Forecasting Realized Volatility with Kernel Ridge Regression

Blake LeBaron *

International Business School

Brandeis University

August 2018

Abstract

This paper explores a common machine learning tool, the kernel ridge regression, as applied to financial volatility forecasting. It is shown that kernel ridge provides reliable forecast improvements to both a linear specification, and a fitted nonlinear specification which represents well known empirical features from volatility modeling. Therefore, the kernel ridge specification is still finding some nonlinear improvements that are not part of the usual volatility modeling toolkit. Various diagnostics show it to be a reliable and useful tool. Finally, the results are applied in a dynamic volatility control trading strategy. The kernel ridge results again show improvements over linear modeling tools when applied to building a dynamic strategy.

Keywords: Machine learning, realized volatility, kernel ridge regression

*415 South Street, Mailstop 32, Waltham, MA 02453 - 2728, blebaron@brandeis.edu, www.brandeis.edu/~blebaron.

1 Introduction

Machine learning (ML) tools have been dramatically changing the world of data analytics. They have been applied in almost all fields with enough data to make improved predictive modeling a possibility. Even though the basic ML toolbox has been around for over 50 years, modern computing, data availability, and improved algorithms have invigorated this area of research. Finance is obviously a major area for application of these tools.[1] It is obvious why primary interest is directed at forecasting returns and developing trading strategies. Economic gains in this area have the potential to be quite large. However, they run up against efficient market limits that imply predictability of market returns should be difficult if not impossible. This makes this modeling space difficult, and more importantly, difficult to perform clear model comparisons in a world where the signal to noise ratio is very low. This paper turns to modeling volatility, where predictability is higher, and selecting good predictive models may have a larger impact.[2]

The conditional variance of returns, or volatility, is very predictable. This feature has been known for a long time, and has led to the development of many useful models and procedures.[3] Early work concentrated on squared and absolute daily returns. In the early 1990's a revolution occurred with the use of high frequency data to estimate variances over a given day using intraday data. Known as realized volatility, it provided more accurate measures of volatility and its dynamics.[4]

The purpose of this paper is to develop and explore predictive models for daily realized volatility time series using a kernel ridge regression. This machine learning tool is capable of capturing a rich set of nonlinear features in the data. There are many methods in the standard machine learning tool set including, nearest neighbors, support vectors, ridge and lasso regressions, regression trees, and random forests. There are also deep neural networks which form much of the basis for modern image classification. A full comparison of all these tools is beyond the scope of this paper. Also, which tool is well suited for time series analysis is still an open question. Analysis here will concentrate on the kernel ridge regression. Recently, kernel ridge has been applied in macro forecasting in Exterkate, Groenen, Heij & van Dijk (2016). It looks like a promising tool for exploring a large rich set of nonlinear features while avoiding model over fitting.

[1]ML research in finance is not new. In the 1990's there was an earlier wave of interest. Several early examples include Diebold & Nason (1990), LeBaron (1992), Meese & Rose (1990), and Mizrach (1992). LeBaron (1998) brings together many tools considered important today. These include neural networks, evolutionary learning, and boostrap/cross-validation systems in the search for improved foreign exchange trading strategies.

[2]Examples of machine learning tools for volatility prediction are Andrada-Felix, Fernandez-Rodriguez & Fuertes (2016), Audrino & Knaus (2016), Chen, Hardle & Jeong (2010), Luo, Zhang, Xu & Wang (2017), and Lux, Hardle & Lessmann (2018).

[3]See Andersen, Bollerslev, Christoffersen & Diebold (2006) for one of many surveys in this large area of research.

[4]This area is now very large, and is a standard for volatility modeling. See Barndorff-Nielsen & Shephard (2010) and Andersen, Bollerslev, Christoffersen & Diebold (2013) for two of the many surveys available.

Electronic copy available at: https://ssrn.com/abstract=3229272

It is closely related to support vector machine regressions, but these use a slightly nonstandard objective function that makes comparisons to other time series results in econometrics more difficult. By utilizing a flexible functional form, kernalized methods in machine learning can theoretically represent any possible nonlinear relationship. This makes them a good approach in searching for possible volatility nonlinearities.

Volatility may be much more predictable than returns, but it provides its own set of challenges. First, the known predictable features mean that the bar is now higher in terms of forecast comparisons. Just being able to forecast volatility is not interesting. The question is whether a model is able to beat a well defined benchmark. This opens the question of exactly what this benchmark should be. This paper searches through some common examples from the recent financial time series literature. Second, volatility has some patterns that suggest either formal long memory processes, or potential regime shifts that might make building smaller parsimonious models difficult.

Section 2 will introduce the data and linear and nonlinear models that will be used. Section 3 estimates the various models and measures the comparisons across the standard models along with the kernel ridge target. Section 4 sets up a basic volatility control strategy to perform some initial comparisons on the economic significance of the results, and section 5 concludes.

