month percentage change in nominal exchange rate: using first order two state Markov chain where 1 represents crisis and 0 represent non-crisis	Markov – Switching Approach	Deviations of RER from trend, Current Account Balance/GDP, Export Growth Rate, M2/Reserves level, M2/ Reserves growth rate, Reserves growth rate, Growth rate of real domestic credit (deflated by nominal GDP), Industrial Production Growth rate, Real GDP growth rate, Stock market performance growth, Real Interest Rate, LIBOR, Bank assets/ GDP growth rate, STD/Reserves, Cumulative non FDI flows/GDP, Portfolio flows, Bank reserves/total bank assets, CB credit to banks/ Total bank-liabilities, Bank Deposits/ M2 level, Bank Deposits/ M2 growth rate, Loans/deposits level, Loans/deposits Growth rate,
Yu et al. (2006)	2 mos.		Generalized Neural Network	currency exchange price, rate of change of price (yt), ten-day moving stochastic oscillator of price (vt), ten-day moving averaging of yt (yt bar), ten-day moving variance of (yt), moving variance ratio (rt).
Busierre and Fratzscher (2006) -	1yr	EMP: Weighted average of indicators, where weights	Multinomial Logit: Controlling for pre, during, and post-crisis periods using	REER overvaluation, Current account (%GDP), Short-term debt/reserves, Real GDP growth rate, Domestic credit to private and government sector (level and

		correspond to inverse variance of vars. 2SD from country specific mean	standardization techniques of respective variance	growth rate), Contagion Measure
Abiad (2003)	3 mos.	Month to month percentage change in nominal exchange rate: using first order two state Markov chain where 1 represents crisis and 0 represent non-crisis	Markov Switching Process	Deviations of RER from trend, Current Account Balance/GDP, Export Growth Rate, M2/Reserves level, M2/ Reserves growth rate, Reserves growth rate, Growth rate of real domestic credit (deflated by nominal GDP), Industrial Production Growth rate, Real GDP growth rate, Stock market performance growth, Real Interest Rate, LIBOR, Bank assets/ GDP growth rate, STD/Reserves, Cumulative non FDI flows/GDP, Portfolio flows, Bank reserves/total bank assets, CB credit to banks/ Total bank-liabilities, Bank Deposits/ M2 level, Bank Deposits/ M2 growth rate, Loans/deposits level, Loans/deposits Growth rate,
Yu et al. (2006)	2 mos.	EMP: Three-factor Equal weights	Generalized Neural Network	currency exchange price, rate of change of price (yt), ten-day moving stochastic oscillator of price (vt), ten-day moving averaging of yt (yt bar), ten-day moving variance of (yt), moving variance ratio (rt).

Appendix C : Data Transformation (Adapted from Bussiere and Fratzscher ,2006)

1. Current Account to GDP

   Sources: IMF International Financial Statistics (IFS). Data are quarterly and have been trans- formed into monthly frequency by using a cubic spline.

2. Short-term debt to reserves ratios

   Sources: IMF IFS for reserves and the IMF, World Bank, OECD, BIS joint table for the debt statistics for short-term debt. Data are mostly quarterly and have been transformed into monthly frequency using a cubic spline.

3. Overvaluation of the exchange rate

   Sources: IMF International Financial Statistics. The real effective exchange rate (REER) is a valuable concept to look at since it takes into account the competitiveness of the home country, as well as competition with third countries. As a definition of overvaluation, we use the REER deviation from a linear trend computed over the full sample period for each country. The motivation for using such a trend as the benchmark is that one would expect the currencies of emerging markets to appreciate over time due to the Balassa-Samuelson effect. However, a potential problem arising from this definition is that the trend is strongly influenced by the first few and last few observations in the sample.

4. Real GDP growth

   Source: IMF (IFS). Data are quarterly and have been transformed into monthly frequency by taking a moving average.

5. Financial sector fragilities

   Source: IMF (IFS). The measure first suggested by Sachs et al. (1996), who define a lending boom index in a country as the growth rate of domestic credit to the private sector over the past four years was used.