Online Appendix A: Comparing the forecasting ability of commonly employed econometric techniques

We examine the forecasting ability of three of the most widely used econometric methods for forecasting financial prices, namely, AR, ARIMA, GARCH (Charles, Darné, & Kim, 2011). ARIMA is also one of the most widely used benchmarking methods in machine learning studies (Atsalakis & Valavanis, 2009). We follow Pai and Lin's (2005) settings for ARIMA and, hence, we do not include technical indicators. We follow Awartani & Corradi (2005) by using GARCH(1, 1) and we employ the AR(1) model with the innovation distribution defined as Gaussian with constant variance. Econometric methods are generally employed to predict the change in value of an index. However, to compare the results with those studies employing ML techniques, we convert the predictions of change in value to predictions of change in direction.

To achieve these objectives, we estimate the following regression models for predictive accuracy and ROI:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{ARIMA}} ARIMA + \beta_{\text{GARCH}} GARCH + \varepsilon, \tag{16}$$

$$\text{ROI} = \alpha + \beta_{\text{MI}} MI + \beta_{\text{ST}} ST + \beta_{\text{T}} T + \beta_{\text{D}} D + \beta_{\text{ARIMA}} ARIMA + \beta_{\text{GARCH}} GARCH + \varepsilon, \tag{17}$$

where *MI, ST, T, D, ARIMA* and *GARCH* are dummy variables, taking the value 1 for middle income markets, a static model simulation, when using technical indicators alongside basic price covariates, a daily forecast horizon, ARIMA prediction and GARCH prediction, respectively, and 0 otherwise. That is, AR is the base model.

The results relating to predictive accuracy and ROI are shown in Tables A.1 and A.2, respectively. The estimated coefficients of ARIMA (-0.0162) and GARCH (-0.0072) are negative and statistically significant in the regression reported in Table A.1, indicating that AR outperforms ARIMA and GARCH in terms of accuracy. The coefficient of ARIMA is negative (-0.072) and statistically significant and the coefficient of GARCH is not statistically significant in the regression reported in Table A.2. These indicate that in terms of ROI, AR outperforms ARIMA and there is no difference between AR and GARCH. Taking both accuracy and ROI, into account, AR out-performs GARCH and ARIMA. Hence, we compare AR with SVM in our main analysis.

TABLE A.1: REGRESSION ANALYSIS OF PREDICTIVE ACCURACY

Predictive accuracy	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5081	0.0028	181.21	**< 10^{-16}**
Market maturity (MI)	0.0034	0.0029	1.1702	0.2426
Model simulation methodology (ST)	0.0114	0.0024	4.6725	**<10^{-5}**
Forecast horizon (D)	0.0037	0.0024	1.5307	0.1266
Prediction method (ARIMA)	-0.0162	0.003	-5.4513	**< 10^{-7}**
Prediction method (GARCH)	-0.0072	0.003	-2.4173	**0.0161**
Residual standard error	0.02458	df	402	
R^2	0.1211	Adjusted R^2	0.1102	
F-statistic	11.08	(on 5 and 402 DF)		
p-value	**< 10^{-9}**			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

TABLE A.2: REGRESSION ANALYSIS OF ROI

ROI	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	1.0537	0.0268	39.337	**< 10^{-16}**
Market maturity (MI)	-0.0023	0.0274	-0.0851	0.9322
Model simulation methodology (ST)	0.1791	0.0233	7.7007	**< 10^{-12}**
Forecast horizon (D)	-0.0324	0.0233	-1.3934	0.1643
Prediction method (ARIMA)	-0.072	0.0285	-2.5276	**0.0119**
Prediction method (GARCH)	0.0007	0.0285	0.0258	0.9794
Residual standard error	0.2348	df	402	
R^2	0.148	Adjusted R^2	0.1374	
F-statistic	13.97	(on 5 and 402 DF)		
p-value	**< 10^{-11}**			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

