Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD) Gray Calhoun (Iowa State)

May 2, 2012

The Motivating Problem

(a) Typical comments about future of forecasting imply a large role for nonlinear models.
(b) Many studies show a limited role in providing forecasts from nonlinear models.
(c) Reviews of macroeconomic forecasting for variables of interest tend to ignore nonlinear models.

No real examination of why these models have not appeared to perform well in practice.

We step back and try to explain this 'lack' of results.

What we do

Seek to understand why standard models do not perform well.

We suggest a number of possibilities
(a) There is no evidence of nonlinearity in the data
(b) Common functional form choices are poor approximations to nonlinearity
(c) Functional forms are ok however gains too small to be interesting
(d) Estimation issues may overwhelm the methods
(e) Performance issues may be due to choice of evaluation period.

Outline of the Talk

1. Models and Previous Results
2. Results with updated data
3. Potential Explanations
4. Concluding Remarks

Models Examined

We examine models of the form

$$y_{t+1} = \phi_0 + \sum_{j=1}^{k} \phi_j y_{t-j+1} + \left(\theta_0 + \sum_{j=1}^{k} \theta_j y_{t-j+1} \right) G(s_t; \gamma, c) + \varepsilon_{t+1}$$

for various specifications of $G(s_t; \gamma, c)$ where s_t is an observed variable and $\{\phi_i, \theta_i\}_{j=0}^{k}$ and $\{\gamma, c\}$ are unknown parameters.
Choose:
- $s_t = y_t$ for all models
- k selected by BIC for linear model.
- estimation by concentrated ML, constrain the parameter spaces for nonlinear coefficients
- forms

LSTAR	$(1 + \exp(-\gamma(s_t - c)))^{-1}$
ESTAR	$1 - \exp\{-\gamma(s_t - c)^2\}$
SETAR	$1(s_t > c)$

Models Examined

Primary justifications for these models
(a) Take 'single' regime linear and make them 'two' regime so is a straightforward extension
(b) Nest linear models

Some issues:
(a) Endogeneity of nonlinearity has a big impact on the behavior of y_{t+1}.
(b) γ and c poorly identified for various ranges (if γ near zero or large).

Tests for nonlinearity

Why do nonlinear models provide poor macroeco-nomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

The tests often considered are LM tests on coefficients from regressions of the form

$$y_{t+1} = \delta_0 + \sum_{j=1}^{k} \delta_j y_{t-j+1} + \sum_{j=1}^{k} \beta_j y_{t-j+1}^2 + u_{1t}$$

or

$$y_{t+1} = \delta_0 + \sum_{j=1}^{k} \delta_j y_{t-j+1} + \sum_{j=1}^{k} \beta_{1j} y_{t-j+1}^2 + \sum_{j=1}^{k} \beta_{2j} y_{t-j+1}^3 + u_{2t}.$$

The test examines $H_0 : \beta_{ij} = 0$ for all i,j vs $H_a : \beta_{ij} \neq 0$ for some $\{i, j\}$. Nonlinearity is detected when these tests reject.

Approximate Nonlinear models

The suggested approach (van Dijk et. al (2002)) for modelling and forecasting with STAR type models is to pretest for nonlinearity, then on rejection use the nonlinear models.

Hence the basic approach is to check whether the squared and possibly cubed terms are useful in modelling y_{t+1} and if so discard them and use one of the earlier mentioned functional forms instead. Since many nonlinear forms each result in a similar LM test we suggest also examining the test statistic regressions directly as potential forecasting models.

Pros:
(a) We know if the test rejects that they are picking up something
(b) They might be more robust to various true models
Cons:
(a) They are not going to be stationary models.

Previous Empirical Results

Stock and Watson (2001). 215 monthly US macro variables from 1959 to 1996, apply many techniques including STAR and ANN models, find very little role for either nonlinear model.

Marcellino (2004), Similar study with Euro area macro variables, better evidence for STAR and ANN, best at short horizons. When dominated by linear models the nonlinear models are often very poor predictors.

