# What univariate models can tell us about multivariate macroeconomic models

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

December 14, 2014

**Abstract**

A longstanding feature of macroeconomic forecasting has been that a wide variety of multivariate models have struggled to out-predict univariate representations. We seek an explanation of this feature in terms of population properties. We show that if we know the univariate properties of a time-series $y_t$ these constrain tightly both the dimensions and the predictive power of the multivariate macroeconomic model that generated $y_t$. We illustrate our analysis using data on U.S. inflation. We find that, especially in recent years, univariate properties of inflation dictate that: a) any efficient predictor of inflation must itself be near white noise; and b) all multivariate models for inflation will struggle to out-predict a univariate model.

**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

Data-consistent macroeconomic models are all, to a greater or less extent, built upon the goodness-of-fit of predictive regressions. The point is obvious for vector autoregressions (VARs). But it also applies to the rapidly growing literature on estimated Dynamic Stochastic General Equilibrium (DSGE) models; since, even for Bayesians with strong priors, the likelihood function of a state space model is central to (posterior) estimation, and can be expressed in terms of the one-step ahead forecast errors.[1]

But 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). 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), since these effectively give greater weight in estimation to finite order univariate autoregressive representations.

A celebrated example of the struggle to out-predict a univariate model is that of U.S. inflation. As an extensive literature has demonstrated, and as we illustrate below in Section 2, it is extremely hard to find predictor variables for U.S. inflation that have more than (at best) marginal predictive power, relative to a univariate benchmark. A similar story holds for stock returns, exchange rates and consumption.[2] The common feature of all of these literatures is that the model-to-beat has always been, and remains, a univariate model; and the margin by which it can be beaten is frequently quite narrow.

In this paper we seek an explanation of this common feature in terms of population properties. We do so by taking a backwards look at the relationship between multivariate and univariate properties. While our results are econometric in nature, they have strong implications for macroeconomics.

We analyse a stationary time series process, $y_t$, which is generated by a generic multivariate macroeconomic model. The state variables of this system can be captured by a vector of (possibly unobserved) predictors, $\mathbf{x}_{t-1}$, that predict $y_t$ up to a serially independent error, $u_t$. We ask: if we just observed the history of $y_t$, what would its univariate properties, as captured by a finite order ARMA representation, tell us about the proper-

[1]E.g., see Canova (2007), Section 6.2. For some estimation approaches (e.g., matching impulse response functions; see Hall *et al.*, 2012) the link is less direct, but is still there, since the impulse responses that are matched usually derive from estimated VARs.

[2]For recent surveys see, for example, Rapach and Zhou (2013) (stock returns), Rossi (2013) (exchange rates) and Lahiri *et al.* (2013) (consumption).

ties of the underlying multivariate model? We show that, for some $y_t$ processes, univariate properties alone tightly constrain the properties of the structural multivariate model.

Our first key result - which exploits existing results, but views them from a different angle - is about the dimensions of the underlying macroeconomic model. If $y_t$ is an ARMA$(p, q)$ in population then this tells us that, unless we wish to impose specific coefficient restrictions on the model that generates the data, the univariate MA order $q$ must be identical to the number of predictors that arise from the multivariate system (although these predictors may themselves be composites of multiple underlying state variables).

Our second key result is, as far as we are aware, entirely new. It can be viewed narrowly, from an econometrician's point of view, as being about the predictive power of the underlying macroeconomic model. But from the macroeconomist's perspective, in light of the growing literature on "hidden" structural shocks (e.g., Fernández-Villaverde *et al.*, 2007; Leeper *et al.*, 2013) it tells us how much it actually *matters* whether these shocks are observed. We show that the one-step-ahead predictive $R^2$ of the predictive regression that arises from the macroeconomic model must lie between bounds, $R^2_{\min}$ and $R^2_{\max}$, both of which can be derived from ARMA parameters alone, and hence can be derived solely from the history of $y_t$. The nature of these bounds also provides crucial insights. The lower bound $R^2_{\min}$ is simply the $R^2$ of the ARMA representation. The upper bound, $R^2_{\max}$, can also be calculated solely from these ARMA parameters: it 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.[3]

While one can improve the (univariate) prediction by using the true predictor vector, the upper bound implies that there is a limit to this improvement, which is determined solely by the univariate properties of $y_t$. For some time series the gap between $R^2_{\min}$ and $R^2_{\max}$ can be quite narrow. For the macroeconomist this implies that, for such series, it may not *matter* very much, at least in predictive terms, if structural shocks are "hidden" or not.

While our results for the general ARMA$(p, q)$ case are derived on the assumption of time-invariant models, we show that at least for an important special case our results extend straightforwardly to cases where the predictive model and the univariate model both have time-varying coefficients and (co)variances. Hence our core results do not rely on the assumption of structural stability.

To illustrate, we use our analysis to shed light on Stock and Watson's (2007, 2010) conclusion that U.S. inflation has become harder to forecast. We show that in recent data

[3] While nonfundamental representations are nonviable predictive models their *properties* can be derived straightforwardly from the parameters of the fundamental representation.

their preferred univariate representation implies that the upper and lower bounds for $R^2$ (which in their case will vary over time) are very close to each other. Thus univariate population properties *dictate* the feature that, however well the structural model predicts, it can at best only marginally out-predict a univariate model. As a direct corollary we also show that this should lead to univariate and (efficient) multivariate forecasts of inflation being strongly correlated. The univariate properties of inflation also imply that the predicted value of *changes* in inflation from a multivariate model should themselves be near-white noise. Yet both empirical predictors and most structural models share the common feature that predicted changes in inflation are strongly persistent. We argue that this helps to explain the long and fruitless search by macroeconomists for predictor variables that consistently forecast inflation. On a more positive note, for those seeking to find predictor variables for inflation, we note that one potential source of independent and identically distributed (IID) predictions would be in 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 U.S. inflation.

The rest of the paper is structured as follows. In Section 2 we motivate our analysis by providing an empirical illustration of the problems multivariate models have in out-predicting a univariate representation. Section 3 sets out the links between the ARMA representation and the multivariate model. Section 4 describes the $R^2$ bounds and their implications. In both Section 3 and 4 we illustrate our results with reference to the case of an ARMA(1,1). Section 5 shows that our core results can be generalised to accommodate time variation. Section 6 presents our empirical application to U.S. inflation, and Section 7 concludes. Appendices provide proofs and derivations.

