Democritus University of Thrace
DEPARTMENT OF ECONOMICS
Excellence. Science. Innovation.

Economic and Financial Forecasting Using Machine Learning Methodologies

Periklis Gogas
Associate Professor

European Union
European Social Fund

OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS, CULTURE & SPORTS
MANAGING AUTHORITY

NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND

Co-financed by Greece and the European Union

Presentation Outline

- Intuition for using Machine Learning in Forecasting
- Forecasting applications:
  - Exchange rate forecasting
  - Forecasting House Prices in the U.S.
  - Forecasting Bank Failures and stress testing in the U.S.
  - Forecasting Recession

The framework

- Research Grant “Thales” – awarded to T. Papadimitriou
- Started 2012
- Concluded November 2015

Why Machine Learning?

- Renewed **interest** in machine learning
- **Higher** volumes and **varieties** of available data
- **Affordable** data storage.
- More importantly: computer processing is **cheaper** and more **powerful**

Application 1: Forecasting FX Rates

Journal of Forecasting, 2015, vol. 34 (7), pp. 560-573

- 2 exchange rate frequencies
- Theory: different data generating processes
- Is this confirmed?

High frequency (daily)

- Driven by microeconomic factors: Markets, demand & supply

Lower frequency (monthly)

- Driven by Macroeconomics: e.g. Purchasing Power parity, etc.

Methodologies employed

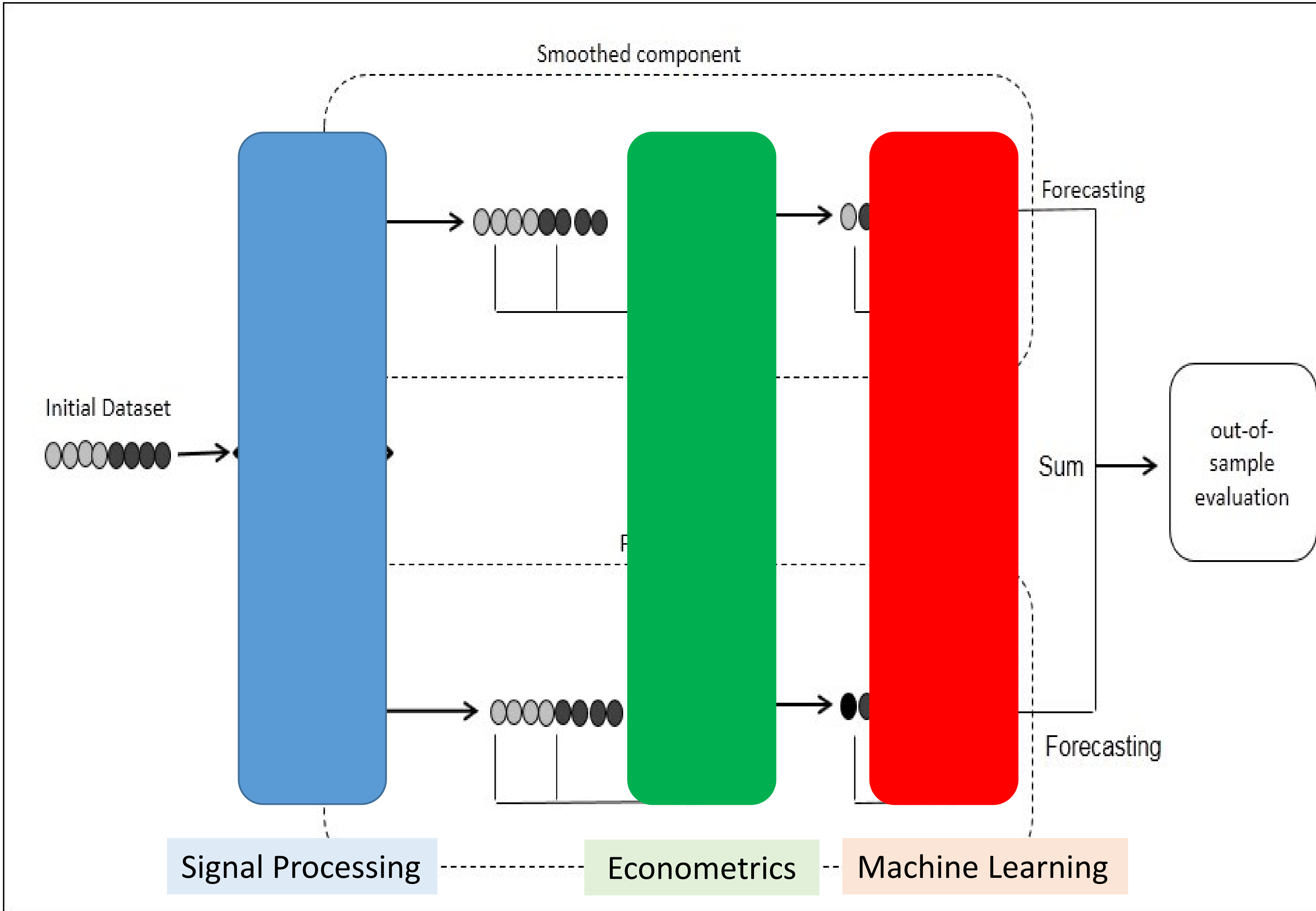
Machine Learning

- Artificial Neural Networks
- Support Vector Regression

Econometrics

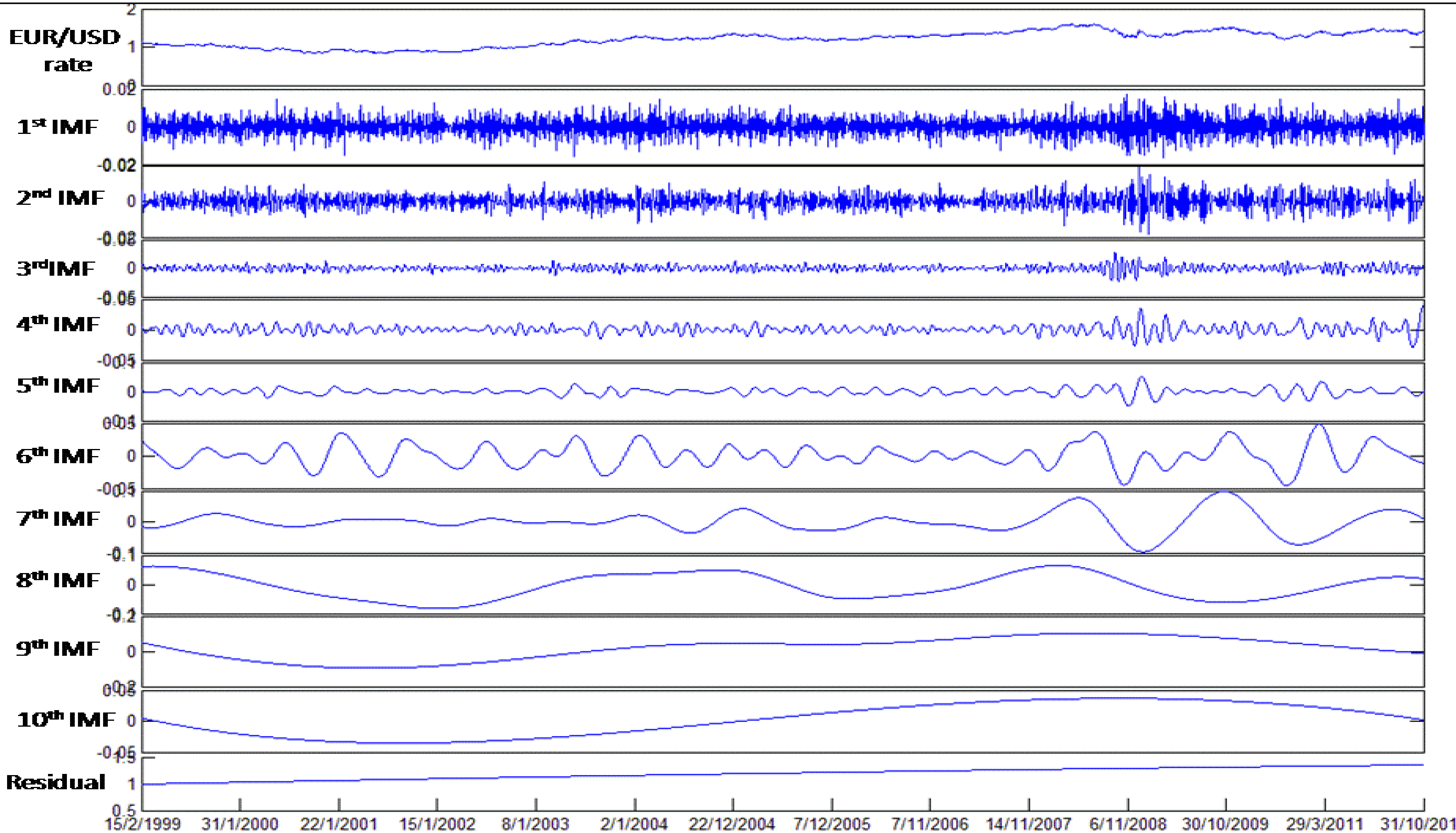
- ARCH
- GARCH
- EGARCH
- AR
- ARMA
- ARIMA
- AFRIMA
- Random Walk

Overview: Hybrid Method



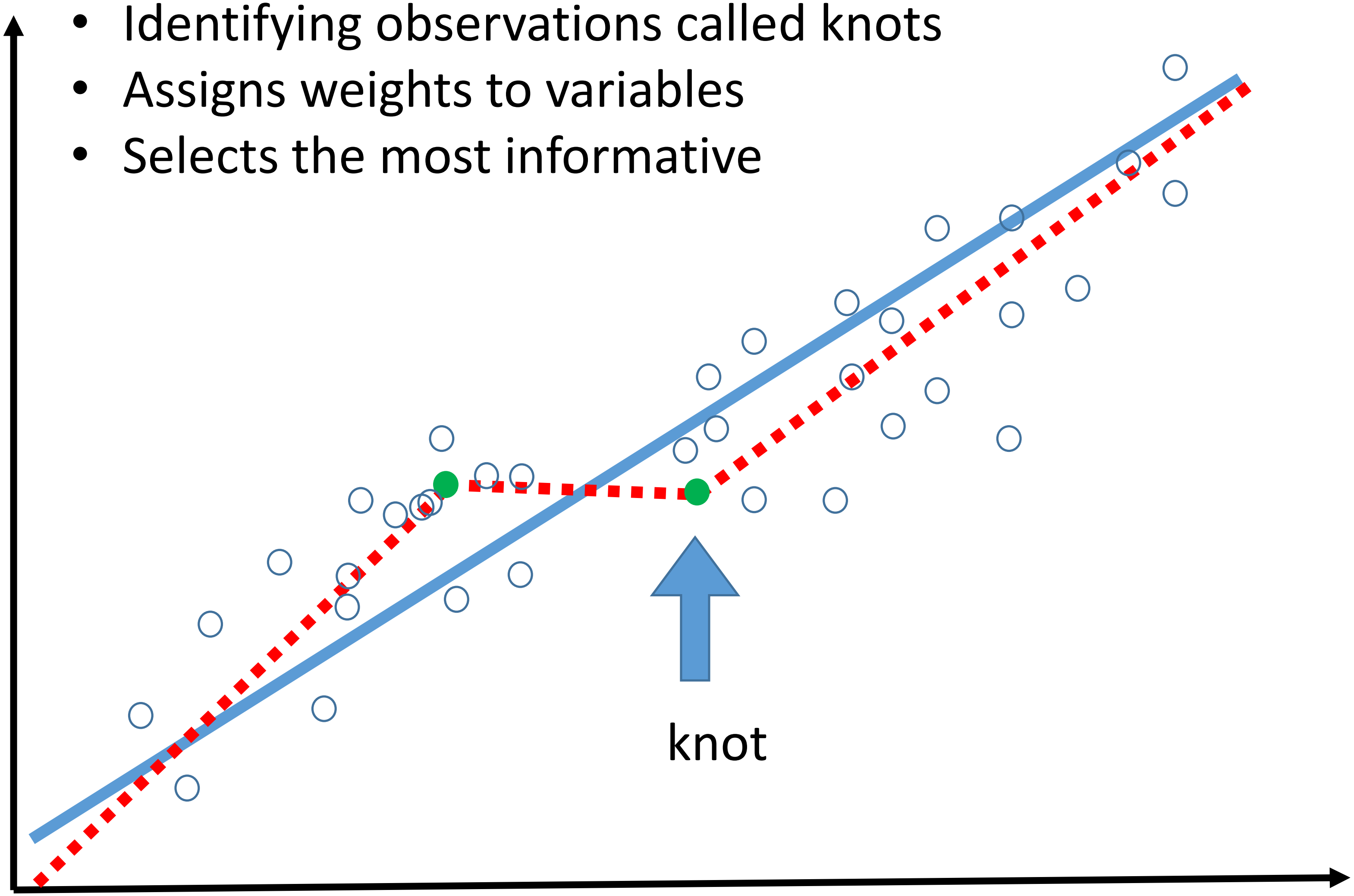
Decomposition: Ensemble Empirical Mode Dec.

