What univariate models tell us about multivariate macroeconomic models

James Mitchell (Warwick), Donald Robertson (Cambridge) and Stephen Wright (Birkbeck)

NBP Workshop on Forecasting 2016 (Warsaw, Poland)

21 November 2016

Puzzle 1: The Predictive Puzzle

- Multivariate macro models often fail to outpredict univariate models
  - Evidence from Nelson (1972, AER) through to Stock & Watson (2015)...
- AR(MA) is a tough benchmark to beat
  - for inflation, exchange rates, stock prices, consumption...
  - it's challenging to find predictor variables that persistently predict

Puzzle 2: The Order Puzzle

- The univariate models that compete with multivariate models in predictive terms tend to be of low order
- But the multivariate models commonly have large numbers of state variables
  - which should imply high-order ARMA reduced forms (see Wallis, 1977 *Econometrica*; Cubbada *et al.*, 2009 *Journal of Econometrics*)

Our Aims

- The paper offers an explanation for both Puzzles - in terms of population properties
- we work backwards, observing just the univariate history of y_t, data for which are assumed to be generated by an (unknown) multivariate macroeconomic model and show that
- univariate properties can tightly constrain the properties of the (true but unknown) multivariate macro model

Our key results - The predictive puzzle

1. We show how much better we could forecast y_t if we conditioned on the true state variables of the underlying multivariate model, rather than just its history, y^t
   - The one-step-ahead predictive R^2 must lie between bounds, $R^2_{\min}$ and $R^2_{\max}$, that are strictly within $[0, 1]$
     - bounds derive from the ARMA parameters based on y^t
     - results generalise to time-varying parameter models
   - the gap between $R^2_{\min}$ and $R^2_{\max}$ can be narrow for some y_t processes
     - searching for reliable predictors is then a thankless task

Our key results - The order puzzle

2. If y_t is ARMA(p, q) in population then (absent tight restrictions) the # unique predictors from the macro model, $r = q$

   - In finite samples, p and q may never be knowable
   - So to explain the order puzzle
   - either, the true multivariate model does have many state variables, but it's close to satisfying parameter restrictions that imply (near) cancellation of AR and MA roots in the reduced form. Why?
   - or, there are only a few distinct eigenvalues driving the macroeconomy - as in factor models

The true multivariate model

- Let y_t be generated by a multivariate macroeconomic model in the n-vector of states $\mathbf{z}_t$

$$\begin{aligned} \mathbf{z}_t &= \mathbf{A}\mathbf{z}_{t-1} + \mathbf{B}\mathbf{s}_t \\ \mathbf{y}_t &= \mathbf{C}\mathbf{z}_{t-1} + \mathbf{D}\mathbf{s}_t \end{aligned}$$

- Under (weak, and standard) Assumptions, ABCD implies a true predictive regression for the first element of $\mathbf{y}_t$, where $r \leq n$ due to (possible) linear dependence between the states

$$\begin{aligned} y_t &= \beta' \mathbf{x}_{t-1} + u_t \\ \underset{(r\times 1)}{\mathbf{x}_t} &= \mathbf{M}\mathbf{x}_{t-1} + \mathbf{v}_t \end{aligned}$$

where $\mathbf{M} = diag(\mu_1, ..., \mu_r)$ contains the distinct eigenvalues of $\mathbf{A}$

Assumptions

A1 *The autoregressive matrix* $\mathbf{A}$ *of the state variables* $\mathbf{z}_t$ *can be diagonalised as* $\mathbf{A} = \mathbf{T}^{-1}\mathbf{M}^*\mathbf{T}$ *where* $\mathbf{M}^*$ *is an* $n \times n$ *diagonal matrix, with first* r *diagonal elements* $\mathbf{M}^*_{ii} = \mu_i$, $i = 1, ..., r$, $r \leq n$ *being the distinct eigenvalues of* $\mathbf{A}$

A2 $|\mu_i| < 1$, $i = 1, ..., r$

A3 $\mathbf{s}_t$ *is an* $s \times 1$ *vector of mutually orthogonal Gaussian IID processes with* $E(\mathbf{s}_t\mathbf{s}_t') = \mathbf{I}_s$