# 2 A motivating empirical example: univariate vs. multivariate models of U.S. inflation

As noted in the Introduction, a celebrated example of the struggle to beat a univariate model is the case of U.S. inflation. Figure 1, to which we return at various points in the paper, illustrates by comparing alternative one-period-ahead in-sample forecasts of the change in the U.S. (GDP deflator) inflation rate, $\Delta\pi_t$, measured on a quarterly basis.[4]

[4] We consider the change in the inflation rate, rather than its level, for comparability with Stock and Watson (2007) who assume a unit root in inflation. We stress that nothing in our results hinges on this specific data transformation - it is for comparability and expository purposes only. Our theoretical analysis below simply requires stationarity of $y_t$. If inflation is believed to be stationary then $\pi_t$ could be

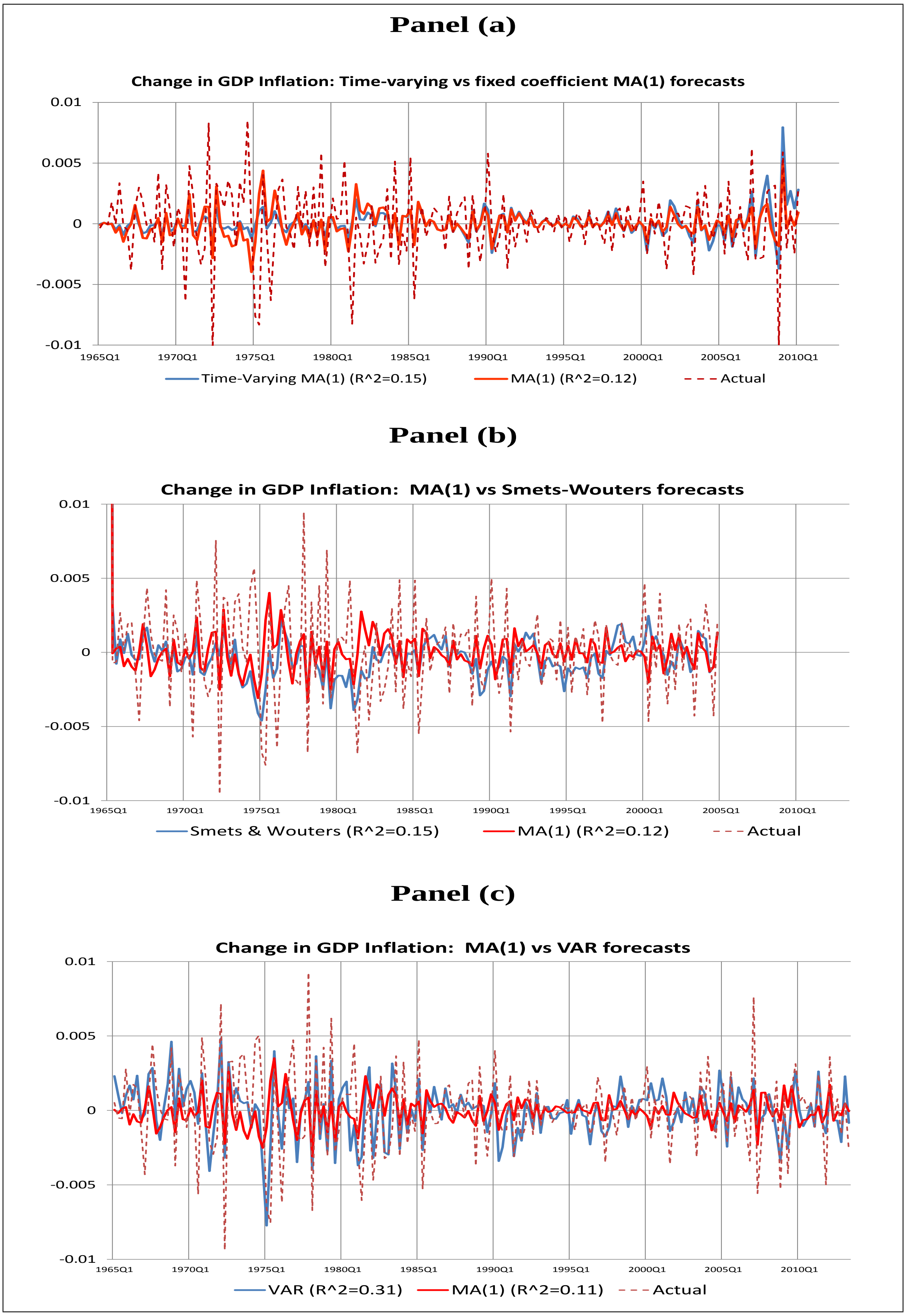


Figure 1: Univariate versus Multivariate In-Sample Forecasts of U.S. Inflation

Our benchmark for comparison is the simplest possible univariate model, a fixed coefficient moving average model of order one, MA(1): $\Delta\pi_t = \varepsilon_t - \theta\varepsilon_{t-1}$, where $E(\varepsilon_t) = \sigma_\varepsilon^2$, which captures the feature that surprise increases in inflation (conditional upon its own past) in one quarter tend to reverse, at least partially, in the next quarter. The forecasting performance of this model is not particularly impressive: over the full sample 1965 to 2013, shown in Figure 1 (panel (c)) the predictive $R^2$ is only around 11%. The issue is whether other forecasting models can do better.

One way of improving the univariate model, without adding to the information set, follows Stock and Watson (2007, 2009, 2010) in letting both the moving average parameter $\theta$ and the variance of the univariate shock $\varepsilon_t$ vary over time. Panel (a) of Figure 1 shows that the fit of this time-varying MA(1) model is improved by having a lower value of $\theta_t$ during the 1970s and a distinctly higher value in more recent data.[5] Strikingly, however, for most of the sample the predicted values are very similar to those of the fixed coefficient MA model.