- Signal processing - Huang et al. (2009)
- Additive, oscillatory, signals: Intrinsic Mode Functions (IMFs)
- High energy signals were treated as noise – we exploit them



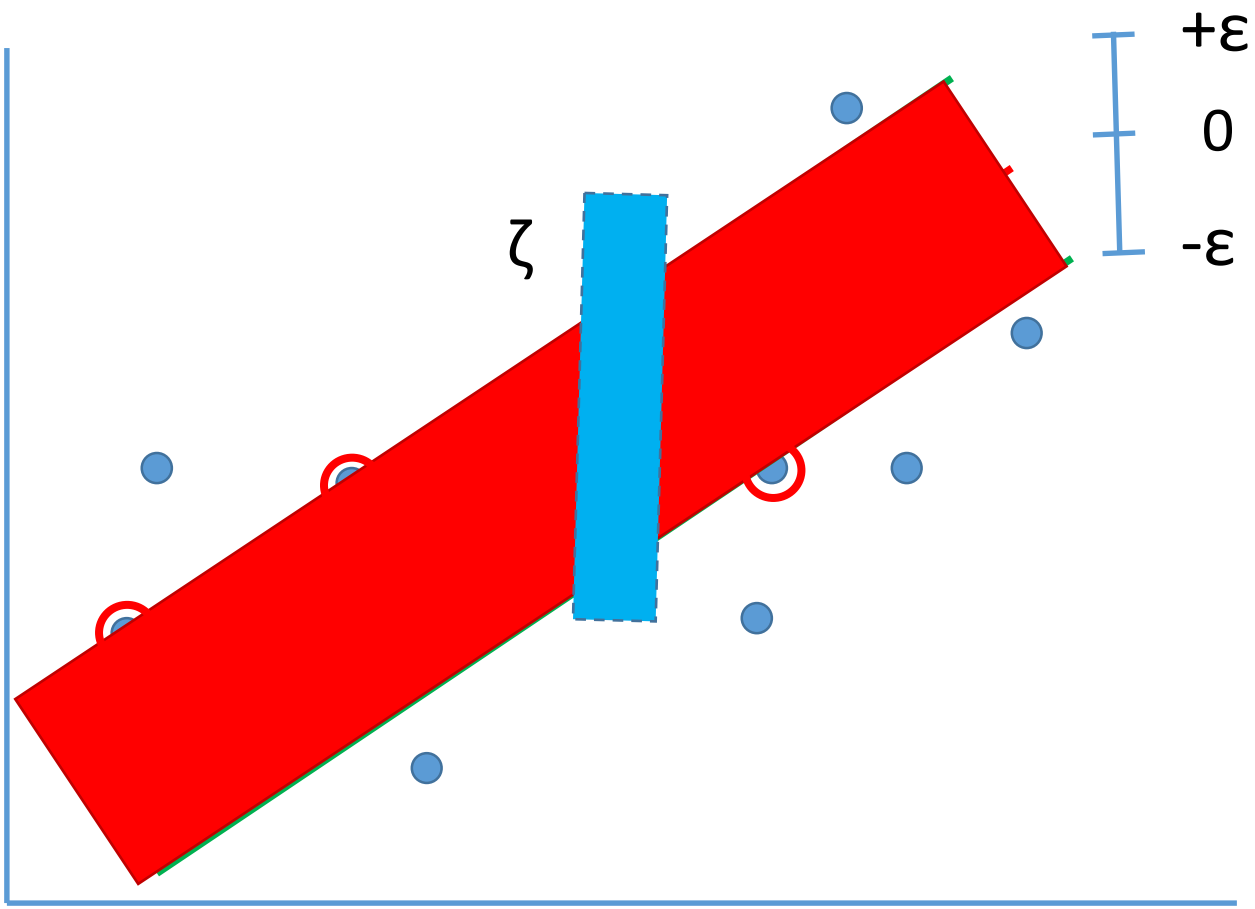
Variable Selection: Multivariate Adaptive Regression Splines (MARS)

- Breaks dataset into subsamples
- Identifying observations called knots
- Assigns weights to variables
- Selects the most informative



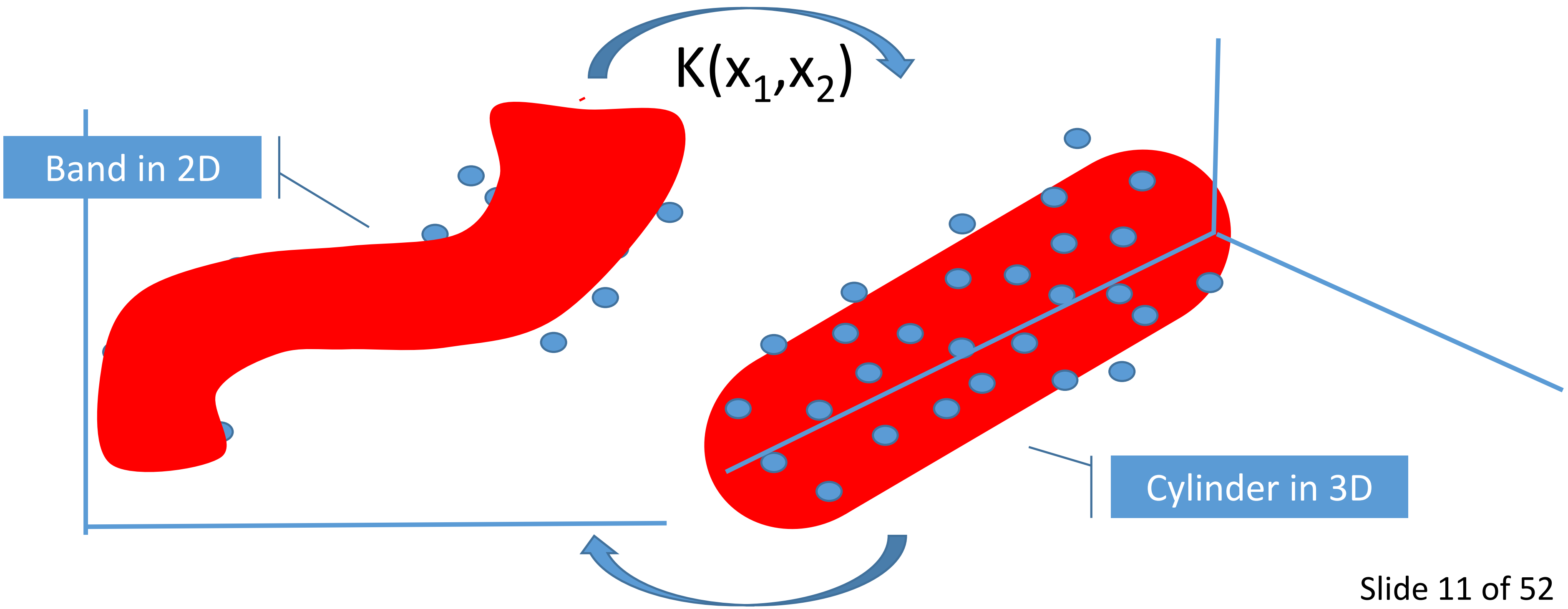
Forecasting: Support Vector Regression

- Fit **error tolerance band:** errors set to zero
- Higher **flexibility** than fitting a simple line
- Position defined according to subset of observations called **support vectors**



The Kernel Projection

- We fit a **linear** model
- Real phenomena usually **non-linear**
- Project to **higher** dimensions
- Find a dimensional space where a **linear error tolerance band** is defined
- Re-project to original space and obtain non-linear error tolerance band



Kernels Used

Linear $$K(x_1, x_2) = x_1^T x_2$$

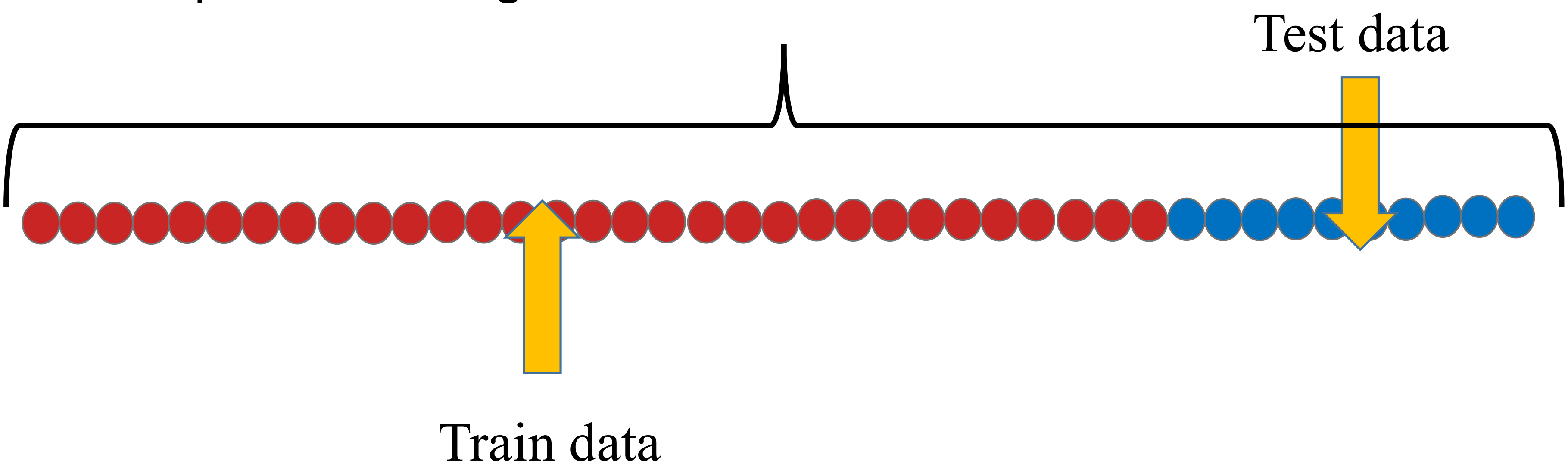
RBF $$K(x_1, x_2) = e^{-\gamma \|x_1 - x_2\|^2}$$

Polynomial $$K(x_1, x_2) = (\gamma x_1^T x_2 + r)^d$$

Sigmoid $$K(x_1, x_2) = \tanh(\gamma x_1^T x_2 + r)$$

Building the Model: Split the dataset into two parts

- Separate data
- **Train** data: used to obtain the **optimum** model
- **Test** data: never seen by the optimum model, used for out-of-sample forecasting



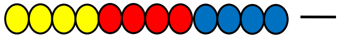
Overfitting and Cross-Validation

- The problem of **overfitting**:
  - **High** accuracy in a specific sample
  - **Low** generalization ability
- Solution: use k-part **Cross-Validation**
- Example of a 3-part CV

Overfitting and Cross-Validation

Initial

Dataset



Forecasting Exchange Rates

The data used:

- 5 rates of various trading volumes:
- 2 **high** volume rates USD/EUR, USD/JPY
- 1 **medium** volume rate NOK/AUD
- 2 **low** volume rates ZAR/ PHP and NZD/BRL

Input Variables (192)

Commodities

Crude Oil
Cotton
Lumber
Cocoa
Coffee
Orange Juice
Sugar
Corn
Wheat
Oats
Rough Rice
Soybean Meal
Soybean Oil
Soybeans
Feeder Cattle
Lean Hogs
Live Cattle
Pork Bellies
Iron Ore

Metals

Gold
Copper
Palladium
Platinum
Silver
Aluminum
Zinc
Nickel
Lead
Tin

Stock Indices

Dow Jones
Nasdaq 100
S & P 500
DAX
CAC 40
FTSE 100
Nikkei 225

Interest rates

T-bill 6 months
T-bill 10 years
Spread MLP-EURIBOR 3M
Spread MLR-Eonia
Spread FF-CP
Spread FF-EFF
EONIA
EURIBOR 1 Week
ECB Interest rate
EURIBOR 1 Month
FED rate

