What univariate models tell us about multivariate macroeconomic models

James Mitchell[*], Donald Robertson[†] and Stephen Wright[‡]

October 27, 2016

Abstract

A longstanding puzzle in macroeconomic forecasting has been that a wide variety of multivariate models have struggled to out-predict univariate representations. We seek an explanation for this puzzle in terms of population properties. We show that if we just know the univariate properties of a time-series, y_t, this can tell us a lot about the dimensions and the predictive power of the true (but unobservable) multivariate macroeconomic model that generated y_t. We illustrate using data on U.S. inflation. We find that, especially in recent years, the univariate properties of inflation *dictate* that even the true multivariate model for inflation would struggle to out-predict a univariate model. Furthermore, predictions of changes in inflation from the true model would either need to be IID or have persistence properties quite unlike those of most current macroeconomic models.

Keywords: Forecasting; Macroeconomic Models; Autoregressive Moving Average Representations; Predictive Regressions; Nonfundamental Representations; Inflation Forecasts

JEL codes: C22, C32, C53, E37

---

[*]Corresponding author: Warwick Business School, University of Warwick, CV4 7AL, U.K. James.Mitchell@wbs.ac.uk

[†]Faculty of Economics, University of Cambridge, CB3 9DD, U.K. dr10011@cam.ac.uk

[‡]Department of Economics, Maths & Statistics Birkbeck College, University of London, W1E 7HX, U.K. s.wright@bbk.ac.uk

1 Introduction

A long-standing and, on the face of it, puzzling feature of macroeconomic forecasting (that goes back at least as far as Nelson, 1972) has been that a wide variety of multivariate models have struggled to out-predict univariate models, particularly in terms of a consistent performance over time (e.g., see D'Agostino and Surico, 2012; Chauvet and Potter, 2013; Rossi, 2013a; Estrella and Stock, 2015; Stock and Watson, 2007, 2009, 2010, 2015). Indirect evidence of the power of univariate models can also be inferred from the relative forecasting success of Bayesian VARs that utilise Minnesota type priors (e.g., see Banbura *et al.*, 2010; Canova, 2007, p. 378), since these effectively give greater weight in estimation to finite order univariate autoregressive representations.

Alongside this puzzle (which we term the Predictive Puzzle) is another long-standing puzzle (which we term the Order Puzzle): the univariate models that compete so successfully with multivariate models in predictive terms are almost invariably of low order; yet multivariate models commonly have large numbers of state variables, which should imply high-order ARMA reduced forms (see Wallis, 1977; Cubbada *et al.*, 2009).

In this paper we seek an explanation of both puzzles in terms of population properties. We do so by taking a backwards look at the relationship between multivariate and univariate properties. We analyse a stationary univariate time series process, y_t, data for which are assumed to be generated by a multivariate macroeconomic model. We ask: on the basis of the observed history of y_t alone, what would its univariate properties, as captured in population by a finite order ARMA representation, tell us about the properties of the true multivariate model that generated the data? We show that, for some y_t processes, univariate properties alone will tightly constrain the properties of the true multivariate model.

Our core results relate to the Predictive Puzzle. We first ask: how much better could we predict y_t if we could condition on the true state variables of the underlying multivariate model, rather than just the univariate history, denoted y^t? We show that the resulting one-step-ahead predictive R^2 must lie between bounds, R^2_{min} and R^2_{max}, that will usually lie strictly within $[0, 1]$. Crucially, we can derive these bounds from ARMA parameters, derived only from the history y^t.

The lower bound R^2_{min} is simply the one-step-ahead R^2 of the fundamental ARMA representation. The rationale is straightforward, and exploits well-known properties: if we could condition on the true state variables that generated the data for y_t, we must be able to predict it at least as well as (and usually better than) the univariate representation. But we show, further, that there is a limit to the improvement, R^2_{max}, that the true multivariate model could achieve, which we can *also* calculate solely from the univariate properties of y_t. We show that $R^2_{\max}$ is the (strictly notional) R^2 of a "nonfundamental" (Lippi and Reichlin, 1994) representation in which all the MA roots are replaced with their reciprocals. While such nonfundamental representations are nonviable as predictive models their *properties* can be derived straightforwardly from the parameters of the fundamental ARMA representation.

For some time series, ARMA properties imply that the gap between $R^2_{\min}$ and $R^2_{\max}$ is quite narrow, in which cases our results show that little improvement in predictive performance would be possible, even if we had the true state variables for y_t. In such cases, as a direct consequence, one-step ahead ARMA forecasts must also be very *similar* to those we would make if we conditioned on the true state variables.

Knowing that both bounds are derived from ARMA models also tells us that, if the true multivariate system were to have a predictive R^2 in the neighbourhood of either of the R^2 bounds, the covariance matrix of predictive errors in the system would need to be very close to singularity, i.e., must be close to being driven by a single structural shock.[1]

Our R^2 bounds extend straightforwardly to situations where the true underlying multivariate model (and the derived univariate model) both have time-varying coefficients and (co)variances. Hence our core results do not rely on the assumption of structural stability.

Our analysis also allows us to address the Order Puzzle. If y_t is an ARMA(p, q) in population then standard results (dating back at least to Wallis, 1977) tell us that both p and q must capture key dimensions of the underlying multivariate system. If r is the number of distinct eigenvalues of $\mathbf{A}$, the autoregressive matrix of the true states, then, in the absence of exact restrictions across the parameters of the model, $r = q$, and p tells us the number of non-zero eigenvalues. This is of course a population result, and the true values of p and q may never be knowable in finite samples. Nonetheless it is striking that, whereas macro models typically have large numbers of state variables[2], the univariate models that compete so successfully with them are almost invariably of low order (Diebold and Rudebusch, 1989; Cubbada *et al.*, 2009).

We discuss two possible explanations. One (noted by Wallis, 1977) is that the true

[1] It is something of a paradox that the stochastic singularity of early real business cycle models, and their inability, in consequence, to match multivariate properties of macro data, gave an impetus to the development of models with multiple structural shocks (as in, e.g., Smets & Wouters, 2007); but our analysis implies that these later models, by construction, cannot approach $R^2_{\max}$ for any individual y_t process.

[2] In the benchmark empirical DSGE of Smets and Wouters (2007), for example, $\mathbf{A}$ has 16 distinct eigenvalues.

multivariate model that generates y_t has multiple state variables, but that the model satisfies, or is very close to satisfying, parameter restrictions that either allow state variables to be aggregated, or that generate cancellation of AR and MA roots in the reduced form. We show that this also imposes restrictions on the nature of the predictions generated by the true multivariate system. So this potential explanation of the Order Puzzle then opens up a new puzzle (as far as we know, unanswered) of why the true multivariate model should satisfy these restrictions. But we note that there is a second, simpler explanation: that there may just be very few distinct eigenvalues driving the dynamics of the macroeconomy.[3] This second explanation is consistent with the empirical finding that a small number of factors can explain a large fraction of the variance of many macroeconomic variables (e.g., see Stock and Watson, 2002).

To illustrate, we consider Stock and Watson's (2007, 2009, 2010, 2015) findings on U.S. inflation. Their results allow us to shed light on both the Predictive and Order Puzzles.

Stock & Watson's (2007) preferred univariate unobserved-components stochastic volatility representation reduces to a time-varying IMA(1) model. In recent data this implies that our upper and lower bounds for R^2(which in this case will vary over time) must be close to each other. Thus the observable univariate properties Stock & Watson uncover *dictate* the feature that, however well the true multivariate model for inflation might predict, it could at best only marginally out-predict a univariate model. As a direct corollary we also show that this must lead to any (efficient) multivariate forecast of inflation being strongly correlated with Stock & Watson's univariate forecasts.

Data on inflation also provide a prime example of the Order Puzzle. The IMA(1,1) univariate property in Stock & Watson's representation implies that, in the absence of cancellation, all predictability of changes in inflation can be captured by a single composite IID predictor.[4] Yet in most structural macroeconomic models predicted changes in inflation arise from multiple, and strongly persistent state variables capturing some form of pressure of demand in the real economy.

Of course, in a finite sample we cannot *know* that the true univariate process for inflation is an IMA(1,1). The true ARMA representation might in principle be higher order than in Stock and Watson's representation, and might arise from a true multivariate system with multiple state variables. But even if this were the case, our analysis provides insights. First, to be consistent with their results, the true multivariate model

[3]Note that this can in principle be consistent with larger numbers of underlying state variables, as long as they have shared eigenvalues.

[4]Time variation in the univariate representation can be captured by a single time varying slope parameter and time variation in the conditional variance of inflation.

for inflation would have to be close to satisfying the parameter restrictions that generate near-cancellation of, or near-zero AR and MA roots in the ARMA reduced form. As noted above, it is unclear why this should be a structural feature of the macroeconomy. Second, we show that, for the true multivariate model for inflation to escape our R^2 bounds would require it to generate predicted changes in inflation that were *not* IID, but with persistence properties quite unlike those generated by most current macroeconomic models.[5]

We argue that this helps to explain the long and fruitless search by macroeconomists for predictor variables that consistently forecast inflation (e.g. see Atkeson and Ohanian, 2001 and Stock and Watson, 2007, 2009, 2010). On a more positive note, for those seeking to find predictor variables for changes in inflation, then assuming Stock and Watson's univariate representation is correct, it could arise from a structural model where the prediction is itself an innovation to some information set: i.e., it is "news". Indeed we argue that univariate properties imply that *only* a variable representing news could predict changes in U.S. inflation.

The rest of the paper is structured as follows. Section 2 sets out the links between the ARMA representation and the multivariate model and describes the R^2 bounds and their implications. In Section 3 we illustrate our results with reference to the case of an ARMA(1,1). Section 4 shows that our core results can be generalised to accommodate time variation. Section 5 presents an empirical application drawing on Stock & Watson's analysis of U.S. inflation, and Section 6 concludes, and draws out further implications of our results. Appendices provide proofs and derivations.

[5] We illustrate using the Smets & Wouters (2007) model.

2 What the ARMA representation tells us about the true multivariate system

2.1 The true multivariate structural model and its implied predictive regression for y_t

Consider a univariate time series y_t that is generated by a linear (or linearised) multivariate macroeconomic model:[6]

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{B}\mathbf{s}_t \tag{1}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{z}_{t-1} + \mathbf{D}\mathbf{s}_t \tag{2}$$

where $\mathbf{z}_t$ is an $n \times 1$ vector of state variables hit by a vector of structural economic shocks, $\mathbf{s}_t$, and $\mathbf{y}_t$ is a vector of observed macroeconomic variables, one element of which is the variable of interest.

We wish to consider what observing y_t in isolation can tell us about the nature of the true underlying system in (1) and (2).

We make the following assumptions:

Assumptions

A1 $\mathbf{A}$ *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* $\mathbf{M}^*_{ii} = \lambda_i$, $i = 1, ..., r$, $r \leq n$, $\lambda_i \neq \lambda_j$, $\forall i, \forall j$ *being the distinct eigenvalues of* $\mathbf{A}$.

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

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

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

Assumption A1, that the ABCD system can be diagonalised, is in most cases innocuous.[7] Assumption A2, that the system is stationary, is also simply convenient. There may in principle be unit roots in some of the states and observables, but if so these can be differenced out to generate a stationary representation.

[6] We use the notation of the generic ABCD representation of Fernández-Villaverde *et al.* (2007). They assume that this system represents the rational expectations solution of a DSGE model (in which cases the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ are usually functions of a lower dimensional vector of deep parameters, $\boldsymbol{\delta}$). But the representation is sufficiently general to capture the key properties of a wide range of multivariate models, including VAR and factor models.

[7] It allows for possibly complex eigenvalues, and hence elements of $\mathbf{z}_t$. It can be generalised completely by letting $\mathbf{M}^*$ take the Jordan form (with 1s on the sub-diagonal). This admits, in terms of the discussion below, ARMA(p, q) representations with $q > p$, but does not otherwise change the nature of our results.

Assumption A3 follows Fernández-Villaverde *et al.* (2007). It is convenient (but not essential) to assume normality to equate expectations to linear projections; while the normalisation of the structural shocks to be orthogonal, with unit variances, is simply an identifying assumption, with the matrices $\mathbf{B}$ and $\mathbf{D}$ accounting for scale factors and mutual correlation.

The assumption that the structural disturbances $\mathbf{s}_t$ are serially uncorrelated, while standard is, however, crucial - as we discuss below in Lemma 1. Finally, Assumption A4 ensures that none of the elements of $\mathbf{y}_t$ are linear combinations of $\mathbf{z}_{t-1}$: this simply rules out the possibility that y_t, the variable of interest, may be a pre-determined variable.

The assumptions of a time-invariant model and of normality of the structural shocks are, it should stressed, *not* crucial; they merely simplify the exposition. In Section 4 we consider generalisations to cases where the parameters of the structural model may vary over time.

These assumptions allow us to derive a particularly simple specification for the true predictive regression for y_t, a single element of $\mathbf{y}_t$. This conditions on a minimal set of AR(1) predictors that are linear combinations of the state variables in the system in (1) and (2):

Lemma 1 ***(The True Predictive Regression)*** *Under A1 to A3 the structural ABCD representation implies the true predictive regression for y_t, the first element of $\mathbf{y}_t$:*

$$y_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t \tag{3}$$

where $\mathbf{x}_t$ is an $r \times 1$ vector of predictors, with $r \leq n$, with law of motion

$$\mathbf{x}_t = \mathbf{M}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{4}$$

where $\mathbf{M} = diag(\lambda_1, ..., \lambda_r)$, so that each element of $\mathbf{x}_t$ is a univariate AR(1). At most one element may have $\lambda_i = 0$ in which case x_{it} is IID.

Proof. See Appendix A. ■

Since (3) is derived from the structural model that generated the data, the r-vector of predictors $\mathbf{x}_{t-1}$, a linear combination of the n underlying states, can be viewed as generating the data for y_t, up to a white noise error, u_t (given Assumption A3).