We compare the prediction performance of the most commonly employed econometrics methods (ARIMA, GARCH and AR: Charles, Darné, & Kim, 2011) with that of ANN and SVM. We adopt ARIMA, GARCH and AR among the well-known econometrics methods. We employ the same settings for the econometrics models indicated above, namely GARCH(1, 1) and AR(1) with the innovation distribution defined as Gaussian with constant variance. Overall, we obtain 136 simulation results for each of the prediction models (AR, ARIMA, GARCH, ANN, SVM), resulting in 680 simulation results in total. We estimate the following regression models to explain predictive accuracy and ROI, respectively:

$$\text{Accuracy} = \alpha + \beta_{\text{MI}}MI + \beta_{\text{ST}}ST + \beta_{\text{D}}D + \beta_{\text{ANN}}ANN + \beta_{\text{AR}}AR + \beta_{\text{ARIMA}}ARIMA + \beta_{\text{GARCH}}GARCH + \varepsilon, \tag{18}$$

$$\text{ROI} = \alpha + \beta_{\text{MI}}MI + \beta_{\text{ST}}ST + \beta_{\text{D}}D + \beta_{\text{ANN}}ANN + \beta_{\text{AR}}AR + \beta_{\text{ARIMA}}ARIMA + \beta_{\text{GARCH}}GARCH + \varepsilon, \tag{19}$$

where *MI, ST, T, D, ANN, AR, ARIMA and GARCH* are dummy variables, taking the value 1 for middle income markets, a static model simulation, a daily forecast horizon, ANN prediction, AR prediction, ARIMA prediction, GARCH prediction (with SVM as reference prediction), respectively, and 0 otherwise.

The results are displayed in Table A.3. In the regressions associated with both accuracy and ROI, all the coefficients of *ANN*, *AR*, *ARIMA* and *GARCH* are negative, indicating that SVM, the reference model, outperforms the other models. In addition, the coefficients of ANN are higher than those of *AR*, *ARIMA* and *GARCH* for both accuracy and ROI. Consequently, our results are consistent with the literature, that suggests machine learning techniques outperform traditional econometric methods when predicting stock prices (Pai & Lin, 2005).

Table A.3: REGRESSION ANALYSIS OF PREDICTION TECHNIQUES

	Predictive accuracy				ROI			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5197	0.0025	204.438	**< 10^{-15}**	1.2439	0.0242	51.379	**< 10^{-15}**
Market maturity (MI)	-0.0016	0.0022	-0.720	0.4718	-0.0582	0.0211	-2.756	**0.006**
Model simulation methodology (ST)	0.0102	0.0019	5.426	**< 10^{-7}**	0.1470	0.0180	8.207	**< 10^{-14}**
Forecast horizon (D)	0.0097	0.0019	5.176	**<10^{-6}**	-0.0268	0.0180	-1.498	0.135
ANN	-0.0108	0.0030	-3.634	**0.0004**	-0.1240	0.0283	-4.380	**<10^{-4}**
AR	-0.0129	0.0030	-4.351	**<10^{-4}**	-0.1638	0.0283	-5.785	**< 10^{-7}**
ARIMA	-0.0292	0.0030	-9.816	**< 10^{-15}**	-0.2358	0.0283	-8.326	**< 10^{-15}**
GARCH	-0.0201	0.0030	-6.774	**< 10^{-10}**	-0.1631	0.0283	-5.759	**< 10^{-7}**
Residual standard error	0.0245	df	672		0.2335	df	672	
R^2	0.1961				0.1847			
Adjusted R^2	0.1877				0.1762			
F-statistic	23.42	(on 7 and 672 DF)			21.75	(on 7 and 672 DF)		
p-value	**< 10^{-15}**				**< 10^{-15}**			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; D: Forecast horizon of one day; ANN: Prediction method employed is ANN;

AR: Prediction method employed is AR; ARIMA: Prediction method employed is ARIMA; GARCH: Prediction method employed is GARCH.

ONLINE APPENDIX B: IDENTIFYING THE PREDICTION METHOD PRODUCING THE BEST ACCURACY AND ROI FOR EACH STOCK INDEX

Table B.1 displays the setting of the experiments with the best performance for each stock index. As we discuss in the results section in the main paper, certain settings are positively influential on performance. For example, daily prediction appears more often than hourly prediction in the following table, indicating that daily predictions are generally more accurate and produce higher ROI than hourly predictions. Similarly, the static setting appears more often than the sliding window setting, indicating that predictions produced using the static setting are generally more accurate and produce higher ROI than predictions derived from a sliding window setting. In Table B.1, mature markets are more often those that display higher accuracy figures; e.g., 0.58 in France and 0.62 in Denmark. It is worth noting that our analysis, discussed in the main paper, suggests that SVM outperforms ANN significantly in terms of accuracy. However, ANN achieves the highest accuracy in more markets than SVM. We can explain the contrast by looking at the ANN and SVM boxplot in Figure 1: the average accuracy of SVM is higher, but there are more outliers for ANN, i.e. SVM performs better than ANN on average, but ANN occasionally achieves good performances across the combination of settings.