Technical Analysis variables

Five Day Index
Moving Average 3 day
Moving Average 5 day
Moving Average 10 day
Moving Average 30 day

Macroeconomic Vars all countries

CPI
Productivity
index
GDP
Trade Balance
Unemployment
Central Bank
Discount rate
Long Term
Interest Rates
Short Term
Interest Rate
Aggregate money
M1, M2
Public Debt
Deficit/Surplus of
Government
Budget

Exchange Rates

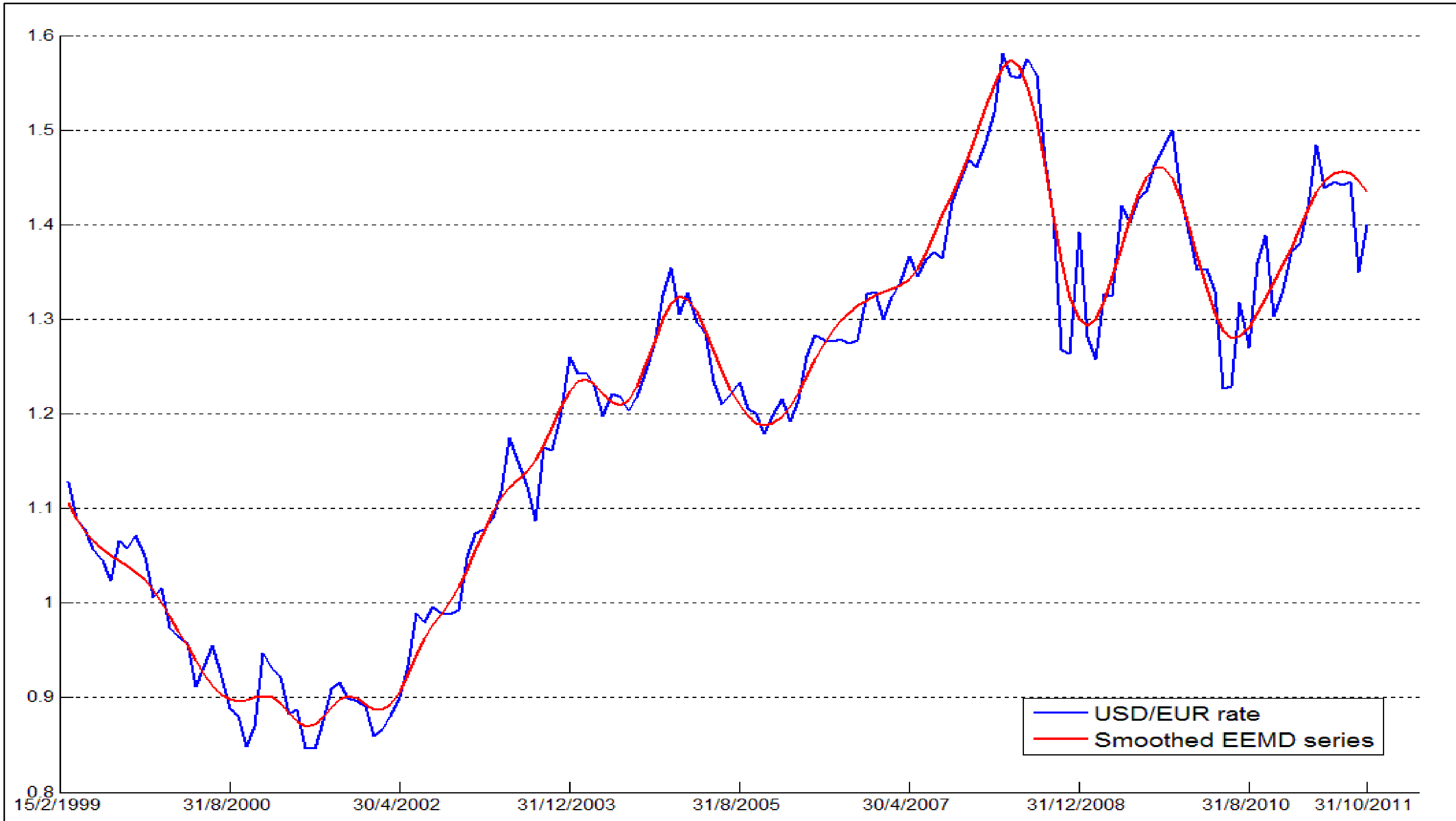
JPY/EUR
JPY/USD
USD/GBP
BRL/NZD
NOK/AUD
PHP/ZAR
EUR/GBP
EUR/USD

USD Trade Weighted Indices

Major partners
Broad Index
Other Partners

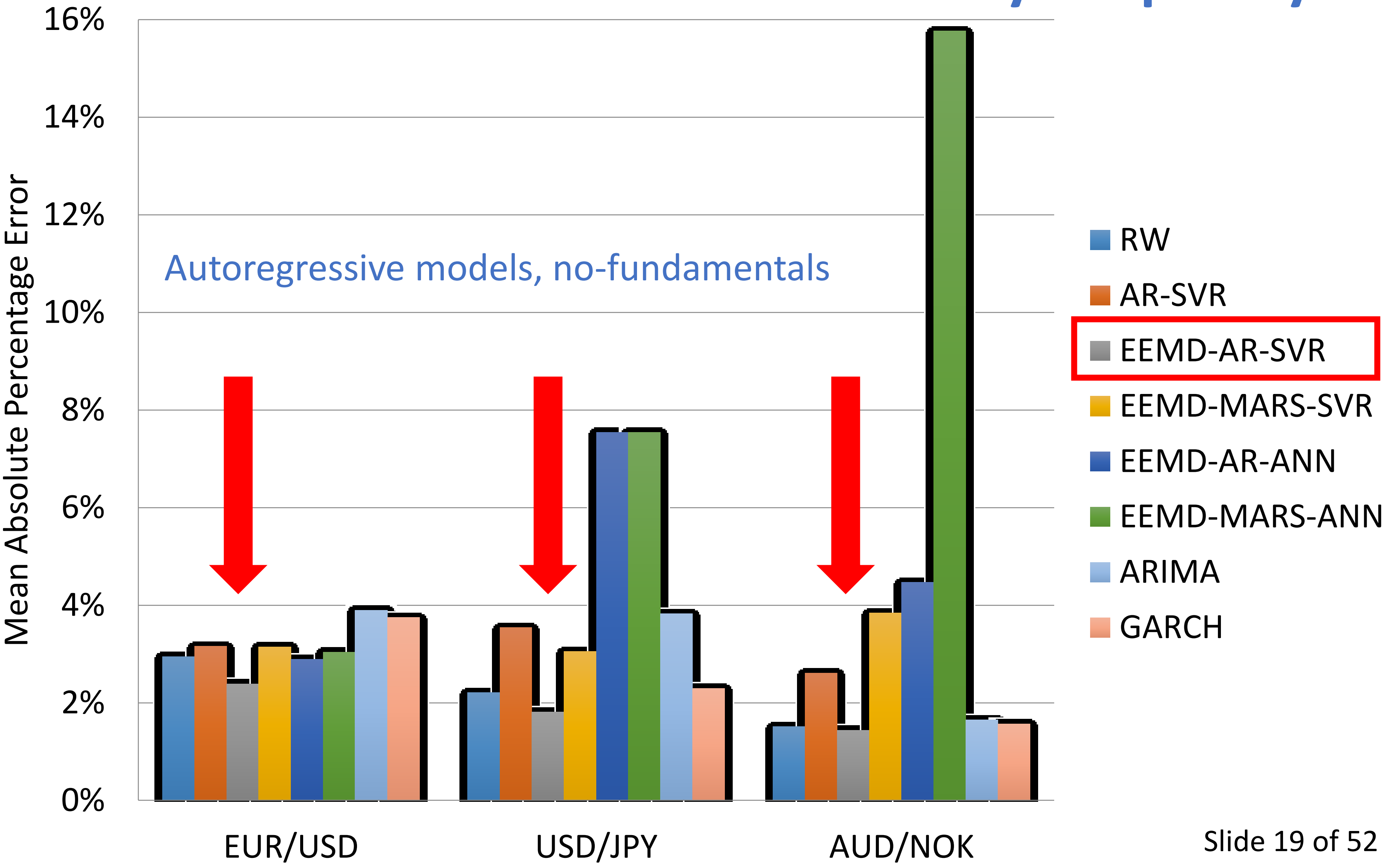
Forecasting Exchange Rates

EEMD smoothed component vs USD/EUR series



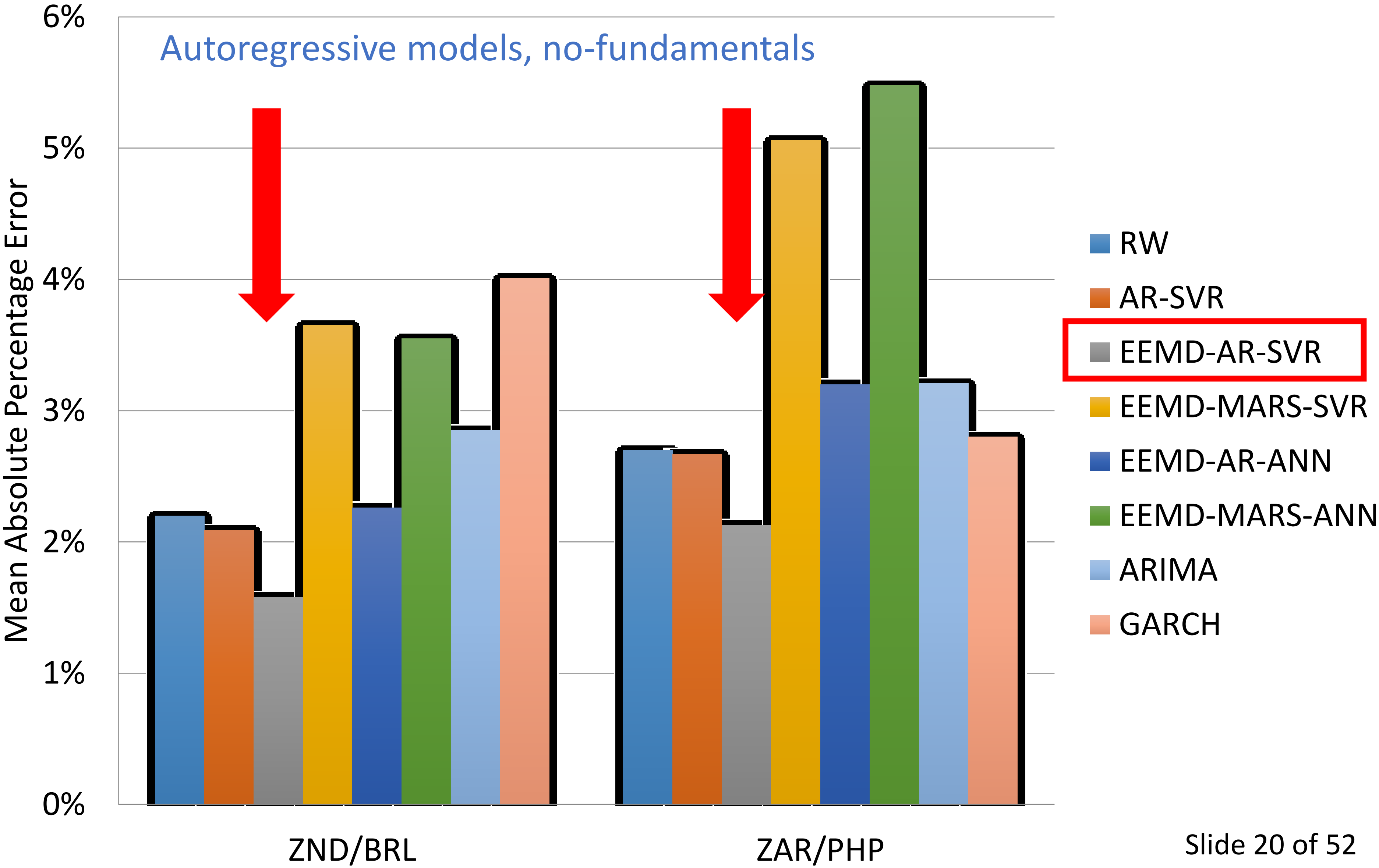
Out-of-sample forecasting results

Daily frequency



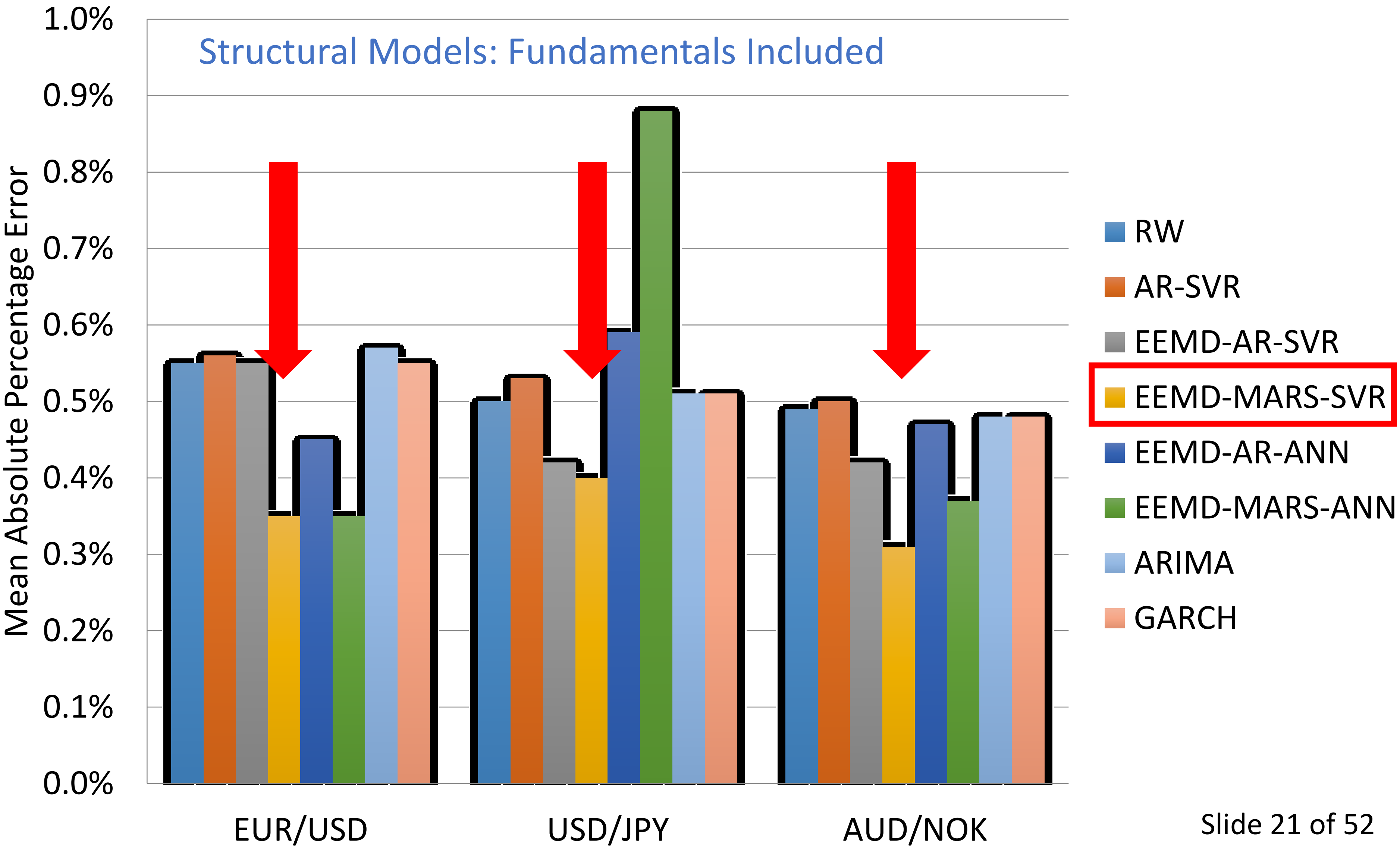
Out-of-sample forecasting results