Results with Updated Stock and Watson (2001) data

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Partially updated data set (115 of the 215 variables)
Updated through to October 2011.
Use exactly the same data transformations as original study.
Variables included include various measures of IP, Consumption, Captial utilization, Employment/Unemployment, Earnings, housing, Interest rates, Money Stock, PPI/CPI.
Estimate on data up to 2000, evaluate on the remainder.

Results with Updated Stock and Watson (2001) data

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Use BIC lag selected AR for the baseline. In terms of improvement over the baseline, we have MSE smaller for the following fractions of series in our dataset.

Method	Unfiltered	Filtered
LSTAR	0.20	0.33
SETAR	0.24	0.34
ESTAR	0.17	0.33

(Filter: If forecasted is more than 3 standard deviations from current outcome, use baseline forecast).

This is far better than SW results, redid for their sample and LSTAR (filtered) best for 7% of series.

Results with Updated Stock and Watson (2001) data

Why do nonlinear models provide poor macroeco-nomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

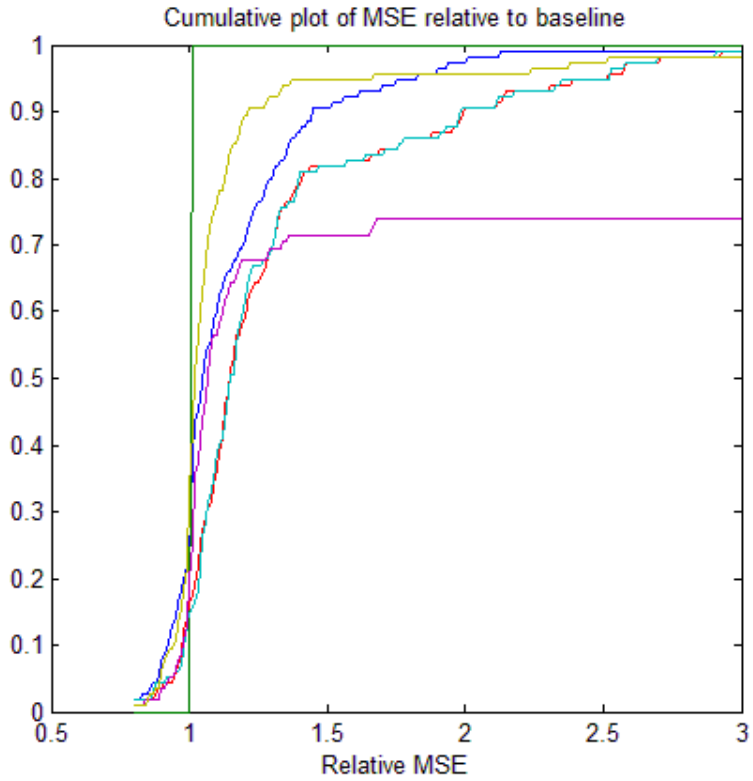
Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Figure: Performance of Forecast Methods



Notes: Green line: baseline model, yellow is the (insanity filtered) STAR, purple the unfiltered STAR. Blue is AR(1) model, red and aqua are the approximate nonlinear models

Results with Updated Stock and Watson (2001) data

Why do nonlinear models provide poor macroeco-nomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