Remark: *Elements of the predictor vector $\mathbf{x}_t$ in the true predictive regression may be aggregates of the elements of the underlying true state vector $\mathbf{z}_t$ if $\mathbf{A}$, the autoregressive matrix of the states, has repeated eigenvalues. Thus r, the dimension of the predictor vector, may be less than n, the dimension of the true underlying states.*

2.2 The Macroeconomist's ARMA

Exploiting standard results (e.g., applying Corollary 11.1.2 in Lütkepohl (2007)),[8] it is straightforward to derive the true univariate reduced form for y_t:

Lemma 2 ***(The Macroeconomist's ARMA)*** *The true predictive regression in (3) and the process for the associated predictor vector (4) together imply that* y_t *has a unique fundamental ARMA*(r,r) *representation with parameters* $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_r)$ *and* $\boldsymbol{\theta} = (\theta_1, ..., \theta_r)$

$$\lambda(L)\, y_t = \theta(L)\, \varepsilon_t \tag{5}$$

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

Proof. See Appendix B. ■

The θ_i are solutions to a set of r moment conditions that match the autocorrelations of y_t, as set out in Appendix B. The condition $|\theta_i| \leq 1$ gives the unique fundamental solution (Hamilton, 1994, pp. 64-67; Lippi and Reichlin, 1994) since it ensures that $\varepsilon_t = \theta(L)^{-1} \lambda(L)\, y_t$ is recoverable as a non-divergent sum of current and lagged values of y_t.[9]

Note that we refer to this representation as the "Macroeconomist's ARMA" because its properties follow directly from those of the underlying macroeconomic model. Thus $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ are functions of the full set of parameters $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ of the underlying system in (1) and (2).

2.3 What the ARMA representation tells us about the true predictive regression

2.3.1 Bounds for the predictive R^2

We have derived the ARMA representation from the underlying structural model. We now look at this process 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 properties of the structural multivariate system that generated the data for y_t?

We first show that the degree of predictability measured by the R^2 of the true predictive regression (3), that conditions on a linear combination of the state variables in the

[8] Drawing on the seminal work of Zellner and Palm (1974) and Wallis (1977).

[9] The limiting case $|\theta| = 1$, which is not invertible but is still fundamental, may in principle arise if y_t has been over-differenced (as in, e.g. Smets & Wouters, 2007). But since this case essentially arises from a mis-specification of the structural model we do not consider it further.

structural model, must lie between bounds that can be defined solely in terms of ARMA parameters.

Proposition 1 ***(Bounds for the Predictive*** R^2*) Let*

$$R^2 = 1 - \sigma_u^2/\sigma_y^2 \tag{6}$$

be the one-step-ahead predictive R^2 *for the predictive regression for* y_t *(3), that is derived from the ABCD representation (1) and (2) of the true multivariate model.* *Under A1 to A4,* R^2 *satisfies*

$$0 \leq R^2_{\min} \leq R^2 \leq R^2_{\max} \leq 1 \tag{7}$$

where

$$R^2_{\min} = 1 - \sigma_\varepsilon^2/\sigma_y^2 \tag{8}$$

is the predictive R^2 *from the ARMA representation (5) and so is a function of* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\theta}$ *alone, and*

$$R^2_{\max} = R^2_{\min} + \left(1 - R^2_{\min}\right)\left(1 - \prod_{i=1}^{r} \theta_i^2\right) \tag{9}$$

so $R^2_{\max}$ *is also a function of* λ *and* θ *alone.*

Proof. See Appendix C. ■

Proposition 1 says that, for the predictive regression for y_t that arises from the true structural model, the predictive R^2 must lie between bounds that can be defined solely in terms of ARMA parameters.

The lower bound for R^2

The intuitive basis for the lower bound, $R^2_{\min}$, is straightforward and derives from known results (e.g., see Lütkepohl (2007), Proposition 11.2). Predictions generated by the fundamental ARMA representation condition only on the history of y_t; so they cannot be worsened by conditioning on the true state variables. Indeed, the true R^2 must be strictly greater than $R^2_{\min}$ except in the limiting case that $u_t = \varepsilon_t$. Furthermore, for any y_t process that is not IID this lower bound is itself strictly positive.

The upper bound for R^2

The upper bound, $R^2_{\max}$, is calculated from the parameters $(\boldsymbol{\lambda}, \boldsymbol{\theta})$ of the minimal ARMA representation. But the proof of the proposition shows that it also has a clear-cut interpretation:

Remark: *If $\theta_i > 0\ \forall i$, the upper bound $R^2_{\max}$ is the notional R^2 from a nonfundamental ARMA representation in which all the θ_i are replaced with their reciprocals:*

$$\lambda(L)\, y_t = \theta^N(L)\, \eta_t \tag{10}$$

where $\lambda(L)$ is as in (5), and $\theta^N(L) = \prod_{i=1}^{r}\left(1 - \theta_i^{-1}L\right)$.

Recall that, in deriving the ARMA from the structural model, we noted that the MA parameters, θ must satisfy r moment conditions to match the autocorrelations of y_t, subject to the constraint that all the θ_i live within $(-1, 1)$. However, there are a further $2^r - 1$ *non*fundamental ARMA representations, in which one or more of the θ_i is replaced by its reciprocal, (Lippi & Reichlin, 1994),[10] each of which also satisfies the moment conditions, and thus generates identical autocorrelations to (5). In the particular nonfundamental representation, (10), relevant to Proposition 1 all the θ_i in (5) are replaced by their reciprocals.[11]

Like all nonfundamental representations (10) is a non-viable predictive model, because its shocks η_t, cannot be recovered from the history of y_t. However, its *properties* can still be calculated from the parameters of the fundamental ARMA representation in (5).

Thus the proposition says that, while we can increase R^2, relative to the lower bound given by the ARMA, by conditioning on the true state variables, there is a limit to the extent that R^2 can be increased. Furthermore, this limit can be calculated solely from the parameters in the univariate model for y_t.

In Section 3.2 we provide further intuition for the existence of an upper bound, in a simple analytical example.

2.4 The R^2 bounds and observable predictors

Our R^2 bounds apply to predictions that condition on the true state variables that generated the data for y_t. In practice, of course, we must make do with what predictors we can actually observe. Suppose, for some observable predictor vector, $\mathbf{q}_t$, we simply run a predictive regression that is just a least squares projection of the form $y_t = \boldsymbol{\gamma}'\mathbf{q}_{t-1} + \xi_t$. If $\mathbf{q}_t \neq \mathbf{x}_t$, but contains elements that are at least somewhat correlated with elements of $\mathbf{x}_t$, any such regression may have predictive power, but we would *not* necessarily expect

[10] Note that as discussed in Lippi & Reichlin (1994) some of the ψ_i may be complex conjugates.

[11] Note that if $\theta_i = 0$ for some i the nonfundamental representation is undefined but we can still use (9) to calculate $R^2_{\max} = 1$.

the resulting predictive R^2 to exceed our lower bound, $R^2_{\min}$.[12]

However, a straightforward corollary of Proposition 1 implies that, at least in population, our R^2 bounds must still apply for *any* predictive regression for y_t in which information from observable predictors is used *efficiently*:

Corollary 1 *Consider some set of estimates $\widehat{\mathbf{x}}_t = E\left(\mathbf{x}_t | \mathbf{q}^t, y^t\right)$ derived by the Kalman Filter, that condition on the joint history of a vector of observable predictors, $\mathbf{q}^t$ and y^t. The predictive R^2 for a predictive regression of the same form as (3), but replacing $\mathbf{x}_t$ with $\widehat{\mathbf{x}}_t$, also satisfies $R^2_{\widehat{\mathbf{x}}} \in \left[R^2_{\min}, R^2_{\max}\right]$, as in Proposition 1.*

Proof. See Appendix D. ■

If the observable predictor vector $\mathbf{q}_t$ has any informational content about the true state variables that is independent of the history y^t, then $R^2_{\widehat{\mathbf{x}}}$ *must* be strictly greater than $R^2_{\min}$, since this comes from a predictive model that conditions only on the history y^t. Clearly the more information $\mathbf{q}_t$ reveals about the true states, the closer the resulting R^2 can get to $R^2_{\max}$. If, in contrast, $\mathbf{q}_t$ reveals no information about $\mathbf{x}_t$ that cannot be recovered from y^t, it is predictively redundant, i.e., $E\left(\mathbf{x}_t | \mathbf{q}_t, y^t\right) = E\left(\mathbf{x}_t | y^t\right)$.[13] But this may not be evident from a least squares regression that does not correctly condition on y^t.

Covariance properties

Since both R^2 bounds are associated with ARMA representations, Proposition 1 also has significant implications for the covariance properties of the underlying structural model and its associated predictive regression.

We focus initially on the case where the true predictive regression is at, or in the neighbourhood of, either bound:

Corollary 2 *If the true predictive regression (3) attains either R^2 bound the error covariance matrix of the predictive system (3) and (4), $\Omega \equiv E\left(\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix} \begin{bmatrix} u_t & \mathbf{v}_t \end{bmatrix}\right)$ must be rank 1.*

Thus not only do the R^2 bounds limit the degree of multivariate predictability, they also tightly constrain the way in which that predictability can arise. Near both bounds

[12] Not least because the predictive errors ξ_t cannot in general be jointly IID with the innovation to a time series representation of $\mathbf{q}_t$ (a point made forcefully by Pastor and Stambaugh, 2009).

[13] This is indeed the null hypothesis of no Granger Causality from $\mathbf{q}_t$, as originally formulated by Granger (1969).

prediction errors for y_t and the predictor vector $\mathbf{x}_t$ must be close to perfectly correlated: i.e., must be close to being generated by a single structural shock.

More generally, the very existence of the bounds also puts a constraint on how *different* the predictions from the true model can be from a univariate forecast:

Corollary 3 ***(The Prediction Correlation)***

Let $\rho = corr\left(E\left(y_{t+1}|\mathbf{x}_t\right), E\left(y_{t+1}|y^t\right)\right)$ be the correlation coefficient between the predictions from the true predictive regression (3) and the predictions from the fundamental ARMA representation

$$\rho = \sqrt{\frac{R^2_{\min}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R^2}} \geq \rho_{\min} = \sqrt{\frac{R^2_{\min}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R^2_{\max}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}} \geq 0 \tag{11}$$

where both inequalities are strong for $r > 0$.

Proof. See Appendix E. ■

Corollary 3 shows that the prediction correlation ρ between the predictions from the fundamental ARMA and the predictions that condition on the true state variables is monotonically decreasing in R^2. Thus the closer is R^2 to its lower bound, then the more the system predictions must resemble those from the ARMA model. But, since Proposition 1 places an upper bound on R^2, it also follows that the narrower is the gap between these R^2 bounds, the closer ρ must be to unity, and hence the more strongly correlated must be the univariate and true multivariate forecasts.

2.5 What the ARMA representation tells us about the dimensions of the true structural model

We have shown that the true structural model implies an ARMA and that there is as a result a link between the ARMA parameters and the degree of predictability we could achieve if we could condition on the true state variables. We can also look at this relationship backwards. We now show that, in the absence of restrictions on the structural model, the dimensions of the ARMA tell us those of the predictive system in (3) and (4) that derives from the true structural model.

Proposition 2 ***(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 A4, the data y_t *must have been generated by a structural model as in (1) and (2) 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 *as in (3) and (4), with* $r = q$ *predictors. Furthermore, for* $q > 0$ *the bounds for* R^2 *from Proposition 1 must lie strictly within* $[0, 1]$

Proof. See Appendix F. ■

Corollary 4 *The predictor vector* $\mathbf{x}_t$ *in the true predictive regression (3) 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*.*

Proposition 2 simply inverts the process by which we derived the macroeconomist's ARMA.[14] If the minimal population ARMA representation has MA order q, this tells us that the data for y_t must have been generated by a predictive system with $r = q$ predictors, and hence by an ABCD system in which $\mathbf{A}$ has r distinct eigenvalues.[15]

Corollary 4 in turn states the only circumstances in which this is *not* the case: if the $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ parameters of the structural model satisfy restrictions that ensure exact cancellation in the univariate reduced form. Under these circumstances, but *only* under these circumstances, the order of the macroeconomist's ARMA (5) may differ from that of the minimal ARMA(p, q) representation. In the absence of such restrictions the orders of the two representations must be identical.

2.6 Time series properties of the predictions

The ARMA also tells us a key characteristic of the predictions that arise from the structural model.

[14] As such we make no claim that it is a new result: it is simply a new way of looking at an old one.

[15] The order of the AR polynomial will also be $p = q$ unless $\lambda_i = 0$ for some i, in which case $p = q - 1$, since, from Assumption A1 there can be at most one such λ_i $(i = 1, ..., q)$. It is reasonably easy to generalise to other values of p (for details see an earlier working paper version of this paper: Robertson and Wright, 2012).

Corollary 5 *If $r = q$ then the predictions $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_t$ have an ARMA$(p, q-1)$ representation.*

The key insight here is that the time series properties of the predictions $\widehat{y}_t$ are inherently different from those of y_t itself. Indeed this inherent difference in time series properties is a crucial part of the explanation of *why* there must be the upper bound for R^2, as in Proposition 1. We discuss this issue further in Section 3.4 below, in relation to our illustrative analytical example. We shall also see that this difference in time series properties provides important insights into our empirical example, discussed in Section 5.

3 An illustrative example: The ARMA(1,1) case

As an illustrative example we explore the following simple (but popular) case.

3.1 The macroeconomist's ARMA with $r = 1$

Consider the case in which data for y_t are generated by an ABCD model with a single state variable and a 2×1 vector of structural shocks. This implies that the counterparts to (3) and (4) in Lemma 1 are

$$y_t = \beta x_{t-1} + u_t \tag{12}$$

$$x_t = \lambda x_{t-1} + v_t \tag{13}$$

with $\mathbf{x}_t = x_t = z_t$, $\mathbf{A} = \lambda$, $\mathbf{C} = \beta$, $v_t = \mathbf{B}\mathbf{s}_t$ and $u_t = \mathbf{D}\mathbf{s}_t$, with $\mathbf{B}$ and $\mathbf{D}$ both 1×2 row vectors that generate a covariance structure for u_t and v_t. While this is an extremely simple system, a specification of this form has, for example, dominated the finance literature on predictive return regressions, with y_t some measure of returns or excess returns, and z_t some stationary valuation criterion. Note that a predictive system of this form can also easily subsume the case of an underlying structural ABCD representation in which the state vector $\mathbf{z}_t$ has $n > 1$ elements, but in which $\mathbf{A}$ has a single common eigenvalue, λ, in which case the single predictor in (12) and (13) would be a composite of the n underlying state variables in $\mathbf{z}_t$.