Daily frequency



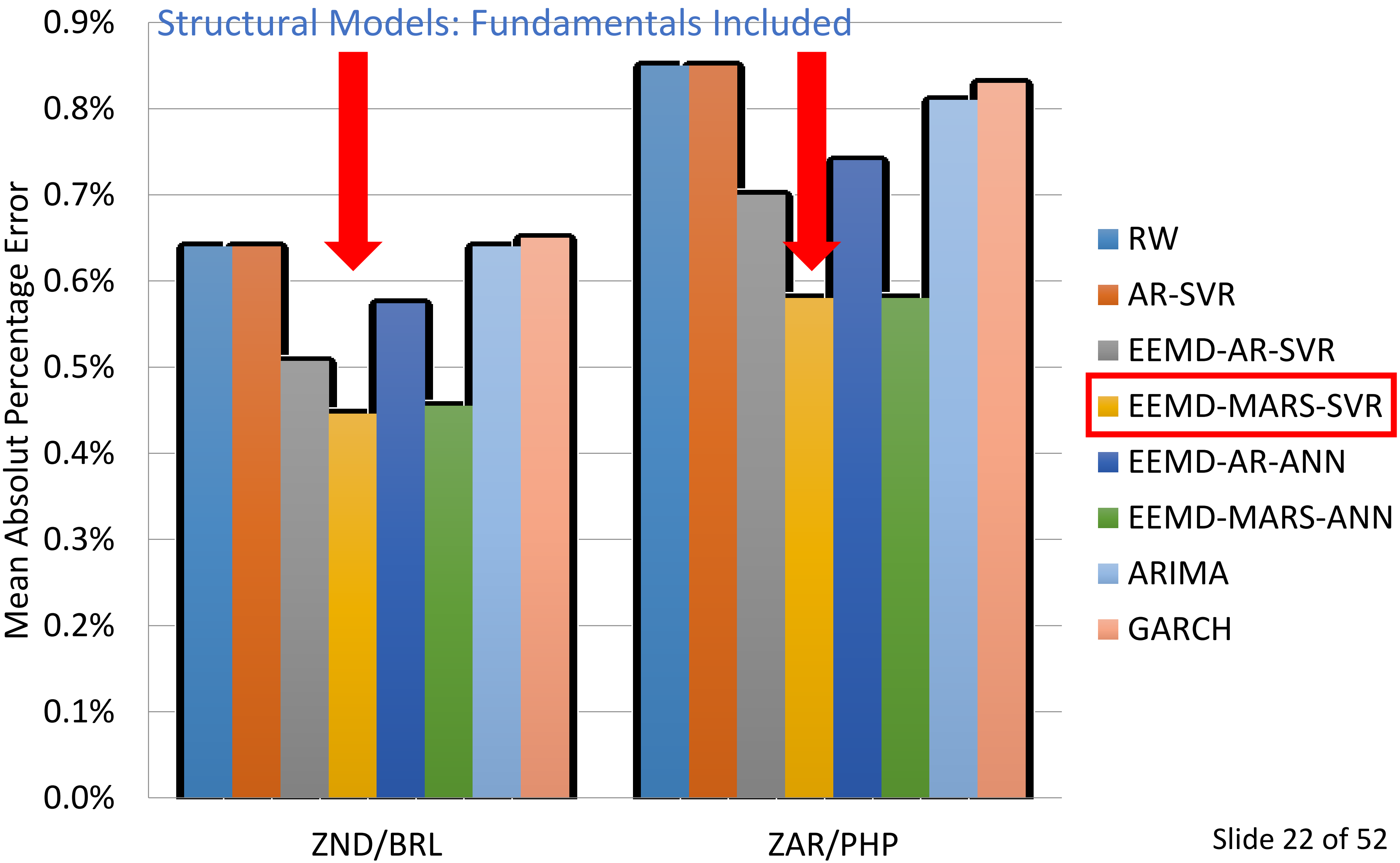
Out-of-sample forecasting results

Monthly frequency



Out-of-sample forecasting results

Monthly frequency



Conclusions

- ✓ Best model **outperforms** the RW model
- ✓ ML: **outperforms** all econometric methodologies
- ✓ **Rejection** even of the weak form of efficiency
- ✓ The model captures the **different** data generating processes that drive exchange rates in **short** and **long** run

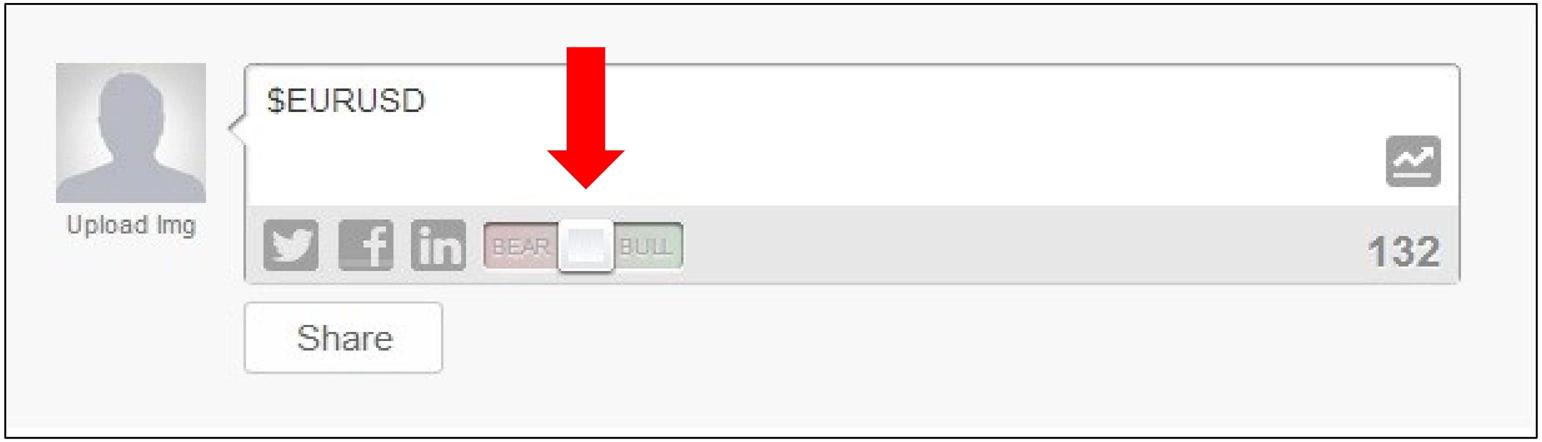
Application 2: Forecasting Exchange Rates Directionally

Algorithmic Finance, vol. 4 (1-2), pp. 69-79.

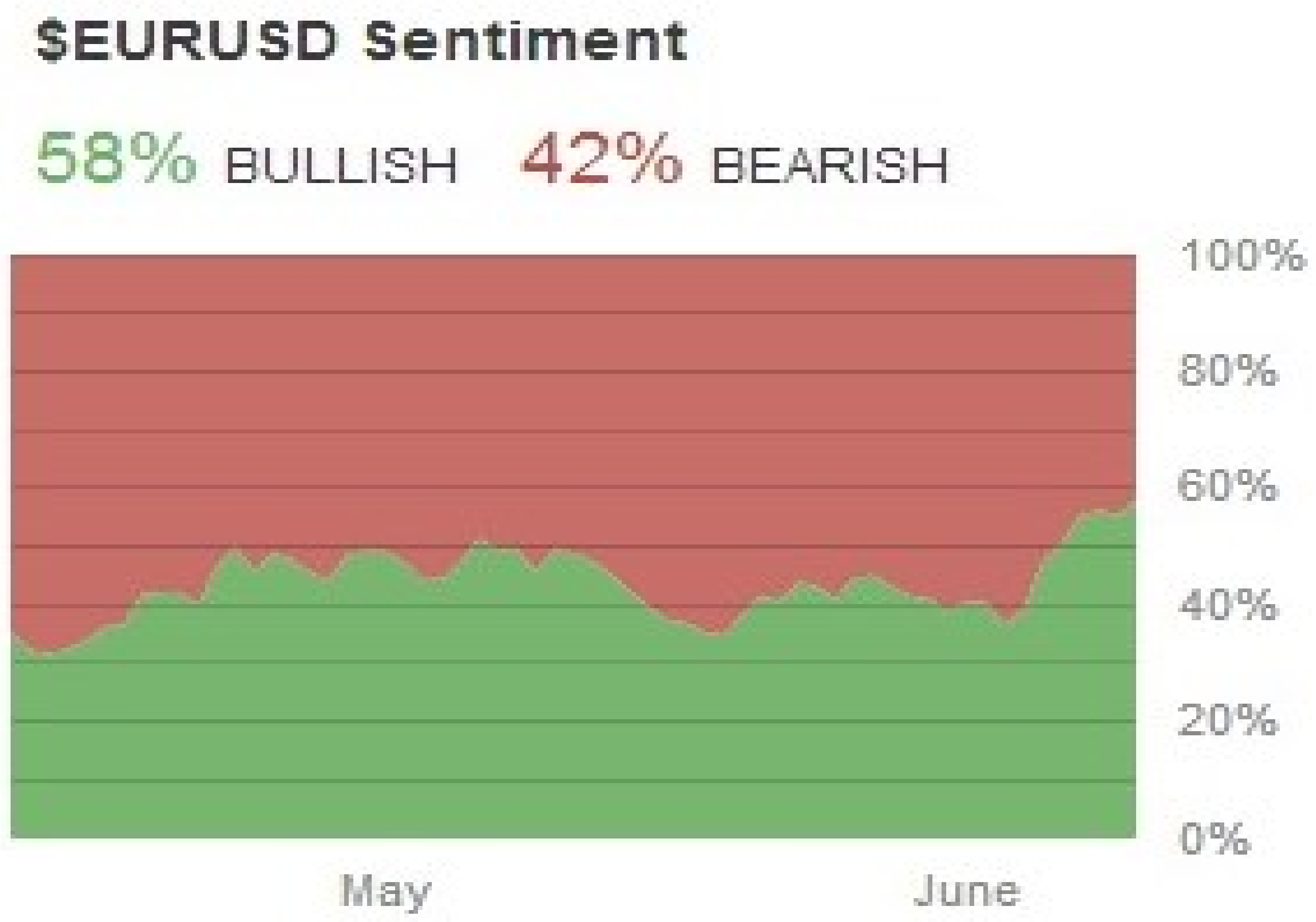
- Forecast direction: UP or DOWN
- Frequency: Daily
- Rates: USD/EUR, USD/JPY, USD/GBP and USD/AUD
- Data span: January 2, 2013 to December 26, 2013
- Training 200
- Test: 51