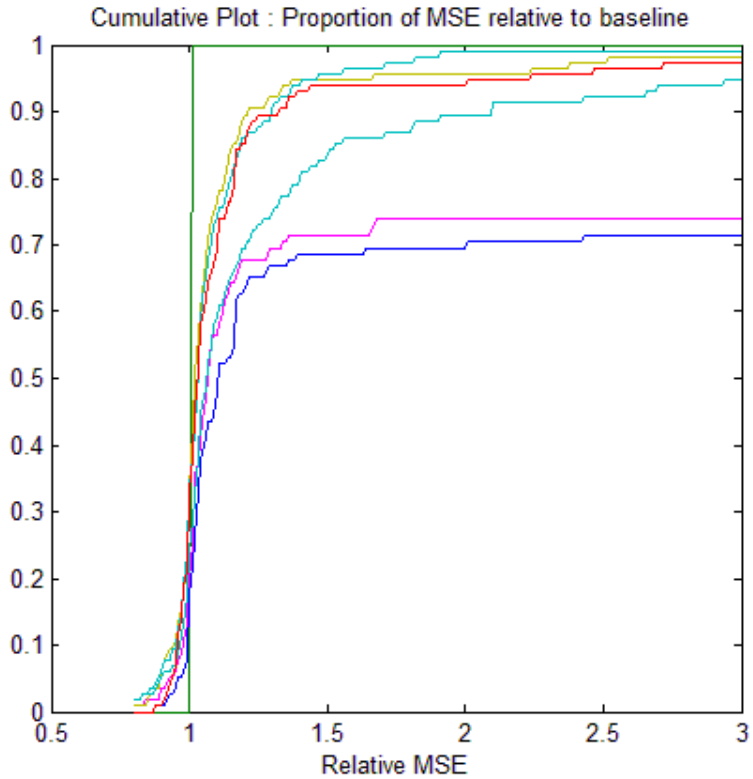
Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Figure: Performance of Forecast Methods



Notes: The green line: baseline model, yellow is the (insanity filtered) STAR, purple the unfiltered STAR. Black is the insanity filtered SETAR model and aqua the unfiltered SETAR blue for unfiltered

Potential Explanations

We suggested earlier five potential explanations:
(a) There is no evidence of nonlinearity in the data
(b) Common functional form choices are poor approximations to nonlinearity
(c) Functional forms are ok however gains too small to be interesting
(d) Estimation issues may overwhelm the methods
(e) Performance issues may be due to choice of evaluation period.

We will take each in turn.

Evidence for Nonlinearity

First, we can ask the question of whether or not tests for nonlinearity are likely to pick up much in the way of nonlinear behavior, i.e. how well do the tests work in practice.
Second, we can ask empirically if the tests are detecting nonlinear behavior in the data.

For the first, assuming an LSTAR model, we run some Monte Carlo experiments to examine the behavior of the tests (there are no analytical results available in the literature to provide guidance).
Design: one lag included although coefficients are zero, various γ with $c = 0$, vary the constant term.

Evidence for Nonlinearity

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

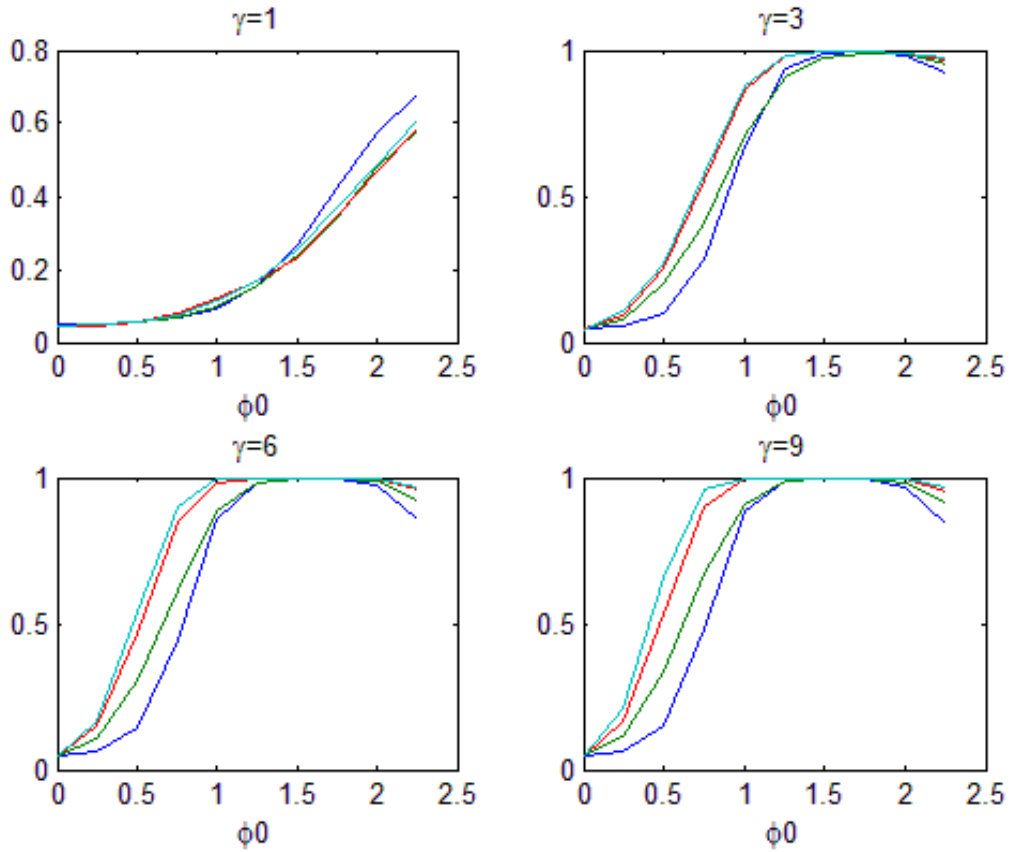
Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Figure: Power of Tests for nonlinearity