2 Data and methodology

The key financial time series used in this paper is a modified realized volatility measure. As with other realized volatility measures it uses intraday data to generate daily volatility estimates. Instead of using data on the order of 5 minute observations, it uses the less frequent sampling in a data set provided by Global Financial Data (GFD). They provide a series with hourly observations going back to 1933, and thirty minute observations starting in 1987. Only the post 1987 data will be used in this study.[5]

This estimated realized volatility is given by,

$$RV_t^1 = \sqrt{N}\sqrt{\frac{\pi}{2}}(\frac{1}{N})\sum_{h=1}^{N}|r_{t,h}|, \tag{1}$$

where $r_{t,h}$ are the log returns across 30 minute intervals within day t. This differs from traditional realized volatility in that for robustness the absolute return replaces usual squared returns. They are adjusted to line up with daily standard deviations by multiplying by an adjustment factor. The application of $\sqrt{N}$ is used

[5] LeBaron (2018) also uses the longer hourly series. In this paper the 30 minute series is used to give a more detailed picture of volatility for the nonlinear methods.

to blow up the 30 minute standard deviations to daily units. As is common in many volatility series, means for the series are taken over different horizons. The approach of Corsi (2009) is used in constructing both a weekly and monthly volatility measure aggregated as

$$RV_t^m = \frac{1}{m}\sum_{j=0}^{m-1} RV_{t-j}^1.$$

As in most of these studies m will be set to 5 and 22 days. Finally, to exclusively concentrate on short term volatility features the values are divided by a lagged 250 day volatility estimate, RV_t^{250}.

Two other measures are used from the GFD data set. One uses the 30 minute trading volume series which measures the total number of shares traded on the Dow in intervals of 30 minutes. This series is converted to a local trend measure that tests whether volume is locally rising, or falling through the day. The first 30 minutes and last 30 minutes are removed to eliminate the well known unusual trading behavior at the open and the close. Then the remaining 30 minute periods are divided by two, and the mean of the second part is divided by the mean of the early part. This gives a simple ratio indicating whether trading volume is rising or falling. The ratio is then logged to give a trend volume measure denoted by TV_t.

A more common technical indicator using daily information is generated from closing and high/low information. It is constructed from,

$$CR_t = 2\left(\frac{C_t - L_t}{H_t - L_t}\right) - 1.$$

This value ranges from -1 to 1 corresponding to a close at the low or the high at the end of the day. It captures some possible persistence in daily price moves, and might be an indicator of increased trading activity, and also volatility in the future.

Table 1 presents summary statistics on all these measures. It is clear that the RV_t measures are highly skewed and leptokurtic. Kurtosis for the daily RV estimate is over 70. Often estimation of volatility models log transforms these to get closer to Gaussian series. These will be given by

$$rv_t^m = \log(RV_t^m)$$

The table shows that this transformation moves the data much closer to Gaussian features which will make model estimation easier.

In nonlinear modeling the actual units matter, and estimation can be hindered by widely different units of variability in different series. At an extreme, researchers can normalize series by ranges, or standard

Electronic copy available at: https://ssrn.com/abstract=3229272

deviations. This study refrains from doing this, but tries to keep the units comparable. The volatility measures are all in units of logged daily standard deviations normalized by a 250 day mean volatility. Looking at the rv_t^m measures in table 1 shows that this gives a reasonably close range in variability of these measures as shown by their standard deviations. There is a clear reduction in variability moving from the high to lower frequency volatility. The other two measures are left alone since their units are meaningful in both cases. The trend volume indicator does show a relatively low standard deviation (0.10) relative to the other measures.

Baseline volatility modeling follows Corsi (2009) by building a model for volatility dynamics and forecasting from,

$$rv_{t+1}^1 = \beta_0 + \beta_1 rv_t^1 + \beta_2 rv_t^5 + \beta_3 rv_t^{22} + \mu_t. \tag{2}$$

This basic model can be estimated with ordinary least squares (OLS), and gives a useful linear comparison with other more complex models that will follow.

Finally, volatility forecasts will be mapped back into normalized standard deviations by taking,

$$\widehat{RV}_{t+1}^1 = e^{\hat{rv}_{t+1}^1 + 0.5 * \sigma_{rv}^2}, \tag{3}$$

where σ_{rv}^2 is a estimate of the variance of the logged RV measures. The latter term in the exponential is the usual bias adjustment when moving in and out of log measures. It would give the true expectation if rv_t were strictly Gaussian, but it is a commonly used approximation for volatility.

The linear model mentioned above will be added to in several ways. The primary purpose of this paper is to see if the linear framework can be improved with kernel ridge. However, before doing this the target benchmark will also be raised a little. First, standard nonlinear features will be added to it. Volatility is well known to have a sign asymmetry in lagged returns. When returns are negative, future volatility tends to be higher.[6] More recently, Wang & Yang (2017) show that volatility persistence itself can depend on past returns in a nonlinear way. Combining both these features into the original linear framework gives,[7]

$$rv_{t+1}^1 = \beta_0 + (\beta_1 + \beta_5 r_t + \beta_6 |r_t|) rv_t^1 + \beta_2 rv_t^5 + \beta_3 rv_t^{22} + + \beta_4 I(r_t > 0) + \mu_t. \tag{4}$$

[6] This feature was discovered in Black (1976), and named in Christie (1982) as being related to the amount of leverage at various firms. It has recently been reexamined in Hasanhodzic & Lo (2011). Glosten, Jagannathan & Runkle (1993) and Nelson (1991) are useful models, and a more modern approach is in Curci & Corsi (2012).