Application 2: Forecasting Exchange Rates Directionally

- Ideally use: Order Flow Analysis
- Problem: data availability
- Alternative: www.Stocktwits.com
- Investor sentiment index
- Investors explicitly provide their sentiment



Application 2: Investor Sentiment Index



Input Variables Sets

- Past values of the exchange rate
- The volume of “Bearish” and “Bullish” posts per day
- The volume of “Bearish”, “Bullish” and total posts per day
- Past values of the exchange rate and the volumes of “Bullish” and “Bearish” posts per day
- Past values of the exchange rate, volumes of “Bullish”, “Bearish” and total posts per day

Application 2: Methodologies

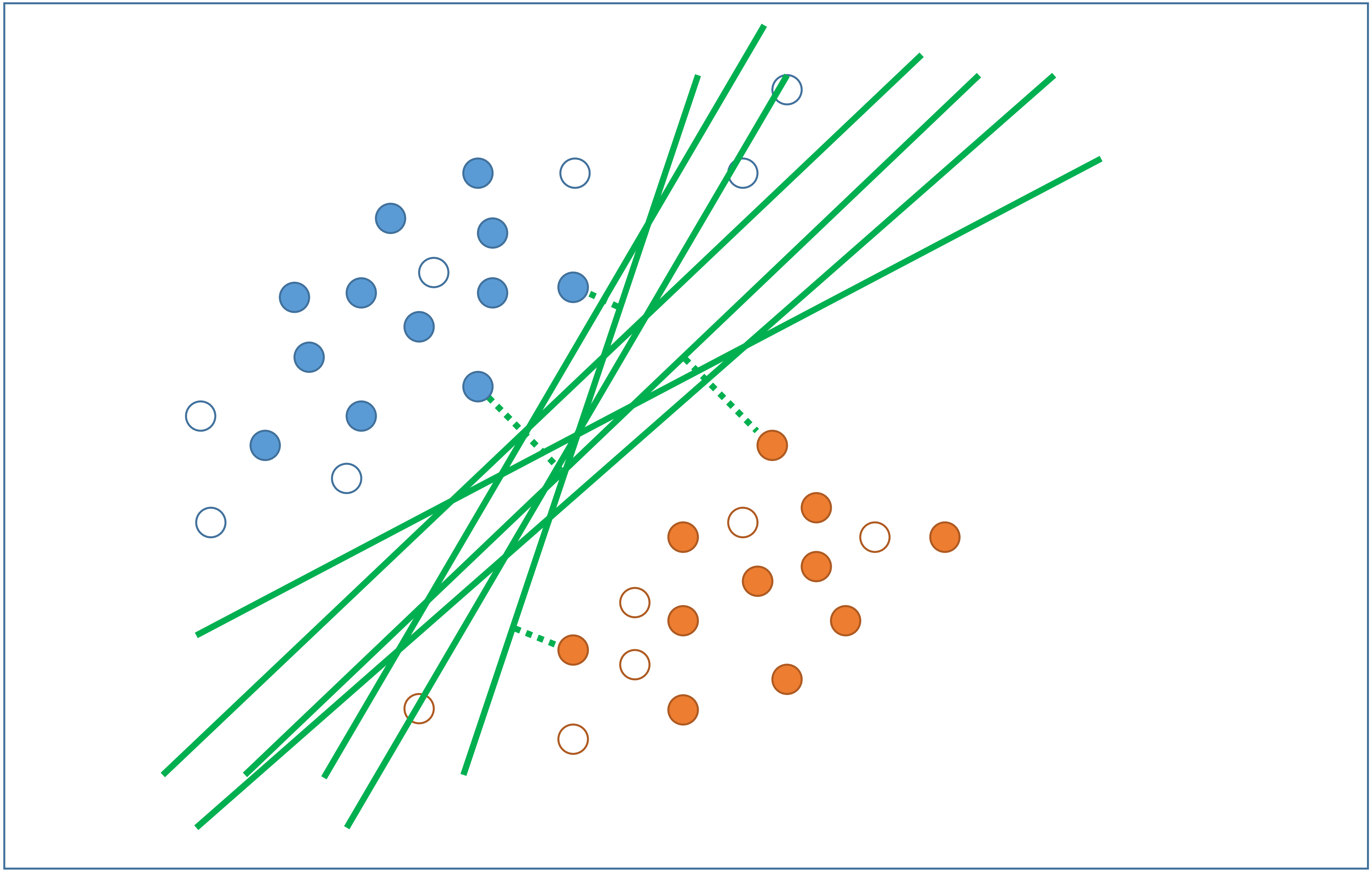
Machine Learning

- Artificial Neural Networks
- Support Vector Machines
- Knn Nearest Neighbors
- Boosted Decision Trees