Notes: Dark blue: infeasible upper bound for the test power, light blue: is the upper bound for power (infeasible test). Green line is the

Evidence for Nonlinearity

From an empirical perspective, we can run these tests for each of the series in our database.

Table 1: Tests for Nonlinearity

	Full Dataset			Up to Dec 1996		
	OLS	OLS(W)	OLS(NW)	OLS	OLS(W)	OLS
Squared terms	47	20	27	32	17	22
Squared,cubed	61	45	52	52	42	43
App. logistic	56	41	43	49	32	39

Applied to the SW period tests reject for 32-52% of the series, so there is reasonable evidence there as well.

Functional Form

Even with nonlinearity present in half the dataset, it might well be that the particular functional forms typically chosen are poor approximations to the true dgps.
Empirically we can examine this through
(a) Examining if the nonlinear models are outperforming when we find evidence of nonlinearity in the data
(b) Examine if the approximate nonlinear models outperform the usual choices for functional form.

Functional Form

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Table 2: Nonlinear Model Forecasts and Nonlinearity Tests

	Rel. MSE<1	Rel. MSE<1 | Reject Nonlinearity
LSTAR	0.33	0.40
SETAR	0.34	0.39
ESTAR	0.33	0.33
Approx 1	0.17	0.19
Approx 2	0.15	0.17

Notes: Test for nonlinearity is a full sample test.

Functional Form

LSTAR for the most part better than the approximations methods, which beat the baseline less often but at the same time do not require use of the insanity filter to any great degree.

Gains from standard models too small?

It may be that standard models are well specified but are just incapable of leading to large gains in forecasting performance.

Theoretically, the gains may be small for reasonable models. Examine in dgp for one lag, T=600, vary constant. Various γ and two choices of AR parameter (alpha).

Empirically, gains might be small when the nonlinear models are better but losses may be big when nonlinear models perform worse than the baseline.

Gains from standard models too small?

Why do nonlinear models provide poor macroeco-nomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