Can multivariate models - whether structural or atheoretical - do better than this? If we wish to compare like with like, then a fixed coefficient univariate model should be compared with fixed coefficient multivariate predictive models. The remaining panels of Figure 1 provide two simple comparisons. Panel (b) shows the in-sample predictions generated by the Smets and Wouters (2007) estimated DSGE model[6]; and panel (c) shows predictions from a 4th order VAR using the same seven variables as Smets and Wouters (2007), over the full data sample, alongside the MA(1) predictions from the relevant dataset.[7] In sample, at least, both multivariate models do (as we would expect) fit better than the univariate benchmark (albeit only very marginally so in the case of the DSGE model); but as the figures illustrate, both still leave most of the variability of inflation changes unexplained, even in sample.

While Figure 1 does not present a particularly favourable picture of multivariate mod-

modelled directly; but given the important role the lagged inflation rate would then play, whether or not there is an exact unit root, from an expository perspective the ensuing inflation forecasts would simply track actual inflation more closely than the forecasts for the change in inflation seen in Figure 1.

[5]We use replication code from Mark Watson's website. Stock and Watson (2007)'s actual estimation procedure is an unobserved components trend-cycle model with stochastic volatility; but they show this is equivalent to a time-varying MA(1).

[6]For Smets and Wouters (2007) we use their dataset, replication code and posterior mean coefficients. Their dataset results in a shorter sample. Note that the three panels of Figure 1 use three different vintages of data. But data revisions have made relatively little difference. The correlation between inflation changes in the common samples of the different datasets is fairly high (above 0.9).

[7]Both multivariate models include terms in the lagged level of inflation and hence in principle allow for predictability of changes in inflation due to mean reversion. Univariate models can also include lagged levels terms but this makes virtually no difference to the results.

els, it is easy to show that it is actually unduly flattering. While Panel (c) suggests a predictive advantage for the VAR versus the univariate model, if these models are both estimated recursively this advantage almost entirely disappears. This result is consistent with past research. A sequence of papers (e.g., Atkeson and Ohanian (2001), Ang *et al.* (2007), Stock and Watson (2007, 2009, 2010) and D'Agostino and Surico (2012)) have shown that, particularly once we consider out-of-sample performance, it is extremely hard to find predictor variables for U.S. inflation that have more than (at best) marginal predictive power, relative to a univariate benchmark.

Two other features of Figure 1 - to which we shall revert at various points in the paper - are worth noting. The first is that both multivariate models generate predicted values for inflation that are reasonably strongly correlated with those from the univariate model.[8] The second is that the predicted values themselves all share the time series property of being very close to IID. We shall show that both of these features arise naturally given the univariate features of inflation. We also show that those same univariate properties would lead us to expect multivariate models: a) to offer only limited additional predictive power for inflation - particularly over the past decade or two; and b) to suffer from chronic structural instability problems when estimated recursively.

# 3 What the ARMA Representation tells us about the Predictive System

## 3.1 The Multivariate Structural Model and the Predictive System for $y_t$

Consider the generic (Fernández-Villaverde *et al.*, 2007) ABCD representation of a multivariate macroeconomic model:

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

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

where $\mathbf{z}_t$ is an $n \times 1$ vector of (possibly unobserved) states hit by a vector of structural economic shocks, $\mathbf{s}_t$, and $\mathbf{y}_t$ is a vector of observed macroeconomic variables. Fernández-Villaverde *et al.* (2007) assume this system to represent the rational expectations solution

[8] In Section 6 we argue that the observed correlations are actually distinctly lower than we would expect if multivariate forecasts were informationally efficient.

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 system in (1) and (2) is also consistent with a wide range of factor and VAR models that are also commonly used to forecast in macroeconomics and econometrics.[9]

In the generic ABCD framework predictability may appear to play a limited role. But indirectly its role is crucial. Both equations in the system are predictive regressions by construction; and in any empirical exercise that attempts to recover the structural parameters of the data, the one-step-ahead covariance matrix of the observables, $\mathbf{D}E\left(\mathbf{s}_t\mathbf{s}_t'\right)\mathbf{D}'$ plays a crucial role in likelihood-based inference.

**Assumptions**

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

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

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

Assumption A1, that the ABCD system can be diagonalised, is in most cases innocuous.[10] Assumption A2, that the system is stationary, is also simply convenient. There may be underlying state space representations with unit roots in some states and observables, but these can be differenced out to generate a stationary representation of both.

Finally Assumption A3 follows Fernández-Villaverde *et al.* (2007). It is convenient 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 Lemma 1.

The assumptions of a time-invariant model (which, given constancy of $\mathbf{B}$ and $\mathbf{D}$, in turn implies a time-invariant distribution of prediction errors), and of normality of the structural shocks are, however, *not* crucial; they merely simplify the exposition. In Section

[9] The former follows since the $\mathbf{z}_t$ could be the observed common factors extracted from a large dimensional dataset, say via principal components (e.g., Stock and Watson, 2002) and/or Kalman filtering methods (e.g., Giannone *et al.*, 2008). The latter follows since (1) can be the companion form of a finite order VAR representation of $\mathbf{y}_t$.

[10] 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 $q > p$.

5 we consider generalisations to cases where the parameters of the structural model may vary over time.

Under these quite minimal assumptions, the system in (1) and (2) allows us to derive a particularly simple specification for the predictive regression for $y_t$, a single element of $\mathbf{y}_t$, which, without loss of generality, we take to be its first element.

**Lemma 1** ***(The Predictive System)*** *Under A1 to A3 the structural ABCD representation implies the 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 $\mathbf{M} = diag\left(\mu_1, ..., \mu_r\right)$, so that each element of $\mathbf{x}_t$ is a univariate AR(1). At most one element may have $\mu_i = 0$ in which case $x_{it}$ is IID.*

**Proof.** See Appendix A. ■

This representation may in principle nest cases in which there are more than $r$ elements of the underlying state vector $\mathbf{z}_t$ (i.e., (1) may not be a minimal state representation; see Komunjer and Ng, 2011).[11] Assumption A3, that the structural disturbances $\mathbf{s}_t$ are serially uncorrelated, implies that (3), being derived from the structural model that generated the data, is not misspecified. Thus $\mathbf{x}_{t-1}$ can be viewed as generating the data for $y_t$, up to a white noise error, $u_t$.