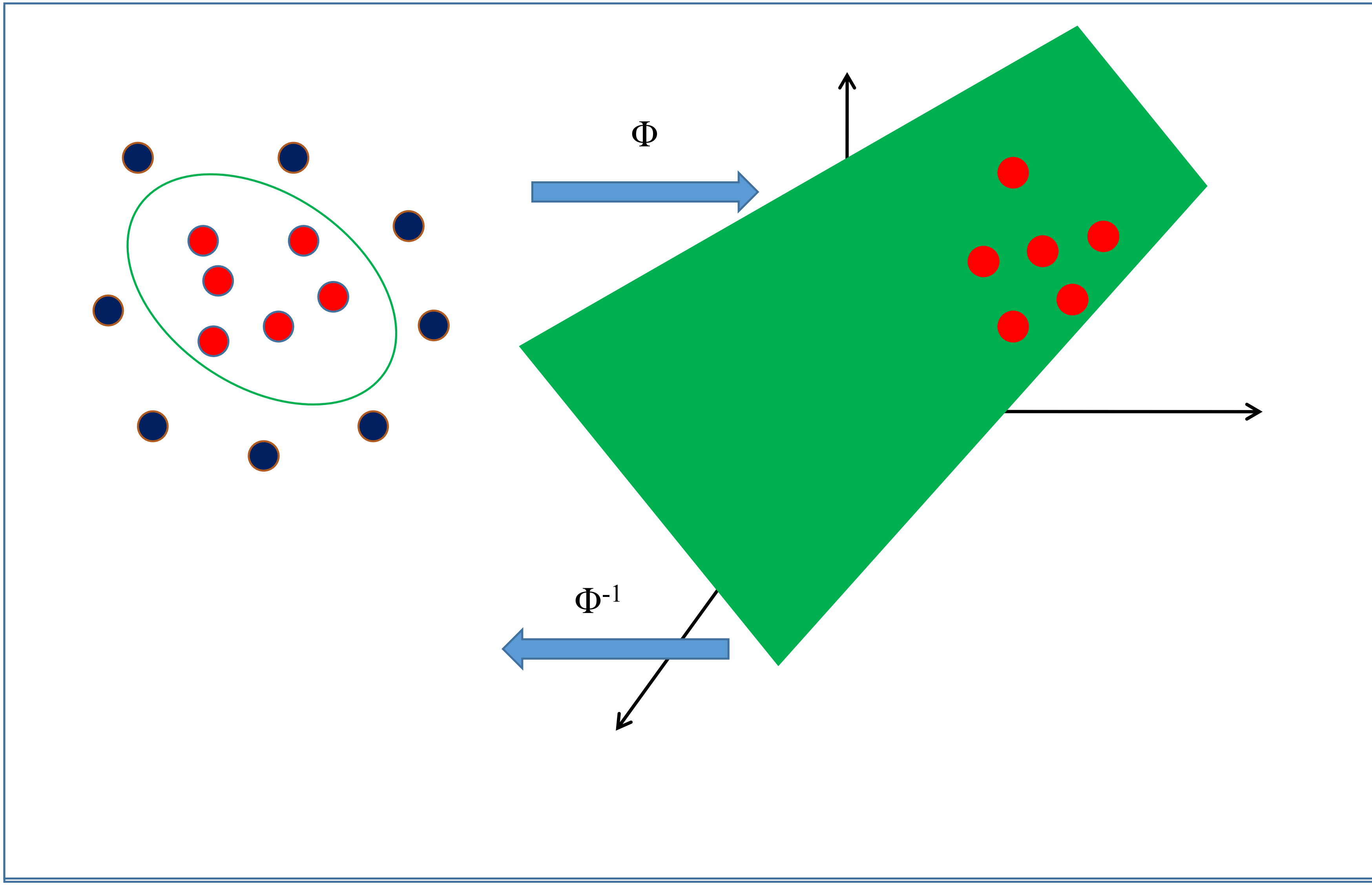
Econometrics

- Logistic Regression
- Naïve Bayes Classifier
- Random Walk

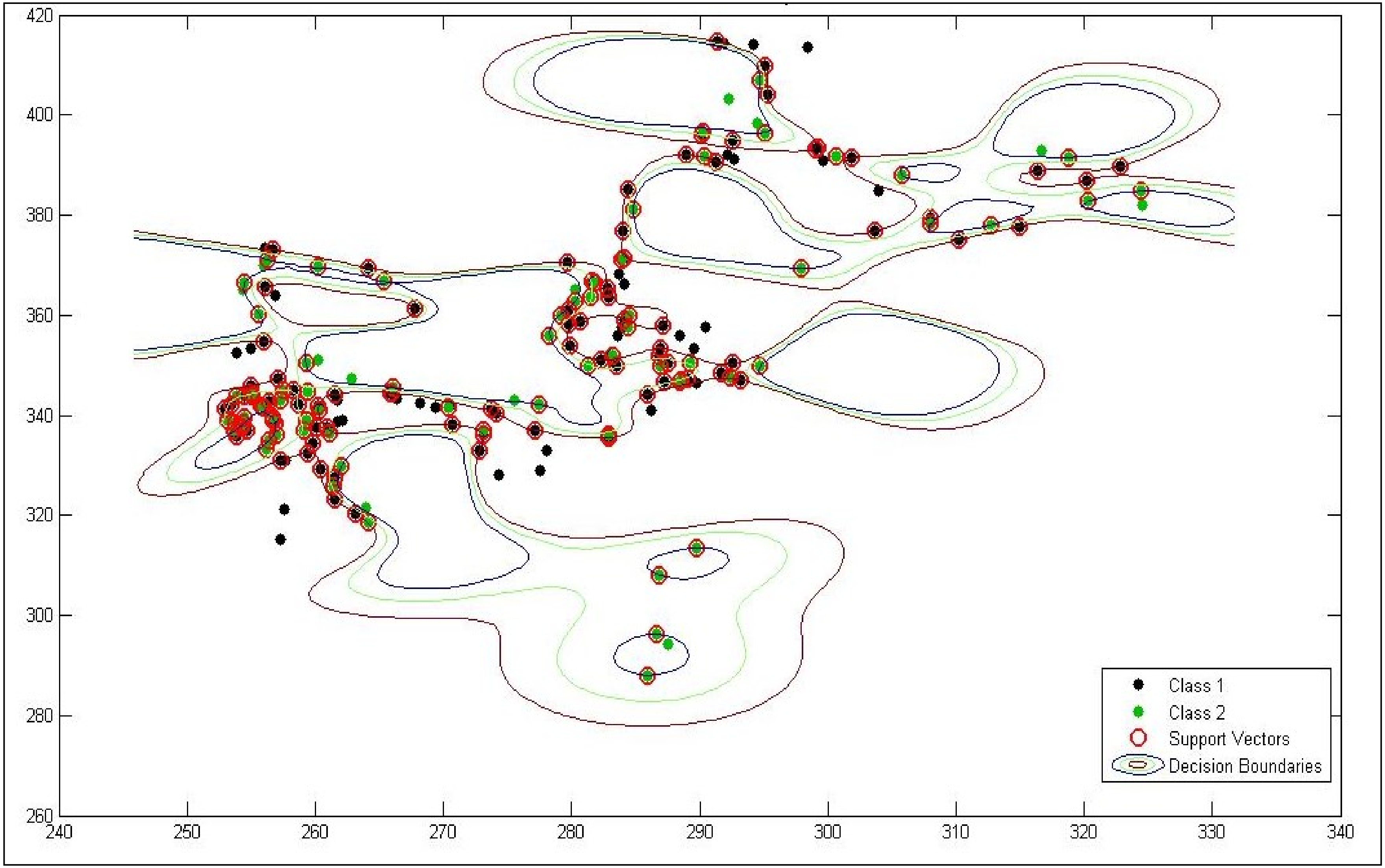
Classification: Support Vector Machines



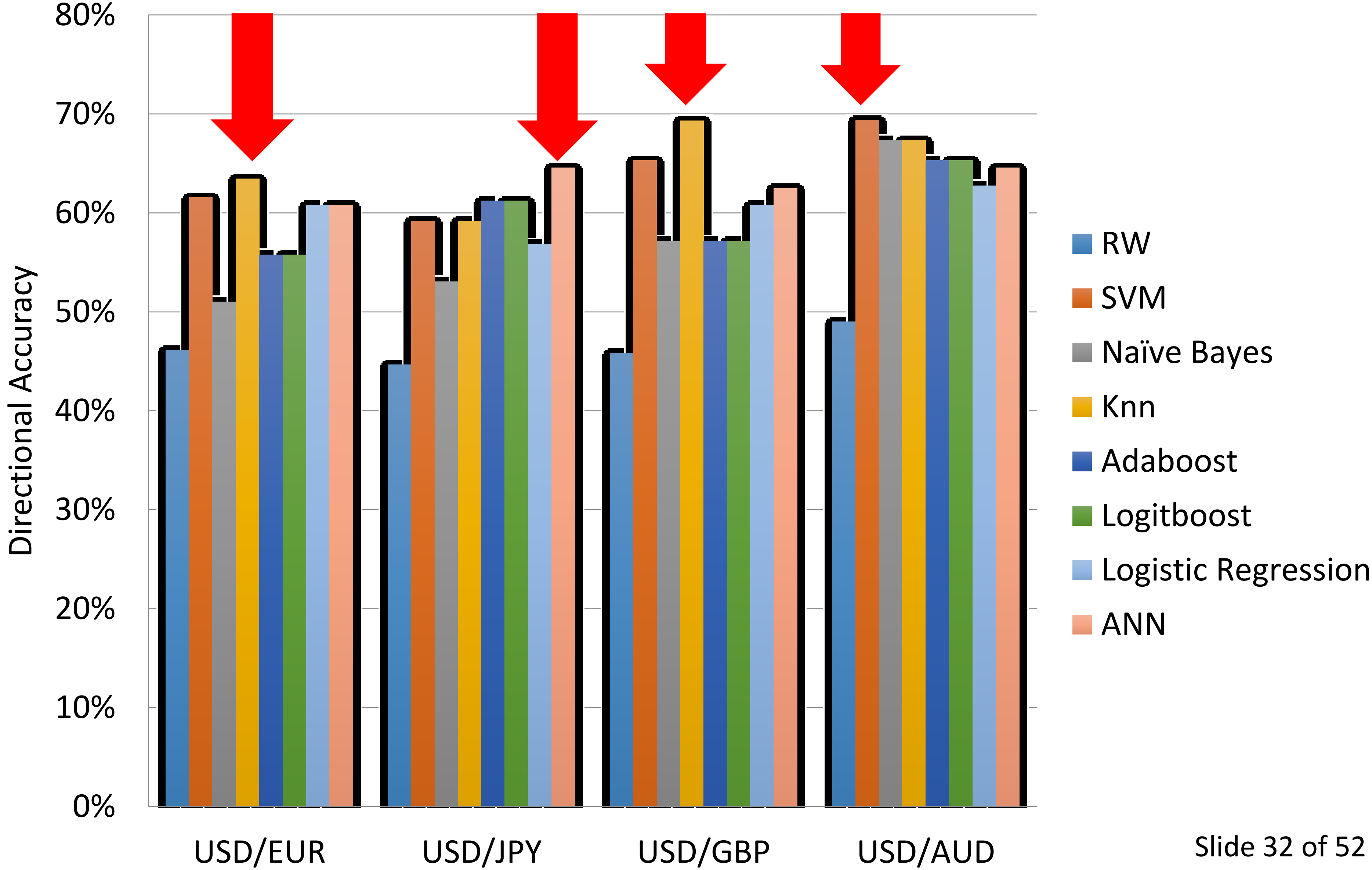
Projection to n+i dimensions



Support Vector Machines



Forecasting Exchange Rates
(Short term– microstructural approach - classification)



Conclusions

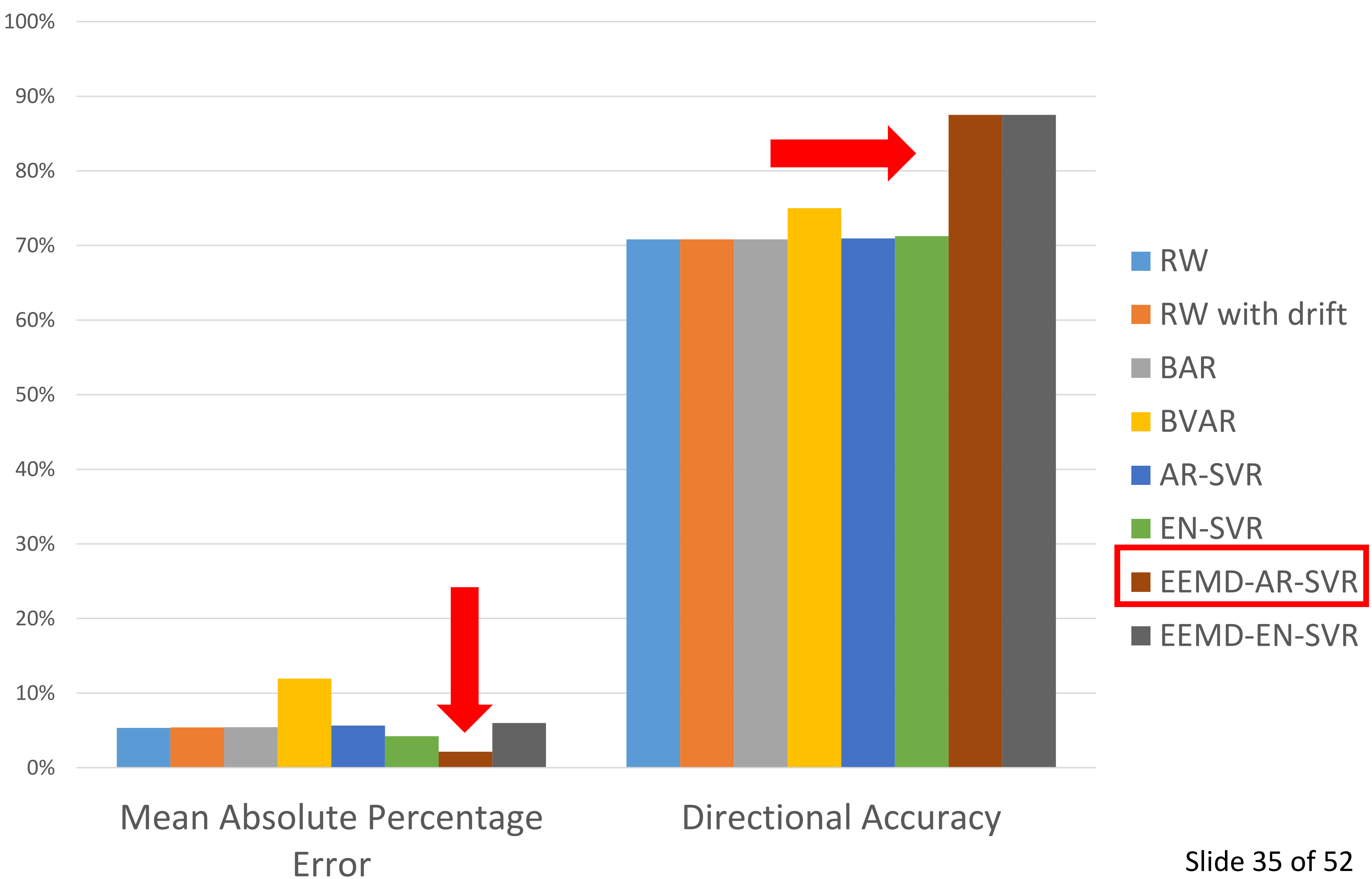
- The machine learning methodologies **outperform** the RW model
- **Machine Learning** techniques forecast more accurately that the econometric methodologies the out-of-sample direction
- **Market hype** expressed through the **volume** (**total number) of posts** improves the total forecasting ability

Application 3: Forecasting the Case & Schiller house price index

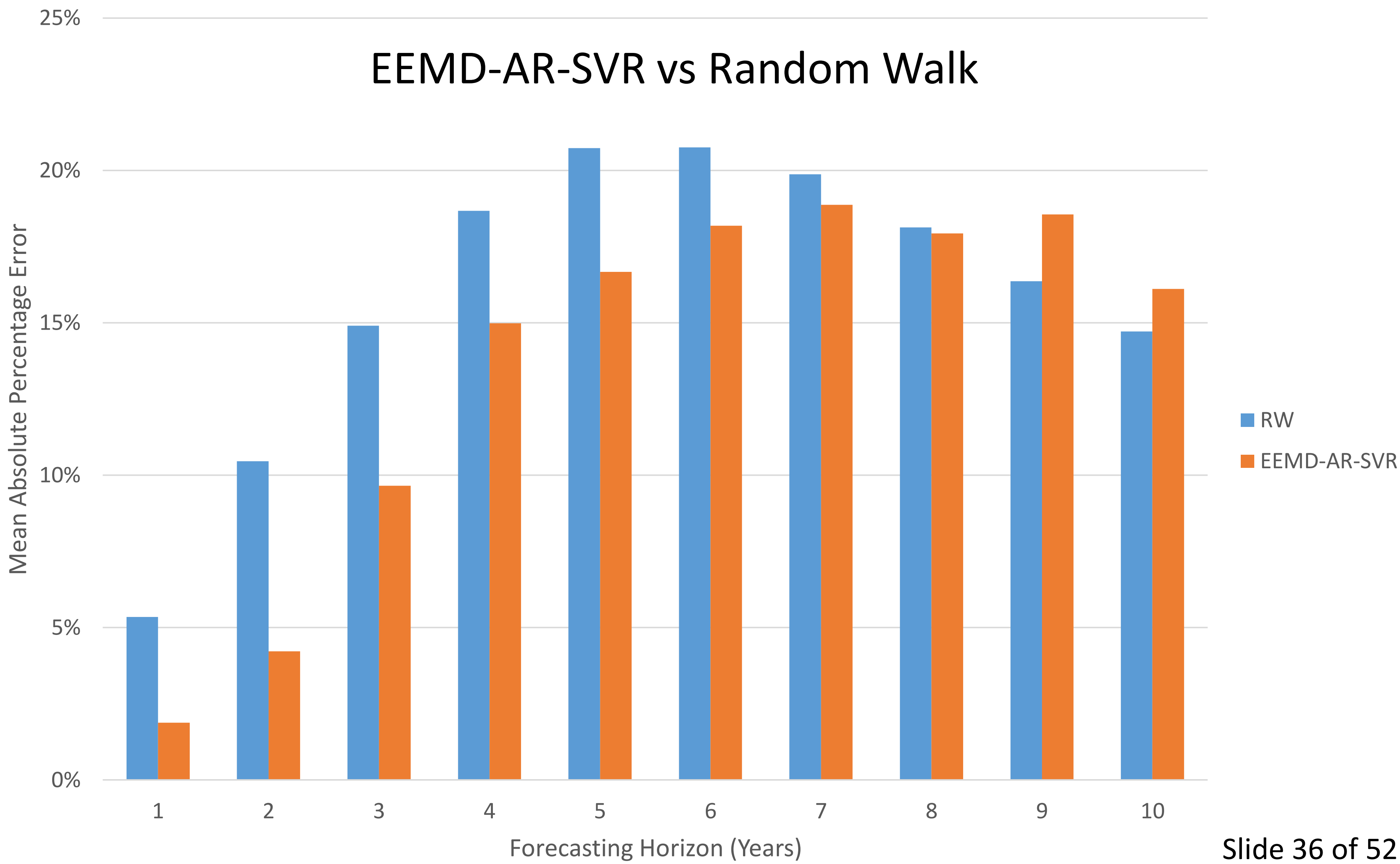
Economic Modelling, 2015, vol. 45, pp. 259-267.