By substitution from (12) into (13) we have

$$(1 - \lambda L)\, y_t = \beta v_{t-1} + (1 - \lambda L)\, u_t \tag{14}$$

The right-hand-side of this expression is an MA(1) so y_t admits a fundamental ARMA(1,1) representation

$$(1-\lambda L)\, y_t = (1-\theta L)\, \varepsilon_t \tag{15}$$

with $|\theta| < 1$. The first order autocorrelation of the MA(1) process on the right-hand side of (15) matches that of the right-hand-side of (14): i.e., the single MA parameter θ is the the solution in $(-1, 1)$ to the moment condition

$$\frac{-\theta}{1+\theta^2} = \frac{\lambda\sigma_u^2 - \beta\sigma_{uv}}{\left(1+\lambda^2\right)\sigma_u^2 + \beta^2\sigma_v^2 + 2\lambda\beta\sigma_{uv}} \tag{16}$$

Since the right-hand-side of (16) is derived from the parameters of the ABCD representation, we have $\theta = \theta\left(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\right)$, as for the general case.

3.2 Proposition 1 in the ARMA(1,1) case: Bounds for R^2

It is straightforward to show that the R^2 of the true predictive regression (12) that conditions on the single true state variable $x_t = z_t$ has a lower bound given by

$$R^2_{\min} = \frac{(\theta-\lambda)^2}{1-\lambda^2+(\theta-\lambda)^2} \tag{17}$$

which is the predictive R^2 of the ARMA representation. The upper bound is

$$R^2_{\max} = R^2_{\min} + \left(1 - R^2_{\min}\right)\left(1-\theta^2\right) = \frac{(1-\lambda\theta)^2}{1-\lambda^2+(\theta-\lambda)^2} \tag{18}$$

which would be the notional R^2 of the nonfundamental representation associated with (15)

$$(1-\lambda L)\, y_t = \left(1-\theta^{-1} L\right)\eta_t \tag{19}$$

which is a special case of (10).[16]

The bounds in (17) and (18) can be used to illustrate limiting cases.

If θ is close to λ, so that y_t is close to being white noise, $R^2_{\min}$is close to zero. If θ is close to zero, $R^2_{\max}$ is close to one. But only if θ and λ are *both* sufficiently close to zero (implying that both y_t and the single predictor z_t are close to white noise), does the inequality for R^2 open up to include the entire range from zero to unity. Thus only in this doubly limiting case is Proposition 1 entirely devoid of content.

[16]Note that the moment condition (16) is satisfied by θ and also by θ^{-1}. While in general, as discussed in Section 2.2 there will be multiple $(2^r - 1)$ nonfundamental representations of the same order, in this particular case, with $r = q = 1$, there is only one.

In marked contrast, as $|\theta|$ tends to unity the range of possible values of R^2 collapses to a single point (which is $\frac{1-sgn(\theta)\lambda}{2}$). This has the important implication that for any ARMA$(1,1)$ process with high $|\theta|$ there would be very little scope for the true predictive regression to outperform the ARMA. Furthemore, in such cases ρ, the correlation between the predictions from the ARMA and those from the true predictive regression must be very close to unity.[17]

It is straightforward to examine the limiting cases in which the true predictive regression attains either bound. Straightforward manipulations show that in these two cases the predictive system (12) and (13) must simply be reparameterisations of the fundamental and nonfundamental ARMA representations, (15) and (19)

$$\begin{aligned} y_t &= \beta_F x_{t-1}^F + \varepsilon_t \\ x_t^F &= \lambda x_{t-1}^F + \varepsilon_t \end{aligned}$$

with $\beta_F = \lambda - \theta$, and

$$\begin{aligned} y_t &= \beta_N x_{t-1}^N + \eta_t \\ x_t^N &= \lambda x_{t-1}^N + \eta_t \end{aligned}$$

with $\beta_N = \lambda - \theta^{-1}$. By construction these two special cases have$R^2 = R^2_{\min}$ and $R^2 = R^2_{\max}$, respectively. Note that in both cases the error in the predictive equation and the innovation to the single state variable are perfectly correlated, illustrating Corollary 2.

Note also that the maximal R^2 would be attained by the state variable $x_t^N = (1-\lambda L)^{-1}\eta_t$. Since the resulting predictive system is a reparameterisation of a nonfundamental representation x_t^N cannot be derived as a convergent sum of past y_t. However we can write

$$x_t^N = \left(1-\theta^{-1}L\right)^{-1} y_t = -\frac{\theta L^{-1}}{(1-\theta L^{-1})} y_t = -\sum_{i=1} \theta^i y_{t+i}$$

thus x_t^N is a convergent sum of *future* values of y_t. Thus the improvement in predictive power comes about because x_t^N acts as a window into the future: the lower is θ, the more it will reveal.[18]

The true state variable x_t will predict y_t better, the more closely it resembles x_t^N; but it *cannot* predict better than x_t^N.

---

[17] Since, using Corollary 5 $\rho_{\min} = \sqrt{\frac{R^2_{\min}}{R^2_{\max}}} = \left|\frac{\theta-\lambda}{1-\lambda\theta}\right|$.

[18] Note that only in the limiting case as $\theta \to 0$ does it actually reveal y_{t+1} perfectly.

3.3 Proposition 2 in the ARMA(1,1) case: why (absent exact parameter restrictions) $r = q = 1$.

Proposition 2 tells us that, if y_t is an ARMA(1,1), we must have $r = q = 1$, unless restrictions hold on the parameters of the structural model that generated the data. To illustrate, we show the restrictions that must be imposed in two cases where this relation would *not* hold

3.3.1 $r = 1, q = 0$

In this case y_t is a pure AR(1) in reduced form. But inspection of the moment condition (16) shows that this requires

$$\lambda\sigma_u^2 = \beta\sigma_{uv} \text{ or equivalently } \mathbf{ADD}' = \mathbf{CDB}' \tag{20}$$

thus for a pure AR(1) y_t process to be generated by a system with a single state variable imposes a restriction between the single eigenvalue λ, and the rest of the model.

3.3.2 $r > q = 1$

Assume, initially, that the underlying structural model reduces to a predictive system with $r = 2$ of the form

$$y_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + u_t \tag{21}$$

$$\mathbf{x}_t = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \tag{22}$$

where (λ_1, λ_2) are the distinct eigenvalues of the autoregressive matrix $\mathbf{A}$ of the underlying states. For the general case, this would imply that the macroeconomist's ARMA is an ARMA$(2, 2)$, i.e.,

$$y_t = \frac{(1 - \theta_1 L)(1 - \theta_2 L)}{(1 - \lambda_1 L)(1 - \lambda_2 L)}\varepsilon_t$$

where both θ_1 and θ_2 are functions of the ABCD parameters of the underlying system. However, for this predictive system to have generated the data for y_t, and for y_t in turn to be an ARMA(1,1) requires that there must be redundancy in the macroeconomist's

ARMA, and hnecce the underlying model must satisfy the restriction

$$\lambda_2 (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \theta_2 (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) \tag{23}$$

Analogous arguments apply for systems for any value of r greater than unity.

In both these examples of cancellation we see that $r \neq q$ can only occur if exact non-linear restrictions hold across the parameters of the underlying structure. These involve both behavioural coefficients and parameters describing the stochastic environment and consequently such exact restrictions seem unlikely to occur by chance, or to arise from a theoretical model.

3.4 The upper bound for R^2 and the time series properties of the predictions

If y_t is an ARMA(1,1), and (in the absence of any restrictions such as those discussed above), if $r = q = 1$, then, by inspection of (13) the single state variable x_t must be an AR(1) (a special case of Corollary 5). As a result x_t can have quite distinct time series properties from those of y_t.

Write the ARMA process for $y_t = C_y (L) \varepsilon_t$ and that for the state variable as $x_t = C_x (L) v_t$. Assume, for example, $\theta > \lambda > 0$. Then $C_y (1) = \frac{1-\theta}{1-\lambda} \in (0, 1)$, a property often attributed to, eg stock returns and changes in real exchange rates. But $C_x (1) = \frac{1}{1-\lambda} \in (1, \infty)$: the Beveridge-Nelson decomposition of the process for x_t (and hence for the predictions $\widehat{y}_t = \beta x_{t-1}$) is distinctly different from that of y_t itself.

This provides additional intuition for the the upper bound for R^2 in Proposition 1. For the true state variable to predict y_t well must ultimately require the the predictions it generates to mimic the time series properties of y_t itself. But if the time series properties of y_t and x_t are so different, this must imply a limit on how well x_t can predict y_{t+1}.

4 Time-varying parameters

Models with time-varying parameters are increasingly used in forecasting (e.g. see Cogley and Sargent (2005); D'Agostino et al. (2013), Rossi, 2013b). In general, if any of the parameters in the true structural model (1) and (2) are non-constant over time, this must translate into time variation in the parameters of the associated predictive regresssion (3) and the process for the predictor variables (4), i.e., the coefficient vector $\boldsymbol{\beta}$, the vector of AR parameters λ and the error covariance matrix $\mathbf{\Omega}$, all of which can be derived as functions of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. This will, in turn, translate into time variation in the parameters of the univariate representation for y_t. However, this does not detract from the insight our analysis provides; it merely complicates the algebra. The proof of the core result, the R^2 bounds in Proposition 1, relies on the assumption that the underlying innovations are independently distributed, not on their having a time-invariant distribution; nor does it rely on the constancy of λ, $\boldsymbol{\beta}$ or $\boldsymbol{\Omega}$.

Before considering an extension of our analysis to time-varying parameters, it is perhaps worth stressing two points. First there are some important forms of parameter variation that *can* be captured by a stationary ABCD representation with constant parameters and IID (but non-Gaussian) shocks. Hamilton (1994, p. 679) shows, for example, that if the conditional mean of y_t shifts due to a state variable that follows a Markov chain this implies a VAR model for the state; this in turn implies stationary ARMA and ABCD representations for y_t but with non-Gaussian shocks.[19] Second, even forms of structural instability that cannot be captured in this way should arguably still imply a time-invariant representation in *some* form. Thus, for example, the unobserved components stochastic volatility model of inflation popularised by Stock and Watson (2007), that is nested in the case we analyse below, has a time-invariant state-space representation - it is simply nonlinear rather than linear.

In what follows we simply assume that there is some model of time variation that results in a sequence $\{\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t, \mathbf{D}_t\}$, and hence time-varying ARMA parameters (including the innovation variance), without considering how this is generated. We show that we can generalise our key result on the R^2 bounds, at least for the special case of a time-varying ARMA(1,1), which we exploit in the empirical example in the next section :

Proposition 3 ***(Bounds for the Predictive R^2 of a Time-Varying ARMA(1,1))*** *Assume y_t is generated by the time-varying parameter structural model*

$$y_t = \beta_t x_{t-1} + u_t \tag{24}$$

$$z_t = \lambda_t x_{t-1} + v_t \tag{25}$$

where x_t, the single state variable has a time-varying AR(1) representation, $w_t = (v_t, u_t)'$ is a serially independent vector process with $E(w_t w_t') = \Omega_t$, all elements of which are potentially time-varying. In reduced form y_t has the time-varying fundamental ARMA(1,1)

[19] Any ARMA model has a state space representation (Hamilton, 1994, chapter 13, pp. 375-6). Permanent mean shifts induce a unit root that can be differenced out to derive a stationary ABCD representation.

representation

$$(1-\lambda_t L)\, y_t = (1-\theta_t L)\, \varepsilon_t \tag{26}$$

with $0 < |\theta_t| < 1$, $0 < |\lambda_t| < 1$, $\lambda_t \neq \theta_t$ *(thus* $r = q = 1$*) and* ε_t *is a serially uncorrelated error orthogonal to* y^t*, with* $E\left(\varepsilon_t^2\right) = \sigma_{\varepsilon,t}^2$*. Letting*

$$R_t^2 = 1 - \sigma_{u,t}^2/\sigma_{y,t}^2 \tag{27}$$

be the time-varying R^2 *for the predictive regression that conditions on the true state variable* x_t *(24) (with* $\sigma_{y,t}^2 = \sigma_{\varepsilon,t}^2 + \theta_t^2 \sigma_{\varepsilon,t-1}^2$*) then*

$$0 < R_{\min,t}^2 \leq R_t^2 \leq R_{\max,t}^2 < 1$$

where $R_{\min,t}^2$ *is the time-varying* R^2 *of (26), and* $R_{\max,t}^2$ *is the time-varying* R^2 *of the associated time-varying nonfundamental representation*

$$(1-\lambda_t)\, y_t = (1-\gamma_t L)\, \eta_t \tag{28}$$

where

$$\gamma_t = \frac{1}{\theta_t} \frac{\sigma_{\varepsilon,t}^2}{\sigma_{\varepsilon,t-1}^2}$$

Proof. See Appendix G ■

The proof of this proposition shows that time-varying parameters introduce simultaneity into the moment conditions for θ_t and $\sigma_{\varepsilon,t}^2$ (whereas in the time-invariant case these can be solved independently). As far as we are aware the exact derivation of the processes for θ_t and $\sigma_{\varepsilon,t}^2$, and of the associated nonfundamental representation, has not been carried out before. While solution of the moment conditions is distinctly more complicated for the time-varying case, once this problem has been solved, the proof of the (time-varying) R^2 bounds follows quite straightforwardly, and analogously to the proof of Proposition 1. All the associated formulae nest the time-invariant results for the ARMA(1,1) model, as given above in Section 3, as a special case.

Note that the case analysed in Proposition 3 nests the Stock Watson (2007) unobserved components stochastic volatility model, discussed in the next section, as a special case with $\lambda_t = 0$, $\forall t$. In this case the inequality constraints on the sequence $\{\theta_t\}$ are imposed by the properties of the (time-invariant) structural state space model (see below), which implies that θ_t is strictly positive and less than unity by construction.

We conjecture that the result can be generalised to higher order time-varying ARMA representations; although in practice estimated versions of such models on macroeconomic data have as far as we are aware only been of low order.

5 An empirical application: U.S. inflation

In this section we look at the implications for multivariate forecasting models of U.S. inflation using the properties of the influential univariate model of Stock and Watson (2007). Like Stock and Watson we focus on inflation as measured by the GDP deflator, although their analysis shows considerable commonality across a range of price indices.