[7] Another nonlinear specification that adjusts volatility measures by using intraday fourth moments is given in Bollerslev, Patton & Quaedvlieg (2016), and also in Buccheri & Corsi (2017). Also, a survey of realized volatility forecasting with nonlinear models is in McAleer & Medeiros (2010).

Electronic copy available at: https://ssrn.com/abstract=3229272

This specification will be raced against a nonlinear kernel representation. The key question is just what sort of nonlinear econometric technology should be used. Modern machine learning provides many approaches to approximating arbitrary nonlinear functions. It is beyond the scope of this paper to test all of them, but some rudimentary decisions and comparisons have been made. The model comparison and search is greatly assisted by the Python Scikit Learn package. Not only does it contain many of the tools mentioned here, but its consistent interface across models allows for easy comparison and cross validation.

For a modern data analytic example this one is not all that standard. Modern "big data" problems usually consist of large cross sections and many potential right hand side variables. In this problem the data set is only moderately sized (about 8000 observations), and the number of right side variables is small (3-7). The sample size makes it unlikely that highly parameterized, deep learning, neural networks will succeed. There is still a set of nonlinear models which may be useful. Particularly important are kernel based approaches. Support vector regressions have been used in several finance applications, but they are designed more for classification, and usually involve an objective function which is nonstandard for time series analysis.[8] A relatively new tool is kernel ridge regression. This is both more directly applicable to standard least square objectives. Also, it has recently seen some early success in macroeconomic time series forecasting.[9] Kernel ridge combines a large sequence of nonlinear kernels on the right-hand side along with the standard ridge regression, (L^2), penalty function to avoid over fitting. This model was the most reliable of the nonlinear models used.

The models are evaluated using standard measures such as mean squared error (MSE), and mean absolute error (MAE). They are estimated in the target space with forecasts of next period conditional standard deviation. For example,

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(RV_t - \widehat{RV}_{t|t-1})^2 \tag{5}$$

where $\widehat{R}V_{t|t-1}$ is the RV forecast determined on the previous day. Also reported is a pseudo out of sample R^2 measure which is given by,

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(RV_t - \widehat{RV}_{t|t-1})^2}{\sum_{t=1}^{T}(RV_t - \overline{RV_t})^2}, \tag{6}$$

where $\overline{RV_t}$ represents the mean RV estimated over a training sample. In most cases all these estimates will be made out of sample as part of a cross validation procedure.

[8]See Chen et al. (2010) for an application of support vector machines for volatility prediction.

[9]See Exterkate et al. (2016) and Exterkate (2013). These papers also include much of the mathematical structure and details of the model that are skipped here. Also, Efron & Hastie (2016) provides a good textbook summary of the kernel methods as applied to support vector machines. Much of the intuition and machinery carries over to kernel ridge regression. They also provide comparisons with kernel smoothing methods which are different from the kernel basis function approach.

Electronic copy available at: https://ssrn.com/abstract=3229272

3 Empirical results

3.1 Benchmark models

This section introduces several baseline models which will be used as comparison with the kernel ridge regression. The time period will be January 1987 through October 2017, yielding a sample of 8005 daily observations. All estimation done here is full sample, and all estimation is by ordinary least squares (OLS). This generates obvious bias in model fitting parameters, but identification is done by minimizing the Bayesian Information Criterion (BIC) which penalizes model complexity.

Table 2 begins by estimating a simple autoregressive model of order 1 (AR(1)) on the volatility process. It generates a large significant coefficient, and a R^2 of nearly 0.327. The numbers in parenthesis are standard errors. This indicates that even the most basic of models generates a fair amount of predictability when it comes to daily realized volatility. The second model fit is the basic Corsi model with the three different lags. All three coefficients are highly significant, the BIC and MSE drop dramatically, and the R^2 moves to 0.390.

The next change adds the impact of the sign of current return through β_4. Model improvement is again observed through a reduction in BIC. The fourth line now adds the parameters from Wang & Yang (2017). Both are highly significant, and BIC drops to 6098. The model now has a MSE which is 10 percent less than the standard Corsi model. Finally, the two extra features are added for the trend volume (TV_t), and the closing ratio (CR_t). Both these coefficients are significant, but there is only a small improvement in MSE. BIC increases, so the optimal model specification would not use these variables. The target nonlinear model is the one given in line 4 which is essentially Wang & Yang (2017)'s format with some extra information for return signs.

It is important to note that while goodness of fit measures do improve, the changes are not enormous. Improvements in MSE beyond the basic Corsi model are only 10 percent.

3.2 Kernel predictors

This section begins estimating nonlinear kernel ridge regressions for different sets of predictors. The object is to determine the optimal kernel framework, and to also get some insight into how well the kernel handles nonlinear functions.