Forecasts	Input Variables	Methodologies
• 1-10 years ahead • 1890-2012 • 80% - 20%	• Real GDP per capita • Long and short term interest rate • Population number • Real asset value • Real construction cost • Unemployment / Inflation • Real oil Prices • Fiscal Policy Indicator	• EEMD – Elastic Net – SVR • Bayesian AR/VAR

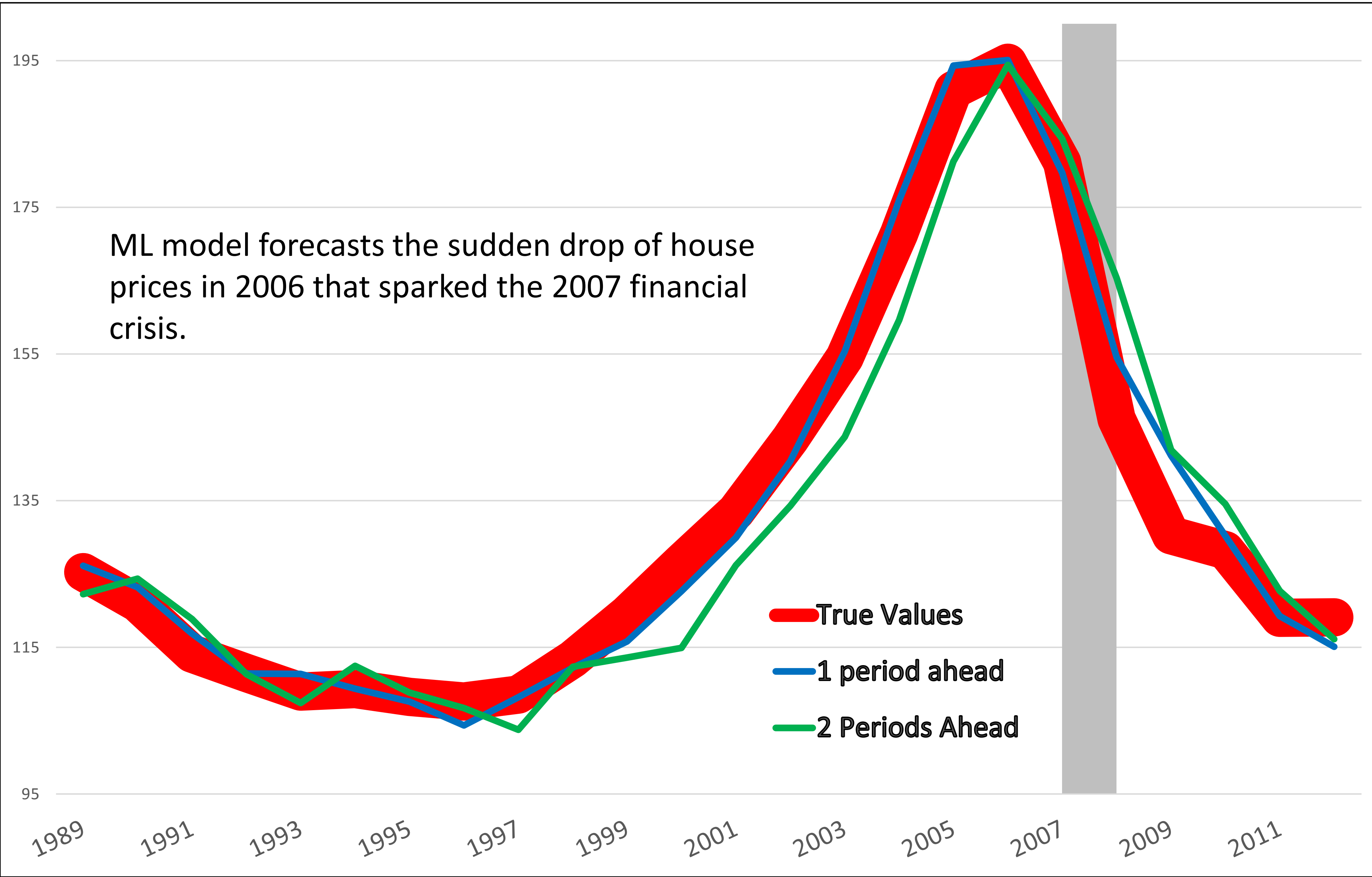
Optimum Model



Comparison in Alternative Forecasting Horizons



The EEMD-AR-SVR and the Collapse of the Housing Market – Actual vs Forecasted



Conclusions

- **EEMD-SVR** forecasts more accurately the evolution of house prices in out-of-sample forecasting
- The proposed model forecasts almost **2 years ahead** the actual 2006 sudden drop in house prices

Application 4: Forecasting bank failures and stress testing

- 1443 U.S. Banks
- 962 solvent
- 481 failed

Period 2003-2013

Data from FDIC (Federal Deposit Insurance Corporation)

The dependent variable

- Financial position of a bank (solvent or insolvent)

The independent variables

- 144 financial variables and ratios for each bank

Variable selection:

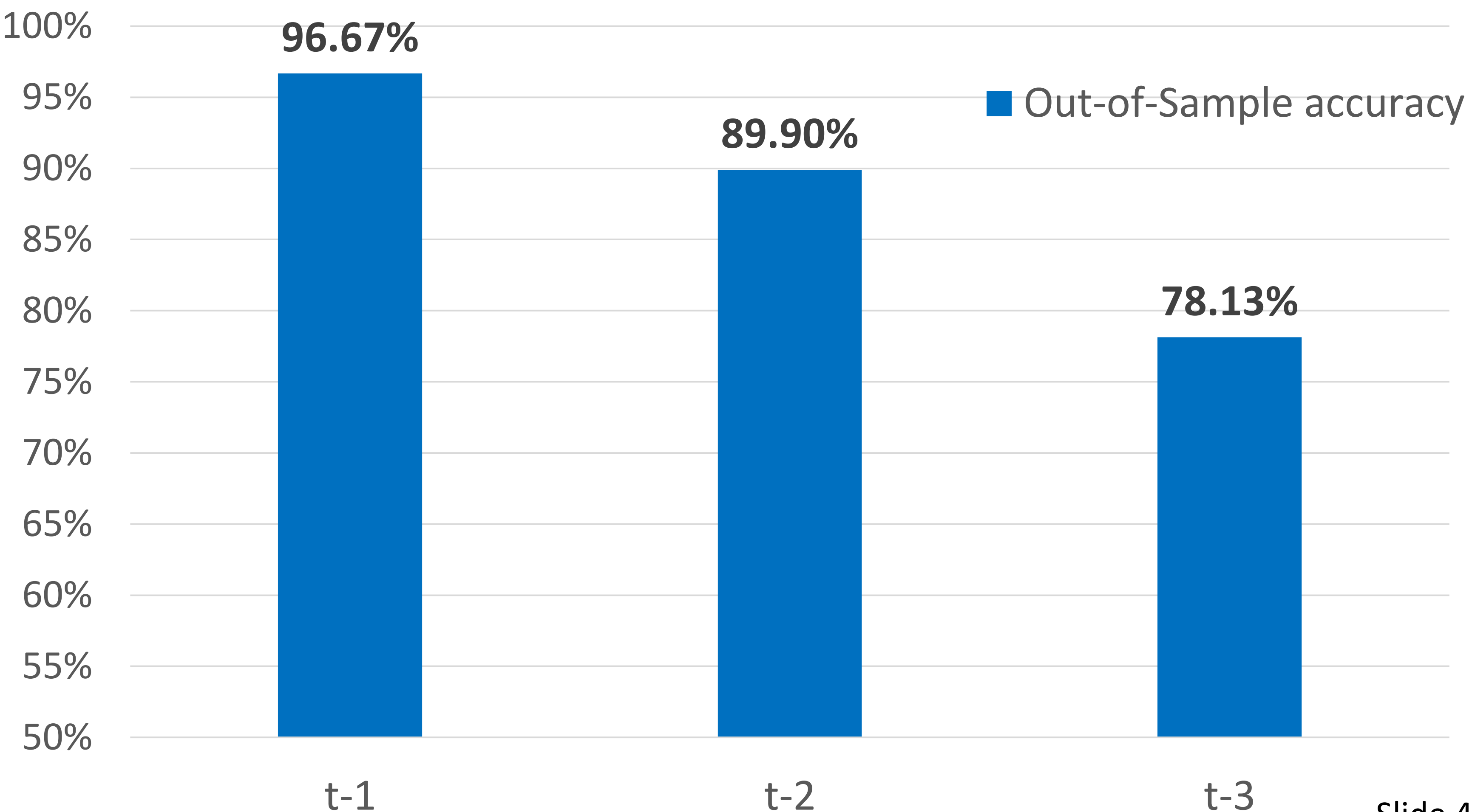
Local-learning for high-dimensional data (Sun et al. in 2010)

Selected variables:

- Tier 1 (core) risk-based capital/total assets year t-1
- Provision for loan and lease losses/total interest income year t-1
- Loan loss allowance/total assets year t-1
- Total interest expense/total interest income year t-1
- Equity capital to assets year t-1

Forecasting bank failures and stress testing

- 3 forecasting windows



Forecasting accuracy

Optimum model: Year t-1	**Real Solvent 228**	**Real Failed 115**
Predicted Solvent	223	3
Predicted Failed	5	112
Accuracy	97.39%	97.81%

Solvent Misclassified as failed 5

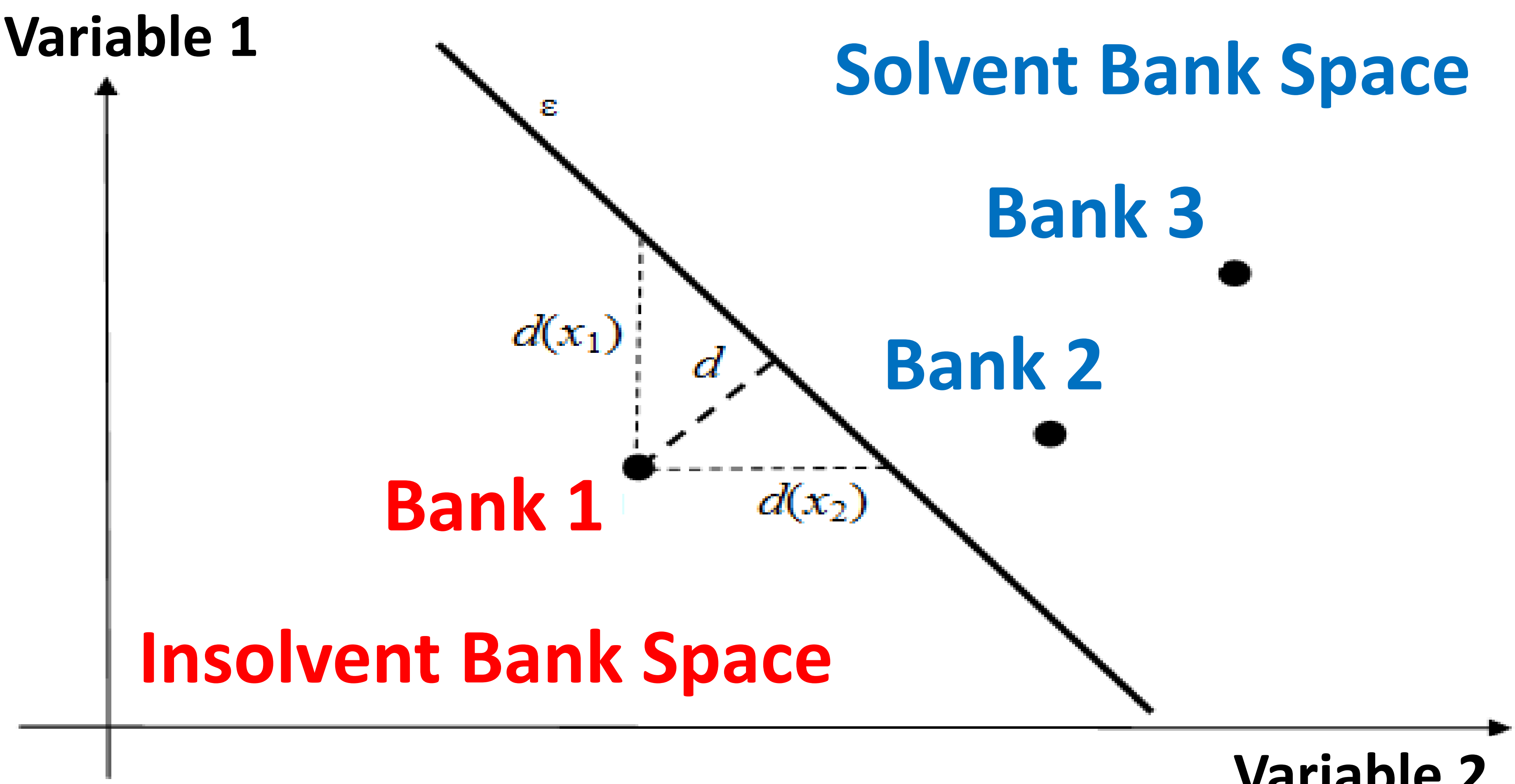
- 4 out of 5 received enforcement action from FDIC
- 1 received financial help from the U.S. Treasury as part of the Capital Purchase Program under the Troubled Asset Relief Program