5.1 What do the univariate properties of inflation tell us about predictability of inflation?

Stock and Watson argue that U.S. inflation, π_t, can be well represented as the sum of a random walk component and a serially uncorrelated component, each of which has stochastic volatility (both volatilities follow logarithmic random walks). Stock and Watson note that this unobserved components-stochastic volatility (UC-SV) representation can also be written as a time varying IMA(1) model; so in our framework we can set $y_t \equiv \Delta\pi_t = (1 - \theta_t L)\,\varepsilon_t$. Stock and Watson document the changing MA structure implied by their representation, and also describe the puzzle that, given a variety of possible predictive variables, it appears that inflation has recently become harder to forecast.

Our analysis shows that the inability to improve (much) on univariate forecasts is exactly what one would expect, given the nature of the univariate properties Stock and Watson report.

Using Proposition 3, Stock and Watson's univariate representation implies that the data for U.S. inflation must have been generated by an underlying multivariate system with time-varying parameters of the form in (24) and (25), in which, in the absence of restrictions across the underlying structural model, the single predictor x_t must be IID[20]. This does not rule out the possibility that x_t may in principle be some aggregate of multiple state variables for inflation, but only if these are all themselves IID.[21]

Figure 1 shows the median and 16.5% and 83.5% quantiles of the posterior distribution of the time varying MA(1) parameter from the Stock Watson representation, which

[20] Since Stock and Watson's model has no AR component, λ_t is set to zero. It is then straightforward to show that stochastic volatility in any predictor can be captured by time variation in β_t.

[21] We consider below, in Section 5.2 cases where there may be one or more persistent predictors of inflation, but which satisfy restrictions that imply cancellation in the reduced form.

displays clearly the gradual drift upwards from the late 1970s as documented by Stock and Watson.[22]

Using these values (and the associated estimates of time varying variances) we obtain in Figure 2 the implied time varying bounds for the predictive R^2. We again plot the the posterior median values of the calculated bounds, using the formulae of Proposition 3.

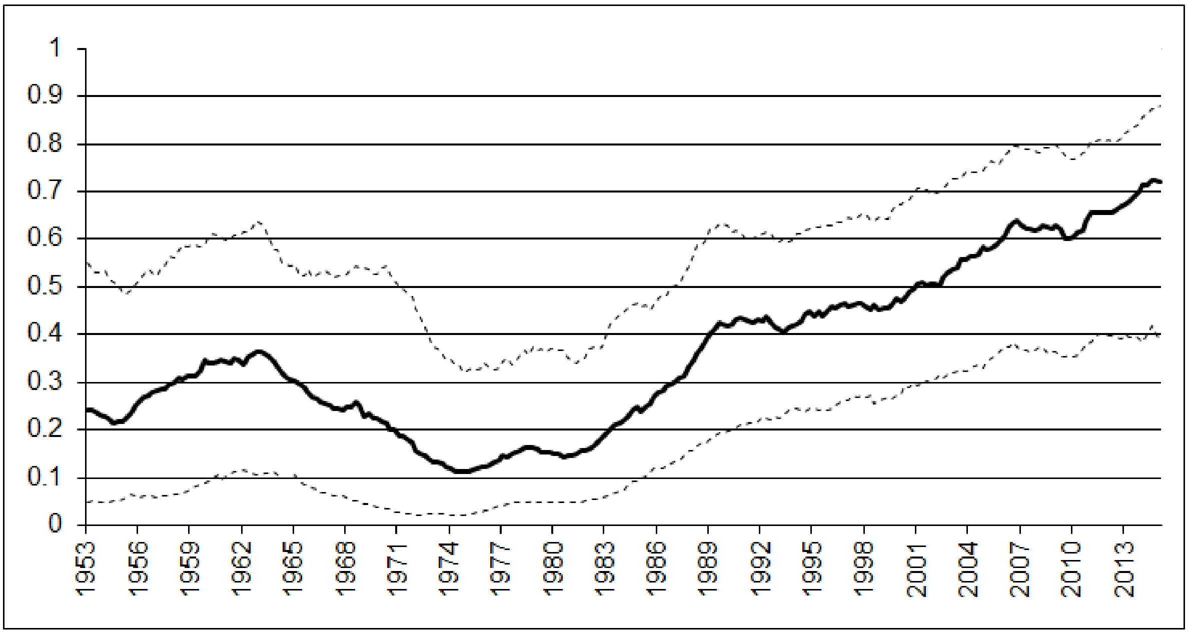

Figure 1: Time varying MA(1) parameter (median and 16.5% & 83.5% quantiles of the posterior distribution)

[22]We estimate, using the code kindly provided on Watson's website, the UC-SV model by Markov Chain Monte Carlo with the same priors as in Stock and Watson (2007).

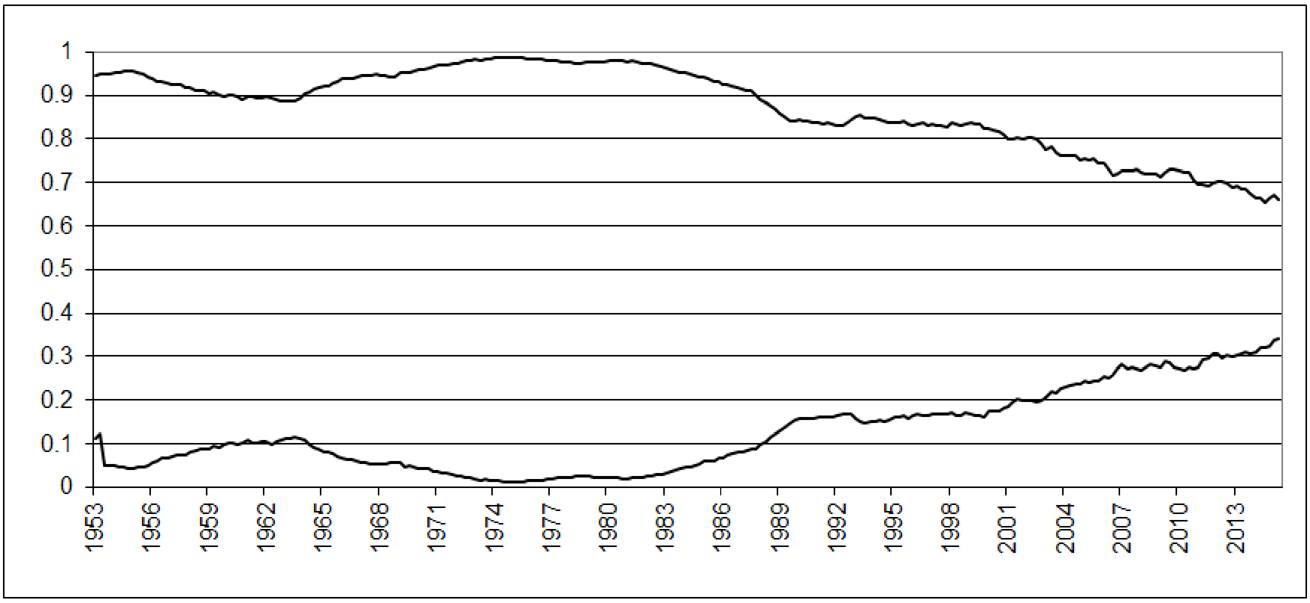

Figure 2: Time varying R^2 bounds: posterior medians

The lower line in Figure 2 gives the predictive R^2 from the time varying MA representation for $\Delta\pi_t$. As the MA parameter has risen the predictability of $\Delta\pi_t$ from its own history has also risen. The upper line shows the bound for predictability from any predictive system consistent with the univariate properties of inflation. This has declined since the 1980s and shows that, even if we could condition on the true state variable for inflation, this would provide only a quite limited improvement in predictability beyond that arising from the history of inflation itself.

Figure 3 shows the 16.5% and 83.5% posterior quantiles of the gap $R^2_{max} - R^2_{min}$; and shows that, even allowing for uncertainty in the estimated parameter, there has been a dramatic decline in the theoretical forecastability of the inflation process since the beginning of the 1980s.[23]

[23] We do not show posterior quantiles for the bounds individually since their posterior distributions are strongly negatively correlated (recall that in the time-invariant case the gap between the two is determined solely by θ).

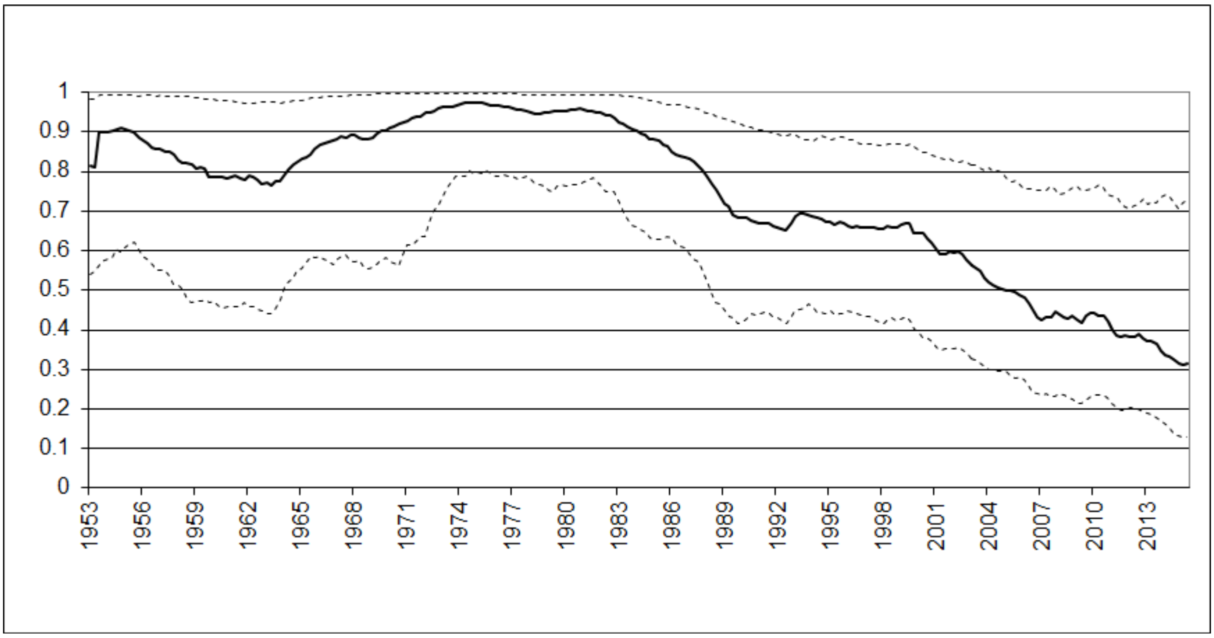

Figure 3: Time varying 16.5% and 83.5% quantiles for $R^2_{max} - R^2_{min}$

The clear conclusion is that the changing nature of the dynamic process for inflation (presumably reflecting changing parameters in some underlying multivariate model that generates the inflation data) is directly reflected in the difficulty of finding variables that can predict better than the simple univariate MA(1).

5.2 If U.S. inflation was generated by a structural model with $r > q = 1$

The illustrations in the previous section of the R^2 bounds implied by Stock and Watson's univariate representation apply to any structural model, the state variables of which can be reduced to a single IID predictor. This raises the obvious question: do these results *only* apply in this case? Is it possible that higher-dimensional systems might escape the upper bound for R^2 that we derive for a single IID predictor?

The simple answer is that, while it is logically possible that the true structural model could indeed escape these bounds, we show that it can only actually do so if it satisfies conditions that are *not* typically satisfied by mainstream macro models.

Inspection of the formula for $R^2_{\max}$ in Proposition 1, shows that, in logic, even a single non-zero eigenvalue in the autoregressive matrix $\mathbf{A}$ of the structural model (implying a single (persistent) predictor in the predictive regression) could result in a value of $R^2_{\max}$ arbitrarily close to unity, if the Macroeconomist's ARMA, (5), has a second MA parameter, θ_2 sufficiently close to zero. But we cannot simply pull structural models with this property out of a hat. As noted in Section 3.4, for any structural model to predict sufficiently well, the time series properties of the predictions it generates must ultimately mimic the time series properties of inflation itself. However, it turns out this is *not* a property that typically arises out of mainstream macreconomic models: indeed the reverse is more likely to be the case.

To keep thing simple, consider initially the possibility that data for inflation were generated by a structural model that can be reduced to a predictive system with two predictors. This would be nested in the system set out in relation to our ARMA(1,1) example in Section 3.3.2.[24] But to be consistent with the univariate properties of inflation as an MA(1) in differences, we would need the AR(1) parameter λ_1 of the first predictor to be zero (or sufficiently close to zero as to be undetectable in the data for y_t). The system would also need to satisfy the restriction in (23) such that $\lambda_2(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) = \theta_2(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ (or, again be sufficiently close to satisfying it as to be undetectable). Under these restrictions, inflation would still be (or be arbitrarily close to) an IMA(1); but, for λ_2 sufficiently close to zero, $R^2_{\max}$ could in principle be arbitrarily close to unity.

However, these restrictions are necessary but *not* sufficient conditions to escape from the bounds we showed in Figure 2. In Appendix H we show analytically that for any value of r, for a predictive model to escape the R^2 bounds in Figure 2 (which are derived on the assumption that $r = q = 1$) requires the predictions it generates to have the distinctive time series property of a variance ratio (Cochrane, 1988) that declines with the forecast horizon.[25]

Most (if not all) multivariate models designed to forecast inflation do *not* have this property. Indeed structural macroeconomic (including DSGE) models usually forecast inflation using some relationship akin to a Phillips Curve, in which the predictor for inflation is some measure of excess demand (e.g., unemployment or the output gap) or marginal costs.[26] The key characteristic of virtually all such predictors is that they have strong persistence, and hence a variance ratio that *in*creases, rather than decreases with the forecast horizon. In Appendix H we show that the benchmark DSGE model of Smets & Wouters (2007) generates predictions for inflation with a variance ratio well above

[24] For simplicity, here we ignore time variation in parameters, which does not affect the nature of the argument.

[25] Or, equivalently, the variance of the Beveridge-Nelson permanent component of $\widehat{y}_t$ must be less than $var(\widehat{y}_t)$.

