Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

A Coherent Framework for Predicting Emerging Market Credit Spreads with Support Vector Regression

Gary Anderson and Alena Audzeyeva

2019-074

Please cite this paper as:
Anderson, Gary, and Alena Audzeyeva (2019). "A Coherent Framework for Predicting Emerging Market Credit Spreads with Support Vector Regression," Finance and Economics Discussion Series 2019-074. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2019.074.

A Coherent Framework for Predicting Emerging Market Credit Spreads with Support Vector Regression*

Gary Anderson†
Board of Governors of the Federal Reserve System, USA

Alena Audzeyeva‡
Keele Management School, Keele University, UK

October 15, 2019

*Correspondence to: Alena Audzeyeva, Keele University. The analysis and conclusions in this paper are those of the authors and do not indicate concurrence by members of the research staff or the Federal Reserve Board of Governors.

†Email: cemar.llc@gmail.com

‡Email: a.audzeyeva@keele.ac.uk

Abstract

We propose a coherent framework using support vector regression (SVR) for generating and ranking a set of high quality models for predicting emerging market sovereign credit spreads. Our framework adapts a global optimization algorithm employing an hv-block cross-validation metric, pertinent for models with serially correlated economic variables, to produce robust sets of tuning parameters for SVR kernel functions. In contrast to previous approaches identifying a single "best" tuning parameter setting, a task that is pragmatically improbable to achieve in many applications, we proceed with a collection of tuning parameter candidates, employing the Model Confidence Set test to select the most accurate models from the collection of promising candidates. Using bond credit spread data for three large emerging market economies and an array of input variables motivated by economic theory, we apply our framework to identify relatively small sets of SVR models with superior out-of-sample forecasting performance. Benchmarking our SVR forecasts against random walk and conventional linear model forecasts provides evidence for the notably superior forecasting accuracy of SVR-based models. In contrast to routinely used linear model benchmarks, the SVR-based models can generate accurate forecasts using only a small set of input variables limited to the country-specific credit-spread-curve factors, lending some support to the rational expectation theory of the term structure in the context of emerging market credit spreads. Consequently, our evidence indicates a better ability of highly flexible SVR to capture investor expectations about future spreads reflected in today's credit spread curve.

KEYWORDS: Machine Learning; Support Vector Machine Regressions; Sovereign credit spreads; Emerging Markets; Out-of-sample predictability; Model Confidence Set.

JEL CLASSIFICATIONS: G17; G15; C53.

Contents

1 Introduction

The emerging economies bond market has become a core global asset class in the past decade, playing an increasingly important role in portfolio allocation and risk management decisions of international investors. Moreover, measures of sovereign credit spreads serve as benchmarks for pricing other domestic assets such as corporate bonds and credit derivatives. Yet, very little is known about their predictability in real-time, in sharp contrast to the voluminous literature on *riskless* bond yields that has evolved in recent years; some prominent examples include Fama and Bliss (1987), Diebold and Li (2006), Ang and Piazzesi (2003), and Ludvigson and Ng (2009). Furthermore, a key role played by emerging debt markets in global financial stability provides additional important motivation for developing accurate predictions of emerging market sovereign credit spreads.

However, recent empirical studies of sovereign emerging market credit spreads tend to focus on their predictability in-sample; see, for example, Longstaff et al. (2011) and Comelli (2012). Such evidence does not directly extend to predictions in real time. Furthermore, the few studies that attempt such forecasts document that credit spreads are difficult to predict out-of-sample (OOS). In particular, in a recent study Audzeyeva and Fuertes (2018) show that predictive models employing country-specific credit-spread-curve factors, namely, the level, slope and curvature, known to contain useful information for future yields in the context of *riskless* debt, cannot beat a random walk for *emerging market* bonds. They find that employing additional global and country-specific predictors improves the model predictive ability, but that even with these additional predictors the model-based-forecasts cannot always outperform a random walk.

However, these forecasting results rely on assuming that expectations about future credit spreads are a linear function of the level, slope and curvature factors of today's credit spread curve. While this is a plausible assumption, also advocated, for example, in Diebold and Li (2006) for U.S. Treasury yields, it represents only one of many possible expectation mechanisms that, in fact, is likely to be more complex in nature for globally-traded emerging market bonds. In particular, it is well known that emerging market bond prices tend to be noisy and that their distributional properties change over time, triggered, for example, by a domestic or external credit event, or a global economic or financial market crisis. This complexity motivates the application of SVR for modeling and forecasting of such series. The key advantages of SVRs, which are data-driven, non-parametric

models, are that (a) unlike linear models, they do not require strong a-priori assumptions about the relationship between the target variable and predictors, and (b) they are, by design, better able to allay the issue of over-fitting inherent in the standard multivariate linear regression techniques. Furthermore, SVRs have demonstrated superior performance in time series prediction relative to both conventional modeling approaches and alternative machine learning techniques such as neural networks; see Cao and Tay (2001) and Stasinakis et al. (2016) for comprehensive overviews of SVR financial modeling applications.

This study contributes, first, to the sparse literature on sovereign credit spread prediction by applying SVR for the OOS forecasting of credit spreads of large emerging market borrowers. To the best of our knowledge, this is the first study that applies the SVR methodology in the context of emerging bond markets. In doing so, we extend the analysis of linear predictive models underpinned by the assumption of linearly formed expectations in extant emerging market bond studies by permitting more complex expectation mechanisms afforded by highly-flexible SVR model specifications. Furthermore, we go one step beyond many SVR forecasting studies of financial market time series that, similar to technical analysis traders, tend to rely on predictive content in the historical data of the target variable alone for constructing input variables. Examples are Law and Shawe-Taylor (2017) who forecast the U.K. and U.S. based stock market indices, commodity futures, government bond yields and corporate CDS, Stasinakis et al. (2016) and Sermpinis et al. (2017b) who focus on predicting U.S. based commodity exchange traded funds (ETF) and European stock market ETFs, respectively. In contrast, our predictive models are motivated by economic theory and as such are aligned with investment strategies of more sophisticated fundamental traders, employing as input variables an expanded array of predictors containing both global and domestic fundamentals.

Second, we contribute to the forecasting methodology by proposing a coherent framework using SVR for generating and ranking a robust *set* of high quality forecasting models. This contrasts with extant studies advocating the selection of *one* "best" predictive model, a task that is pragmatically improbable to achieve in many applications. To illustrate the issue, consider the methodology for developing an SVR model that entails two important stages. During the first stage, the modeler selects specific SVR kernels and must set a small number of tuning parameters that determine how well the model produced by the subsequent SVR optimization stage will characterize the data

(Cao and Tay, 2001). While the methodology for the second, optimization stage is theoretically sound and relatively straight forward to implement, Stasinakis et al. (2016), Law and Shawe-Taylor (2017), Sermpinis et al. (2017a) and Sermpinis et al. (2017b) among others emphasize that SVR forecasting performance is highly sensitive to the kernel, or tuning, parameters selected at the first stage, pointing out that there is little formal guidance in the literature on how to set these parameters.

With lacking compelling theoretical guidance for setting tuning parameters, modelers tend to resort to applying an optimization technique, with grid search being the most frequent choice, in conjunction with a metric characterizing the goodness of fit, typically based on the forecast RMSE; see, for example, Min and Lee (2005), Ding et al. (2008) and Gunduz and Uhrig-Homburg (2011). Among few studies employing global optimization techniques with their measure of fitness, Stasinakis et al. (2016) apply the Krill-Herd meta-heuristic optimization method to introduce Krill-Herd SVR whereas Sermpinis et al. (2017a) incorporate opposition based optimization to develop the reverse adaptive Krill global search method in a search for the best tuning parameter setting. Law and Shawe-Taylor (2017) apply a Bayesian approach that follows Gao (2002) in assuming the existence of a most likely SVR prediction function generated by a Gaussian process. However, the mapping from the data to the tuning parameters is highly non-linear and extremely challenging to compute. To simplify the problem, the authors employ a Taylor-series approximation for the relationship, performing global quasi-optimization to find the most likely values of tuning parameters.