Note that we make no assumptions about whether the true predictor vector, $\mathbf{x}_t$, and the structural shocks, $\mathbf{s}_t$, are observable.

## 3.2 The Macroeconomist's ARMA

Exploiting standard results (e.g., applying Corollary 11.1.2 in Lütkepohl (2007)), substitute from (4) and rewrite (3) as

$$\mu\left(L\right)y_t = \boldsymbol{\beta}' adj\left(\mathbf{I} - \mathbf{M}L\right)\mathbf{v}_{t-1} + \mu\left(L\right)u_t \tag{5}$$

[11] For example, in the case of the benchmark Smets and Wouters (2007) model, the underlying state vector $\mathbf{z}_t$ is $20 \times 1$, but 4 of the underlying states can be derived as linear combinations of other states. With no common eigenvalues amongst the remaining states the diagonalised predictor vector $\mathbf{x}_t$ in (4) is $16 \times 1$.

where $\mu(L) \equiv \prod_{i=1}^{r}(1-\mu_i L) \equiv \det(I-\mathbf{M}L)$ is of order $r$ given Assumption A1. The right-hand-side of (5) is an MA($r$) process, since the the highest power of $L$ in $adj\,(\mathbf{I}-\mathbf{M}L)$ is $r-1$. This implies that $y_t$ has an ARMA($r,r$) representation

$$\mu(L)\, y_t = \psi(L)\, \varepsilon_t \tag{6}$$

where $\psi(L) \equiv \prod_{i=1}^{r}(1-\psi_i L)$, and the $|\psi_i| < 1$, $\forall i$, so (6) is "fundamental" (Hamilton,1994, pp. 64-67; Lippi and Reichlin, 1994) since $\varepsilon_t = \psi(L)^{-1}\mu(L)\, y_t$ is recoverable as a non-divergent sum of current and lagged values of $y_t$.

The $\psi_i$ are solutions to a set of $r$ moment conditions such that the autocorrelations of $\psi(L)\,\varepsilon_t$, the moving average error on the right-hand-side of (6), match those of the right-hand-side of (5) (which is entirely described by its first $r$ autocorrelations). These are set out in Appendix B. The condition $|\psi_i| < 1$ gives the unique fundamental solution. The $\psi_i$ and $\mu_i$ are functions of the full set of parameters of the predictive system (3) and (4) and hence ultimately of the structural model (1) and (2). Therefore we have

$$\boldsymbol{\psi} = \boldsymbol{\psi}(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})\,;\ \boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) \tag{7}$$

where $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_r)'$; $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_r)'$.

Note that for any given $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ there will be $2^r$ solutions to the moment conditions, but the condition $|\psi_i| < 1$, $\forall i$, ensures uniqueness. There are $2^r - 1$ other solutions, in which one or more of the $\psi_i$ is replaced with its reciprocal (Lippi and Reichlin, 1994).[12] Each of these implies a nonfundamental ARMA representation for the same $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$; but the shocks to these alternative nonfundamental ARMA representations cannot be recovered from the history $y^t = \{y_t, y_{t-1}, y_{t-2}, ...\}$. They are therefore non-viable as predictive models, but the *properties* of these representations can be calculated straightforwardly from those of the unique fundamental representation. We shall show below that the properties of one particular nonfundamental representation, in which *all* the $\psi_i$ are replaced with their reciprocals, plays a crucial role in our analysis.

## 3.3 The macroeconomist's ARMA: a simple illustration

To illustrate, we introduce an example to which we shall revert at various stages in the paper.

Consider an ABCD model with a single state variable $z_t$ and a $2\times1$ vector of structural

[12] Note that, as discussed in Lippi and Reichlin (1994), some of the $\psi_i$ may be complex conjugates.

shocks. This implies the predictive system

$$y_t = \beta x_{t-1} + u_t \tag{8}$$

$$x_t = \mu x_{t-1} + v_t \tag{9}$$

with $\mathbf{A} = \mu$, $\mathbf{C} = \beta$, $v_t = \mathbf{B}\mathbf{s}_t$ and $u_t = \mathbf{D}\mathbf{s}_t$, with $\mathbf{B}$ and $\mathbf{D}$ both $1 \times 2$ 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 $x_t$ some stationary valuation criterion. Note that a predictive system of this form also subsumes the case of an underlying structural ABCD representation with $n > 1$ states, but a single common eigenvalue, in which case the single predictor in (8) and (9) will be a composite of the $n$ underlying state variables in $\mathbf{z}_t$.

By substitution from (8) into (9) we have

$$(1 - \mu L)\, y_t = \beta v_{t-1} + (1 - \mu L)\, u_t \tag{10}$$

The right-hand-side of this expression is an MA(1) process (the counterpart to the right-hand-side of (5) above). So $y_t$ admits a fundamental ARMA(1,1) representation (the counterpart to (6))

$$(1 - \mu L)\, y_t = (1 - \psi L)\, \varepsilon_t \tag{11}$$

with $|\psi| < 1$. The first order autocorrelation of the MA(1) process on the right-hand side of (11) matches that of the right-hand-side of (10): i.e., the single MA parameter $\psi$ is the the solution in $(-1, 1)$ to the moment condition

$$\frac{-\psi}{1 + \psi^2} = \frac{\mu\sigma_u^2 - \beta\sigma_{uv}}{(1 + \mu^2)\, \sigma_u^2 + \beta^2\sigma_v^2 + 2\mu\beta\sigma_{uv}} \tag{12}$$

Since all the parameters on the right-hand-side of (12) can in turn be derived from the parameters of the ABCD representation, we have $\psi = \psi\left(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\right)$, as for the general case. Note that, as discussed for the general case, this is not the unique solution to the moment condition. It is also satisfied by $\psi^{-1}$; but this results in a nonfundamental representation, with a different IID shock process, which cannot be recovered from the history $y^t$ (we discuss this representation below).

## 3.4 The macroeconomist's ARMA and structural shocks

Much empirical macroeconomic modelling focusses on the attempt to extract structural shocks (here, the $s \times 1$ vector $\mathbf{s}_t$ in (1) and (2)) from the multivariate model. It should be fairly evident that in general univariate models will *not* allow us to identify structural shocks.