TABLE B.1: SETTING OF BEST ACCURACY AND ROI FOR EACH STOCK INDEX

		Setting of Highest Accuracy					Setting of Highest ROI				
Market	Index	Accuracy	Model	Method	Tech. Indicators	Horizon	ROI	Model	Method	Tech. Indicators	Horizon
Netherland	AEX	0.56	ANN	Static	No	Daily	1.3	SVM	Static	No	Hourly
Austria	ATX	0.56	SVM	Static	No	Daily	1.79	SVM	Static	No	Hourly
Belgium	BEL20	0.55	SVM	Static	Yes	Daily	1.37	ANN	Static	No	Daily
Brazil	Brazilian Bovespa Futures	0.54	ANN	Sliding Window	No	Daily	1.4	ANN	Sliding Window	No	Daily
Hungary	BUX	0.53	SVM	Sliding Window	Yes	Hourly	1.38	SVM	Sliding Window	No	Hourly
France	CAC 40	0.58	SVM	Static	No	Daily	1.79	SVM	Static	No	Daily
Germany	DAX	0.56	ANN	Static	Yes	Daily	1.39	ANN	Static	No	Daily
US	Dow Jones Industrial Average	0.55	ANN	Static	No	Daily	1.29	SVM	Static	No	Hourly
UK	FTSE 100	0.56	ANN	Static	No	Daily	1.27	SVM	Static	Yes	Hourly
Finland	OMXH25	0.53	ANN	Static	No	Hourly	1.59	ANN	Static	No	Hourly
Hong Kong	Hang Seng Index	0.53	ANN	Static	No	Daily	1.22	ANN	Sliding Window	Yes	Daily
Spain	IBEX 35	0.54	ANN	Static	No	Daily	2.19	SVM	Static	Yes	Hourly
Italy	FTSE MIB Index	0.54	ANN	Static	No	Daily	1.62	ANN	Static	Yes	Hourly
Indonesia	Jakarta Composite Index	0.56	ANN	Static	No	Daily	1.7	ANN	Sliding Window	No	Hourly
Denmark	OMX Copenhagen Index	0.62	ANN	Static	No	Daily	1.49	ANN	Static	No	Daily
Malaysia	FTSE Bursa Malaysia KLCI Index	0.54	ANN	Static	No	Daily	1.22	ANN	Sliding Window	Yes	Hourly
Korea	KOSPI 200 Index	0.57	SVM	Static	Yes	Daily	1.22	SVM	Static	Yes	Daily

US	NASDAQ-100	0.57	ANN	Sliding Window	No	Hourly	1.41	ANN	Static	No	Daily
Japan	Nikkei 225	0.55	ANN	Static	No	Daily	1.77	ANN	Static	Yes	Hourly
Norway	OSE All Share Index	0.55	ANN	Sliding Window	No	Daily	1.36	ANN	Static	Yes	Hourly
Portugal	PSI-20	0.54	ANN	Static	No	Daily	1.63	ANN	Static	Yes	Hourly
Czech	Prague Stock Exchange Index	0.53	SVM	Static	Yes	Daily	1.2	SVM	Static	Yes	Daily
Latvia	OMX Riga Index	0.54	ANN	Static	No	Daily	2.11	SVM	Sliding Window	No	Hourly
China	ShangHai SE Composite Index	0.54	SVM	Static	No	Daily	1.67	SVM	Static	No	Daily
Sweden	OMX ALL-SHARE Stockholm Index	0.56	ANN	Static	No	Daily	1.36	ANN	Static	No	Hourly
US	S&P 500	0.57	ANN	Static	No	Daily	1.35	ANN	Static	No	Daily
Singapore	Straits Times Index	0.55	ANN	Sliding Window	No	Daily	1.2	ANN	Sliding Window	Yes	Daily
Switzerland	Swiss Market Index	0.57	ANN	Static	No	Daily	1.34	ANN	Static	No	Daily
Estonia	OMX Tallinn Index	0.53	ANN	Sliding Window	No	Daily	1.49	SVM	Static	No	Hourly
Thailand	Thai Stock Exchange MAI Securities Index	0.56	ANN	Static	No	Daily	1.74	ANN	Sliding Window	No	Hourly
South Africa	FTSE/JSE Africa Top40	0.55	ANN	Static	No	Daily	1.43	ANN	Static	No	Hourly
Canada	SP TSX Composite Index	0.58	ANN	Static	No	Daily	1.23	ANN	Static	No	Daily
Turkey	ISE-100	0.54	ANN	Static	No	Daily	2.4	ANN	Sliding Window	Yes	Hourly
Lithuania	OMX Vilnius Index	0.54	ANN	Static	No	Daily	1.31	ANN	Static	No	Daily