[26] See, for example, the list of variables analysed by Stock & Watson (2007).

unity. If it were indeed the case that U.S. inflation is generated by a system of this kind, the upper bound for R^2 calculated in the previous section would actually *over*state the scope for a multivariate model to outpredict a univariate model. Indeed, this may help to explain why Stock & Watson (2007) find that similar predictors, far from outperforming their univariate model, typically do worse.

If we are looking for multivariate models of inflation that could do better than Stock & Watson's univariate model, this suggests that we are currently looking in the wrong place. Our results suggest that *either* we should be looking for predictors that are IID (if we restrict ourselves to single predictor models), *or* if we are to consider multiple predictors, that these should result in predicted values that are near-IID processes, but with low persistence.

6 Conclusions

The analysis of this paper shows that what we know about the univariate time-series properties of a process can tell us a lot about the properties of the true but unknown multivariate model that generated the data, even before we look for data on predictors or run a single predictive regression.

We motivated our analysis with reference to two empirical puzzles: the Predictive Puzzle, that multivariate models struggle to outpredict univariate representations; and the Order Puzzle, that ARMA representations are typically of much lower order than would be implied by most multivariate models.

Our analysis of population properties suggests at least a partial resolution of the Predictive Puzzle. We show that, for some y_t processes, even if we could condition our predictions on the true state variables of the true multivariate system that generated the data, we would not be able to out-predict the ARMA by a very wide margin. A prime example is inflation: we show, using the univariate representation of inflation of Stock & Watson (2007), that in recent years the gap between our R^2 bounds has narrowed very significantly.

We must acknowledge that this is only a partial resolution of the Predictive Puzzle. Our analysis suggests that, even when we cannot observe the true state variables of the macroeconomy, *efficient* forecasts using imperfect indicators of the true state variables should allow us to outperform univariate representations by at least a narrow margin. Yet, as we noted at the outset, the evidence for outperformance of ARMA representations by multivariate forecasts is quite limited. (This lack of evidence is particularly marked for, but not limited to, the particular example of inflation that we analyse in our empirical application.)

A common response to this feature of the data is to conclude that the macroeconomy is simply unforecastable. But this would be incorrect. There is plenty of evidence of at least a modest degree of univariate predictability (again, our empirical example of inflation is not an isolated example), with AR(MA) benchmarks commonplace across the macroeconomic forecasting literature: the challenge is to find additional *mult*ivariate predictability that improves on the univariate performance.

Our analysis of the Order Puzzle suggests one avenue that future researchers might pursue. We have shown that resolutions of the Order Puzzle that rely on cancellation or near-cancellation of AR and MA roots in high-order ARMA reduced forms raise new puzzles that are not easily resolved. The alternative resolution is to take the implications of univariate representations seriously. In the case of our empirical example, this suggests that future researchers looking for predictors of inflation should be focussed on IID predictors (or "news"). More generally, low-order ARMA reduced forms suggest that there may just be very few distinct eigenvalues that drive the dynamics of the macroeconomy. Researchers should be looking for multivariate predictability consistent with this, and hence with the univariate evidence.[27]

[27] Our ongoing parallel research provides support for this conjecture. Macqueen & Wright (2015) show that the predictive power of the benchmark Smets-Wouters (2007) DSGE model for the real economy can be replicated by a pair of very simple near-univariate predictive regressions; and Mitchell, Robertson & Wright (2015) show that a tightly restricted and highly parsimonious VARMA that implies ARMA(1,1) univariate forms for all its component time series also captures virtually all the predictability of much larger and more complex macroeconomic models.

References

[1] Atkeson, A. and L.E. Ohanian (2001), "Are Phillips Curves Useful for Forecasting Inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), 2-11.

[2] Banbura, M., D. Giannone and L. Reichlin (2010), "Large Bayesian vector auto regressions", *Journal of Applied Econometrics*, 25(1), 71-92.

[3] Beveridge, S. and C.R. Nelson (1981), "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'", *Journal of Monetary Economics*, 7(2), 151-174.

[4] Canova, F. (2007), *Methods for Applied Macroeconomic Research.* Princeton University Press.

[5] Chauvet, M. and S. Potter (2013), "Forecasting Output". In *Handbook of Economic Forecasting, Volume 2A*, (eds. Elliott, G. and A. Timmermann), Amsterdam: Elsevier, pp. 1–56.

[6] Cochrane, J.H. (1988), "How big is the random walk in GDP? ", *Journal of Political Economy*, 96(5), 893-92.

[7] Cogley, T. and T. Sargent (2005), "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.", *Review of Economic Dynamics*, 8, 262-302.

[8] Cubadda, G., A. Hecq and F.C. Palm (2009), "Studying co-movements in large multivariate data prior to multivariate modelling", *Journal of Econometrics*, 148(1), 25-35.

[9] D'Agostino, A. and P. Surico (2012), "A Century of Inflation Forecasts", *Review of Economics and Statistics*, 94(4), 1097-1106.

[10] D'Agostino, A, L. Gambetti and D. Giannone (2013), "Macroeconomic forecasting and structural change", *Journal of Applied Econometrics*, 28(1), 82-101.

[11] Diebold, F.X. and G.D. Rudebusch (1989), "Long memory and persistence in aggregate output", *Journal of Monetary Economics*, 24(2), 189-209.

[12] Estrella, A. and J.H. Stock (2015), "A State-Dependent Model for Inflation Forecasting", with James H. Stock, in *Unobserved Components and Time Series Econometrics*, Oxford University Press, Ch 3, 14-29.

[13] Fernández-Villaverde, J., J.F. Rubio-Ramírez, T.J. Sargent and M.W. Watson (2007), "ABCs (and Ds) of Understanding VARs", *American Economic Review*, 97(3), 1021-1026.

[14] Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37(3), 424-38.

[15] Hamilton, J.D. (1994), *Time Series Analysis.* Princeton University Press.

[16] Hansen, L.P. and T.J. Sargent (2013), *Recursive Models of Dynamic Linear Economies*, Princeton University Press.

[17] Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

[18] Lippi, M. and L. Reichlin (1994), "VAR analysis, nonfundamental representations, Blaschke matrices", *Journal of Econometrics*, 63, 307-325.

[19] Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*, Springer.

[20] Macqueen, R. and S. Wright (2015), "DSGE Models and predictability in the real economy", *Working Paper.*

[21] Mitchell, J., D. Robertson and S. Wright (2015), "The predictable economy is simpler than you thought", *Working Paper.*

[22] Nelson, C.R. (1972), "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review*, 62, 902-917.

[23] Pástor, L. and R.F. Stambaugh (2009), "Predictive Systems: Living with Imperfect Predictors", *Journal of Finance*, 64(4), 1583-1628.

[24] Robertson, D. and S. Wright (2012), "The Predictive Space: If $\mathbf{x}$ predicts y, what does y tell us about $\mathbf{x}$?", Birkbeck College Working Paper.

[25] Rossi, B. (2013a), "Exchange Rate Predictability", *Journal of Economic Literature*, 51(4), 1063-1119.

[26] Rossi, B. (2013b), "Advances in Forecasting under Model Instability". In: G. Elliott and A. Timmermann (eds.) *Handbook of Economic Forecasting*, Volume 2B, Amsterdam: Elsevier-North Holland.

[27] Smets, F. and R. Wouters (2007), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach", *American Economic Review*, 97(3), 586-606.

[28] Stock, J.H and M.W. Watson (1998), "Testing For Common Trends", J*ournal of the American Statistical Association*, 83, 1097-1107.

[29] Stock, J.H and M.W. Watson (2002), "Forecasting Using Principal Components From a Large Number of Predictors", *Journal of the American Statistical Association*, 97, 1167-1179.

[30] Stock, J.H and M.W. Watson (2007), "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money Credit and Banking*, 39, s1, 13-33.

[31] Stock, J.H. and M.W. Watson (2009), "Phillips Curve Inflation Forecasts". In *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective*, Federal Reserve Bank of Boston.

[32] Stock, J.H. and M.W. Watson (2010), "Modeling Inflation After the Crisis", FRB Kansas City symposium, Jackson Hole, Wyoming, August 26-28, 2010

[33] Stock, J.H. and M.W. Watson (2015), "Core Inflation and Trend Inflation", working paper Princeton University.

[34] Wallis, K.F. (1977), "Multiple time series analysis and the final form of econometric models", *Econometrica*, 45, 1481-1497.

[35] Zellner, A. and F.C. Palm (1974), "Time series analysis and simultaneous equation econometric models", *Journal of Econometrics*, 2, 17-54.

A Proof of Lemma 1.

We can write the state equation (1) as

$$\begin{aligned}\mathbf{z}_t &= \mathbf{T}^{-1}\mathbf{M}^*\mathbf{T}\mathbf{z}_{t-1} + \mathbf{B}\mathbf{s}_t \\ \mathbf{T}\mathbf{z}_t &= \mathbf{M}^*\mathbf{T}\mathbf{z}_{t-1} + \mathbf{T}\mathbf{B}\mathbf{s}_t \\ \mathbf{x}_t^* &= \mathbf{M}^*\mathbf{x}_{t-1}^* + \mathbf{v}_t^*\end{aligned}$$

where $\mathbf{v}_t^* = \mathbf{T}\mathbf{B}\mathbf{s}_t$, with observables equation (2) as

$$\begin{aligned}\mathbf{y}_t &= \mathbf{C}\mathbf{T}^{-1}\mathbf{T}\mathbf{z}_{t-1} + \mathbf{D}\mathbf{s}_t \\ &= \mathbf{C}\mathbf{T}^{-1}\mathbf{x}_{t-1}^* + \mathbf{D}\mathbf{s}_t\end{aligned}$$

where $\mathbf{x}_t^* = \mathbf{T}\mathbf{z}_t$ is $n \times 1$. Then, letting

$$\boldsymbol{\beta}^{*\prime} = \begin{bmatrix} 1 & \mathbf{0}_{1\times(n-1)} \end{bmatrix} \mathbf{C}\mathbf{T}^{-1}$$

we can write

$$y = \boldsymbol{\beta}^{*\prime}\mathbf{x}_{t-1}^* + u_t$$

where

$$u_t = \begin{bmatrix} 1 & \mathbf{0}_{1\times(n-1)} \end{bmatrix} \mathbf{D}\mathbf{s}_t$$

But this representation may in principle have state variables with identical eigenvalues (e.g., multiple IID states). To derive a minimal representation define a $r \times n$ matrix $\mathbf{K}$, such that

$$\mathbf{x}_t = \mathbf{K}\mathbf{x}_t^*$$

where $\mathbf{x}_t$ is $r \times 1$ and

$$\mathbf{K}_{ij} = 1\left(\mathbf{M}_{ii}^* = \mathbf{M}_{jj}^*\right)\frac{\beta_j^*}{\beta_i^*};\ i = 1, ..., r;\ j = 1, ..., n$$

that is, each element of $\mathbf{x}_t^*$ with a common eigenvalue is weighted by its relative β. Then we can write

$$y_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t \tag{29}$$

as in Lemma 1 with

$$\boldsymbol{\beta} = \begin{bmatrix} \overset{r\times r}{\mathbf{I}} & \overset{r\times(n-r)}{\mathbf{0}} \end{bmatrix} \boldsymbol{\beta}^*$$

so that $\boldsymbol{\beta}$ contains the first r elements of $\boldsymbol{\beta}^*$, with a law of motion for the minimal state vector, as in Lemma 1

$$\mathbf{x}_t = \mathbf{M}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{30}$$

where $\mathbf{M} = diag\left(\lambda_1, ..., \lambda_r\right)$ and $\mathbf{v}_t = \mathbf{K}\mathbf{v}_t^*$.

To derive (30), partition $\mathbf{K}$, $\mathbf{x}_t^*$ and $\mathbf{M}^*$ conformably as

$$\mathbf{K} = \left[\begin{array}{cc} \underset{r\times r}{\mathbf{I}} & \underset{r\times(n-r)}{\mathbf{K}_{12}} \end{array}\right]; \quad \mathbf{x}_t^* = \left[\begin{array}{c} \underset{r\times 1}{\mathbf{x}_{1t}^*} \\ \underset{(n-r)\times 1}{\mathbf{x}_{2t}^*} \end{array}\right]; \quad \mathbf{M}^* = \left[\begin{array}{cc} \underset{r\times r}{\mathbf{M}} & \underset{r\times(n-r)}{\mathbf{0}} \\ \underset{(n-r)\times(n-r)}{\mathbf{0}} & \underset{(n-r)\times r}{\mathbf{F}'}\ \mathbf{M}\ \underset{r\times(n-r)}{\mathbf{F}} \end{array}\right]$$

with $\mathbf{F}_{ij} = 1\left(\mathbf{K}_{(r+i)j} > 0\right)$, so $\mathbf{F}'\mathbf{M}\mathbf{F}$ selects the repetitions of eigenvalues in $\mathbf{M}^*$. Note that both $\mathbf{K}_{12}$ and $\mathbf{F}$ have a single non-zero element in each column, and each $1 \times (n-r)$ row of $\mathbf{F}$ is a vector of modulus zero or unity The non-zero elements of $\mathbf{F}$ occupy the same positions as the non-zero elements of $\mathbf{K}_{12}$, since $\mathbf{F}$ is recording which elements of $\mathbf{x}_{2t}^*$ have the same eigenvalues and $\mathbf{K}_{12}$ constructs the appropriate weighted aggregates of those variables.