A4 $\mathbf{DD}'$ *has non-zero diagonal elements.*

Lemma

(The Macroeconomist's ARMA) *The true predictive regression and the process for the associated true predictor vector* $\mathbf{x}_{t-1}$ *together imply that* y_t *has a unique fundamental ARMA*(r, r) *representation with parameters* $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_r)$ *and* $\boldsymbol{\theta} = (\theta_1, ..., \theta_r)$

$$\lambda(L)\, y_t = \theta(L)\, \varepsilon_t$$

where $\lambda(L) \equiv \prod_{i=1}^{r}(1 - \lambda_i L) \equiv \det(I - \mathbf{M}L)$ *and* $\theta(L) \equiv \prod_{i=1}^{r}(1 - \theta_i L)$, $|\theta_i| \leq 1$, $\forall i$

The macroeconomist's ARMA - a simple illustration

- Consider an ABCD model with a single state variable z_t and a 2×1 vector of structural shocks. This implies the predictive system

$$
\begin{aligned}
y_t &= \beta x_{t-1} + u_t \\
x_t &= \mu x_{t-1} + v_t
\end{aligned}
$$

  with $\mathbf{A} = \mu$, $\mathbf{C} = \beta$, $v_t = \mathbf{Bs}_t$, $u_t = \mathbf{Ds}_t$

- This system is common in predictive return regressions
  - with y_t some measure of returns or excess returns and x_t some stationary valuation criterion

The macroeconomist's ARMA(1,1) - a simple illustration

- By substitution

$$(1-\mu L)\, y_t = \beta v_{t-1} + (1-\mu L)\, u_t$$

- i.e. y_t admits a *fundamental* ARMA(1,1) representation

$$(1-\mu L)\, y_t = (1-\psi L)\, \varepsilon_t$$

with $|\psi| < 1$, where $\psi = \psi\left(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\right)$ is the solution to

$$\frac{-\psi}{1+\psi^2} = \frac{\mu\sigma_u^2 - \beta\sigma_{uv}}{\left(1+\mu^2\right)\sigma_u^2 + \beta^2\sigma_v^2 + 2\mu\beta\sigma_{uv}}$$

- ψ^{-1} is also a solution
  - but this results in a *nonfundamental* representation, with a different IID shock process, which cannot be recovered from the history y^t (we revert to this representation in due course)

What the ARMA tells us about the true predictive regression

- We now work backwards, and ask: if we only observed the history of y_t, what would its univariate properties (as captured by $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$) tell us about the underlying multivariate model?
  - They do not, in general, tell us about the structural shocks $\mathbf{s}_t$
    - $s_t \propto \varepsilon_t$ only when $n = 1$ and the ARMA is fundamental
- But the ARMA is informative about **some** properties of the underlying multivariate model
  1. The predictive power of the multivariate model, expressed in terms of its one-step-ahead predictive R^2, must lie between bounds defined w.r.t. the ARMA parameters

Bounds for the predictive R-sq

Under A1-A4, the predictive $R^2(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = 1 - \sigma_u^2/\sigma_y^2$ of the regression derived from the underlying ABCD macro model

$$0 < R^2_{\text{min}} \leq R^2 \leq R^2_{\text{max}} < 1$$

where R^2_{min} is the predictive R^2 from the ARMA representation and so is a function of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ alone, and

$$R^2_{\text{max}} = R^2_{\text{min}} + \left(1 - R^2_{\text{min}}\right)\left(1 - \prod_{i=1}^{r} \theta_i^2\right)$$

so R^2_{max} is also a function of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ alone

The fundamental ARMA(1,1)

$$(1-\lambda L)\,y_t = (1-\theta L)\,\varepsilon_t$$

- λ and θ are less than unity in modulus
- ε_t are then derived from the history of y_t

$$\varepsilon_t = \left(\frac{1-\lambda L}{1-\theta L}\right) y_t = \sum_{i=0}^{\infty} \theta^i \left(y_{t-i} - \lambda y_{t-1-i}\right)$$

and

$$R^2_{\min} = R^2_F = 1 - \sigma^2_\varepsilon / \sigma^2_y$$

The non-fundamental ARMA(1,1)

$$y_t = \left(\frac{1 - \theta^{-1} L}{1 - \lambda L} \right) \eta_t$$

Now the (structural) shocks *cannnot* be derived from the history of y_t. Instead