Given that the macroeconomist's ARMA, (6), is a fundamental representation, we can write the univariate innovations $\varepsilon_t$ in terms of the history $y^t$ and hence in terms of the history of the structural shocks. This follows by equating the right-hand-sides of (5) and (6), giving

$$\varepsilon_t = \psi(L)^{-1} \left[\boldsymbol{\beta}' adj\left(\mathbf{I} - \mathbf{M}L\right)\mathbf{v}_{t-1} + \mu(L) u_t\right] = \mathbf{f}(L)\,\mathbf{s}_t \tag{13}$$

where $\mathbf{f}(L) = \mathbf{f}_0 + \mathbf{f}_1 L + \mathbf{f}_2 L^2 + .....$ is a $1 \times s$ row vector of non-divergent lag polynomial functions in non-negative powers of $L$. This can be written in terms of the structural shocks $\mathbf{s}_t$ since, from Lemma 1, $\mathbf{v}_t$ and $u_t$ are both linear combinations of elements of $\mathbf{s}_t$.

Equation (13) makes clear why, for the general case, the univariate innovation does not allow us to identify the structural shocks. The problem can be partitioned into two sub-problems: of dimensionality and of fundamentalness. Trivially, for $s > 1$, we cannot identify individual shocks. But in the special case that $\mathbf{f}_i = 0$ for $i > 0$ we can at least recover some linear combination of shocks from the univariate innovation; i.e., in that case $\varepsilon_t = \mathbf{f}_0 \mathbf{s}_t$ so that a particular linear combination of structural shocks is fundamental for $y^t$. If $s = 1$, we do not have a problem of dimensionality, but we still may have a problem of fundamentalness. Since for the general case $\mathbf{f}(L)$ is convergent in positive powers of $L$, even in the scalar case we cannot invert it to derive the single structural shock $s_t$ from the history of $\varepsilon_t$, and hence from $y^t$. In such cases (in an unfortunate clash of terminologies between macroeconomics and time series analysis) $s_t$ is structural but not fundamental. Only if we have both $\mathbf{f}_i = 0$ for $i > 0$, *and* $s = 1$ - thus for a very restricted class of structural models - can we recover the single structural shock as a scaling of the fundamental univariate innovation.[13]

For macroeconomists who seek to identify structural shocks, systems with multiple observables are therefore typically a requirement. But, as we now go on to show, univariate models can still provide us with important insights into the nature of the structural model that generated the data. Furthermore in some cases our results will show that even if we

[13] Only in this last case does the structural model satisfy Fernández-Villaverde *et al.*'s (2007) "Poor Man's Invertibility Condition" in terms of the single observable, $y_t$. In this case, factoring $f(L)$ as $f(L) = \psi^{-1}(L)\, g(L)$ the condition can by inspection be re-expressed as the requirement that, up to a scaling factor, $g(L) = \psi(L)$.

*could* observe the structural shocks, the resulting improvement in predictive power for $y_t$ might be quite limited.

### 3.5 The Econometrician's ARMA

We have shown that the structural model, and hence the predictive system, implies an ARMA. We now show that this can be looked at backwards; that is, an ARMA must imply a predictive system, and hence a structural model.

From the perspective of an econometrician the predictive model for $y_t$ (conditioning only on the history $y^t$) will be a univariate ARMA$(p, q)$ representation. For a sufficiently long history we assume that this converges to the minimal population representation

$$\lambda(L) y_t = \theta(L) \varepsilon_t \tag{14}$$

where $\lambda(L) \equiv \prod_{i=1}^{p} (1 - \lambda_i L)$, $\theta(L) \equiv \prod_{i=1}^{q} (1 - \theta_i L)$; and "minimal" implies there is no redundancy in the representation, so that none of the $\theta_i$ or $\lambda_i$ are zero, and there is no cancellation of the MA and AR polynomials. From the econometrician's point of view, the reciprocal roots of the AR and MA polynomials, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_p)$ and $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)$, are simply parameters.

For the error $\varepsilon_t$ in (14) to be recoverable from the history $y^t$ the representation must be "fundamental", i.e. we must have $|\theta_i| < 1$, $|\lambda_j| < 1$, $\forall i, \forall j$. $\varepsilon_t$ must therefore be identical, in population, to the error in the macroeconomist's ARMA, (6).

For now, consistent with our treatment of the structural model, we assume that (14) is a representation with time-invariant coefficients. Later in the paper (see Sections 5 and 6) we consider extensions to cases where the coefficients may evolve over time.

### 3.6 What the Econometrician's ARMA representation tells us about the dimensions of the predictive system and the structural model

**Proposition 2** ***(ARMA order and the Predictive System)*** *If $y_t$ admits an ARMA$(p, q)$ representation, under A1 to A3 this implies a predictive system as in (3) and (4) with $r$ predictors, and hence an ABCD system as in (1) and (2) in which* $\mathbf{A}$ *has $r$ distinct eigenvalues, with*

$$r = q$$

*and the econometrician's ARMA and the macroeconomist's ARMA are identical* ($\mu_i = \lambda_i$, $\psi_i = \theta_i, \forall i, i = 1, ..., q$) ***unless*** *the parameters* $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ *are such that* $\psi_i (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = 0$ *for some* $i$*,* ***or*** $\psi_j (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = \mu_k (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$*, for some* $j$ *and* $k$*.*

**Proof.** See Appendix C. ■

Proposition 2 says that the MA order of the econometrician's ARMA, $q$, must reveal $r$, the number of predictors of the predictive system, and hence the dimension of the minimal state representation of the ABCD system that generated the data for $y_t$. This can only *not* be the case if there are specific restrictions across the parameters of the structural model, and hence of the predictive system. We provide illustrations of the nature of the required restrictions in the following two sub-sections.