Since the kernel ridge is our target model for study, performance measurement now becomes more critical. The various forecast performance metrics are measured with a randomized 5 fold cross validation.

Electronic copy available at: https://ssrn.com/abstract=3229272

In a time series situation it is important to see how this is done. All lags for various regressions are formed initially. For example, rv_t^{22} which goes back 22 periods will be estimated for each target, RV_{t+1}. These can be thought of as lined up for a cross sectional regression as in,

$$rv_{t+1}^1 = f(rv_t^1, rv_t^5, rv_t^{22}). \tag{7}$$

Then 4/5's of these points are chosen at random as the training set for model estimation, and 1/5 are used for testing model performance. Randomized cross-validation is important in time series situations, since in makes sure that randomized test sets are spread across the entire time series. Obviously, this implicitly is imposing strong stationarity assumptions.

Table 3 reports the results across 250 randomized train/test pairs, and estimates MSE, R^2, and MAE for several different selections of predictors. Estimation is done with the Python Scikit Learn kernel ridge regression. It uses the Radial Basis Function kernel with a penalty $\alpha = 0.1$, and a bandwidth $\gamma = 0.05$. These optimal parameters were estimated through a grid search procedure using randomized 5 fold cross validation, and 250 draws. The columns on the left side of the table indicate the predictors that were included.

The testing begins with the standard Corsi three lagged model. The table reports a mean MSE of 0.190 which is slightly better than the linear specification from the last table. The number in parenthesis estimates the standard error on the MSE based on the 250 simulations. We know that the sign of the previous return matters in volatility forecasts, but for a nonlinear model, we can let the kernel model decide how to use this, so in row two the return, r_t, is added. This leads to a large drop in MSE, and an increase in R^2. Small forecast improvements continue as the volume trend and the close ratio are added in rows 3 and 4.

The final two rows are diagnostic, and designed to understand if the kernel ridge is able to understand various nonlinear features detected in previous research. Both $|r_t|$ and the sign of r_t are redundant information given that the kernel should be able to approximate any nonlinear function for the current return, r_t. This does seems to be the case since the addition of neither function generates any forecast improvement.

From these results, the model from row 4 with r_t, trend volume, and the closing ratio, along with the Corsi lags will be used as the optimal kernel ridge combination. In the next subsection it will be raced against some of the benchmark comparison models.

Electronic copy available at: https://ssrn.com/abstract=3229272

3.3 Kernel ridge comparisons

This section will compare the various volatility forecasting models. All comparisons will be done using 250 randomized 5 fold cross validations where 4/5 of the data is randomly chosen as a training set, and the final 1/5 is the testing set. As before, time ordering is randomized in this setup. Models that will be used for comparison are the simple 3 factor linear model from Corsi (2009), the augmented nonlinear model with return and absolute return interactions as in Wang & Yang (2017), and finally, a linear ridge regression is also run on the same information set as the kernel ridge. The last test is a comparison that tests the importance of the nonlinear radial basis terms in the ridge regression.

Results are given in table 4. Models are compared using mean squared error (MSE), R^2, and mean absolute error (MAE).[10] The first line presents the results from the nonlinear kernel ridge using the optimal specification from table 3, row 4. It presents the MSE, R^2, and MAE means from the 250 cross-validation experiments. Numbers in parenthesis are the standard errors for these estimated means. Both the MSE and R^2 are improvements over the various models fit in table 2. However, the improvement does get smaller as the comparison models are made more complex.[11] The MSE is also slightly larger than the value estimated in table 3, but this is within the range of sampling error as indicated by the standard errors.

The second line in table 4 refers to the basic three factor linear volatility model estimated on line two of table 2. The table presents values relative to the nonlinear kernel ridge models. For example, the MSE value of 0.873 indicates that the mean MSE for the kernel model is about 87 percent the MSE for the linear model, a roughly 13 percent improvement. Similarly, the R^2 for the kernel (showing a value of 1.204 is over 20 percent larger than the linear model. MAE improvements are not as dramatic, but appear significant. Standard errors for the mean ratios are given in parenthesis. Finally, the values in brackets present the fraction of cross-validation runs where the kernel ridge model beats the comparison model in terms of the respective goodness of fit estimate. For the linear model these values are all well above 0.95.

The third line in the table moves to the augmented linear model with nonlinear interaction terms as given in line 4 from table 2. The improvements of the kernel ridge relative to this more complex model decline. The MSE ratio now shows a value of 0.937, in other words the kernel gives an improvement of about 6.5 percent. The small standard errors still show that statistically this is a dramatic improvement. A similar result is given for the R^2 where there is an 8 percent improvement for the kernel ridge model. Quantitatively the MAE reduction is again small, but significant. Finally, the values in brackets are above

[10] The R^2 measure is not a true R^2. See earlier section for discussions.

[11] This is interesting since the early comparison table was estimated on the full sample.

or near 0.95 indicating that 5 percent or fewer of the nonlinear model runs beat the richer kernel ridge specification.