A common assumption in all of these approaches is that they can identify with some degree of certainty one “best" set of tuning parameter values producing a single best SVR forecasting model for a given kernel. However, the complexity of the objective function and the existence of many local optima makes it unattainable for even global optimization techniques to distinguish with certainty between a number of potentially “best" solutions. The problem is further exacerbated by typically limited data sets available to the modelers. Law and Shawe-Taylor (2017) highlight the issue, proposing to calculate error bounds for their parameter values as a way to resolve it. However, their error bound calculations depend on a number of approximations, with the resulting accuracy requiring further analysis. As a result, there is no persuasive evidence that extant approaches can successfully identify a unique global solution among perhaps many plausible local alternatives.

Another issue faced by the modeler is that the standard SVR forecasting methodology may be unsuitable for applications involving serially correlated data series that tend to generate serially

correlated forecast errors as is the case in our bond market application. In particular, Brabanter et al. (2011) show that standard cross-validation schemes can interpret the serial correlation as a high frequency relationship with small variance, leading to spurious parameter choices.[1]

To address the first methodological issue, contrary to previous SVR literature requiring the modeler to select one best parameter setting, our forecasting framework permits that a chosen SVR kernel can generate multiple profitable forecasts. The multiplicity of forecasts arises because a SVR kernel can accommodate a set of viable choices for tuning parameters. Next, we propose a coherent 3-step strategy for managing such multiplicity. Our approach first uses a robust global optimization algorithm, namely, multi-sequential number-theoretic optimization (MSNTO) put forward by Xu et al. (2005), in conjunction with *hv-block* cross validation of Racine (2000) to generate a set of promising model candidates. The advantage of the global optimization routine is that it is more robust than alternative methods such as a grid search against choosing bad local optima and does not require much guidance on what would be a good initial guess. Our choice of *hv-block* cross validation as a measure of fitness is motivated by its robustness for serially correlated series and forecast errors that are common in many economic and financial time series applications, addressing the second methodological issue in the SVR forecasting literature caused by presence of serial correlation in the data. In step two, predictor models from step one are estimated using the standard SVR optimization methodology. In step three, we select a robust set of most accurate models from the collection of promising candidates from the previous step by applying the Model Confidence Set test of Hansen et al. (2011), known to be robust to data snooping, that permits identifying a subset of a group of models with the best forecast accuracy.

We apply our framework to a quarter-ahead OOS forecasting of credit spreads for three large and relatively mature emerging markets of Brazil, Mexico and Turkey. All three sovereign borrowers are major emerging market economies and members of the G20. In our predictive analysis, we employ the data set of Audzeyeva and Fuertes (2018). Our analysis employs SVR with linear, sigmoid, RBF and polynomial kernels to generate credit spread forecasts using sets of input variables that are motivated by economic theory. We find that our framework delivers a relatively small robust set of

[1]Among few studies aiming to address the issue of serial correlation, Bergmeir et al. (2018) argue that *k-fold* cross-validation that they apply to both generated and real-life commodity data sets can be adequate for some applications using non-parametric estimation that satisfy certain conditions. However, we find that our bond market application does not meet their requirements.

high-quality SVR forecasting models for each sovereign borrower. The model sets are characterized by country-specific preferred kernel functions. Our results provide evidence of notably superior forecasting accuracy of the highly-ranked SVR models relative to both alternative SVR specifications and standard benchmarks. We further find that our SVR models can deliver accurate credit spread forecasts with only a small set of input variables limited to the credit-spread-curve factors for Mexico and Turkey and the credit-spread-curve factors augmented by global yield-curve factors for Brazil, performing as well as or better than both other SVR and benchmark models using extended input sets.

2 Data variables

2.1 Emerging market credit spreads and spread curve factors

The target variable in our predictability analysis is the credit spread on sovereign bonds of an emerging market country c

$$y_{c,t}(\tau) \equiv Y_{c,t}(\tau) - Y_{US,t}(\tau), \tag{1}$$

where $Y_{c,t}(\tau)$ and $Y_{US,t}(\tau)$ are the time-t yield to maturity on τ-maturity zero-coupon bonds of emerging-market country c and the U.S. Treasury, respectively. In our predictive analysis we set $\tau = 5$ years. The weekly frequency dataset of Audzeyeva and Fuertes (2018) contains the yields to maturity on zero-coupon U.S. Treasuries and U.S. dollar-denominated Eurobonds of each country we used to calculate the country-specific credit spreads.[2]

Adopting the Nelson and Siegel (1987) representation permits expressing the credit spread on τ-maturity zero-coupon bond of the emerging-market sovereign c as a parsimonious function of its *credit-spread-curve* factors:

$$y_{c,t}(\tau) = \beta_{c0,t} + \beta_{c1,t}\left(\frac{1-e^{-\lambda_{c,t}\tau}}{\lambda_{c,t}\tau}\right) + \beta_{c2,t}\left(\frac{1-e^{-\lambda_{c,t}\tau}}{\lambda_{c,t}\tau} - e^{-\lambda_{c,t}\tau}\right) \tag{2}$$

[2]Audzeyeva and Fuertes (2018) extract yields on risky and riskless bonds from cross-sections of bond market prices for a given emerging market country and U.S. Treasury bonds, respectively, by following an established methodology that builds on seminal work of Fama and Bliss (1987), Svensson (1994) and Diebold and Li (2006).

where $\beta_{c0,t}$, $\beta_{c1,t}$ and $\beta_{c2,t}$, are the *level*, *slope* and *curvature* factors, respectively. The country credit-spread-curve factors are available at a weekly frequency in our dataset.[3]

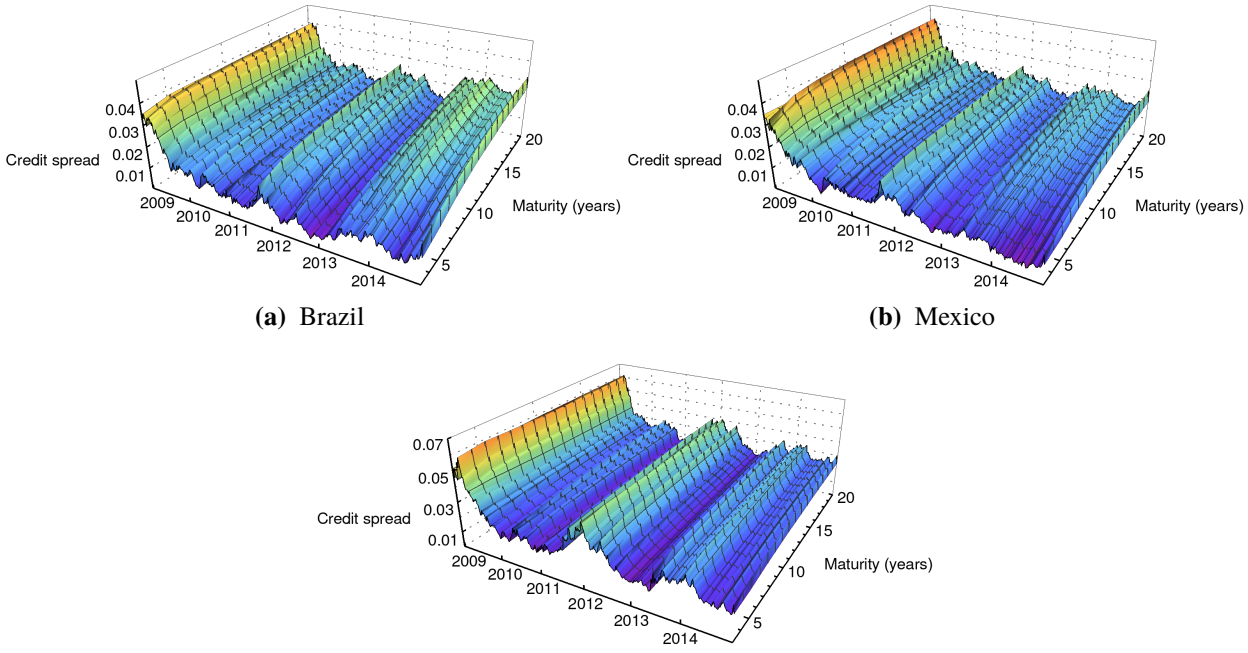
Figure 1 plots the evolution of credit spread curves for Brazil, Mexico and Turkey during our December 2, 2008 to December 29, 2015 sample period while Table A1 in Appendix A gives summary statistics for the target variable, 5-year credit spreads, and the credit-spread-curve factors.[4] Credit spreads and credit-spread-curve factors of all countries exhibit stylized persistence. Figure 1 further shows that although the dynamics of credit spreads exhibit common trends across the three emerging market economies, driven by various global market factors, country-specific variations are, nevertheless, apparent. Such variations primarily reflect differences in country-specific creditworthiness. In particular, in December 2008 Brazil and Mexico, both investment-grade BBB-rated by the S&P credit rating agency, exhibited lower credit spreads than Turkey that was rated as speculative-grade BB throughout the data sample period. While Mexico's credit rating remained stable, with its spreads experiencing a long-term downward trend, Brazil's spreads rose sharply with the S&P issuing a negative rating outlook for Brazil on June 6, 2013, followed by several negative rating changes that led to Brazil's downgrade to speculative-grade rating BB on September 9, 2015. In Table A1, these differences in creditworthiness are reflected in the lower mean credit spread and spread volatility at 139 and 65 basis points, respectively, for Mexico, with the corresponding figures reaching 159 and 81 basis points for Brazil and even higher, 253 and 93 basis points, for Turkey.