Failed misclassified as solvent 3

- 1 discovered with unreported losses and was closed
- 2 small banks with total assets $44.5 & $26.3 million respectively

Solvency elasticity – Stress testing

- Banks are mapped on feature space
- Distance from the separating hyperplane provides the sensitivity of its state
- This property may be used as an auxiliary stress testing methodology



Application 5: Yield Curve and Recession Forecasting

- The behavior of the yield curve is associated with the business cycles.
- Indicator of future economic activity.

Application 5: The Yield Curve and EU Recession Forecasting

International Finance, vol. 18 (2), pp. 207-226

- **Forecasting**: GDP deviations from trend
- positive > Inflationary gaps
- negative > Unemployment gaps
- **Literature:** Pairs of interest rates, other variables

Application 5: The Yield Curve and EU Recession Forecasting

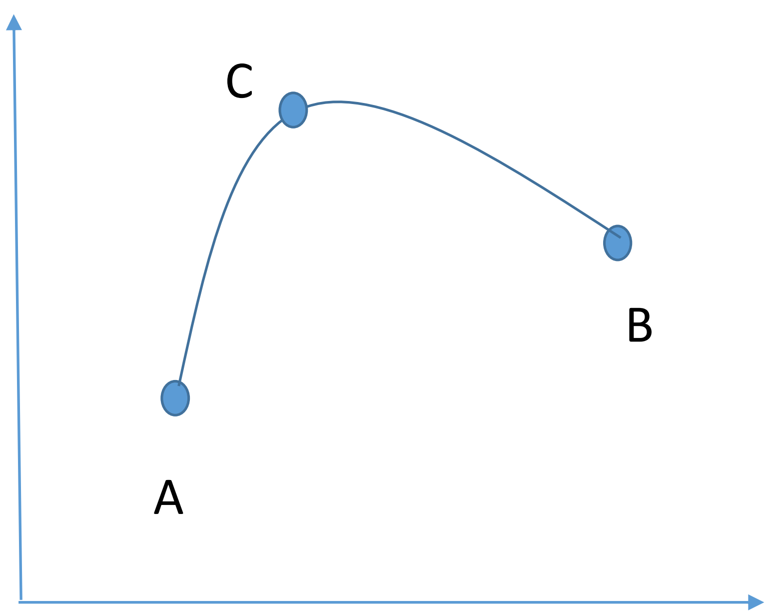
International Finance, vol. 18 (2), pp. 207-226

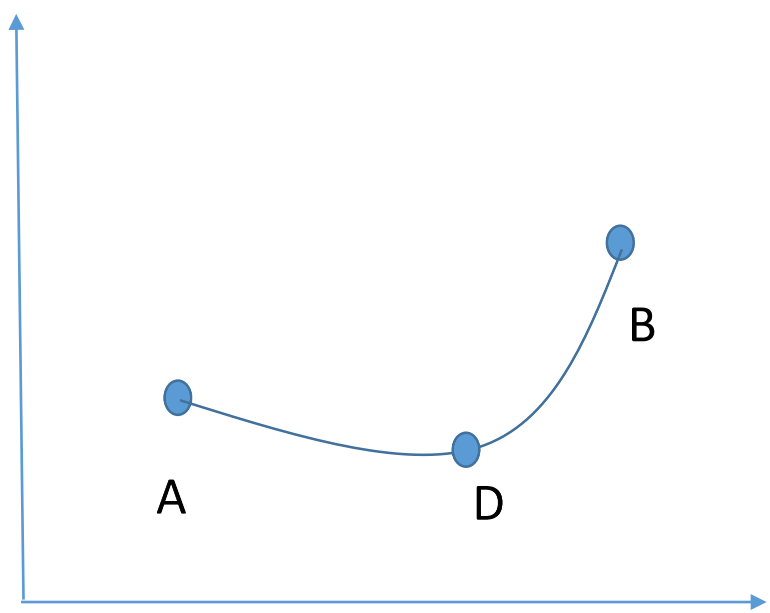
- **Data span**: September 2009 - October 2013
- **Methodologies**: SVM vs Probit vs ANN
- **Variables**: Eurocoin and interest rates: 3 & 6 months and 1, 2, 3, 5, 7, 10, 20 years.
- **We tested**:
  - 27 interest rates in pairs
  - 27 interest rates in triplets
  - All interest rates

Application 5: The Yield Curve and EU Recession Forecasting

International Finance, vol. 18 (2), pp. 207-226

- Motivation for interest rate triplets: exploit the curvature
- AB same slope
- ACB and ADB different curvature (concave vs convex)





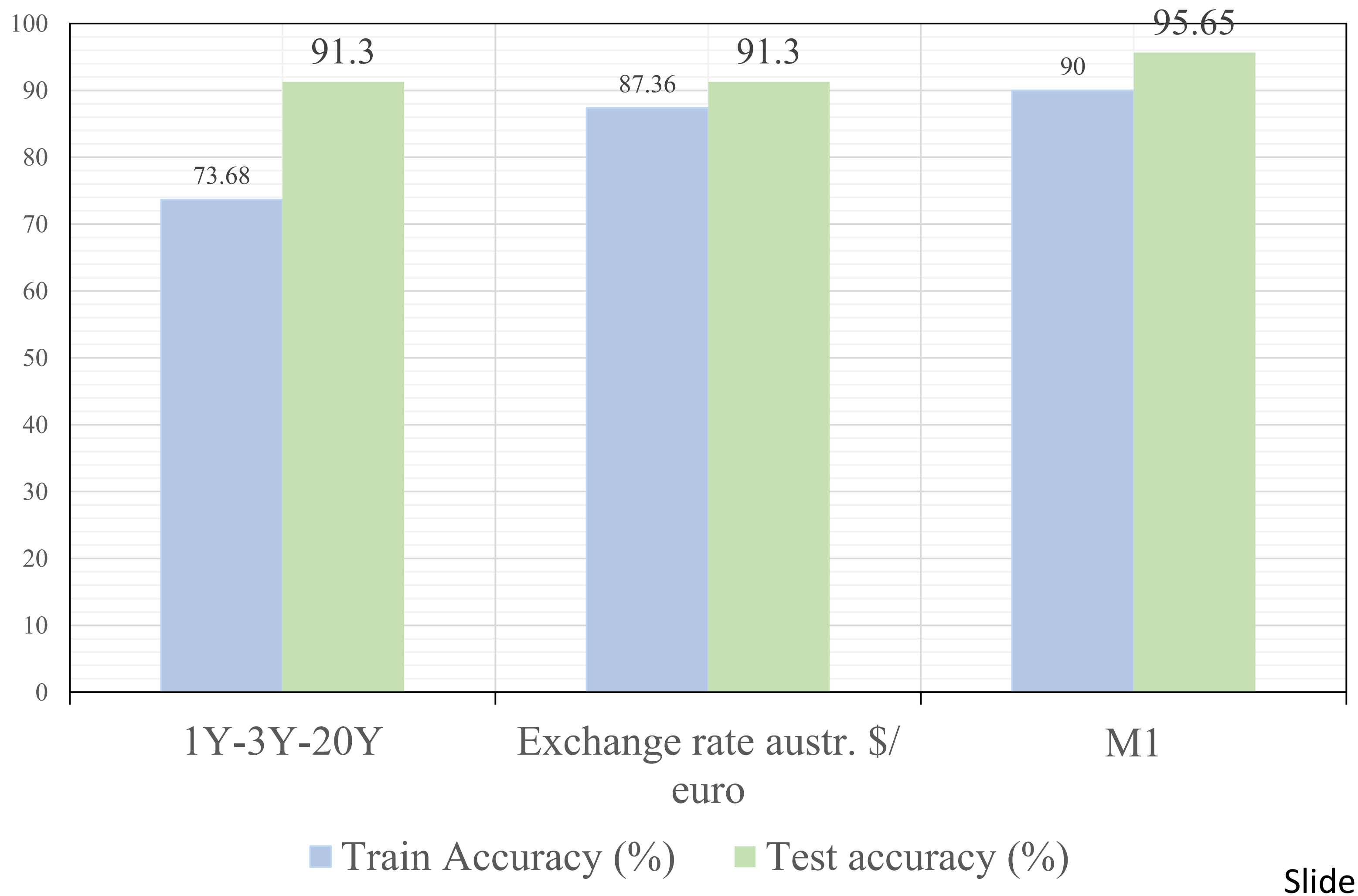
Forecasting EU recessions

	Methodology	Kernel	Combination	Train accuracy (%)	Test accuracy (%)	Growth accuracy (%)	Recession Accuracy (%)
Pairs	**Probit**		2Y-10Y	82,00	65,00	57,00	78,00
	SVM	**RBF**	2Y-20Y	73,68	78,26	93,00	56,00
	SVM	**RBF**	3Y-20Y	72,63	78,26	93,00	56,00
	SVM	**Polynomial**	3M-5Y	78,94	69,56	50,00	100,00
	SVM	**Polynomial**	1Y-2Y	77,89	69,56	50,00	100,00
	NN		3M-5Y	79,59	65,21	42,85	100,00
Triplets	**Probit**		1Y-5Y-10Y	74,00	61,00	43,00	89,00
	SVM	**RBF**	1Y-3Y-20Y	73,68	91,30	86,00	100,00
	SVM	**RBF**	3M-2Y-20Y	73,68	78,26	64,00	100,00
	SVM	**Polynomial**	1Y-3Y-7Y	74,73	65,22	50,00	89,00
	SVM	**Polynomial**	3M-5Y-20Y	71,57	65,22	50,00	89,00
	NN		6M-5Y-7Y	81,63	65,21	57,14	77,78
All interest rates	**Probit**			83,00	65,00	57,00	78,00
	SVM	**Linear**		58,94	39,13	0,00	100,00
	SVM	**RBF**		74,73	69,56	79,00	56,00
	SVM	**Polynomial**		74,73	56,52	50,00	67,00
	NN			75,51	69,56	78,57	55,56

Augmenting with Fundamentals

- Australian dollar/euro
- Canadian dollar/euro
- Czech corone-euro
- Chinese yuan-euro
- Dollar-euro
- CPI
- Oil price
- Unemployment
- Inflation
- M1
- M2
- M3
- Industrial Index
- PPI
- External liabilities
- External demand

Forecasting EU recessions



Thank you

Acknowledgments

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: **THALES**. Investing in knowledge society through the European Social Fund.

European Union
European Social Fund

OPERATIONAL PROGRAMME
EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS, CULTURE & SPORTS
MANAGING AUTHORITY

NSRF 2007-2013
programme for development
EUROPEAN SOCIAL FUND

Co-financed by Greece and the European Union