The last row removes the nonlinear components from the kernel ridge regression. The same predictor variables are used, and a new penalty function, $\alpha = 3$, is estimated using a grid search on a 250 length cross-validation. This experiment directly tests the value added of the nonlinear kernel component of the ridge regression. The results show that the nonlinear kernel components still add value to the forecasts. Linear ridge regression on its own does well relative to the other models. It is especially interesting, that it seems to be roughly equivalent to the nonlinear model specification. However, it still falls short of the full kernel ridge specification. The latter specification shows an MSE improvement of nearly 7 percent, and a R^2 improvement of 9 percent. It does show one of the weakest values for comparisons with other models in that only 87 percent of the kernel ridge runs beat the linear kernel in terms of MSE. This result is a little curious since the other comparisons, R^2 and MAE, are still well above 0.90.

4 Volatility control strategies

This section begins to explore the economic significance of the forecast comparisons. For volatility forecasts, a common application is a volatility control strategy. In these strategies the volatility forecast is used to dynamically adjust a portfolio between cash and equity to come as close as possible to a target level of volatility. The portfolio problem can be stated as,

$$\alpha_t = \frac{\sigma^T}{\hat{\sigma}_{t+1}}, \tag{8}$$

where σ^T is a target standard deviation, and $\hat{\sigma}_{t+1}$ is the one period ahead forecast standard deviation. It is easy to show that this strategy applied to an equity and risk free asset gives,

$$r^p_{t+1} = \alpha_t r^e_{t+1} + (1 - \alpha_t) r^f,$$

delivering a portfolio that tries to hit the target level of volatility. It is important to note that strategies of this type completely avoid trying to forecast expected returns. For the runs here, the target will be set to 10 percent in annual standard deviation units. Also, the strategy is implemented on each day, and it is assumed to buy at the open, and sell at the close since our volatility estimates purposefully ignore overnight returns.

Electronic copy available at: https://ssrn.com/abstract=3229272

All the results report the values from a 5 fold randomized cross-validation as used in the previous sections. The estimated portfolio strategies, α_t, are bias adjusted in the training samples. This means that a proportional increase or decrease in α_t is applied, $\lambda\alpha_t$, to bring the portfolio standard deviation to the exact target in the training data. Then λ is applied to the testing data. The reason for doing this is to adjust for many bias levels that have been added when estimating the final α_t target.

Results will again be compared using a similar cross validation strategy as in the last section. Forecasting models are estimated on training data, and implemented on test data. Strategy summaries are given in table 5. For these simulations the risk free rate is assumed to be 2 percent per year with 250 trading days in a year. The first column is the most important performance measure. For each cross validation run, the portfolio returns are generated, and their standard deviations and means are estimated. The first column displays the MSE difference between the estimated portfolio standard deviation and the target. This is the most critical value for understanding how well the strategy is performing. The second column, $\text{std}(\sigma_t^p)$, displays the standard deviation of the portfolio volatility estimates. This gives an idea of how well volatility is being controlled even though it may be biased off its target. The final two columns show the expected return and the Sharpe ratio for the dynamic portfolios. Both are reported as annualized returns.

The first line in the table, labeled "naive", shows the results for a static portfolio fraction. The fraction, α, is estimated using the standard deviation in the training set, and no attempt is made to forecast conditional variances or dynamically adjust the strategy. This is an important benchmark comparison model. It should be compared to the next line, "linear", which refers to the three lag linear model. It can be seen here that the variability of the portfolio variances around their target drops by nearly a factor of 3. Overall variability, as given in the Std measure, falls by a factor of 2. It is clear that the dynamic strategy is having a large impact on the smoothness of the volatility controlled portfolios.

The third row in table 5 moves to the nonlinear augmented volatility model. In terms of variability around the target volatility there is another large reduction of nearly 1/3 relative to the linear forecast model. There is a small reduction in variability as the std moves from 0.00026 to 0.00023. The final line moves to the kernel ridge model. Given that it has been the best performer in terms of prediction, it is not surprising that it shows further model improvement in the first two columns. However, the gains are now getting smaller. For the MSE from the target, the value has only fallen by about 10 percent. The std value only falls from 0.00023 to 0.00022. These gains do not appear that dramatic. In some ways they might have been predicted by thinking about the overall improvement in R^2 from 0.440 to 0.469 when moving from the nonlinear model to the kernel ridge. Further work will be necessary to determine if these improvements are really important to investors.

The last two columns are presented only for some extra information on the strategies. They demonstrate that the returns often move in unexpected ways. They actually rise for the nonlinear strategies. This is not expected, and may indicate something interesting about the nonlinearities that the strategies are finding. Also, it should be noted that the modest Sharpe ratios are partially a result of the fact that the strategy is only implemented during the open to close period each day, and does not contain the overnight returns.

Figure 1 compares the distributions of the estimated standard deviations for the different strategies from table 5. The blue bars represent the kernel ridge based portfolios. The target standard deviation is 0.10 in annual standard deviation units. The other extreme, the naive strategy, is shown with the green bars. The difference between the two is quite dramatic, and visually repeats the result from the table. It is clear why investors who desire a smoother portfolio return would be interested in implementing these strategies. The benchmark nonlinear model (orange) is very close to the kernel in terms of distribution. It is difficult to think that any set of investor preferences would be able to tell the difference between these two strategies.