2.2 Input variable selection

The selection of predictive variables for the input vector in our analysis is based on prior research on predictability of sovereign emerging market spreads. We employ predictive variables, motivated by economic theory, that have been previously analyzed in Audzeyeva and Fuertes (2018). Accordingly, our *Baseline* model is rooted in the expectations theory of the term structure of interest rates of Sargent (1972) and Roll (1970). The main idea is that investor expectations about future credit spreads, reflected in today's forward spreads, exploit all the available information. Consequently,

[3] Audzeyeva and Fuertes (2018) extract the credit curve level, slope and curvature factors from weekly cross-sections of country-specific bond prices; see their paper for details.

[4] We use the dataset constructed for Audzeyeva and Fuertes (2018) for Figure 1, Table A1 and all our training and validation calculations.



(a) Brazil

(b) Mexico

(c) Turkey

Figure 1 Emerging market credit spreads

today's credit spread curve, which embeds forward credit spreads, ought to contain all relevant information for predicting future spreads. In line with this theory, the *Baseline* model employs as input variables *level*, *slope* and *curvature* of the country-specific *credit spread curve*, known to summarize the information content in today's spread curve. Furthermore, as in Audzeyeva and Fuertes (2018), we test two alternative model specifications that augment the input vector in the *Baseline* model by (a) global macroeconomic variables, model G, and (b) both global and domestic macroeconomic variables, model GEM. Table 1 lists input variables in each predictive model and Table A1 in Appendix A provides summary statistics of the data variables.

3 Forecasting framework

Our forecasting framework entails a three-step process. Step one involves specifying predictive models by selecting SVR kernel functions and setting tuning parameters. In step two, predictive models obtained in step one are estimated using the standard SVR methodology, and their OOS forecasting accuracy is subsequently assessed, permitting the selection of a subset of best performing models in step three. For the OOS predictability analysis, we allocate for training the first $2/3N_{obs}$

Table 1
Input variables in the forecasting models

Input Variables	Models		
	Baseline	G	GEM
Country spread curve factors (level, slope, curvature)	✓	✓	✓
Global predictors			
US yield curve factors (level, slope, curvature)		✓	✓
Volatility of US short-term rate		✓	✓
Domestic predictors			
Country risk rating			✓
Trade balance level			✓
Trade balance volatility			✓
Terms-of-trade growth level			✓
Terms-of-trade growth volatility			✓

consecutive weeks of the N_{obs} = 367 weekly observations available in the data sample period, setting aside remaining $1/3N_{obs}$ weeks for OOS forecast evaluation (Hansen et al., 2011). The forecasts are generated using rolling regressions.

In what follows we first introduce the standard SVR methodology (step two) for which kernel specifications serve as inputs, subsequently outlining our methodology for setting SVR tuning parameters (step one) and evaluating model forecasts for model selection (step three).

3.1 Support vector regression

SVR implements structural risk minimization, an important tenet of statistical learning theory, with the purpose of constructing models with reliable OOS performance (Vapnik, 1995). Instead of empirical risk minimization, which minimizes the error on observed data, as in linear regression and most other conventional estimation methodologies, SVR seeks to minimize an upper bound on the generalization error. As a consequence of its documented superior OOS predictive performance, the technique has found wide acceptance in financial series forecasting; see, for example, Cao and Tay (2001), Min and Lee (2005), Stasinakis et al. (2016), Law and Shawe-Taylor (2017), Sermpinis et al. (2017a), and Sermpinis et al. (2017b).

Using a training sample containing I observations of a scalar target variable y and a vector of

predictor variables $x \in \mathcal{R}^Q$, SVR constructs predictive models for q-step-ahead forecasts of y of the form:

$$y_{t+q} = f(x_t) + u_t = w_0 + \sum_{i=1}^{I} w_i k(x_i, x_t; \Upsilon) + u_t \tag{3}$$

We set $q = 13$ weeks, or one quarter, in our predictive analysis, as in Audzeyeva and Fuertes (2018). Kernel function $k(x_i, x; \Upsilon)$ in Eq.3 effectively maps the input data vector x into a higher dimensional space representing a w_i-weighted linear sum of terms that can better predict the target variable y due to its superior flexibility. The vector of tuning parameters Υ provides a way for varying aspects of the nonlinear mapping of the data. This non-linear mapping, often referred to as "kernel trick", provides a variety of parsimoniously parameterized nonlinear functional forms (Hofmann et al., 2008). Our analysis employs four kernel functions described in Table 2: linear, polynomial, radial basis and sigmoid functions.

SVR chooses a vector of weights w_i that minimizes the *regularized empirical risk function*:

$$\begin{aligned} &\min \frac{1}{2} \|w\| + C \sum_{i=1}^{I} (\xi_i + \xi_i^*), \quad \text{such that} \\ &y_{i+q} - f(x_i) \leq \epsilon + \xi_i, \;\; f(x_i) - y_{i+q} \leq \epsilon + \xi_i^*, \\ &\qquad \xi_i, \xi_i^* \geq 0 \;\; \text{for all } i. \end{aligned} \tag{4}$$

Here the bandwidth parameter, ϵ, determines an ϵ-*insensitive* region for characterizing empirical risk, and the regularization parameter, C, determines the trade-off between a measure of flatness of the function, $\|w\|$, and the level of empirical risk.

The predictions of the parameterized function $f(x_i)$ can violate the constraint, but at a cost proportional to C. With this so called "double-hinged" loss function, the loss will be zero when $|y_{i+q} - f(x_i)| \leq \epsilon$, and increase linearly at the rate C for points where the predicted value falls outside the ϵ-insensitive region. Based on statistical learning theory, this ϵ-SVR formulation explicitly provides robustness against parameter-driven model over-fitting through the judicious choice of the regularization and bandwidth parameters (Smola and Schölkopf, 2004).[5]

[5]In an alternative ν-SVR formulation, the modeler specifies parameter ν determining an upper bound on the fraction observations which can fall outside the ϵ-insensitive region and a lower bound of the fraction of support vectors (Scholkopf et al., 2000). Then, ν is used to determine C and ϵ. Since there is no strong prior available regarding an

Table 2
Kernel functions

Name	Kernel function $k(x_i, x)$	Kernel-specific parameters[1]
Linear	$({x_i}^T x)$	—
Radial basis functions[2]	$e^{-\frac{1}{\psi^2 \|\|x_i - x\|\|^2}}$	ψ
Sigmoid[3]	$\tanh(\gamma({x_i}^T x) + s)$	s, γ
Polynomial[4]	$(\gamma({x_i}^T x) + s)^g$	s, g, γ

[1] Each kernel function requires a choice of the regularization parameter, C, and the bandwidth parameter, ϵ, along with any kernel-specific parameters, altogether comprising the vector of tuning parameters Υ.

[2] Parameter ψ controls the radius of influence of individual observations: larger ψ reduces the radius of influence.

[3] The sigmoid kernel retains some of the properties of a logistic curve but here the values range between ± 1. Reducing γ makes for a more gradual transition between the extreme values and for the response to appear more linear for intermediate values. Whereas s determines the location of the point where the kernel function value crosses zero.

[4] The polynomial kernel generalizes the linear kernel function by providing a nonlinear response to the dot product value: the larger the g, the more nonlinear the response. Here γ moderates the sensitivity to the nonlinear interaction term while s determines the location of the zero response.