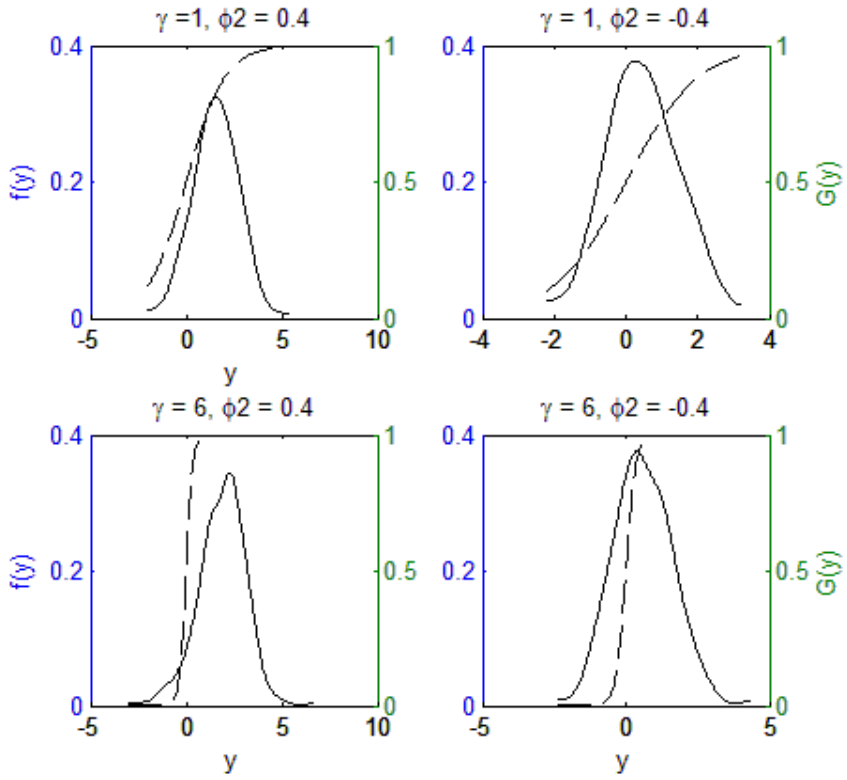
Models and Previous Results

Results with Updated Date

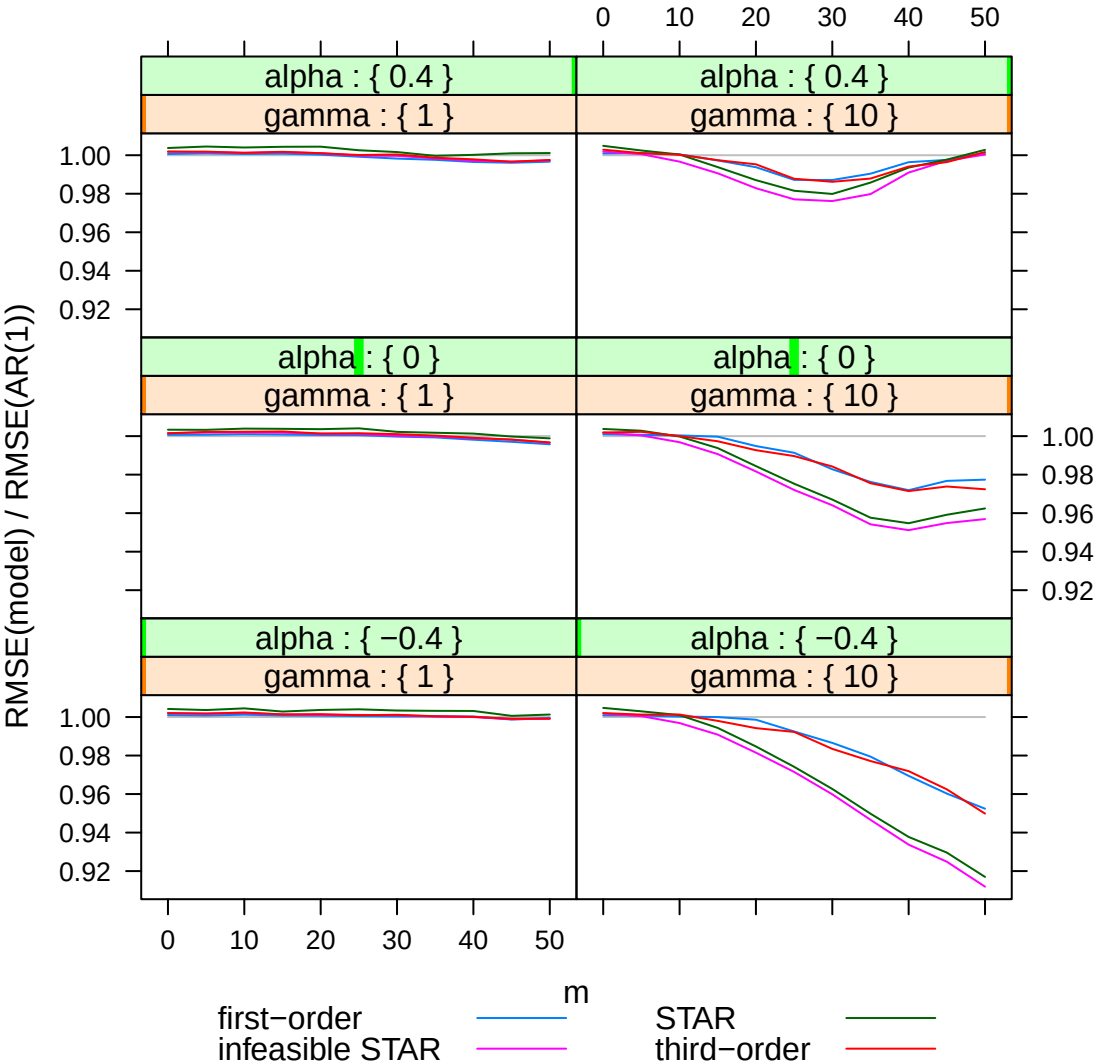
Potential Explanations

What explains these results?

Figure: Support of yt for LSTAR models and transition function



Notes: The solid lines give estimates of the density of y when $T = 600$ for various models (left hand scale). Dashed lines give


alpha : { 0.4 }
gamma : { 1 }
alpha : { 0.4 }
gamma : { 10 }
alpha : { 0 }
gamma : { 1 }
alpha : { 0 }
gamma : { 10 }
alpha : { −0.4 }
gamma : { 1 }
alpha : { −0.4 }
gamma : { 10 }
RMSE(model) / RMSE(AR(1))
m
first−order
infeasible STAR
STAR
third−order

Gains from standard models too small?