Now

$$\mathbf{x}_t = \mathbf{K}\mathbf{x}_t^* = \left[\begin{array}{cc} \mathbf{I}_r & \mathbf{K}_{12} \end{array}\right]\left[\begin{array}{c} \mathbf{x}_{1t}^* \\ \mathbf{x}_{2t}^* \end{array}\right] = \mathbf{x}_{1t}^* + \mathbf{K}_{12}\mathbf{x}_{2t}^*$$

and

$$\mathbf{x}_t = \mathbf{K}\mathbf{x}_t^* = \mathbf{K}\mathbf{M}^*\mathbf{x}_{t-1}^* + \mathbf{K}\mathbf{v}_t^*$$

so

$$\mathbf{K}\mathbf{M}^* = \left[\begin{array}{cc} \mathbf{I}_r & \mathbf{K}_{12} \end{array}\right]\left[\begin{array}{cc} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}'\mathbf{M}\mathbf{F} \end{array}\right] = \left[\begin{array}{cc} \mathbf{M} & \mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F} \end{array}\right]$$

Thus

$$\begin{aligned} \mathbf{x}_t &= \left[\begin{array}{cc} \mathbf{M} & \mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F} \end{array}\right]\left[\begin{array}{c} \mathbf{x}_{1t-1}^* \\ \mathbf{x}_{2t-1}^* \end{array}\right] + \mathbf{K}\mathbf{v}_t^* \\ &= \mathbf{M}\mathbf{x}_{1t-1}^* + \mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F}\mathbf{x}_{2t-1}^* + \mathbf{K}\mathbf{v}_t^* \end{aligned}$$

Now $\mathbf{x}_t = \mathbf{x}_{1t}^* + \mathbf{K}_{12}\mathbf{x}_{2t}^*$ so we can write this as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{M}\left(\mathbf{x}_{t-1} - \mathbf{K}_{12}\mathbf{x}_{2t-1}^*\right) + \mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F}\mathbf{x}_{2t-1}^* + \mathbf{K}\mathbf{v}_t^* \\ &= \mathbf{M}\mathbf{x}_{t-1} + \mathbf{K}\mathbf{v}_t^* + \left(\mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F} - \mathbf{M}\mathbf{K}_{12}\right)\mathbf{x}_{2t-1}^* \end{aligned}$$

so we require $\mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F} - \mathbf{M}\mathbf{K}_{12} = \mathbf{0}$ for (30) to be correct.

To show this, first note that $\mathbf{K}_{12}\mathbf{F}'$ and $\mathbf{F}\mathbf{F}'$ are both diagonal $r \times r$ matrices (hence symmetrical) with non-zero elements on the diagonal corresponding to the rows of $\mathbf{K}_{12}$ (or $\mathbf{F}$) that have non-zero elements. For $\mathbf{F}\mathbf{F}'$ these elements are unity. The number of 1's on the diagonal of $\mathbf{F}\mathbf{F}'$ equals the column rank of $\mathbf{F}$. Thus

$$\mathbf{K}_{12}\mathbf{F}'\mathbf{M}\mathbf{F} - \mathbf{M}\mathbf{K}_{12} = \mathbf{F}\mathbf{K}_{12}'\mathbf{M}\mathbf{F} - \mathbf{M}\mathbf{K}_{12} = \mathbf{F}\mathbf{F}'\mathbf{M}\mathbf{K}_{12} - \mathbf{M}\mathbf{K}_{12} = \left(\mathbf{F}\mathbf{F}' - \mathbf{I}_r\right)\mathbf{M}\mathbf{K}_{12}$$

since $\mathbf{K}_{12}'\mathbf{M}\mathbf{F}$ is also a symmetric matrix. $\mathbf{F}\mathbf{F}'$ is a diagonal matrix with zeros or ones on the leading diagonal. The non-zero elements are in the rows corresponding to non-zero rows of $\mathbf{F}$ (and hence also of $\mathbf{K}_{12}$ and $\mathbf{M}\mathbf{K}_{12}$). So $\mathbf{F}\mathbf{F}'$ acts as an identity matrix on anything in the column space of $\mathbf{F}$ and therefore $\mathbf{F}\mathbf{F}'\mathbf{M}\mathbf{K}_{12} = \mathbf{M}\mathbf{K}_{12}$ ($\mathbf{F}\mathbf{F}'$ picks unchanged the non-zero rows of $\mathbf{M}\mathbf{K}_{12}$ and multiplies the remaining rows by zero). Hence $\left(\mathbf{F}\mathbf{F}' - \mathbf{I}_r\right)\mathbf{M}\mathbf{K}_{12} = 0$ so we obtain the transition equation (30) above.■

B Proof of Lemma 2

After substitution from (4) the predictive regression (3) can be written as

$$\det\left(\mathbf{I} - \mathbf{M}L\right) y_t = \boldsymbol{\beta}'\operatorname{adj}\left(\mathbf{I} - \mathbf{M}L\right)\mathbf{v}_{t-1} + \det\left(\mathbf{I} - \mathbf{M}L\right) u_t \tag{31}$$

Given diagonality of $\mathbf{M}$, from A1, we can rewrite this as

$$\widetilde{y}_t \equiv \prod_{i=1}^{r}\left(1 - \lambda_i L\right) y_t = \sum_{i=1}^{r}\beta_i \prod_{j \neq i}\left(1 - \lambda_j L\right) L v_{it} + \prod_{i=1}^{r}\left(1 - \lambda_i L\right) u_t \equiv \sum_{i=0}^{r}\boldsymbol{\gamma}_i' L^i \mathbf{w}_t \tag{32}$$

wherein $\widetilde{y}_t$ is then an MA(r), $\mathbf{w}_t = \begin{pmatrix} u_t & \mathbf{v}_t' \end{pmatrix}'$, and the final equality implicitly defines a set of $(r+1)-$vectors, $\boldsymbol{\gamma}_i\left(\boldsymbol{\beta}, \lambda\right)$, for $i = 0, \ldots, r$ each of which is $(r+1) \times 1$.

Let acf_i be the ith order autocorrelation of $\widetilde{y}_t$ implied by the predictive system. Write $\boldsymbol{\Gamma} = E\left(\mathbf{w}_t\mathbf{w}_t'\right)$ and we have straightforwardly

$$acf_i\left(\boldsymbol{\beta}, \lambda, \boldsymbol{\Omega}\right) = \frac{\sum_{j=0}^{r-i}\boldsymbol{\gamma}_j'\boldsymbol{\Omega}\boldsymbol{\gamma}_{j+i}}{\sum_{j=0}^{r}\boldsymbol{\gamma}_j'\boldsymbol{\Omega}\boldsymbol{\gamma}_j} \tag{33}$$

To obtain explicitly the coefficients of the MA(r) representation write the right hand side of (32) as an $MA(r)$ process $\sum_{i=0}^{r}\boldsymbol{\gamma}_i' L^i \mathbf{w}_t = \prod_{i=1}^{r}\left(1 - \theta_i L\right)\varepsilon_t = \theta\left(L\right)\varepsilon_t$ for some white noise process ε_t and r^{th} order lag polynomial $\theta\left(L\right)$.

The autocorrelations of $\theta\left(L\right)\varepsilon_t$ are derived as follows. Define a set of parameters c_i

by

$$\prod_{i=1}^{r}(1-\theta_i L)=1+c_1L+c_2L^2+...+c_rL^r \tag{34}$$

Then the ith order autocorrelation of $\theta(L)\varepsilon_t$is given by (Hamilton, 1994, p.51)

$$\frac{c_i+c_{i+1}c_1+c_{i+2}c_2+....+c_rc_{r-i}}{1+c_1^2+c_2^2+...+c_r^2} \quad i=1,\ldots,r \tag{35}$$

Equating these to the ith order autocorrelations of $\widetilde{y}_t$ we obtain a system of moment equations

$$\frac{c_i+c_{i+1}c_1+c_{i+2}c_2+....+c_rc_{r-i}}{1+c_1^2+c_2^2+...+c_r^2}=acf_i\left(\boldsymbol{\beta},\lambda,\boldsymbol{\Omega}\right), \quad i=1,\ldots,r \tag{36}$$

which can be solved for c_i, $i=1,\ldots,r$, and hence for θ_i. The solutions are chosen such that $|\theta_i|<1$, $\forall i$.■

C Proof of Proposition 1

We start by establishing the importance of two of the set of possible ARMA representations.

Lemma 3 *In the set of all possible nonfundamental ARMA(r,r) representations consistent with (5) in which $\theta_i>0$ $\forall i$, and θ_i is replaced with θ_i^{-1} for at least some i, the moving average polynomial $\theta^N(L)$ in (10) in which θ_i is replaced with θ_i^{-1} for all i, has innovations η_t with the minimum variance, with*

$$\sigma_\eta^2=\sigma_\varepsilon^2\prod_{i=1}^{q}\theta_i^2 \tag{37}$$

Proof. Equating (5) to (10) the non-fundamental and fundamental innovations are related by

$$\varepsilon_t=\prod_{i=1}^{r}\left(\frac{1-\theta_i^{-1}L}{1-\theta_iL}\right)\eta_t=\sum_{j=0}^{\infty}c_j\eta_{t-j} \tag{38}$$

for some square summable c_j. Therefore, since η_t is itself IID,

$$\sigma_\varepsilon^2=\sigma_\eta^2\sum_{j=0}^{\infty}c_j^2 \tag{39}$$

Now define

$$c(L) = \sum_{j=0}^{\infty} c_j L^j = \prod_{i=1}^{r} \left(\frac{1 - \theta_i^{-1} L}{1 - \theta_i L} \right) \tag{40}$$

so

$$c(1) = \prod_{i=1}^{r} \left(\frac{1 - \theta_i^{-1}}{1 - \theta_i} \right) = \prod_{i=1}^{r} \left(\frac{-1}{\theta_i} \right) \tag{41}$$

and

$$c(1)^2 = \prod_{i=1}^{r} \frac{1}{\theta_i^2} = \left(\sum_{j=0}^{\infty} c_j \right)^2 = \sum_{j=0}^{\infty} c_j^2 + \sum_{k \neq j} c_j c_k \tag{42}$$

Since ε_t is IID we have

$$E(\varepsilon_t \varepsilon_{t+k}) = 0 \ \ \forall k > 0$$

implying

$$\sum_{j=0}^{\infty} c_j c_{j+k} = 0 \ \forall k > 0 \tag{43}$$

Hence we have

$$\sum_{j \neq k} c_j c_k = 2 \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+k} = 0 \tag{44}$$

thus

$$\sum_{j=0}^{\infty} c_j^2 = c(1)^2 = \prod_{i=1}^{r} \frac{1}{\theta_i^2} \tag{45}$$

Thus using (45) and (39) we have (37).

To show that this is the nonfundamental representation with the minimum innovation variance, consider the full set of nonfundamental ARMA(r, r) representations, in which, for each representation k, $k = 1, \ldots, 2^r - 1$, there is some ordering such that, θ_i is replaced with θ_i^{-1}, $i = 1, \ldots, s(k)$, for $s \leq r$. For any such representation, with innovations $\eta_{k,t}$, we have

$$\sigma_{\eta,k}^2 = \sigma_{\varepsilon}^2 \prod_{i=1}^{s(k)} \theta_i^2 \tag{46}$$

This is minimised for $s(k) = r$, which is only the case for the single representation in which θ_i is replaced with θ_i^{-1} for all i, and thus this will give the minimum variance nonfundamental representation. Note it also follows that the fundamental representation itself has the maximal innovation variance amongst all representations.■

We now define the R^2 of the (maximal innovation variance) fundamental and this (minimal innovation variance) non-fundamental representations as follows

$$R_F^2 = R_F^2(\lambda, \theta) = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2} \tag{47}$$

and

$$R_N^2 = R_N^2(\lambda, \theta) = 1 - \frac{\sigma_\eta^2}{\sigma_y^2} \tag{48}$$

and note that immediately from the above we have

$$R_N^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) = 1 - \left(1 - R_F^2(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \prod_{i=1}^{r} \theta_i^2 \tag{49}$$

Also for the predictive model $y_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t$ we have

$$R^2 = \frac{\sigma_{\widehat{y}}^2}{\sigma_{\widehat{y}}^2 + \sigma_u^2} \tag{50}$$

where

$$\sigma_{\widehat{y}}^2 = \boldsymbol{\beta}' E\left(\mathbf{x}_t \mathbf{x}_t'\right) \boldsymbol{\beta} \tag{51}$$

We now show that we can recast the macroeconomist's ARMA (which we now write as $\lambda(L)y_t = \theta(L)\varepsilon_t$) into fundamental and nonfundamental predictive representations.

Start from these two ARMA representations

$$\prod_{i=1}^{r}(1 - \lambda_i L)\, y_t = \prod_{i=1}^{r}(1 - \theta_i L)\, \varepsilon_t$$
$$\prod_{i=1}^{r}(1 - \lambda_i L)\, y_t = \prod_{i=1}^{r}\left(1 - \theta_i^{-1} L\right) \eta_t$$

Define $r \times 1$ coefficient vectors $\boldsymbol{\beta}_F = \left(\beta_{F,1}, \ldots, \beta_{F,r}\right)'$ and $\boldsymbol{\beta}_N = \left(\beta_{N,1}, \ldots, \beta_{N,r}\right)'$ that satisfy respectively

$$1 + \sum_{i=1}^{r} \frac{\beta_{F,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^{r}(1 - \theta_i L)}{\prod_{i=1}^{r}(1 - \lambda_i L)} \tag{52}$$

$$1 + \sum_{i=1}^{r} \frac{\beta_{N,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^{r}\left(1 - \theta_i^{-1} L\right)}{\prod_{i=1}^{r}(1 - \lambda_i L)} \tag{53}$$

We can then define two $r \times 1$ vectors of "univariate predictors" (which we label as fundamental (F) and nonfundamental (N)) by

$$\mathbf{x}_t^F = \mathbf{M}\mathbf{x}_{t-1}^F + \mathbf{1}\varepsilon_t \tag{54}$$

$$\mathbf{x}_t^N = \mathbf{M}\mathbf{x}_{t-1}^N + \mathbf{1}\eta_t \tag{55}$$

where by construction we can now represent the (fundamental and nonfundamental) ARMAs for y_t as predictive regressions

$$y_t = \boldsymbol{\beta}_F'\mathbf{x}_{t-1}^F + \varepsilon_t \tag{56}$$

$$y_t = \boldsymbol{\beta}_N'\mathbf{x}_{t-1}^N + \eta_t \tag{57}$$

The predictive regressions in (56) and (57), together with the processes for the two univariate predictor vectors in (54) and (55), are both special cases of the general predictive system in (3) and (4), but with rank 1 covariance matrices, $\Omega^F = \sigma_\varepsilon^2\mathbf{11}'$, and $\Omega^N = \sigma_\eta^2\mathbf{11}'$.[28] We shall show below that the properties of the two special cases provide us with important information about *all* predictive systems consistent with the history of y_t. We note that, since these predictive regressions are merely rewrites of their respective ARMA representations, the R^2 of these predictive regressions must match those of the underlying ARMAs (each of which can be expressed as a function of the ARMA coefficients). That is:

1. The fundamental predictive regression $y_t = \boldsymbol{\beta}_F'\mathbf{x}_{t-1}^F + \varepsilon_t$ has $R^2 = R_F^2(\lambda, \theta)$.