The loss function, Eq. 4, has a dual Lagrangian of the form

$$\begin{aligned}\mathcal{L} = \frac{1}{2}||w||^2 + C\sum_{i=1}^{I}(\xi_i + \xi_i^*) - \sum_{i=1}^{I}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \\ \sum_{i=1}^{I}\upsilon_i(\epsilon + \xi_i - y_i + k(x_i, x; \Upsilon) + w_0) - \\ \sum_{i=1}^{I}\upsilon_i^*(\epsilon + \xi_i^* + y_i - k(x_i, x; \Upsilon) - w_0)\end{aligned} \tag{5}$$

Eq. 5 constitutes a quadratic programming problem (QPP) with well known properties and solution algorithms. The optimization produces a set of nonzero weights υ_i, υ_i^* which identifies a collection of "support vectors", influential observations that determine the optimal set of weights – deleting

appropriate value for ν in our application, we have applied ϵ-SVR.

the other observations and again solving the QPP produces the same set of optimal weights.[6]

Note that in minimizing Eq. 4, we are only free to adjust w, ξ_i and ξ_i^*. The aim is to obtain w needed for constructing predictive models of credit spreads, Eq. 3. Two basic tuning parameters C and ϵ and any other parameters in Υ defining the kernel function $k(x_i, x; \Upsilon)$ serve here as inputs. Consequently, the SVR model forecasting performance will depend upon the choice of the kernel function, parameters C and ϵ, and additional tuning parameters, if any, in Υ. Given a kernel function and a set of tuning parameters, the SVR approach will produce a single forecasting model. However, changing tuning parameters would typically produce a different forecasting model. In Sections 3.2 and 3.3 we show how to accommodate this multiplicity.

3.2 Setting tuning parameters

This section outlines the procedure we propose for selecting tuning parameter candidates, Υ, that serve as input into Eq.3. This procedure is nontrivial for two reasons. First, there is little theoretical guidance for guessing appropriate tuning parameter values. Second, as in our bond market application, when searching for appropriate tuning parameters, the modeler is likely to encounter many local minima that are difficult to rank with any degree of certainty. Consequently, to mitigate the risk of neglecting favorable tuning parameters we employ MSNTO, a robust global optimization technique put forward by Xu et al. (2005). Furthermore, as predictive models in our bond market application generate serially correlated forecast errors, we apply the *hv-block* cross-validation algorithm of Racine (2000) together with MSNTO, as an appropriate metric for evaluating the predictive performance of a given set of tuning parameters in this context.

We do not use any of the validation set for tuning parameter selection, employing the training dataset alone at this stage. The resulting tuning parameter values are then fixed during our subsequent OOS forecasting exercise. We will first describe the hv-Block cross-validation metric and then move on to describe the global optimization routine.

[6]For our calculations, we have used the "C++" version of **libsvm**, a widely used open source SVR implementation.

3.2.1 hv-Block cross-validation

Since SVR has been developed primarily for OOS prediction, researchers have routinely used cross-validation procedures, estimating the magnitude of expected OOS forecast errors, to ensure the robustness of forecasts. However, the presence of serial correlation in the data and forecast errors can complicate the estimation process, making routinely used cross-validation techniques like k-fold cross validation potentially unsuitable. This is because the shape and smoothness of the fitted SVR functions depend critically upon the kernel bandwidth ϵ. Many cross-validation techniques interpret the serial correlation as a high frequency relationship with small variance which can result in setting ϵ too low for tracking the function variation (Brabanter et al., 2011). In other words, serial correlation can cause the algorithm to favor narrower than appropriate bandwidths in order to more effectively limit smoothing of the fitted function, erroneously taking the correlated observations as part of variation in the function value instead of part of the error.

The hv-block cross-validation approach put forward by Racine (2000) provides a way to estimate the accuracy of forecasts in the presence of autocorrelated forecast errors. Consider training data set $\mathcal{X} = \begin{bmatrix} x \\ y \end{bmatrix}$ pictured in Figure 2.[7] To apply hv-block cross-validation, for each observation i we collect v observations on either side of the observation $\mathcal{X}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$ to construct a local validation set for time $t = i$ of size $n_v = 2v + 1$. Then h observations are removed from either side of the local validation set, with the remaining $n_c = n - 2v - 2h - 1$ observations forming the local training set for time i.[8] The algorithm then trains on each local training set of size n_c, and computes errors on each local validation set of size n_v. Here h limits the impact of autocorrelation, maintaining "near" independence between validation and training sets by h-blocking, whereas v ensures consistency of the estimated forecast error variance.

Denote the validation data subset, designated "For Validation" in Figure 2, by $\mathcal{X}_{(i:v)} =$

[7]Note that all of the observations in Figure 2 come from the original training data set, so that the cross-validation only uses data from within this data set.

[8]The actual count remaining will increase when the observation near the beginning or end of the available data as one of the holdout sections shrinks.

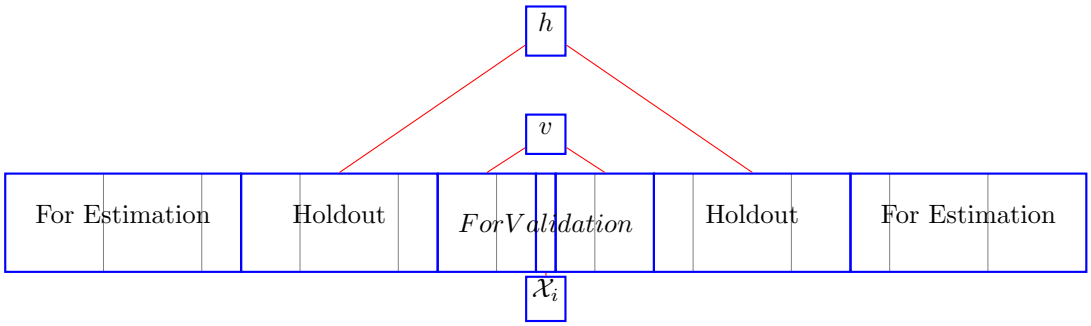


Figure 2 hv-Block cross-validation

$(y_{(i:v)}, x_{(i:v)})$. The hv-block cross-validation estimate of the forecast error variance is given by:

$$\Phi(\Upsilon) = \frac{1}{(n-2v)n_v} \sum_{i-v+1}^{n-v} ||y_{(i:v)} - \widehat{y_{(i:v)}}(\Upsilon)||^2 \tag{6}$$

The guidelines in Racine (2000) support $h = 56, v = 28$ for our training data set of 244 observations.

3.2.2 Multi-Sequential Number-Theoretic Optimization

When searching for optimal tuning parameter values, the modeler faces an optimization problem that can have many local minima. To illustrate the problem, Figure 3 plots the $-\Phi(\Upsilon)$-surface given by Eq.6 for Mexico, as an example, using the Baseline forecasting model with a linear kernel. The figure shows many widely dispersed peaks representing local minima of the objective function, with no clear dominating, or "best", point among the parameter values. Furthermore, we do not have a way to guess, with any precision, the regions where good parameter values might lie. Consequently, this type of problem calls for a robust global optimization routine like MSNTO that can search broadly yet will not be confounded by bad potential parameter choices.[9]

Consider a multivariate domain containing all of the parameter values we care to entertain $\pi = [a, b] \in \mathcal{R}^l$, $a_\iota \leq \Upsilon_\iota \leq b_\iota$, for all $\iota \in \{1, \ldots, l\}$ and a function, $\Phi(\Upsilon)$, continuous on π. We

[9]Employing derivatives seems impractical due to excessive computational complexity in the optimization routine for obtaining cross-validation values.