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

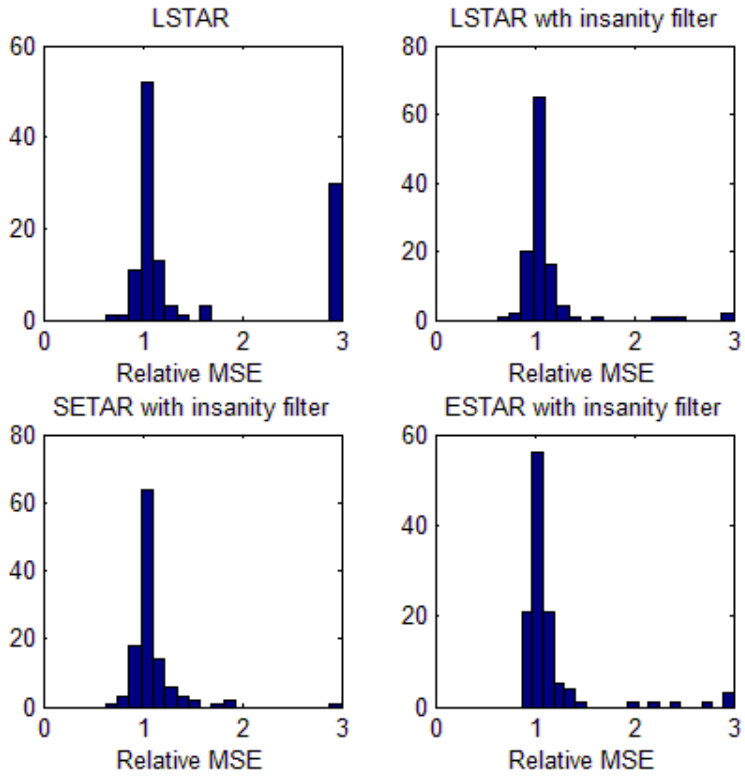
Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Figure: Relative Performance of STAR models



Notes: The histograms show the average MSE relative to the baseline autoregressive model (top coded to 3).

Estimation Issues

A major reason for expecting that nonlinear models might perform poorly is estimation issues, that the functional form is correct but the estimates so imprecise as to render them unuseful for forecasting. For example over the SW dataset the results for the nonlinear models are far inferior (7% for the unfiltered method, 16% for the filtered).