2. The nonfundamental predictive regression $y_t = \boldsymbol{\beta}_N'\mathbf{x}_{t-1}^N + \eta_t$ has $R^2 = R_N^2(\lambda, \theta)$.

We now proceed by proving two results that lead straightforwardly to the Proposition itself.

Lemma 4 *In the population regression*

$$y_t = \boldsymbol{\nu}_{\mathbf{x}}'\mathbf{x}_{t-1} + \boldsymbol{\nu}_F'\mathbf{x}_{t-1}^F + \xi_t \tag{58}$$

where the true process for y_t is as in (3), and $\mathbf{x}_t^F$ is the vector of fundamental univariate predictors defined in (54), all elements of the coefficient vector $\boldsymbol{\nu}_F$ are zero.

[28]Note that we could also write (56) as $y_t = \boldsymbol{\beta}'\widehat{\mathbf{x}}_{t-1} + \varepsilon_t$; where $\widehat{\mathbf{x}}_t = E\left(\mathbf{x}_t | \{y_i\}_{i=-\infty}^t\right)$ is the optimal estimate of the predictor vector given the single observable y_t and the state estimates update by $\widehat{x}_t = \boldsymbol{\Lambda}\widehat{x}_{t-1} + \mathbf{k}\varepsilon_t$, where $\mathbf{k}$ is a vector of steady-state Kalman gain coefficients (using the Kalman gain definition as in Harvey, 1989). The implied reduced form process for y_t must be identical to the fundamental ARMA representation (Hamilton, 1994) hence we have $\beta_{F,i} = \beta_i k_i$.

Proof. The result will follow automatically if we can show that the x^F_{it-1} are all orthogonal to $u_t \equiv y_t - \beta' \mathbf{x}_{t-1}$. Equalising (5) and (3), and substituting from (4), we have (noting that $p = q = r$)

$$y_t = \frac{\prod_{i=1}^r (1-\theta_i L)}{\prod_{i=1}^r (1-\lambda_i L)} \varepsilon_t = \frac{\beta_1 v_{1t-1}}{1-\lambda_1 L} + \frac{\beta_2 v_{2t-1}}{1-\lambda_2 L} + \ldots + \frac{\beta_r v_{rt-1}}{1-\lambda_r L} + u_t \tag{59}$$

So we may write, using (54),

$$\begin{aligned} x^F_{jt-1} &= \frac{\varepsilon_{t-1}}{1-\lambda_j L} \\ &= \left(\frac{L}{1-\lambda_j L}\right) \frac{\prod_{i=1}^r (1-\lambda_i L)}{\prod_{i=1}^r (1-\theta_i L)} \times \\ &\left(\frac{\beta_1 L v_{1t-1}}{1-\lambda_1 L} + \frac{\beta_2 L v_{2t-1}}{1-\lambda_2 L} + \ldots + \frac{\beta_r L v_{rt-1}}{1-\lambda_r L} + u_t\right) \end{aligned} \tag{60}$$

Given the assumption that u_t and the v_{it} are jointly IID, u_t will indeed be orthogonal to x^F_{jt-1}, for all j, since the expression on the right-hand side involves only terms dated $t-1$ and earlier, thus proving the Lemma. ■

Lemma 5 *In the population regression*

$$y_t = \boldsymbol{\phi}'_{\mathbf{x}} \mathbf{x}_{t-1} + \boldsymbol{\phi}'_N \mathbf{x}^N_{t-1} + \zeta_t \tag{61}$$

where $\mathbf{x}^N_t$ is the vector of nonfundamental univariate predictors defined in (55), all elements of the coefficient vector $\boldsymbol{\phi}_{\mathbf{x}}$ are zero.

Proof. The result will again follow automatically if we can show that the x_{it-1} are all orthogonal to $\eta_t \equiv y_t - \beta'_N \mathbf{x}^N_{t-1}$. Equating (10) and (3), and substituting from (4), we have

$$y_t = \frac{\prod_{i=1}^r (1-\theta_i^{-1} L)}{\prod_{i=1}^r (1-\lambda_i L)} \eta_t = \beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_r \frac{v_{rt-1}}{1-\lambda_r L} + u_t \tag{62}$$

Using

$$\frac{1}{1-\theta_i^{-1} L} = \frac{-\theta_i F}{1-\theta_i F}$$

where F is the forward shift operator, $F = L^{-1}$, we can write

$$\eta_t = F^r \prod_{i=1}^{r}(-\theta_i)\left(\frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right)\left(\beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_r \frac{v_{rt-1}}{1-\lambda_r L} + u_t\right) \tag{63}$$

Now

$$\begin{aligned} F^r \frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)} \frac{v_{kt-1}}{(1-\lambda_k L)} &= F^r \left(\frac{\prod_{i\neq k}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right) v_{kt-1} \\ &= v_{kt} + c_1 v_{kt+1} + c_2 v_{kt+2} + \ldots \end{aligned}$$

for some $c_1, c_2, ...$ since the highest order term in L in the numerator of the bracketed expression is of order $r-1$, and

$$F^r \left(\frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right) u_t = u_t + b_1 u_{t+1} + b_2 u_{t+2} + \ldots$$

for some $b_1, b_2, \ldots$, since the highest order term in L in the numerator of the bracketed expression is r. Hence η_t can be expressed as a weighted average of current and forward values of u_t and v_{it} and will thus be orthogonal to $x_{it-1} = \frac{v_{it-1}}{1-\lambda_i L}$ for all i, by the assumed joint IID properties of u_t and the v_{it}, thus proving the Lemma. ■

Now let $R_1^2 = 1 - \sigma_\xi^2/\sigma_y^2$ be the predictive R^2 of the regression (58) analysed in Lemma 4. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^F$ in (56) are both nested within (58) we must have $R_1^2 \geq R^2$ and $R_1^2 \geq R_F^2$. But Lemma 4 implies that, given $\boldsymbol{\nu}_F = 0$ we must have $R_1^2 = R^2$, hence $R^2 \geq R_F^2$.

By a similar argument, let $R_2^2 = 1 - \sigma_\zeta^2/\sigma_y^2$ be the predictive R^2 of the predictive regression (61) analysed in Lemma 5. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^N$ in (57) are both nested in (61) we must have $R_2^2 \geq R^2$ and $R_2^2 \geq R_N^2$. But Lemma 5 implies that, given $\boldsymbol{\phi}_{\mathbf{x}} = 0$ we must have $R_2^2 = R_N^2$, hence $R_N^2 \geq R^2$. We show below that R_F^2 and R_N^2 give the minimum and maximum values of R^2 from all possible (fundamental and non-fundamental) ARMA representations for y_t. Thus writing $R_F^2 = R_{min}^2(\boldsymbol{\lambda}, \boldsymbol{\theta})$ and $R_N^2 = R_{max}^2(\boldsymbol{\lambda}, \boldsymbol{\theta})$ we have

$$R_{min}^2(\lambda, \theta) \leq R^2 \leq R_{max}^2(\boldsymbol{\lambda}, \boldsymbol{\theta})$$

as given in the Proposition.

Moreover these inequalities will be strict unless the predictor vector $\mathbf{x}_t$ matches either the fundamental predictor $\mathbf{x}_t^F$ or the nonfundamental predictor $\mathbf{x}_t^N$ in which case the innovations to the predictor variable match those in the relevant ARMA representation. In the **A**,**B**,**C**,**D** system this occurs only if $rank \begin{bmatrix} \text{B} \\ \text{C} \end{bmatrix} = 1$.

This completes the proof of the Proposition. ■

D Proof of Corollary 1

The proof follows as a direct consequence of efficient filtering, given some observation equation for the observables, $\mathbf{q}_t$: the vector of state estimates, $\widehat{\mathbf{x}}_t$, will have the same autoregressive form as the process in (4) for the true predictor vector (Hansen & Sargent, 2013, Chapter 8), with innovations, $\widehat{\mathbf{v}}_t$, that, given efficient filtering, are jointly IID with the innovations to the associated predictive regression $y_t = \boldsymbol{\beta}'\widehat{\mathbf{x}}_{t-1} + \widehat{u}_t$, which takes the same form as (3). Given that the resulting predictive system is of the same form, the proof of Proposition 1 must also apply.■

E Proof of Corollary 3

Define $\rho = corr\left(E\left(y_{t+1}|\mathbf{x}_t\right), E\left(y_{t+1}|y^t\right)\right) = corr\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right)$. We have

$$R_F^2 = corr\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, y_{t+1}\right)^2 = \frac{\left[Cov\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t + u_{t+1}\right)\right]^2}{Var\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F\right) Var\left(y_{t+1}\right)} \tag{64}$$

$$= \frac{\left[Cov\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right)\right]^2}{Var\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F\right) Var\left(\boldsymbol{\beta}'\mathbf{x}_t\right)} \frac{Var\left(\boldsymbol{\beta}'\mathbf{x}_t\right)}{Var\left(y_{t+1}\right)} \tag{65}$$

$$= \left[corr\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right)\right]^2 R^2 \tag{66}$$

where we use as derived in the Proof of Proposition 1 above that $Cov\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, u_t\right) = 0$.

This then gives, exploiting the inequality in the Proposition,

$$\rho = corr\left(\boldsymbol{\beta}_F'\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right) = \sqrt{\frac{R_F^2}{R^2}} \geq \sqrt{\frac{R_{\min}^2\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R_{\max}^2\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}} \geq 0 \tag{67}$$

completing the proof of the Corollary.■

F Proof of Proposition 2

Given that the macroeconomist's ARMA representation in (5) is ARMA(r,r) the minimal ARMA(p,q) representation will only be of lower order if we have *either* cancellation of some MA and AR roots, *or* an MA or AR coefficient precisely equal to zero. Thus we have

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

thus unless $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ satisfy exact restrictions such that there are zero coefficients or cancellation in the macroeconomist's ARMA we have $r = p = q$. Furthermore for $q > 0$ we have $R_F^2 > 0$ and $R_N^2 < 1$. hence the bounds lie strictly within $[0,1]$.■

G Proof of Proposition 3

Consider first the special case with $\lambda_t = 0$, hence $x_t = v_t$. Without loss of generality we can set $\beta_t = 1$, and re-write (24) as

$$y_t = v_{t-1} + u_t \tag{68}$$

with

$$\Omega_t = \begin{bmatrix} \sigma_{v,t}^2 & \sigma_{uv,t} \\ \sigma_{uv,t} & \sigma_{u,t}^2 \end{bmatrix}$$

This nests both (26) and (28) as special cases, with

$$\Omega_t^F = \sigma_{\varepsilon,t}^2 \begin{bmatrix} 1 & -\theta_t \\ -\theta_t & \theta_t^2 \end{bmatrix}$$

$$\Omega_t^N = \sigma_{\eta,t}^2 \begin{bmatrix} 1 & -s_t/\theta_t \\ -s_t/\theta_t & (s_t\theta_t)^2 \end{bmatrix}$$

where θ_t and $\sigma_{\varepsilon,t}^2$ jointly satisfy the moment conditions

$$\sigma_{y,t}^2 = \sigma_{\varepsilon,t}^2 + \theta_t \sigma_{\varepsilon,t-1}^2 = \sigma_{v,t-1}^2 + \sigma_{u_t}^2$$
$$cov_t\left(y_t, y_{t-1}\right) = -\theta_t \sigma_{\varepsilon,t-1}^2 = \rho_{t-1}\sigma_{v,t-1}\sigma_{u,t-1}$$

where $\rho_t = corr\,(u_t, v_t) \equiv \sigma_{uv,t}/\left(\sigma_{u,t}\sigma_{v,t}\right)$ which can be solved recursively for some initial values $\sigma^2_{\varepsilon,0}$, θ_0 (the same conditions are satisfied substituting $\sigma^2_{\eta,t}$ and $\gamma_t = s_t\theta_t^{-1}$). These two conditions taken together imply that the time-varying autocorrelation satisfies

$$acf_{1,t} \equiv \frac{cov_t\left(y_t, y_{t-1}\right)}{\sigma_{y,t}\sigma_{y,t-1}} = \frac{-\theta_t}{s_t + \theta_t^2} = \rho_{t-1}\sqrt{\left(1 - R^2_{t-1}\right)R^2_t} \tag{69}$$

We then have

$$R^2_{\min,t} = \frac{\theta_t^2}{s_t + \theta_t^2}$$

and we can derive $R^2_{\max,t}$ from (69), setting $\rho_{t-1} = 1$ throughout, and solving recursively backwards. The proof of the inequality follows analogously to the proof of Proposition 1, since this only requires serial independence, it does not require that w_t is drawn from a time-invariant distribution. To see this, by recursively substituting we have

$$\varepsilon_t = y_t + \sum_{i=1}^{\infty} \Pi_{j=1}^{i}\theta_{t-j}y_{t-j}$$

$$\eta_t = -\sum_{i=1}^{\infty} \Pi_{j=1}^{i}\frac{\theta_{t+j}}{s_{t+j}}y_{t+j}$$

so ε_t is a combination of current and lagged y_t, whereas η_t is a combination of strictly future values of y_t. Thus η_t must have predictive power for all possible predictors (except itself), but not vice versa.

To extend to the ARMA(1,1) case, substitute from (25) into (24), giving

$$\left(1 - \lambda_t L\right) y_t = v_{t-1} + \left(1 - \lambda_t L\right) u_t$$

which we can rewrite as

$$\widetilde{y}_t = \widetilde{v}_{t-1} + u_t$$

with $\widetilde{v}_t = v_t - \lambda_t u_t$, an error term with potentially time-varying variance and covariance with u_t. Since this takes the same form as (68) we can then apply the same arguments as for the MA(1) case.$\blacksquare$

H The time series properties of predictions when y_t is an MA(1) generated by a structural model with $r > q = 1$.