In the absence of such restrictions, the AR and MA parameters, $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, of the econometrican's ARMA (14) must reveal the AR and MA parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$, of the macroeconomist's ARMA, (6), which themselves tell us about $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. The order of the AR polynomial will be $p = q$ unless $\mu_i = 0$, in which case $p = q - 1$. By assumption there can be at most one such $\mu_i$ $(i = 1, ..., q)$.[14]

## 3.7 The ARMA(1,1) case revisited

In Section 3.3 we showed that a simple predictive system with a single predictor as in (8) and (9) implies an ARMA(1,1) representation (the macroeconomist's ARMA). Proposition 2 implies that we can also look at this relationship backwards, starting from the econometrician's ARMA. Thus assume that in population $y_t$ is ARMA(1,1):

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

Then, absent restrictions on the parameters $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ of the structural model, we must have $r = q = 1$; hence the predictive system *must* take the form in (8) and (9).

The AR parameter $\lambda$ of the econometrician's ARMA (15) tells us the same parameter $\mu$ in the macroeconomists's ARMA, (11), and hence directly reveals the AR parameter of the single (possibly composite) AR(1) predictor, the single eigenvalue of $\mathbf{A}$. The MA parameter $\theta$ of the econometrican's ARMA must equal the same parameter $\psi$ in the macreconomist's ARMA; which in turn must satisfy $\psi = \psi (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. Since the underlying structural parameters are (usually) of higher dimension, with the exception of $\mu$ $(= \mathbf{A})$, the structural parameters are only set-identified from the econometrician's

[14] It is reasonably easy to generalise to other values of $p$ (see Robertson and Wright, 2012).

ARMA. We shall see shortly that this set-identification can still allow us to derive non-trivial constraints on the predictive system from the ARMA; but at this stage the primary message of Proposition 2 is what the ARMA tells us about dimensions.

Thus, absent restrictions on $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, if $y_t$ is an ARMA(1,1) Proposition 2 rules out a predictive system with $r = 2$ or larger. Any ABCD system with, for example, 2 distinct non-zero eigenvalues, and hence 2 predictors would imply that the macroeconomist's ARMA, (6) would be an ARMA(2,2). Only if the parameters of the structural system satisfied restrictions such that $\psi_2(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ was equal to either $\mu_1$ or $\mu_2$ (both also functions of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$) could a 2 predictor system generate an ARMA(1,1) reduced form for the econometrician. Put another way, an ARMA(1,1) *either* implies a single predictor model ($q = r = 1$) *or* imposes restrictions on the parameter space of a more general macroeconomic system with $r > q$.

### 3.8 A simple special case: the stock return with $p = q = 0$

To demonstrate the implications of Proposition 2 in their most stark form consider the simplest possible case, in which $y_t$ is IID, hence $p = q = 0$. Then Proposition 2 says that there can be *no* predictive system for $y_t$ (i.e., $r = 0$) unless we believe that there are restrictions across the parameter space of the structural model.

To illustrate, suppose that our IID variable, $y_t$, is the stock return. In this case it is a standard result that the IID property results from the joint assumption of informational efficiency, constant expected returns (Samuelson, 1965), and homoscedastic shocks to the market information set. Is it nonetheless possible that there may be a predictor variable for IID stock returns? The answer is yes; but only if we impose tight restrictions on the structural model.

There is clearly one case (albeit a trivial one). If we have a direct line to God, and therefore know the data in advance, then trivially we can predict perfectly, hence $\sigma_u = 0$. The "predictor" $\beta x_{t-1}$ would simply be equal to $y_t$.

Now suppose instead that we have a direct line to Warren Buffett, who, in period $t-1$, provides us with a forecast of the stock return, $\widehat{y}_t = \beta x_{t-1}$. For simplicity, we initially assume that Warren Buffett's single predictor variable, $x_t$, is itself IID, hence $\mu = 0$. (Note that this is a special case of the example analysed in Section 3.3, with $\mu = 0$, and hence $x_t = v_t$). Unless Warren Buffett is God, this will not be a perfect forecast; hence $y_t = \beta v_{t-1} + u_t$, with $\sigma_u^2 > 0$. For the general case this would imply that $y_t$ must be an MA(1), hence the econometrician's ARMA representation of $y_t$ would have $p = 0$, $q = 1$, which would correctly point to a single IID predictor. But if $y_t$ is itself IID, this case is ruled out.

The only way there *could* be a predictive system for $y_t$ with an IID predictor would be if we imposed one or more additional restrictions on the $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ parameters of Warren Buffett's predictive system.

For the simplest case, of an IID predictor, we require $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (\mathbf{0}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ to satisfy $\psi(\mathbf{0}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = 0$ in the macroeconomist's ARMA. Given the simplicity of the model, by inspection of the moment condition (12), for $y_t$ still to be IID requires $u_t$ and $v_t$ to be orthogonal ($\sigma_{uv} = 0$). This is a strong informational restriction. It implies that Warren Buffett's predictor, $v_t$ must capture innovations to an information set such that new data on $y_t$ itself is then informationally redundant.[15] The structural model in turn must satisfy $\mathbf{A} = \mu = 0$ and $\mathbf{BD}' = 0$.

A predictive system with a single IID predictor is not, it should be stressed, the only possible system that could generate an IID process for $y_t$. Thus it is possible in principle that Warren Buffett's predictive system might resemble the system more commonly used in the literature on return predictability, in which, as in the example of Section 3.3, the single predictor (typically some valuation criterion) is usually assumed to have nontrivial persistence, and hence $\mu > 0$. But, if $y_t$ is IID this means that $\mu$ cannot be a free parameter; we would require cancellation of the MA and AR polynomials in the macroeconomist's ARMA, i.e., we require $\mu = \psi(\mu, \mathbf{B}, \mathbf{C}, \mathbf{D})$.[16]

By extension, if there were 2 (or $r$) predictors, this could only be consistent with the stock return being IID if we also imposed 2 (or $r$) appropriate restrictions across the parameter space of the structural model.

---

[15] Since in a filtering framework, if $y_t$ itself is *not* informationally redundant, $u_t$, the innovations to $y_t$, conditional upon state estimates in period $t-1$ would usually have an impact on state estimates in period $t$, via the Kalman gain. Hence $u_t$ and $v_t$ would be correlated. Note that, since Stambaugh (1999), an extensive literature has shown that strong innovation correlation is also a feature of virtually all empirical predictors of returns.

