Nonlinear Forecasting with Many Predictors Using Kernel Ridge Regression

by Peter Exterkate, Patrick Groenen, Christian Heij and Dick van Dijk

Discussant: Francesco Ravazzolo (Norges Bank and BI Norwegian Business School)

Friday May 4, 2012

Summary

- Propose nonlinear forecasting by constructing **nonlinear factors** from many predictors, but keeping a **linear** relationship between dependent variables and factors.
- Application to simulation exercises and an empirical exercise (extended Stock and Watson (2002) database).
- The methodology seems fast and produces accurate **point** forecasts.

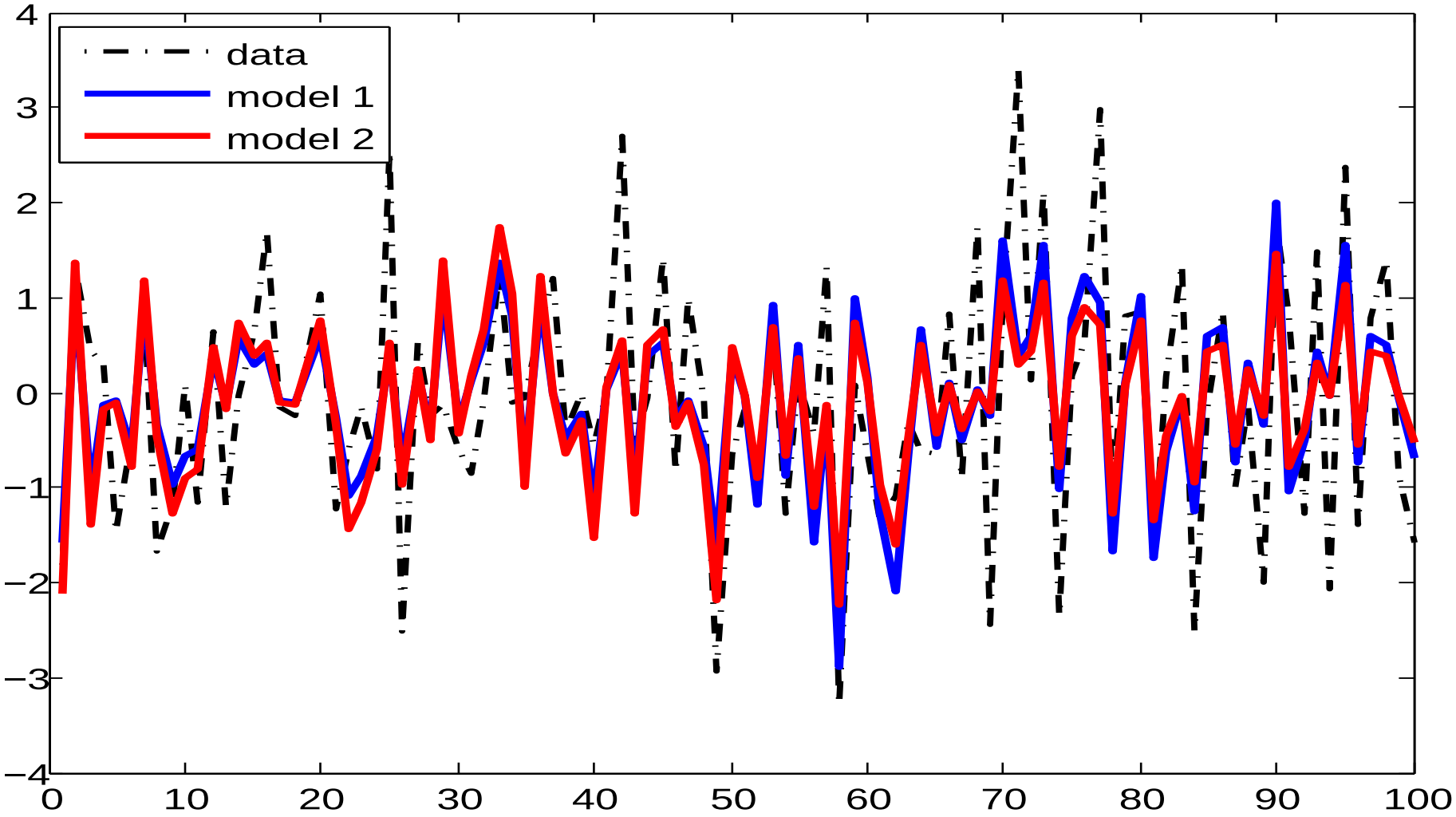
Nonlinearity

- Economists often thinks nonlinearity in the relationships $y = f(x)$ by assuming the function $f()$ to be nonlinear.
- Examples: nonlinear parameter β; smooth transition models.
- Time-varying parameters and stochastic volatility in dynamic factor models (Del Negro and Otrok (2008)).
- Authors follow a different approach and compute factors $z = \varphi(x)$ and then the assume the prediction is linear $\hat{y} = z'\hat{\gamma}$.

Example

- DGP:
  $y_t = \beta_t * x_t + e_t,\ x_t\ e_t\ N(0,1),\ t = 1, ..., 100,\ \beta^5_{s=1}0 = 0.5,\ \beta^1_{s=51}00 = 1.$
- Model 1:
  $y_t = \beta_a * x_t + \beta_b * (d_t * x_t) + e_t,\ \beta_1 = (\beta_a, \beta_b)'.$
- Model 2 (Bai and Ng (2008)):
  $y_t = \beta_a * x_t + \beta_b * x_t^2 + e_t,\ \beta_2 = (\beta_a, \beta_b)'.$
- Authors propose to use a kernel function $\kappa(a, b)$.

Example, in-sample fit



Interpretation

- Model 1: structural change in the coefficient β_1 (e.g., crisis).
- Model 2: the variable x causes y differently depending on the volatility of x.
- How can I interpret the kernel $\kappa(a, b)$?
  - More accurate summary of the information in x, which is useful for forecasting.
  - Can you interpret (a, b) as groups of variables with different volatility?
  - Can you interpret $z \propto \kappa(a, b)$? Check correlation to individual variables is not the best option to underline the nonlinearity.

Point forecasting

- The paper focuses on point forecast evaluation.
- Nonlinearity is probably even more relevant for higher moments (e.g. volatility point on previous slides).
- Apply the methodology to non-quadratic loss function (density forecasting), gains could be substantial.
- Maybe parameters (a, b) capture different moments.

Other minor issues

- Linear PCA performs the best for employment.
- Employment is often considered a highly-non linear series.
- Point forecasting not the right loss function? Or?

- Data is transformed to stationarity (and unit variance?). Would it be possible to work with rough data? Part of nonlinearity can be lost in the transformation.

- Tuning parameters (λ, σ) are selected via "leave-one-out cross-validation". Why not BIC as linear factor models?

Conclusion

- Nice paper, easy to read and to understand.
- The methodology produces accurate point forecasts.
- Interpretation of the factors.
- Full advantage of the flexibility of the methodology.