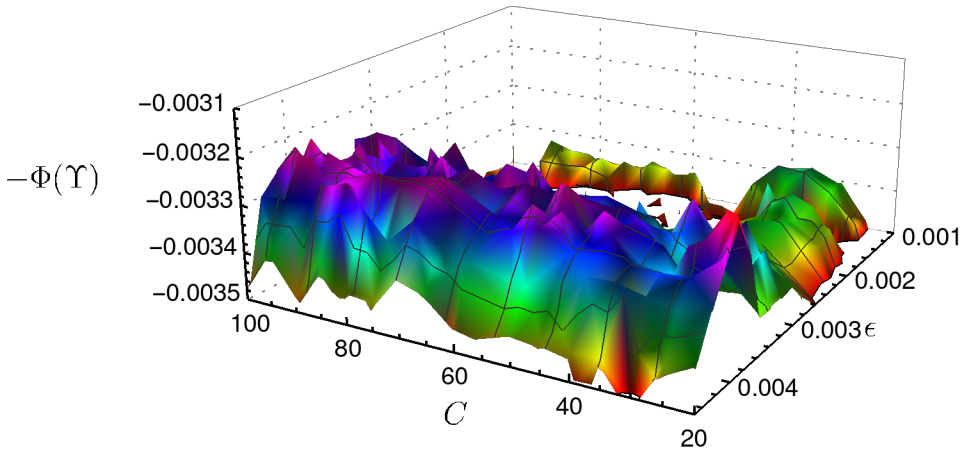


Figure 3 $-\Phi(\Upsilon)$-surface, Eq. 6, for Mexico: *Baseline* model with a linear kernel

to seek to find

$$\Upsilon^* \in \pi \text{ such that } \Upsilon^* = \Phi(\Upsilon^*) = \min_{\Upsilon \in \pi} \Phi(\Upsilon)$$

We choose a set $\boldsymbol{\Upsilon} = \{\Upsilon_\iota, \iota = 1, \ldots, \Lambda\}$ widely dispersed on π, approximating Υ^* by $\widehat{\Upsilon^*_\Lambda} \in \boldsymbol{\Upsilon}$ such that

$$\Phi(\widehat{\Upsilon^*_\Lambda}) = \min_{\iota \in \{1,\ldots,\Lambda\}} \Phi(\Upsilon_\iota).$$

This simple strategy is known to converge slowly. Recursively contracting the search region produces dramatic speed improvements. However, this sequential best choice scheme might find only a local optimum. To avoid getting stuck at a local optimum, various authors propose implementing multiple local searches based on clustering. We employ the MSNTO algorithm that generates sample points using a number-theoretic sequence that maximizes *dispersion*. Then, at each step the algorithm retains a small fraction, Π, of these points for clustering. Since the exact dispersion evaluation would be costly, the algorithm uses an approximation for dispersion,

$$\widehat{\mathcal{B}(\rho, \pi)} = \max_{1 \le \iota \le \Lambda} \min_{1 \le \theta \le \Lambda, \iota \ne \theta} \|\Upsilon_\iota - \Upsilon_\theta\|,$$

and calculates $\rho = \rho_\Phi \widehat{\mathcal{B}(\rho, \pi)}$, a small multiple of the estimate of dispersion. Here ρ determines the range of points that should be included in a single cluster. The algorithm applies a user supplied

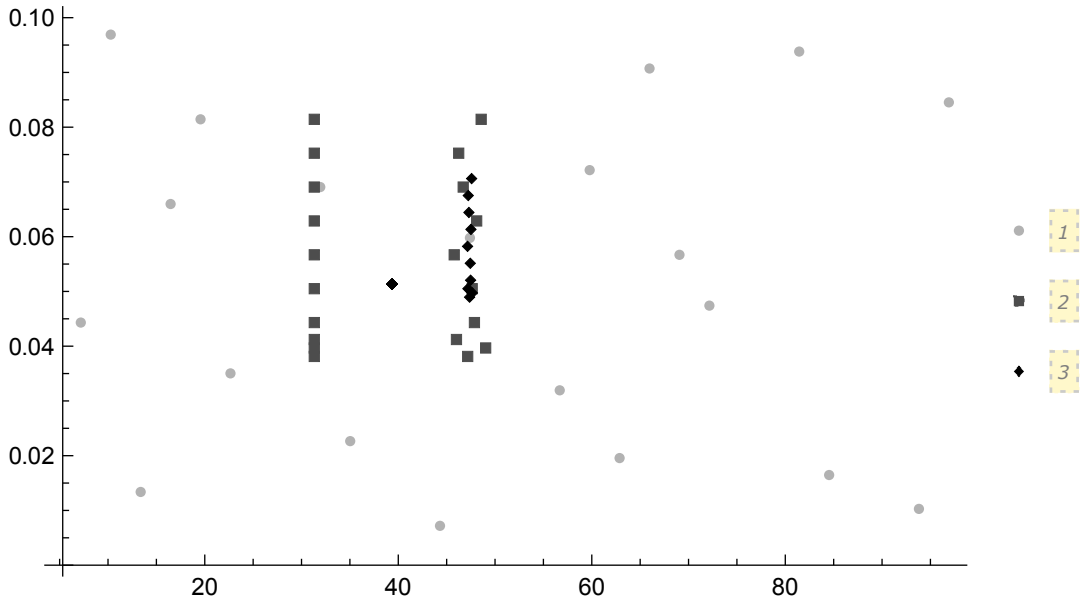


Figure 4 MSNTO $(-\Phi(C, \epsilon))$ evaluation point clustering and contraction: An example from an SVR with a linear kernel

parameter to determine the maximum amount of contraction toward the local minimum to apply to each of the cluster regions.

These new cluster regions replace the old search regions during the next recursion. Xu et al. (2005) show that starting out using a number theoretic sequence with a very large set of points, Λ_1, and continuing the recursion with a somewhat smaller set of points in the sequence, Λ_2, improves the algorithm's performance. We used $\Lambda_1 = 100, \Lambda_2 = 15, \Pi = 0.1, \rho_\Phi = 1.2, \sigma = 0.5$ and generated points using a Niedrreiter sequence (Niederreiter et al., 1983). Table B1 in Appendix B gives the ranges for search for the parameters of each kernel. Figure 4 further illustrates this process for a two dimensional $\Phi(C, \epsilon)$ parameter search using a SVR with a linear kernel as an example. The algorithm constructs clusters of points around each local minimum. The first set of points is widely dispersed, covering a large region of the search space. The second iteration has identified two clusters and retains the points associated with the two clusters. The third iteration further refines the search to clusters that encompass a smaller region thereby increasing the precision of the approximation for the location of the two local minima.

For a given kernel/model combination, applying the MSNTO algorithm to minimize the cross-validation error produces a collection of candidate tuning parameter settings. Table B2 in Appendix B shows our tally of distinct promising parameter settings for each kernel/model combination.

3.3 Model selection and evaluation

We employ the expected value of the squared error loss to evaluate the predictive ability of the quarter-ahead OOS forecasts generated by various SVR candidate models. However, when choosing among a collection of forecasting models using a fixed data set, the "data snooping" issue arises: model results that outperform may just be the result of luck. To overcome this issue, we evaluate statistical significance of gains in predictive accuracy by means of the Model Confidence Set (MCS) test of Hansen et al. (2011). The procedure uses forecast errors to identify a subset of a group of models whose members likely have the best forecasting accuracy.

This subset of models, that the authors call a model confidence set, is constructed to contain all the superior models with a specified level of confidence. The MCS test has a number of important advantages over widely used alternative tests put forward by Diebold and Mariano (1995), White (2000) and Hansen (2005). First, the MCS test provides additional information that is useful to the modeler: a measure of uncertainty surrounding model selection. The second advantage is related to the MCS test sensitivity to the utility of information in the available data such that informative data produces a small collection of good models whereas uninformative data generates large model confidence sets. Third, unlike alternative tests of pair-wise model comparisons, the MCS test does not require a benchmark, facilitating direct comparisons of forecasting accuracy among multiple competing model candidates which is particularly useful in our case.

Let the squared error loss function for the model $j^{'}$ prediction $\hat{y}_{j^{'},t}$ of y_t to be given by $L_{j^{'},t} = L(y_t, \hat{y}_{j^{'},t}) = (\hat{y}_{j^{'},t} - y_t)^2$. Define the measure of relative model performance as $\mu_{j^{'}j^{''}} \equiv E(L_{j^{'},t} - L_{j^{''},t})$. Thus, model $j^{'}$ is preferred to model $j^{''}$when $\mu_{j^{'}j^{''}} < 0$. The authors assume that $\mu_{j^{'}j^{''}}$ is finite and independent of t.