To examine this we re-estimate the main results using rolling regressions (which fix the number of datapoints for estimation) and check to see if the results are different. The rolling regressions use 493 months to estimate the model. The STAR model with the insanity filter outperforms the baseline for 31% of the series (down from 33% for the full sample models).

Period of Evaluation

The period of evaluation might affect the relative performance of linear and nonlinear models. It may well be that the forecasting performance is unstable.

To evaluate this, we compute the proportion of series for which nonlinear models (here STAR) outperform the baseline in each time period. We can then see how this varies over time.

We also compute the averages above over subsamples relating to the pre and post financial crisis periods.

Period of Evaluation

Why do nonlinear models provide poor macroeco-nomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

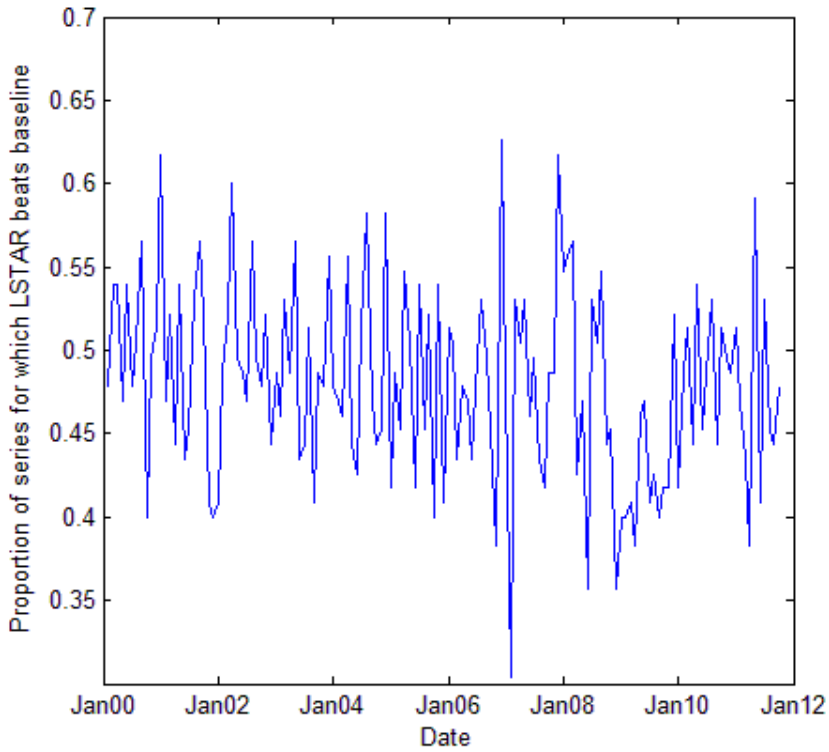
Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations

Figure: Proportion of Series for which LSTAR beats baseline



Notes: At each date we report the average number of series for which LSTAR outperforms the baseline.

Period of Evaluation

In terms of average losses over subperiods for the various models, we have

Table 3: Outperformance of Nonlinear models relative to Baseline.

Period	LSTAR	SETAR
2000:1-2007:6	0.42	0.38
2007:7-2011-11	0.37	0.36
Full period	0.33	0.34

Notes: Reports number of series for which average loss is less than the baseline model.

Summary.

We suggested earlier five potential explanations:
(a) There is no evidence of nonlinearity in the data
(b) Common functional form choices are poor approximations to nonlinearity
(c) Functional forms are ok however gains too small to be interesting
(d) Estimation issues may overwhelm the methods
(e) Performance issues may be due to choice of evaluation period.

Summary

(1) There is extensive evidence of nonlinearity in the data.
(2) Standard functional forms pick up much fewer series, and more often than not series for which the tests fail to reject.
(3) Approximate nonlinear models perform worse than popular models, counter to MC results.
(4) Theoretically unlikely that typical parametric models do outperform baseline models much, because of restrictions on their form and the features of the data generated by reasonable models.
(5) Standard parametric nonlinear models outperform baseline by small amounts in the data, but can be very poor.
(6) Estimation error appears small at reasonable sample sizes.
(7) Some sample evaluation effects, but these appear small.