5 Conclusions

This paper has explored several nonlinear forecasting tools as applied to forecasting a new daily realized volatility series for the Dow Industrials. The results show that a common machine learning tool, kernel ridge regression, was able to find nonlinear features which generated statistically significant improvements in out of sample forecasts. It was able to improve on a basic linear model, and also a nonlinear model using recent results from financial econometrics. In volatility space, where there is a lot of structure, the kernel ridge model is able to detect and utilize some of this structure.

It should be noted that the kernel ridge approach dramatically improved on the basic linear framework, and did so with no complex model specification. There was no need to coax the model with constructed nonlinear feature variables. It was able to learn them itself. The nonlinear specification used here was the result of many years of model explorations. It was even demonstrated that adding extra helper functions of predictors, like absolute values, was completely superfluous to the kernel ridge specification. It had already figured out how to use these, demonstrating its power as a general nonlinear function fitter.

The final results show that most of these improvements are significant to a trader using a dynamic volatility control strategy. However, the marginal gains of kernel ridge versus the nonlinear specification may not be economically important. On the other hand, kernel ridge did not require and careful model

Electronic copy available at: https://ssrn.com/abstract=3229272

prespecification, and its usefulness may be large in areas where lots of nonlinear model exploration has not been done. Further analysis of simple trading systems, such as volatility control, is an important issue for future research.

For a realized volatility forecasting problem, kernel ridge regression is a reliable way to include nonlinear features in modeling. It also appears to be fitting nonlinear features which have not yet been discovered in standard econometric models. While its forecast improvements beyond nonlinear specifications are significant, their economic gains relative to these models may not be that large.

Electronic copy available at: https://ssrn.com/abstract=3229272

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2006), Volatility and correlation forecasting, *in* G. Elliott, C. W. J. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', North Holland, pp. 778–878.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2013), Financial risk measurement for financial risk management, *in* G. Constantinedes, M. Harris & R. Stulz, eds, 'Handbook of the Economics of Finance', Vol. 2, Elsevier, pp. 1127–1220.

Andrada-Felix, J., Fernandez-Rodriguez, F. & Fuertes, A. M. (2016), 'Combining nearest neighbor predictions and model-based predictions of realized variance: Does it pay?', *International Journal of Forecasting* **32**(3), 695–715.

Audrino, F. & Knaus, S. D. (2016), 'Lassoing the HAR model: A model selection perspective on realized volatility dynamics', *Econometric Reviews* **35**(8-10), 1485–1521.

Barndorff-Nielsen, O. E. & Shephard, N. (2010), Measuring and modeling volatility, *in* R. Cont, ed., 'Encyclopedia of Quantitative Finance', John Wiley, pp. 1898–1901.

Black, F. (1976), 'Studies of stock price volatility changes', *Proceedings of the American Statistical Association, Business and Economics Statistics Section* pp. 177–181.

Bollerslev, T., Patton, A. J. & Quaedvlieg, R. (2016), 'Exploiting the errors: A simple approach for improved volatility forecasting', *Journal of Econometrics* **192**, 1–18.

Buccheri, G. & Corsi, F. (2017), HARK the SHARK: Realized volatility modelling with measurement errors and nonlinear dependence, Technical report, University of Pisa.

Chen, S., Hardle, W. K. & Jeong, K. (2010), 'Forecasting volatility with support vector machine-based garch model', *International Journal of Forecasting* **29**(4), 406–433.

Christie, A. A. (1982), 'The stochastic behavior of common stock variances: Value, leverage and interest rate effects', *Journal of Financial Economics* **10**, 407–432.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Electronic copy available at: https://ssrn.com/abstract=3229272

Curci, G. & Corsi, F. (2012), 'Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling', *Journal of Business and Economic Statistics* **30**(3), 368–380.

Diebold, F. X. & Nason, J. M. (1990), 'Nonparametric exchange rate prediction?', *Journal of International Economics* **28**, 315–332.

Efron, B. & Hastie, T. (2016), *Computer Age Statistical Inference*, Cambridge University Press.

Exterkate, P. (2013), 'Model selection in kernel ridge regression', *Computational Statistics and Data Analysis* **64**, 1–16.

Exterkate, P., Groenen, P. J. F., Heij, C. & van Dijk, D. (2016), 'Nonlinear forecasting with many predictors using kernel ridge regression', *International Journal of Forecasting* **32**, 736–753.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**, 1779–1801.

Hasanhodzic, J. & Lo, A. W. (2011), Black's leverage effect is not due to leverage, Technical report, Boston University.

LeBaron, B. (1992), 'Forecast improvements using a volatility index', *Journal of Applied Econometrics* **7**, S137–S150.

LeBaron, B. (1998), An evolutionary bootstrap method for selecting dynamic trading strategies, *in* A. P. N. Refenes, A. N. Burgess & J. Moody, eds, 'Advances in Computational Finance', Kluwer Academic Press, pp. 141–160.