ONLINE APPENDIX C: ANALYSIS OF MARKETS WITH INTRADAY DATA AVAILABLE PRIOR TO 2008

In selecting the sample period for the main analysis, we wanted to maximize the number of markets we could use with sufficient daily and intraday data to test the hypotheses. However, the availability of intraday data was limited and, for many markets, was only available from 2008 onwards. To include as many markets as possible, we choose the period 2008 to 2014, where intraday data is available for most markets. The selection of the sample period may introduce sample selection bias. In order to examine whether this bias affected our conclusion, we conducted additional experiments with the thirteen stock indexes for which we could access intraday data for longer periods (listed in table C.1).

Table C.1 : STOCK INDICES WITH INTRADAY DATA AVAILABLE BEFORE 2008

NO.	Economy	World Bank Income Level	Index	Start Date	End Date
1	US	High	S&P 500	1/2/1983	19/2/2014
2	Japan	High	Nikkei 225	1/7/2003	19/2/2014
3	Korea	High	KOSPI 200 Index	1/2/2004	19/2/2014
4	Hong Kong	High	Hang Seng Index	1/12/2006	19/2/2014
5	France	High	CAC 40	1/7/2003	19/2/2014
6	UK	High	FTSE 100	1/7/2003	19/2/2014
7	Italy	High	FTSE MIB Index	1/7/2003	19/2/2014
8	Germany	High	DAX	1/7/2003	19/2/2014
9	Hungary	Middle	BUX	1/7/2003	19/2/2014
10	Switzerland	High	Swiss Market Index	1/7/2003	19/2/2014
11	Spain	High	IBEX 35	1/7/2003	19/2/2014
12	US	High	Dow Jones Industrial Average	1/4/1993	19/2/2014
13	US	High	NASDAQ-100	2/1/1997	19/2/2014

We estimate regressions based on equations (8) and (9) to explain prediction accuracy and ROI, and the results are presented in Table C.2. The results relating to prediction accuracy are consistent with the results for the larger group of markets for which data was available for 2008-14 (reported in Table 6). We observe some differences in terms of the results relating to ROI compared to those for the larger group of markets for which data was available for 2008-14 (reported in Table 7). In particular, the coefficient of market maturity (MI), which is negative and significant at the 5% level in the results reported in Table 7, is still negative but the p-value increases to 0.0851. This may arise because only one stock index of the 13 indices included in this additional experiment is from a middle income market, i.e. a less mature market; the sample size may therefore be insufficient to observe a significant difference. In addition, the coefficient of forecast horizon (D) is positive in these additional experiments and is not statistically significant. As we discussed earlier, this can be regarded as evidence of an imperfect link between predictive accuracy and profitability. The remaining factors show similar results to those reported in Table 7. Overall, the result from the markets with earlier intraday data is in line with our earlier conclusion and suggests that selection bias did not influence the results from the main analyses reported in the paper.

Table C.2: REGRESSION ANALYSIS OF THE MARKETS WITH INTRADAY DATA BEFORE 2008

	Predictive accuracy				ROI			
	Estimated Coefficient	Std. Error	t value	p value	Estimated Coefficient	Std. Error	t value	p value
(Intercept)*	0.5047	0.0026	197.01	**< 10^{-15}**	1.0772	0.0541	19.901	**< 10^{-15}**
Market maturity (MI)	-0.0186	0.0043	-4.360	**<10^{-4}**	-0.1559	0.0901	-1.730	0.0851
Model simulation methodology (ST)	0.0047	0.0022	2.073	**0.0394**	0.1345	0.0480	2.802	**0.0056**
Covariate composition (T)	0.0001	0.0022	0.042	0.9663	-0.0149	0.0480	-0.310	0.7566
Forecast horizon (D)	0.0074	0.0022	3.258	**0.0013**	0.0216	0.0480	0.451	0.6528
Prediction method (SVM)	0.0080	0.0022	3.512	**0.0005**	0.1486	0.0480	3.094	**0.0023**
Residual standard error	0.0164	df	202		0.3462	df	202	
R^2	0.1863				0.0930			
Adjusted R^2	0.1662				0.0706			
F-statistic	9.251	(on 5 and 202 DF)			4.143	(on 5 and 202 DF)		
p-value	**<10^{-7}**				0.0013			

* Base model indicated by values of experimental factors given in brackets: MI: Middle income markets; ST: Static simulation methodology; T: Technical indicators included amongst the covariates; D: Forecast horizon of one day; SVM: Prediction method employed is SVM