To define the MCS test procedure, consider a finite initial collection of forecasting models, $\mathcal{M}^0$. Let $\mathcal{M}^*$ denote the set of best models for a specified metric of model assessment:

$$\mathcal{M}^* \equiv \{j^{'} \in \mathcal{M}^0 : \mu_{j^{'}j^{''}} \leq 0 \text{ for all } j^{''} \in \mathcal{M}^0\}.$$

A model confidence set at level α, $\mathcal{M}^*_{1-\alpha}$, is then a subset of $\mathcal{M}^0$ containing all of $\mathcal{M}^*$ with a probability $(1-\alpha)$. For a given loss function and confidence level, the test uses sample information about each model's relative performance OOS to sequentially eliminate the poorest performing

models, producing p-values for each model in $\mathcal{M}^0$. Small p-values indicate low probability that the model is actually among the best. The test procedure estimates $\widehat{\mathcal{M}}^*_{1-\alpha}$ via a sequence of significance tests with null hypothesis $H_{0,\mathcal{M}} : \mu_{j'j''} = 0$ for all $j', j'' \in \mathcal{M}; \mathcal{M} \subset \mathcal{M}^0$ and alternative, $H_{A,\mathcal{M}} : \mu_{j'j''} \neq 0$ for some $j', j'' \in \mathcal{M}$. The sequential elimination from the model confidence set continues until doing so reduces the coverage ratio, $(1-\alpha)$, below the specified confidence level.

The authors show that, since the procedure uses the same significance level in all tests, all models with p-values greater than α are in $\widehat{\mathcal{M}}^*_{1-\alpha}$. When the test assigns relatively high p-values to only one or few models, this serves as compelling evidence of their superior predictive accuracy relative to the competitor models. Alternatively, when the evidence does not support a few strong candidates, there may be many models with similarly high p-values. We employ the Hansen et al. (2011) maximum t-statistic test for assessing model forecasting accuracy using OOS forecast errors from rolling regression estimation.[10]

To contrast the accuracy of the SVR forecasts with that of standard benchmarks from the credit spread forecasting literature, we add a variety of benchmark forecasts to the list of competing forecasts in our MCS test. A random walk (RW) model makes the first natural benchmark as it is documented to be difficult to beat in Audzeyeva and Fuertes (2018). To represent another set of benchmarks, commonly utilized linear regression models, we employ three time-series OLS regression models from Audzeyeva and Fuertes (2018) that use the same sets of input variables as our SVR forecasts. Thus, we consider three additional *OLS* benchmark models. The first model, *OLS-Baseline*, generates forecasts that utilize the predictive content in the credit-spread-curve factors:

$$y_{c,t+q} = \psi_c + \kappa_{c0}\hat{\beta}_{c0,t} + \kappa_{c1}\hat{\beta}_{c1,t} + \kappa_{c2}\hat{\beta}_{c2,t} + \nu_{c,t+q} \tag{7}$$

The second OLS benchmark, *OLS-G*, augments *OLS-Baseline* with the vector of global macroeconomic input variables, $\boldsymbol{G}_t$:

$$y_{c,t+q} = \psi_c + \kappa_{c0}\hat{\beta}_{c0,t} + \kappa_{c1}\hat{\beta}_{c0,t} + \kappa_{c2}\hat{\beta}_{c2,t} + \boldsymbol{\theta}^G_c \boldsymbol{G}_t + \nu_{c,t+q} \tag{8}$$

[10]We have used the *ForecastEval* package, a Julia implementation of the MCS.

In the third OLS model, ***OLS-GEM***, the input set is further augmented with the vector of domestic macroeconomic variables, $\boldsymbol{EM}_{c,t}$:

$$y_{c,t+q} = \psi_c + \kappa_{c0}\hat{\beta}_{c0,t} + \kappa_{c1}\hat{\beta}_{c0,t} + \kappa_{c2}\hat{\beta}_{c2,t} + \boldsymbol{\theta}_c^G \boldsymbol{G}_t + \boldsymbol{\theta}_c^{EM} \boldsymbol{EM}_{c,t} + \nu_{c,t+q} \tag{9}$$

The predictive horizon here $q = 13$ weeks as before. The OLS benchmark predictions are obtained using rolling regressions, based on the same training and OOS evaluation windows as respective SVR predictions.

4 Empirical forecasting results

4.1 Predictive ability of SVR models

Accuracy of the quarter-ahead OOS SVR forecasts generated using models with various sets of input variables: $Baseline$, G, and GEM, employing Linear, RBF, Sigmoid and Polynomial SVR kernels, is evaluated by running a horse race among competing models. Table 3 reports the results for twenty SVR-based forecasts that have the lowest RMSE for a given country, contrasting them with forecast RMSE of benchmarks utilized in the literature. To gauge the statistical significance of gains in forecast accuracy, we report Model Confidence Set (MCS) p-values, identifying model forecasts in $\widehat{\mathcal{M}}^*_{75\%}$, $\widehat{\mathcal{M}}^*_{50\%}$ and $\widehat{\mathcal{M}}^*_{25\%}$.[11,12] The interpretation of the MCS test confidence level is analogous to that of a confidence interval for a parameter where MCS identifies from a collection of model candidates a subset of models that contain the best model with a given level of confidence. Model forecasts are ordered by MCS p-value, with those more likely to generate the most accurate forecasts listed first. The wide range of reported p-values across models provides evidence of high information utility in each country's data unambiguously identifying a sub-group of most accurate forecasts.

The OOS forecasting evidence confidently identifies superior predictive accuracy of the SVR-based forecasts over the benchmarks. In particular, ten SVR-based forecasts but none of the

[11] We report levels of confidence that are more conservative than 90% and 75% reported in the forecasting exercise by Hansen et al. (2011).

[12] Given the size of the data sample, increasing the size of the model set beyond 25 may affect the reliability of the MCS test results (Hansen et al., 2011).

Table 3
Predictive ability of the SVR-based models [a, b]

Brazil			Mexico			Turkey		
Model	RMSE	P_{MCS}	Model	RMSE	P_{MCS}	Model	RMSE	P_{MCS}
SVR-RBF-G-10	0.467	1.000***	SVR-Linear-Baseline-5	0.183	1.000***	SVR-Poly-Baseline-50	0.347	1.000***
SVR-RBF-G-6	0.468	0.367*	SVR-Linear-Baseline-59	0.186	0.909***	SVR-Poly-Baseline-63	0.347	0.787***
SVR-RBF-G-102	0.471	0.336*	SVR-Linear-Baseline-52	0.187	0.909***	SVR-RBF-GEM-6	0.349	0.787***
SVR-RBF-G-229	0.475	0.336*	SVR-Poly-Baseline-4	0.187	0.909***	SVR-Sigmoid-Baseline-4	0.348	0.787***
SVR-RBF-G-115	0.472	0.336*	SVR-Linear-Baseline-47	0.187	0.909***	SVR-RBF-GEM-2	0.349	0.137
SVR-RBF-G-107	0.473	0.336*	SVR-Linear-Baseline-70	0.187	0.909***	SVR-RBF-GEM-3	0.349	0.137
SVR-RBF-G-113	0.473	0.336*	SVR-Poly-Baseline-66	0.188	0.909***	SVR-Poly-Baseline-61	0.349	0.119
SVR-RBF-G-241	0.476	0.336*	SVR-Sigmoid-Baseline-22	0.189	0.909***	SVR-RBF-GEM-7	0.349	0.119
SVR-RBF-G-7	0.474	0.332*	SVR-Poly-Baseline-21	0.189	0.909***	SVR-Poly-Baseline-62	0.349	0.119
SVR-RBF-G-199	0.476	0.332*	SVR-Linear-Baseline-42	0.188	0.791***	SVR-RBF-GEM-8	0.349	0.119
SVR-RBF-G-95	0.474	0.189	SVR-Linear-Baseline-9	0.189	0.791***	SVR-RBF-GEM-5	0.349	0.119
SVR-RBF-G-99	0.474	0.189	SVR-Poly-Baseline-26	0.190	0.725**	SVR-RBF-GEM-4	0.349	0.119
SVR-RBF-G-4	0.474	0.189	SVR-Poly-Baseline-22	0.190	0.725**	SVR-Poly-Baseline-49	0.349	0.119
SVR-RBF-G-220	0.477	0.189	SVR-Linear-Baseline-56	0.191	0.621**	SVR-RBF-GEM-9	0.349	0.119
SVR-RBF-G-105	0.475	0.189	SVR-Linear-Baseline-55	0.190	0.621**	SVR-RBF-GEM-1	0.350	0.119
SVR-RBF-G-56	0.475	0.189	SVR-Poly-Baseline-27	0.191	0.621**	SVR-Poly-Baseline-56	0.350	0.119
SVR-RBF-G-134	0.476	0.189	SVR-Linear-Baseline-10	0.191	0.621**	SVR-Poly-Baseline-57	0.351	0.119
SVR-RBF-G-108	0.476	0.189	SVR-Linear-Baseline-15	0.191	0.166	SVR-Poly-Baseline-55	0.351	0.119
SVR-RBF-G-168	0.476	0.189	SVR-Linear-Baseline-34	0.191	0.166	SVR-Poly-Baseline-52	0.352	0.119
SVR-RBF-G-133	0.476	0.189	SVR-Linear-Baseline-72	0.191	0.166	SVR-Sigmoid-G-4	0.353	0.119
SVR-RBF-G-57	0.475	0.189	SVR-Linear-Baseline-50	0.191	0.166	SVR-Sigmoid-G-3	0.357	0.119
SVR-RBF-G-170	0.477	0.189	SVR-Linear-Baseline-7	0.191	0.166	SVR-Poly-Baseline-53	0.355	0.119
SVR-RBF-G-112	0.476	0.189	SVR-Linear-Baseline-16	0.192	0.166	SVR-Linear-Baseline-5	0.355	0.119
SVR-RBF-G-202	0.477	0.189	SVR-Linear-Baseline-48	0.192	0.166	SVR-Sigmoid-Baseline-10	0.358	0.119
SVR-RBF-G-111	0.476	0.189	SVR-Poly-Baseline-25	0.192	0.166	SVR-Sigmoid-Baseline-1	0.357	0.119
OLS-GEM	0.496	0.189	OLS-GEM	0.215	0.166	OLS-Baseline	0.416	0.119
OLS-Baseline	0.546	0.189	OLS-G	0.227	0.166	RW	0.504	0.119
OLS-G	0.520	0.189	RW	0.225	0.048	OLS-GEM	0.462	0.116
RW	0.630	0.189	OLS-Baseline	0.225	0.048	OLS-G	0.483	0.039