LeBaron, B. (2018), A long history of realized volatility, Technical report, Brandeis International Business School.

Luo, R., Zhang, W., Xu, X. & Wang, J. (2017), A neural stochastic volatility model, Technical report, University College London.

Lux, M., Hardle, W. K. & Lessmann, S. (2018), Data driven Value-at-Risk forecasting using a SVR-GARCH-KDE hybrid, Technical report, Humboldt University.

McAleer, M. & Medeiros, M. C. (2010), 'Forecasting realized volatility with linear and nonlinear univariate models', *Journal of Economic Surveys* **25**(1), 6–18.

Electronic copy available at: https://ssrn.com/abstract=3229272

Meese, R. A. & Rose, A. K. (1990), 'Nonlinear, nonparametric, nonessential exchange rate estimation', *American Economic Review* **80**, 192–196.

Mizrach, B. (1992), 'Multivariate nearest-neighbor forecasts of EMS exchange rates', *Journal of Applied Econometrics* **7**, S151–63.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* **59**, 347–370.

Wang, J. & Yang, M. (2017), Conditional volatility persistence, Technical report, University of Technology Sydney.

Electronic copy available at: https://ssrn.com/abstract=3229272

Table 1: Summary Statistics

	Mean	Std	Skewness	Kurtosis
RV^1	1.004	0.56	4.87	72.5
RV^5	1.003	0.43	3.71	36.0
RV^{22}	1.000	0.33	1.96	10.3
rv^1	−0.108	0.46	0.14	4.35
rv^5	−0.065	0.35	0.47	4.59
rv^{22}	−0.045	0.29	0.37	4.20
Trend Volume (TV)	−0.202	0.10	−1.10	7.69
Close Ratio (CR)	0.051	0.54	−0.10	2.09

Summary statistics for realized volatility measures. Values are normalized by 250 day moving averages, which generates means near 1. $rv_t = \log(RV_t)$. Trend volume is the ratio of trading volume in the second half versus the first half of the day. Close ratio refers to where the market closes relative to the current high/low range.

Electronic copy available at: https://ssrn.com/abstract=3229272

$$rv_{t+1}^1 = \beta_0 + (\beta_1 + \beta_5 r_t + \beta_6 |r_t|) rv_t^1 + \beta_2 rv_t^5 + \beta_3 rv_t^{22} + \beta_4 I_t(r_t > 0) + \beta_7 TV_t + \beta_8 CR_t$$

Table 2: Benchmark model fitting

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	BIC	R^2	MSE
0.55 (0.009)								7480	0.327	0.220
0.23 (0.013)	0.38 (0.023)	0.28 (0.023)						6351	0.390	0.200
0.21 (0.013)	0.40 (0.023)	0.28 (0.023)	-0.05 (0.004)					6193	0.414	0.191
0.14 (0.015)	0.38 (0.023)	0.28 (0.023)	-0.05 (0.004)	-3.89 (0.809)	8.35 (0.996)			6098	0.440	0.183
0.16 (0.014)	0.39 (0.023)	0.28 (0.023)	-0.05 (0.004)	-1.03 (0.474)	3.12 (0.540)	0.25 (0.051)	0.04 (0.014)	6102	0.443	0.182

Model estimation: Fitted parameters, Jan 1987-Oct 2017, full sample. Estimation is by OLS, and numbers in parenthesis are standard errors on the parameter estimates. For this table only, R^2 is the traditional, in sample, R^2 measure. BIC refers to the Bayesian information criterion, or Schwarz criterion used for in sample model selection.

Electronic copy available at: https://ssrn.com/abstract=3229272

Table 3: Kernel variable selection

rv^1	rv^5	rv^{22}	r_t	TV_t	CR_t	$\|r_t\|$	$I(r_t > 0)$	MSE	R^2	MAE
X	X	X						0.190 (0.003)	0.411 (0.001)	0.292 (0.001)
X	X	X	X					0.169 (0.003)	0.474 (0.005)	0.284 (0.001)
X	X	X	X	X				0.168 (0.002)	0.477 (0.005)	0.283 (0.001)
X	X	X	X	X	X			0.167 (0.002)	0.479 (0.005)	0.284 (0.001)
X	X	X	X	X	X	X		0.168 (0.002)	0.479 (0.005)	0.284 (0.001)
X	X	X	X	X	X		X	0.168 (0.002)	0.478 (0.005)	0.284 (0.001)

Kernel ridge model estimates using randomized 5-fold cross validation to generate out of sample forecast performance. X's represents subsets of predictor variables used. Values in parenthesis are estimated standard errors across the 250 cross validation trials.