$$\begin{aligned} \eta_t &= \left(\frac{1 - \lambda L}{1 - \theta^{-1} L} \right) y_t = -\theta F \left(\frac{1 - \lambda L}{1 - \theta F} \right) y_t \\ &= -\sum_{i=1}^{\infty} \theta^i \left[y_{t+i} - \lambda y_{t+i-1} \right] \end{aligned}$$

where $F = L^{-1}$

- η_t is a linear combination of current and future values of y_t

Bounds for the predictive R-sq: ARMA(1,1)

$$R^2_{\min} = R^2_F = \frac{(\theta-\lambda)^2}{1-\lambda^2+(\theta-\lambda)^2}$$

$$R_{\max} = R^2_{\min} + \left(1-R^2_{\min}\right)\left(1-\theta^2\right) = \frac{(1-\lambda\theta)^2}{1-\lambda^2+(\theta-\lambda)^2}$$

- If θ is close to λ, y_t is close to white-noise, and $R^2_{\min}$ is close to 0
- If θ is close to 0, $R^2_{\max}$ is close to 1
- Iff θ and λ are *both* sufficiently close to 0 (y_t and x_t are white-noise) the bound opens up to $[0,1]$
- As $|\theta| \to 1$, both $R^2_{\min}$ and $R^2_{\max} \to \frac{1-sgn(\theta)\lambda}{2}$: i.e. the R^2 bound collapses to a single point

Intuition

- Lower bound: the predictions from the fundamental ARMA condition only on the history of y_t
  - they cannot be worsened by increasing the information set to include the true predictor
- Upper bound: η_t has minimum innovation variance
  - η_t are a linear combination of the history and the *future* of y_t
- While we can improve our prediction of y_t by using the true state variable, there is a limit to this improvement
  - and this limit is determined by the univariate properties of y_t

What the ARMA tells us about the dimensions of the true multivariate model

Proposition

(ARMA order and the ABCD Model) *Let y_t admit a minimal ARMA(p, q) representation in population, with*

$$q = r - \#\{\theta_i = 0\} - \#\{\theta_i = \lambda_i \neq 0\}$$
$$p = r - \#\{\lambda_i = 0\} - \#\{\theta_i = \lambda_i \neq 0\}$$

Under A1 to A3, the data y_t must have been generated by a multivariate model in which, in the absence of exact restrictions over the parameters $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, $\mathbf{A}$ has $r = q$ distinct eigenvalues. Thus there must be a true predictive system for y_t with $r = q$ predictors

What the ARMA tells us about the dimensions of the true multivariate model

Corollary

The predictor vector $\mathbf{x}_t$ *in the true predictive regression can only have* $r > q$ *elements if the parameters of the structural model* $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ *satisfy* $r - q$ *exact restrictions that each ensure* $\theta_i (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = 0$ *for some* i*, or* $\theta_j (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \lambda_k (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$*, for some* j *and* k*.*

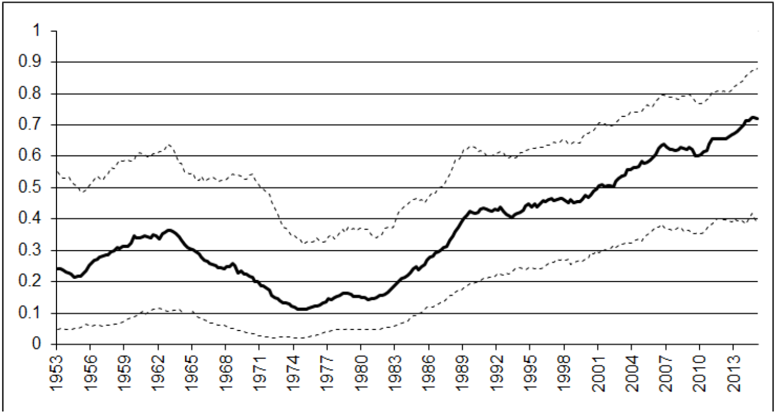
Stock & Watson's (2007) model of US inflation

- Find little evidence that it's possible to out-forecast a univariate model, particularly in recent data
- Their preferred model lets the variances of the permanent and transitory innovations vary over time: a UC-SV model
- Their UC-SV is equivalent to a time-varying IMA(1,1) model of inflation, $y_t = \Delta\pi_t$