Conclusion

Since there is ample evidence of nonlinearity, yet this is unrelated to when the methods outperform, and also that the approximate models are outperformed by the parametric nonlinear models (in MC they are very similar), it would seem that a large part of the story is that the functional forms are indeed poor at capturing the types of nonlinearity in the data. It also seems unlikely that even if these functional forms were good approximations to the dgp, that they would be capable of providing more than a minor improvement in forecasting performance. This is because of the endogeneous nature of the nonlinearity. This was borne out in the data.

Caveats

- hetereoskedasticity
- statistical significance of the differences.
- choices in forms of the nonlinear models.

Nonlinearity over different periods

Proportion of series for which we reject for indicated periods.

Period	Squared	Cubed	App
60's	0.23	0.36	0.36
70's	0.17	0.46	0.39
80's	0.12	0.43	0.36
90's	0.15	0.32	0.23
00's	0.35	0.58	0.35
Full	0.28	0.52	0.42
SW	0.22	0.43	0.39

Set s to be IP

If we set s_t to be Industrial production, we break somewhat the effect this has on the distribution of y_t for each of the models. Doing so results in

(a) Slightly better results (39% of the eseries beat the baseline)

(b) still essentially independent of the tests for nonlinearity

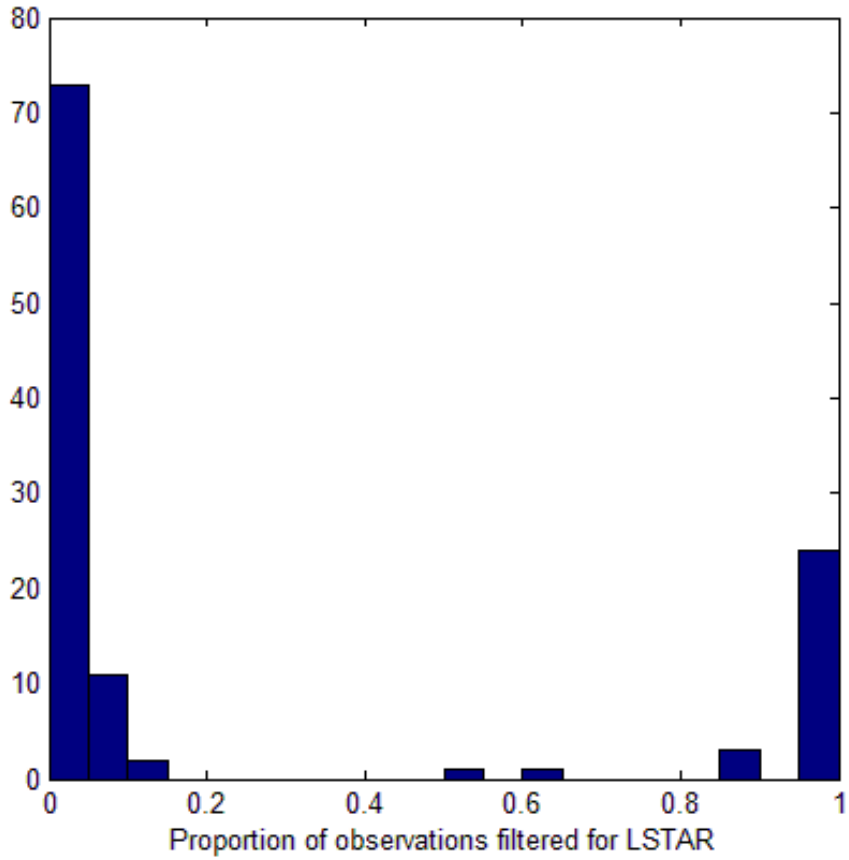
LSTAR and filtering

Why do nonlinear models provide poor macroeconomic forecasts?

Graham Elliott (UCSD), Gray Calhoun (Iowa State)

Motivating Problem

Models and Previous Results

Results with Updated Date

Potential Explanations