[a] For each country, the first column gives the model name, the second column reports forecast RMSE and the third column reports MCS p-values for a model at hand, with *, **, and *** identifying the forecasts in $\widehat{\mathcal{M}}^{*}_{75\%}$, $\widehat{\mathcal{M}}^{*}_{50\%}$ and $\widehat{\mathcal{M}}^{*}_{25\%}$, respectively.

[b] Model names characterize the estimation method (SVR or OLS), the kernel (Linear, RBF, Sigmoid, or Poly) and the vector of economic input variables (Baseline, G or GEM). The trailing number provides a unique identifier for each tuning parameter setting generated in the MSNTO optimization process.

benchmarks enter $\widehat{\mathcal{M}}^*_{75\%}$ for Brazil. Gains in forecast accuracy are equally sizable in economic terms as borne out by a substantial reduction in forecast errors afforded by the SVR models: the ten $\widehat{\mathcal{M}}^*_{75\%}$ models, all G-specifications of SVR, deliver, on average, a 9.1% and 4.7% reduction in RMSE $(1 - RMSE_{SVR}/RMSE_{OLS})$ over the benchmark employing the same set of input variables, OLS-G, and the best performing benchmark, OLS-GEM, respectively. The evidence is even more striking for Mexico and Turkey: eleven SVR-based forecasts for Mexico and four SVR-based forecasts for Turkey enter $\widehat{\mathcal{M}}^*_{25\%}$; the superior model set is dominated by $Baseline$ SVR specifications for both countries. At the same time, similar to Brazil, all benchmark forecasts exhibit relatively low p-values, clearly signaling their inferior predictive accuracy to SVR-based forecasts for both countries. Evidence of substantive economic gains confirms this result: the reductions in forecast RMSE afforded by the $\widehat{\mathcal{M}}^*_{25\%}$ SVR models are 16.8% relative to OLS-$Baseline$ and 12.9% relative to the best performing benchmark, OLS-GEM, for Mexico. The respective gains are equally sizable at 16.4% for Turkey, with OLS-$Baseline$ being the same input-set-based benchmark and best performing benchmark at the same time.

4.2 Further empirical observations

Table 3 further shows that there is no persuasive evidence for singling out a kernel function that may be most suited for modeling credit spreads of various countries. Nevertheless, country-specific evidence suggests that some kernel functions may be particularly well suited for modeling credit spreads of a given country. For example, there is a clear preferred kernel, RBF, for Brazil, with all models in $\widehat{\mathcal{M}}^*_{75\%}$ being RBF-based. In contrast, Linear is the preferred kernel, used in 7 out of 11 best performing models, for Mexico and Poly, featuring in 2 out of 4 best models, is the preferred kernel for Turkey. Interestingly, Poly and Linear-based SVR appear among top performers when employed in conjunction with a small input set such as $Baseline$ whereas RBF-based SVR perform well in conjunction with extended input sets like G for Mexico and GEM in Turkey's case. However, this observation requires further more conclusive evidence.

Furthermore, the results reveal that SVR-based models require only a relatively small set of input variables to deliver accurate forecasts across all three countries. In particular, the best performing SVR employ the $Baseline$ input set for Mexico and Turkey and G, an extension with global but not domestic variables, for Brazil. This finding contrasts with the results for benchmark

models where the benchmark using the largest set of input variables, GEM, containing both global and domestic fundamentals, generates the lowest RMSE for Brazil and Mexico, with $Baseline$ delivering the lowest RMSE only for Turkey. Thus, our findings provide evidence of the SVR models ability to deliver accurate forecasts when employing even smaller input sets than those used by the benchmarks. Furthermore, adding global variables (G) or both global and domestic variables (GEM) to the input set does not deliver improvements in forecast accuracy over SVR $Baseline$ that exploits predictive content only in the credit spread curve for Mexico and Turkey. This SVR-based finding contrasts with evidence for linear-model-based forecasts in our study and also those reported in Audzeyeva and Fuertes (2018), suggesting that the $Baseline\ OLS$ specification cannot always outperform random walk and that its forecasting accuracy can be improved by additional predictors. Only for Brazil, adding global (but not domestic) macroeconomic variables to the $Baseline$ input set helps improve forecast accuracy.[13]

Taken together, the SVR-based evidence lends some support to the rational expectation theory for the term structure of credit spreads when the credit-spread-curve factors are not bounded to a strictly linear relationship with future credit spreads as in the previous literature. A further investigation into the factors that produce supportive evidence for some countries (Mexico and Turkey) but not for others (Brazil) presents a fruitful direction for further research.

5 Conclusions

This paper proposes a coherent framework for producing a set of highly accurate SVR models for forecasting credit spreads of emerging markets. In our main methodological contribution, we put forward a systematic approach for setting robust parameter values for SVR kernel functions, addressing a gap in the SVR literature. In contrast to previous studies aiming to select one "best" kernel setting that serves as input into the "best" SVR predictive model, our approach generates a robust set of viable tuning parameter values feeding into a set of SVR model candidates. We manage this model multiplicity by adopting the MCS test to select a subset of most accurate mod-

[13]The significance of predictive content in global factors for Brazil is likely to be related to a sharp deterioration in Brazil's creditworthiness, forcing Brazil out of investment-grade BBB and into speculative-grade BB rating category in September 2015 while credit ratings of Mexico and Turkey remained relatively stable. A consequent steep rise in Brazil's exposure to global risks may have been reflected in its current credit spread curve with some delay, enabling global factors to add predictive content beyond the spread curve for future credit spreads.

els. Furthermore, our approach accommodates novel economic and financial market applications characterized by serially correlated data.

In the empirical analysis part, the evaluation of a quarter-ahead OOS performance of SVR forecasts that our approach generates for three large sovereign emerging market borrowers using various kernel functions and sets of input variables motivated by economic theory provides evidence that our approach identifies a relatively small set of SVR models with a notably superior OOS forecasting ability in economic and statistical terms relative to both other SVR specifications and standard benchmarks utilized in the credit spreads literature. Moreover, our evidence confirms a finding in Sermpinis et al. (2017b) for European stock market ETF disproving a widely-held belief that the RBF kernel is the optimal choice for modeling financial market series, indicating that the kernel choice may be country-specific for emerging market credit spreads.