H.1 Analytical Results

Proposition 4 *Let y_t be an MA(1) with $\theta > 0$. Assume that the true data-generating process for y_t is a structural ABCD model that implies a predictive regression with r predictors, with $r > q = 1$, and hence must satisfy $r - 1$ restrictions as in Corollary 4 such that $p = 0, q = 1$. Then, for any predictive model of this form,*

$$R^2 > R^2_{\max}(1) \Rightarrow V_{\widehat{y}} < 1$$

where $R^2_{\max}(1)$ is the calculated upper bound from Proposition 1, setting $r = q = 1$ and $p = 0$, and $V_{\widehat{y}}$ is the limiting variance ratio (Cochrane, 1988) of the predictions $\widehat{y}_t$.

Remark 1 *If $r > 1$ there must be at least one persistent predictor (recalling that, from Assumption A1, r is the number of distinct eigenvalues of $\mathbf{A}$).*

To prove the proposition, we first demonstrate a necessary linkage between the variance ratio V_y of the predicted series y_t and three summary features of any multivariate system.

Lemma 6 *Let V_y be the limit of the variance ratio (Cochrane,1988) of the predicted process y_t, defined by*

$$V_y = \lim_{h\to\infty} VR_y(h) = 1 + 2\sum_{i}^{\infty} corr(y_t, y_{t-i}) \tag{70}$$

where $VR_y(h) = var\left(\sum_{i=1}^{h} y_{t+i}\right)/h\sigma_y^2$. The parameters $\mathbf{\Psi} = (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ of the predictive system must satisfy

$$g(\mathbf{\Psi}) = V_y \tag{71}$$
$$\text{where } g\left(R^2, V_{\widehat{y}}, \rho_{BN}\right) = 1 + R^2\left(V_{\widehat{y}} - 1\right) + 2\rho_{BN}\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)}$$

where $R^2(\mathbf{\Psi})$ is the predictive R^2 from (3); $\rho_{BN}(\mathbf{\Psi}) = corr(u_t, \boldsymbol{\delta}'\mathbf{v}_t)$, with $\boldsymbol{\delta}' = \boldsymbol{\beta}'[I - \Lambda]^{-1}$; and $V_{\widehat{y}}(\mathbf{\Psi})$ is the variance ratio of the predicted value $\widehat{y}_t \equiv \boldsymbol{\beta}'\mathbf{x}_{t-1}$, calculated by replacing y_t with $\widehat{y}_t$ in (70).

Proof: The predictive system in Lemma 1 implies the multivariate Beveridge Nelson(1981)/Stock Watson (1988) decomposition

$$\begin{aligned}
\begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} &= \mathbf{C}(L)\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\beta}' L\left[I-\mathbf{\Lambda} L\right]^{-1} \\ 0 & \left[I-\mathbf{\Lambda} L\right]^{-1} \end{bmatrix}\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix} \\
&= \left[\mathbf{C}(1) + \mathbf{C}^*(L)(1-L)\right]\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix} \\
&= \left\{\begin{bmatrix} 1 & \boldsymbol{\beta}'\left[I-\mathbf{\Lambda}\right]^{-1} \\ 0 & \left[I-\mathbf{\Lambda}\right]^{-1} \end{bmatrix} + \begin{bmatrix} 0 & \boldsymbol{\beta}'\left(L\left[I-\mathbf{\Lambda} L\right]^{-1} - \left[I-\mathbf{\Lambda}\right]^{-1}\right) \\ 0 & \left[I-\mathbf{\Lambda} L\right]^{-1} - \left[I-\mathbf{\Lambda}\right]^{-1} \end{bmatrix}\right\}\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix}
\end{aligned} \tag{72}$$

The equation for y_t in the last line can be written, as

$$\begin{aligned} y_t &= y_t^P + y_t^T \\ \text{where } y_t^P &= u_t + \boldsymbol{\delta}'\mathbf{v}_t \end{aligned} \tag{73}$$

and $\boldsymbol{\delta} = \boldsymbol{\beta}'\left[I-\mathbf{\Lambda}\right]^{-1}$ is a vector of "long-run multipliers" that generate the shocks to the Beveridge-Nelson permanent component. Thus $\rho_{BN} = corr\left(u_t, \boldsymbol{\delta}'\mathbf{v}_t\right)$ is the correlation between innovations to y_t and to long-horizon forecasts of (cumulated) y_t.

Cochrane (1988, equation (10)) shows that $V_y = \lim_{h\to\infty} VR_y(h)$ as defined in (70) must, letting $\sigma_p^2 \equiv var\left(y_t^P\right)$, satisfy

$$V_y = \frac{\sigma_p^2}{\sigma_y^2} \tag{74}$$

since σ_p^2 must be equal in population whether it is derived from the univariate or multivariate representation. Hence, using (73)

$$V_y = \frac{var\left(u_t + \boldsymbol{\delta}'\mathbf{v}_t\right)}{var\left(u_t + \boldsymbol{\beta}'\mathbf{x}_{t-1}\right)} \tag{75}$$

By setting u_t to zero in (75) we can also derive the variance ratio of the predicted value for y_t, $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_{t-1}$ i.e.

$$V_{\widehat{y}} \equiv \frac{var\left(\boldsymbol{\delta}'\mathbf{v}_t\right)}{var\left(\beta'\mathbf{x}_t\right)} \equiv \frac{\sigma_{\delta' v}^2}{\sigma_{\widehat{y}}^2} \tag{76}$$

Using these definitions, and partitioning $\Omega = E\left(\mathbf{w}_t\mathbf{w}_t'\right)$ as

$$\Omega = \begin{bmatrix} \sigma_u^2 & \Omega_{uv}' \\ \Omega_{uv} & \Omega_v \end{bmatrix} \tag{77}$$

we have

$$\begin{aligned}
V_y = \frac{var\left(u_t + \boldsymbol{\delta}'\mathbf{v}_t\right)}{\sigma_y^2} &= \frac{\sigma_u^2 + \sigma_{\boldsymbol{\delta}'\mathbf{v}}^2 + 2\boldsymbol{\delta}'\Omega_{uv}}{\sigma_y^2} \\
&= \frac{\sigma_u^2}{\sigma_y^2} + \frac{\sigma_{\boldsymbol{\delta}'\mathbf{v}}^2}{\sigma_{\widehat{y}}^2}\frac{\sigma_{\widehat{y}}^2}{\sigma_y^2} + \frac{2.\boldsymbol{\delta}'\Omega_{uv}}{\sigma_{\boldsymbol{\delta}'\mathbf{v}}.\sigma_u}.\frac{\sigma_{\boldsymbol{\delta}'\mathbf{v}}.\sigma_u}{\sigma_y^2} \\
&= 1 - R^2 + V_{\widehat{y}}R^2 + 2\rho\frac{\sigma_{\boldsymbol{\delta}'\mathbf{v}}}{\sigma_{\widehat{y}}}.\frac{\sigma_{\widehat{y}}}{\sigma_y}\frac{\sigma_u}{\sigma_y} \\
&= 1 + R^2\left(V_{\widehat{y}} - 1\right) + 2\rho\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)}
\end{aligned}$$

as in Lemma 6.■

From Corollary 5 the true predictions $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_{t-1}$ admit an ARMA$(r, r-1)$ representation. However consider a special case in which there is cancellation in the reduced form for $\widehat{y}_t$ such that the predictions are IID. The following result then follows straightforwardly:

Lemma 7 *Assume that, for some $r > q = 1$ the ABCD model generates IID predictions: $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} = \omega_{t-1}$ for some composite innovation ω_t (hence $V_{\widehat{y}} = 1$). Then the upper bound for R^2 from Proposition 1 applies, setting $r = q = 1$, and $\lambda = 0$, hence $R^2_{\max} = \frac{1}{1+\theta^2}$.*

Proof. *By substituting for $\boldsymbol{\beta}'\mathbf{x}_{t-1}$, and setting $\beta x_{t-1} = \omega_{t-1}$ in a single predictor model.* ■

It follows that, to escape the R^2 upper bound implied by an MA(1) (i.e., as shown in Figure 2) requires that the predictions *not* be IID. We now show that this departure from IID predictions requires $V_{\widehat{y}}$, the variance ratio of the predictions, to be below unity.

The restriction on the parameters of the predictive system in (71), in Lemma 6, implicitly defines a function $V_{\widehat{y}} = h\left(R^2, \rho_{BN}\right)$: the value of $V_{\widehat{y}}$ required to be consistent with the given univariate property V_y, for any given $\left(R^2, \rho_{BN}\right)$ combination.

Letting $K = V_y - 1$ and $W = V_{\hat{y}} - 1$, we can rewrite (71) in Lemma 6 as

$$K - WR^2 = 2\rho_{BN}\sqrt{(1+W)R^2\left(1 - R^2\right)}$$

which implies that W must be a solution to the quadratic equation

$$aW^2 + bW + c = 0 \tag{78}$$

where

$$a = R^4$$
$$b = -2R^2\left(K + 2\rho_{BN}^2\left(1 - R^2\right)\right)$$
$$c = K^2 - 4\rho_{BN}^2 R^2\left(1 - R^2\right)$$

A real solution for W and hence for $V_{\hat{y}} = h\left(R^2, \rho_{BN}\right)$ requires the discriminant to be positive which reduces to the condition

$$R^2(1 - \rho_{BN}^2) + \rho_{BN}^2 > -K = 1 - V_y \tag{79}$$

which will *not* automatically hold: this tells us that certain (R^2, ρ_{BN}) combinations are simply impossible given the univariate properties of y_t.[29]

We therefore restrict our analysis to (R^2, ρ_{BN}) combinations that do satisfy (79) and thus generate real roots. For $V_{\hat{y}} > 1$ we would need at least one positive real root, which would require either $c > 0$ or $c > 0$ and $b < 0$. We now show that the assumptions of the Proposition rule out both cases.

Case 1: $c < 0$

In this case real roots must be of opposite signs so this would be a sufficient condition for a positive real root.

From Proposition 4 y_t is an MA(1) in reduced form, with MA parameter $\theta > 1$. We thus have $V_y = \frac{(1-\theta)^2}{1+\theta^2}$, so $K = V_y - 1 = \frac{-2\theta}{1+\theta^2}$. We also require $R^2 \in \left[R_{max}^2\left(1\right), 1\right]$ where $R_{max}^2 = \frac{1}{1+\theta^2}$. with $|\rho_{BN}| \leq 1$.

Let $R^2 = \frac{1+\varepsilon}{1+\theta^2}$ for some $\theta^2 \geq \varepsilon > 0$. Then $1 - R^2 = \frac{\theta^2 - \varepsilon}{1+\theta^2}$ also positive. Hence

$$c < 0 \Longleftrightarrow \frac{4\theta^2}{\left(1+\theta^2\right)^2} < 4\rho_{BN}^2\left(\frac{1+\varepsilon}{1+\theta^2}\right)\left(\frac{\theta^2 - \varepsilon}{1+\theta^2}\right)$$

$$\Longleftrightarrow \theta^2(1 - \rho_{BN}^2) < \rho_{BN}^2\left(\varepsilon\left(\theta^2 - 1\right) - \varepsilon^2\right)$$

but the rhs is negative for $\theta < 1$ and $\varepsilon > 0$ so this cannot happen. Thus there can be no solution with one positive and one negative root.

Case 2: $c > 0$

For this case we would require $b < 0$ which would then imply two positive real roots.

[29] This is a particular example of the more general property outlined by Robertson & Wright (2012) that the parameters of the underlying structural model are constrained to live in a "Predictive Space" defined by the univariate properties of y_t.

We then have

$$b < 0 \iff K > -2\rho_{BN}^2(1 - R^2)$$

or, substituting (again, setting $R^2 = \frac{1+\varepsilon}{1+\theta^2}$)

$$b < 0 \iff \frac{-2\theta}{1+\theta^2} > -2\rho_{BN}^2 \left(\frac{\theta^2 - \varepsilon}{1+\theta^2} \right)$$

which simplifies to give

$$b < 0 \iff \theta < \underbrace{\rho_{BN}^2}_{<1} \underbrace{(\theta^2 - \varepsilon)}_{<\theta^2} < \theta^2$$

which is ruled out for $1 \geq \theta \geq 0$.

We can thus rule out both Case 1 and Case 2, and so we can rule out one or two positive real roots. Thus both roots must be negative implying $V_{\hat{y}} < 1$ for any value of $R^2 > R_{\max}^2(1)$. This completes the proof of Proposition 4.■

H.2 An illustration: time series properties of the inflation predictions from the Smets-Wouters (2007) DSGE model

To illustrate the contrast between the restrictions implied by Proposition 4 and the time series properties of inflation predictions in a benchmark macroeconomic forecasting model, we examine the DSGE model of Smets-Wouters (2007). Based on their own Dynare code, we generate 100 artificial samples of quarterly data for the 16 state variables and 7 observables in the Smets-Wouters model, using posterior modes of all parameter estimates as given in their paper, and generate one-step-ahead predictions of changes in inflation from the simulated data using the appropriate line of (2). Since we do not wish the results of this exercise to be contaminated by small sample bias we set $T = 1,000$, in an attempt to get a reasonably good estimate of the true implied population properties.

Table A1 summarises the results.

Table A1: Time Series Properties of Simulated Inflation Predictions, $\widehat{y}_t \equiv \Delta\widehat{\pi}_t$, in the Smets-Wouters (2007) model at various forecast horizons

	First Order	Sample Variance Ratio (bias-corrected)			
	Autocorrelation	5 years	10 years	15 years	20 years
Mean	0.49	3.81	3.97	4.05	4.15
Median	0.49	3.81	3.89	3.92	3.98
Minimum	0.42	2.77	1.80	1.47	1.16

The first column of Table A1 shows the first-order autocorrelation coefficient of the simulated predictions; the remaining columns show estimates of $V_{\widehat{y}}$ using sample variance ratios (using the small sample correction proposed by Cochrane, 1988) at a range of finite horizons. Table A1 makes it clear that the Smets-Wouters model is very far from generating IID predictions (as would be required if the reduced form for $\Delta\pi_t$ was actually an MA(1) as in Stock and Watson, 2007): instead the predictions generated by the Smets-Wouters DSGE model have strong positive persistence - as would be expected given that predicted changes in inflation in the model are driven by strongly persistent processes in the real economy. Thus, using Proposition 4, the nature of this deviation from IID predictions would, if it were the true DGP, reduce, rather than increase R^2, relative to the calculated upper bound for R^2 derived from a single predictor model.