[16] This point is made, *inter alia*, in Campbell *et al.* (2007, p.266). From (12) this implies that $\mu$ must satisfy. From (12) this implies that $\mu$ must satisfy

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

Note that in the previous case where $\mu = 0$ this means that the restriction is imposed on the right-hand-side, via the covariance structure of the model. Note also that God's predictor (perfect foresight) satisfies the moment condition (12) since it implies $\mu = \sigma_u = \sigma_{uv} = 0$.

# 4 How the ARMA model constrains the predictive system

So far we have largely restricted ourselves to analysing what the dimensions of the ARMA reduced form for $y_t$ tell us about those of the multivariate system. But Proposition 2 also told us that $(\boldsymbol{\lambda}, \boldsymbol{\theta})$ pin down $(\boldsymbol{\psi}, \boldsymbol{\mu})$, which from (7) are themselves parameters of the multivariate structural system, $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. We now show that this imposes constraints on the predictive power of the multivariate system, which we express in terms of its one-step-ahead predictive $R^2$. We then go on to explore the implications of this result.

## 4.1 Bounds for the predictive $R^2$

**Proposition 3** ***(Bounds for the Predictive*** $R^2$) *Let*

$$R^2 (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = 1 - \sigma_u^2 / \sigma_y^2 \tag{16}$$

*be the $R^2$ for a predictive regression for $y_t$ of the form (3), derived from the ABCD representation (1) and (2). Under A1 to A3, if the Macroeconomist's and Econometrician's ARMA coincide (hence, from Proposition 2, $r = q$), $R^2$ satisfies*

$$0 \leq R^2_{\min} (\boldsymbol{\lambda}, \boldsymbol{\theta}) \leq R^2 \leq R^2_{\max} (\boldsymbol{\lambda}, \boldsymbol{\theta}) \leq 1 \tag{17}$$

*where*

$$R^2_{\min} (\boldsymbol{\lambda}, \boldsymbol{\theta}) = 1 - \sigma_\varepsilon^2 / \sigma_y^2 \tag{18}$$

*is the predictive $R^2$ from the fundamental ARMA$(p, q)$ representation (14), and*

$$R^2_{\max} (\boldsymbol{\lambda}, \boldsymbol{\theta}) = 1 - \left(1 - R^2_{\min} (\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \prod_{i=1}^{q} \theta_i^2 \tag{19}$$

*is the notional $R^2$ from a nonfundamental representation in which all the $\theta_i$ are replaced with their reciprocals. For $q > 0$, the $R^2$ bounds lie strictly within $[0, 1]$; and for any system in which the column rank$\left( \begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \right) > 1 \Rightarrow |corr (u_t, \boldsymbol{\beta}' \mathbf{v}_t)| < 1$, $R^2$ itself lies strictly within $[R^2_{\min}, R^2_{\max}]$.*

**Proof.** See Appendix D. ■

Proposition 3 says that, for the predictive regression for $y_t$ that arises from the structural model, $R^2$ must lie between bounds that can be defined solely in terms of ARMA parameters. We consider both bounds in turn.

## 4.2 The lower bound for $R^2$

The intuitive basis for the lower bound, $R^2_{\min}$, in the inequality in Proposition 3 is quite straightforward and well known (e.g., see Lütkepohl (2007)). The predictions generated by the fundamental ARMA representation condition only on the history of $y_t$; so they cannot be worsened by increasing the information set to include the true predictor vector. Indeed, $R^2$ 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, i.e., for which either $p > 0$ or $q > 0$, this lower bound is itself strictly positive.

### 4.2.1 $R^2_{\min}$ and imperfect predictors

While the rationale for the lower bound is straightforward, it is worthwhile considering its implications. It tells us that $\mathbf{x}_t$, the vector of (perhaps unobserved) predictors for $y_t$ in the correctly specified predictive regression (derived from the structural ABCD system that actually generated the data) must predict at least as well as the univariate representation. But it does *not* tell us that if for some arbitrary observable predictor vector, $\mathbf{w}_t$, we simply run a predictive regression that is just a least squares projection of the form $y_t = \boldsymbol{\gamma}'\mathbf{w}_{t-1} + \xi_t$, then this must imply $R^2_{\mathbf{w}} \geq R^2_{\min}$. If $\mathbf{w}_t \neq \mathbf{x}_t$, but contains elements that are at least somewhat correlated with elements of $\mathbf{x}_t$, any such regression may have predictive power, but will in general be misspecified. Hence the predictive errors $\xi_t$ cannot in general be jointly IID with the innovation to a time series representation of $\mathbf{w}_t$ (a point made forcefully by Pastor and Stambaugh, 2009). There is also no guarantee that any such regression will predict as well as the ARMA.

However, $R^2_{\min}$ *will* be a lower bound for any predictive regression in which information from $\mathbf{w}_t$ is used *efficiently.* Consider (following Pastor and Stambaugh, 2009) some set of estimates $\widehat{\mathbf{x}}_t = E\left(\mathbf{x}_t|\mathbf{w}^t, y^t\right)$ derived by the Kalman Filter, that condition on the joint history both of the observable predictors *and* $y_t$ itself. It follows as a direct consequence of efficient filtering (Hansen & Sargent, 2013, Chapter 8) that the resulting vector of state estimates, $\widehat{\mathbf{x}}_t$, will have the same autoregressive form as the true state variables, with innovations, $\widehat{\mathbf{v}}_t$, that are jointly IID with the innovations to the associated predictive regression $y_t = \boldsymbol{\beta}'\widehat{\mathbf{x}}_{t-1} + \omega_t$.[17] Hence a predictive system of the same form as (3) and (4), but replacing $\mathbf{x}_t$ with $\widehat{\mathbf{x}}_t$, is also nested within the general predictive system.

If the observable predictor vector $\mathbf{w}_t$ has any informational content about $\mathbf{x}_t$ 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 $y^t$. If, in contrast, $\mathbf{w}_t$ reveals no information about $\mathbf{x}_t$ that cannot be recovered from $y^t$, it is predictively redundant, i.e., $E(\mathbf{x}_t|\mathbf{w}_t, y^t) = E(\mathbf{x}_t|y^t)$. This is indeed the null hypothesis of no Granger Causality from $\mathbf{w}_t$, as originally formulated by Granger (1969). But this may not be evident from a least squares regression that does not correctly condition on $y^t$.