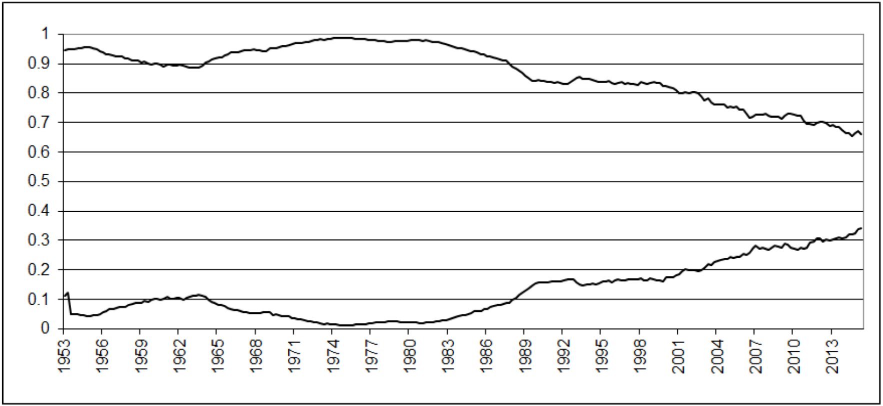
  $$y_t = (1 - \theta_t L)\,\varepsilon_t$$

  where $\lambda = 0$
- θ_t thus determines time-varying R^2 predictive bounds
  - $\lambda = 0$ tells us that any good predictor (in composite) must be close to IID

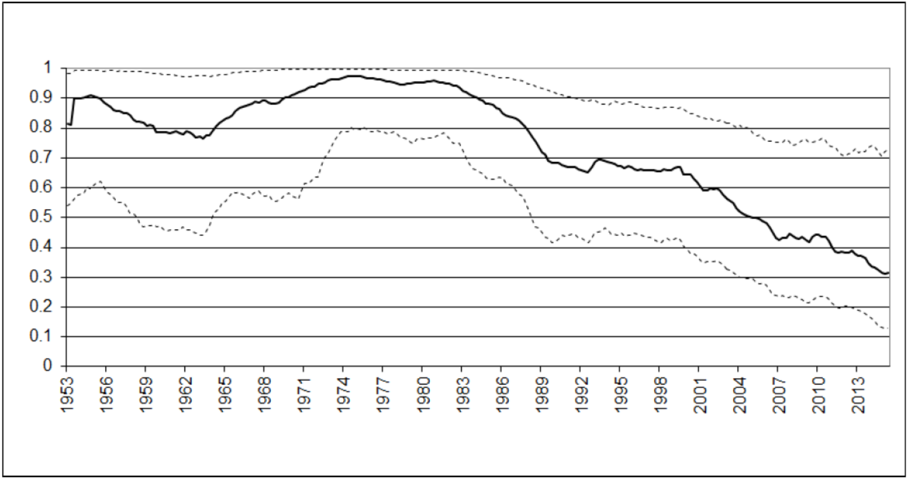
Time-varying MA(1) parameter - median, 16.5% and 83.5% quantiles of the posterior distribution



Time-varying R-sq bounds: posterior medians



Time-varying 16.5% and 83.5% quantiles for R-sq max - R-sq min



Implications of Stock & Watson's (2007) model of inflation

- The changing univariate properties of inflation noted by SW *dictate* that even the true multivariate model would now struggle to out-predict a univariate model
- The IMA(1,1) model implies that, in the absence of cancellation, there is a single IID predictor
  - No wonder macroeconomists struggle to find output gap measures that forecast inflation
  - More promise in looking for "news" type predictors
  - While higher-dimensional multivariate models (with cancellation) can escape the R^2 bounds, why should this be a feature of the macroeconomy? Smets and Wouters' (2007) DSGE model does not escape the bounds, as its predictions for inflation are too persistent

Conclusions

- What we know about the univariate time-series properties of y_t tells us a lot about the properties and performance of the true (but unknown) multivariate predictive model
- Helps us explain why it's often hard to out-forecast a univariate model
  - And why researchers often struggle to find reliable predictors; e.g. why the output gap struggles to forecast inflation
  - Multivariate models with news variables and/or with few distinct eigenvalues may fare better?