Our results further suggests that our SVR approach can deliver accurate credit spread forecasts with a small set of predictors limited to the credit curve level, slope and curvature factors, outperforming the random walk and linear-regression-based benchmarks using even larger predictor sets and performing at least as well as SVR forecasts using extended sets of predictors. Consequently, our findings for SVR-based credit spread forecasts lend support to the rational expectation theory of the term structure in the context of emerging market credit spreads that has been previously rejected for linear-model-based forecasts of emerging-market sovereign credit spreads in Audzeyeva and Fuertes (2018) and mature-market (U.S.) corporate credit spreads in Krishnan et al. (2010). Hence, our results provide indirect evidence that highly flexible SVR models may be better suited than linear models, routinely employed in the literature, for capturing investor expectations about future credit spreads on emerging market bonds. Further direct tests will constitute a fruitful avenue for future research.

References

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term-structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Audzeyeva, A., Fuertes, A.M., 2018. On the predictability of emerging market sovereign credit spreads. Journal of International Money and Finance 88, 140–157.

Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis 120, 70–83.

Brabanter, K.D., Brabanter, J.D., Suykens, J.A.K., Moor, B.D., 2011. Kernel regression in the presence of correlated errors. Journal of Machine Learning Research 12, 1955–1976.

Cao, L., Tay, F., 2001. Financial forecasting using support vector machines. Neural Computing and Applications 10, 184–192.

Comelli, F., 2012. Emerging market sovereign bond spreads: Estimation and backtesting. IMF Working Paper 12/2012, International Monetary Fund, Washington D.C. .

Diebold, F., Li, C., 2006. Forecasting the term-structure of government bond yields. Journal of Econometrics 130, 337–364.

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 13, 253–263.

Ding, Y.S., Song, X.P., Zen, Y.M., 2008. Forecasting financial condition of chinese listed companies based on support vector machine. Expert Systems with Applications 34, 3081–3089.

Fama, E., Bliss, R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.

Gunduz, Y., Uhrig-Homburg, M., 2011. Predicting credit default swap prices with financial and pure data-driven approaches. Quantitative Finance 11, 1709–1727.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hofmann, T., Schölkopf, B., Smola, A.J., 2008. Kernel methods in machine learning. Annals of Statistics 36, 1171–1220.

Law, T., Shawe-Taylor, J., 2017. Practical bayesian support vector regression for financial time series prediction and market condition change detection. Quantitative Finance 17, 1403–1416.

Longstaff, F., Pan, J., Pedersen, L., Singleton, K., 2011. How sovereign is sovereign risk? American Economic Journal: Macroeconomics 3, 75 –103.

Ludvigson, S., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22, 5027–5067.

Min, J.H., Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications 28, 603–614.

Nelson, C., Siegel, A., 1987. Parsimonious modeling of yield curves. Journal of Business 60, 473–489.

Niederreiter, H., Alpár, L., Halász, G., Sárközy, A., 1983. Studies in Pure Mathematics: To the Memory of Paul Turán. Birkhäuser Basel, Basel. chapter 7. pp. 523–529. "A quasi-Monte Carlo method for the approximate computation of the extreme values of a function".

Racine, J., 2000. Consistent cross-validatory model-selection for dependent data: hv -block cross-validation. Journal of Econometrics 99, 39–61.

Roll, R., 1970. The Behavior of Interest Rates: An Application of the Efficient Market Model to U.S. Treasury Bills. Basic Books, Inc., New York.

Sargent, T., 1972. Rational expectations and the term structure of interest rates. Journal of Money, Credit and Banking 4, 74–97.

Scholkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L., 2000. New support vector algorithms. Neural Computation 12, 1207–1245.

Sermpinis, G., Stasinakis, C., Hassanniakalager, A., 2017a. Reverse adaptive krill herd locally weighted support vector regression for forecasting and trading exchange traded funds. European Journal of Operational Research 263, 540–558.

Sermpinis, G., Stasinakis, C., Rosillo, R., de la Fuente, D., 2017b. European exchange trading funds trading with locally weighted support vector regression. European Journal of Operational Research 258, 372–384.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing 14, 199–222.

Stasinakis, C., Sermpinis, G., Psaradellis, I., Verousis, T., 2016. Krill-herd support vector regression and heterogenous autoregressive leverage: Evidence from forecasting and trading commodities. Quantitative Finance 16, 1901–1915.

Svensson, L., 1994. Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical Report 48871. National Bureau of Economic Research Working Paper.

Vapnik, V.N., 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

Xu, Q.s., Liang, Y.z., Hou, Z.t., 2005. A multi-sequential number-theoretic optimization algorithm using clustering methods. Journal of Central South University of Technology 12, 283–293.

Appendix A Data variables: Summary statistics

Table A1

Summary statistics of emerging market sovereign credit spreads and input variables

Country Variable	Mean	StDev	Min	Max	AR(1)
US					
Yield curve level	0.043	0.008	0.027	0.058	0.986
Yield curve slope	-0.028	0.013	-0.053	-0.004	0.981
Yield curve curvature	-0.079	0.025	-0.139	-0.022	0.978
Short-term rate volatility	0.000	0.001	0.000	0.004	0.977
Brazil					
Credit spread	0.016	0.008	0.004	0.050	0.973
Credit-curve-spread curve level	0.025	0.010	0.012	0.061	0.974
Credit-curve-spread curve slope	-0.011	0.020	-0.061	0.062	0.887
Credit-curve-spread curve curvature	-0.025	0.040	-0.154	0.069	0.925
Country risk rating	39.196	3.146	32.500	45.500	0.978
Trade balance	0.067	0.081	-0.159	0.379	0.978
Trade balance volatility	0.054	0.027	0.021	0.157	0.990
Terms-of-trade growth	0.052	8.790	-13.276	19.361	0.997
Terms-of-trade growth volatility	2.747	1.560	0.389	8.899	0.991
Mexico					
Credit spread	0.014	0.006	0.006	0.047	0.972
Credit-curve-spread curve level	0.021	0.006	0.013	0.044	0.930
Credit-curve-spread curve slope	-0.009	0.014	-0.044	0.024	0.874
Credit-curve-spread curve curvature	-0.021	0.032	-0.086	0.063	0.930
Country risk rating	40.289	1.908	35.500	43.000	0.978
Trade balance	-0.033	0.071	-0.272	0.085	0.958
Trade balance volatility	0.066	0.032	0.024	0.135	0.992
Terms-of-trade growth	-2.778	9.850	-22.243	19.394	0.998
Terms-of-trade growth volatility	3.609	2.973	0.356	14.099	0.994
Turkey					
Credit spread	0.025	0.009	0.013	0.073	0.964
Credit-curve-spread curve level	0.029	0.008	0.013	0.063	0.961
Credit-curve-spread curve slope	-0.007	0.017	-0.055	0.061	0.797
Credit-curve-spread curve curvature	-0.006	0.030	-0.124	0.127	0.796
Country risk rating	33.740	2.550	27.000	37.500	0.981
Trade balance	-0.816	0.227	-1.253	-0.202	0.993
Trade balance volatility	0.110	0.041	0.042	0.209	0.989
Terms-of-trade growth	0.857	4.350	-8.864	9.667	0.995
Terms-of-trade growth volatility	1.732	1.183	0.241	5.320	0.993

Appendix B MSNTO: Implementation details

Table B1
MSNTO tuning parameter search ranges

Kernel	Parameter Value Search Regions
Linear	$C \in (1., 200), \epsilon \in (0.0001, 0.1)$
RBF	$C \in (1., 200), \epsilon \in (0.0001, 0.1), \psi \in (0.0001, 40)$
Sigmoid	$C \in (1., 200), \epsilon \in (0.0001, 0.1), \gamma \in (0.0001, 40), s \in (-13, 13)$
Polynomial	$C \in (1., 200), \epsilon \in (0.0001, 0.1), \gamma \in (-13, 13), s \in (0.01, 20), g \in (1, 6)$

Table B2
Number of unique tuning parameter values

Kernel	MODELS		
	Baseline	G	GEM
	Brazil		
Linear	112	100	3
RBF	88	341	84
Sigmoid	190	291	45
Poly	160	246	10
	Mexico		
Linear	110	63	8
RBF	44	254	100
Sigmoid	34	90	80
Poly	66	199	152
	Turkey		
Linear	21	63	31
RBF	215	1292	38
Sigmoid	37	18	20
Poly	65	256	19