Electronic copy available at: https://ssrn.com/abstract=3229272

Table 4: Kernel Ridge Forecast Comparisons

Comparison	MSE	R^2	MAE
Raw	0.169	0.469	0.283
	(0.002)	(0.003)	(0.001)
Linear	0.873	1.204	0.964
	(0.005)	(0.007)	(0.001)
	[0.976]	[0.976]	[1.000]
Nonlinear	0.937	1.082	0.986
	(0.003)	(0.004)	(0.001)
	[0.936]	[0.936]	[0.972]
Linear Ridge	0.932	1.090	0.988
	(0.004)	(0.005)	(0.001)
	[0.876]	[0.976]	[0.928]

Forecast comparisons with kernel ridge regression. Raw refers to the raw forecast parameters from the benchmark kernel ridge regression (optimal kernel ridge model). All other forecast measures represent the kernel ridge values relative to the other comparison models. For example, MSE= 0.873, means that the kernel ridge MSE/Linear MSE mean ratio is 0.873. Numbers in parenthesis are standard errors across the 250 cross validation trials. Numbers in brackets represent the fraction of trials where the kernel ridge generates a lower forecast error than the target comparison model (given in left column).

Electronic copy available at: https://ssrn.com/abstract=3229272

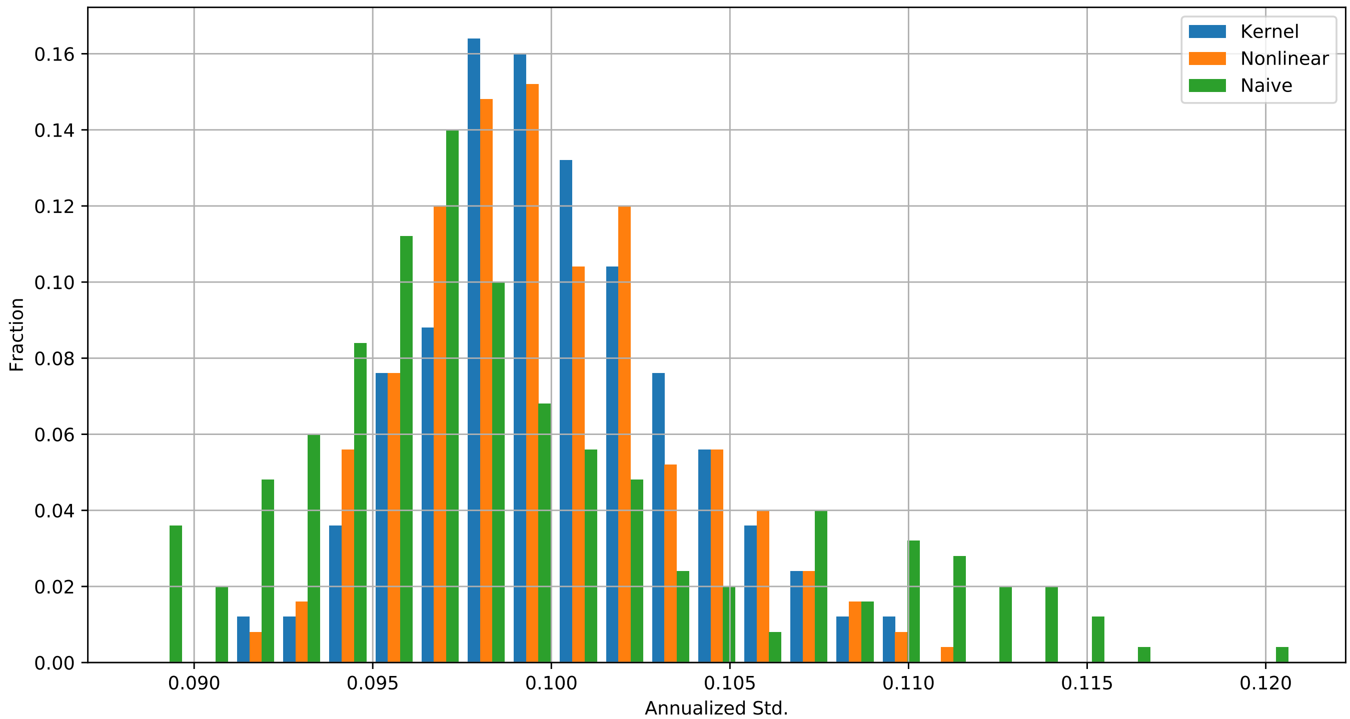
Table 5: Volatility Control

Volatility Model	MSE(target - vol)	Std(σ_p)	$E(r_p)$	Sharpe ratio
Naive	$1.78x10^{-7}$	0.00042	0.046	0.262
Linear	$6.83x10^{-8}$	0.00026	0.045	0.378
Nonlinear	$5.58x10^{-8}$	0.00023	0.058	0.381
Kernel	$5.13x10^{-8}$	0.00022	0.057	0.369

Volatility control strategy results. First column is the MSE of the distance to the target volatility for the simulated portfolios across 250 cross validation runs. Std(σ_p) measures the standard deviation of these volatility estimates. The last two columns are the expected return and Sharpe ratio respectively for the dynamic volatility control portfolios in annualized units. The strategy is implemented only with each day (open to close).

Electronic copy available at: https://ssrn.com/abstract=3229272

Figure 1: Volatility control strategy standard deviations



Electronic copy available at: https://ssrn.com/abstract=3229272