### 4.3 The upper bound for $R^2$ and the minimum variance nonfundamental ARMA representation.

To understand the upper bound $R^2_{\max}$ in Proposition 3 recall that, as discussed in Section 3.2, for $q > 0$ there are $2^q - 1$ nonfundamental representations, in which one or more of the $\theta_i$ is replaced by its reciprocal, each of which satisfies the moment conditions, and thus generates identical autocorrelations to (14). In the particular nonfundamental representation relevant to Proposition 3, all the $\theta_i$ in (14) are replaced by their reciprocals

$$\lambda(L)\, y_t = \theta^N(L)\, \eta_t \tag{20}$$

where $\lambda(L)$ is as in (14), and $\theta^N(L) = \prod_{i=1}^{q}\left(1 - \theta_i^{-1}L\right)$. Like all nonfundamental representations, (20) is a non-viable predictive model because, as discussed above, its shocks $\eta_t$, cannot be recovered from the history of $y_t$. However, the *properties* of this representation can still be calculated from the parameters of the fundamental representation in (14). It is straightforward to show (see Appendix D, Lemma 6) that $\eta_t$, the shock process to this representation, has the minimum innovation variance amongst all fundamental or "basic" (i.e., of the same order) nonfundamental representations. If we calculate the notional $R^2$ for (20) this provides the upper bound for the $R^2$ of (16).

Thus the proposition tells us that, while we can increase $R^2$ by using the predictor vector, relative to the lower bound given by the ARMA, there is a limit to the extent that $R^2$ can be increased; and this limit can be calculated using only the univariate properties of $y_t$.

The intuition for this result arises from the feature that, while the shocks to nonfundamental representations cannot be recovered from the history of $y_t$, they *can* be derived

[17] Given our emphasis on population properties we here assume convergence of the Kalman Filter.

as a linear combination of the history and the *future* of $y_t$. Since future values of $y_t$ can be expressed in terms of current and future values of the true predictor vector $\mathbf{x}_t$, it follows that lags of any nonfundamental shocks would (if we could observe them) have predictive power for $\mathbf{x}_t$, and hence for $y_t$.

Thus far the intuition is relatively straightforward. But the key additional feature of the proof of the proposition follows directly from the distinctive properties of the minimum variance nonfundamental representation (20). The proof shows that the shocks to (20), $\eta_t$, can be expressed as a linear combination of future and current, but *not* lagged values of $y_t$. As a result it follows straightforwardly that there must be one-way Granger Causality from $\eta_t$ to $\mathbf{x}_t$, so information from $\mathbf{x}_t$ cannot improve on the predictions of (20). Furthermore, from (19), for $q > 0$ the upper bound for $R^2$ is strictly less than unity.[18]

Proposition 3 also implies that the better $\mathbf{x}_t$ predicts, the more closely the predicted value from (3) must resemble the predictions we would get from (20) if we could observe $\eta_t$. Indeed, the proof of Proposition 3 implies there can be *no* predictive power from $\mathbf{x}_t$ that would not also be in the predictions from (20). The nonfundamental shocks $\eta_t$ have predictive power because they provide a "window into the future": a particular linear combination of current and future values of $y_t$.[19] It follows that any predictive power from $\mathbf{x}_t$, beyond what the history of $y_t$ itself provides must arise from its resemblance to this particular linear combination of future values of $y_t$. For some $y_t$ processes this best possible forecast may be very good ($R^2_{\max}$ may be close to unity); for others it may be not much better than the ARMA forecast ($R^2_{\max}$ may be close to $R^2_{\min}$). But, crucially, the *properties* of this best possible forecast can be determined entirely from the ARMA properties of $y_t$ itself.

## 4.4 Bounds for $R^2$ in the ARMA(1,1) case

It is straightforward to show that in this case the lower bound, the predictive $R^2$ of the ARMA representation can be derived from the parameters of (15) as

$$R^2_{\min}(\lambda, \theta) = \frac{(\theta - \lambda)^2}{1 - \lambda^2 + (\theta - \lambda)^2} \tag{21}$$

[18] Note that even for $q = 0$ the upper bound (19) is still well defined, and equal to unity, even though the representation in (20) is undefined. Note, however, that in light of the discussion in Section 3.8, this case can only occur if the structural model satisfies parameter restrictions: e.g., if $r = 1$, the MA parameter of the macroeconomist's ARMA, $\psi(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, must be precisely equal to zero.

[19] We illustrate this feature for the special case of the ARMA(1,1) in the next section.

The upper bound for $R^2$, is given by

$$R_{\max}(\lambda,\theta) = 1-\theta^2\left(1-R^2_{\min}(\lambda,\theta)\right) = \frac{(1-\lambda\theta)^2}{1-\lambda^2+(\theta-\lambda)^2} \tag{22}$$

which is the notional $R^2$ of the nonfundamental representation associated with (15)

$$(1-\lambda L)\,y_t = \left(1-\theta^{-1}L\right)\eta_t \tag{23}$$

which is a special case of (20).[20]

The nonfundamental representation (23) can be used to illustrate why, as discussed above in Section 3.2, nonfundamental representations must be non-viable predictive models. Hence their predictive $R^2$ is a strictly notional calculation. To see this we can can compare the properties of the shocks to the two ARMA representations. The innovations to (11) can be derived straightforwardly from the history of $y_t$, giving

$$\varepsilon_t = \left(\frac{1-\lambda L}{1-\theta L}\right) y_t = \sum_{i=0}^{\infty}\theta^i\left(y_{t-i}-\lambda y_{t-1-i}\right) \tag{24}$$

In contrast, the shocks to (23) *cannot* be derived from the history of $y_t$. Instead we have

$$\eta_t = \left(\frac{1-\lambda L}{1-\theta^{-1}L}\right) y_t = -\theta L^{-1}\left(\frac{1-\lambda L}{1-\theta L^{-1}}\right) y_t = -\sum_{i=1}^{\infty}\theta^i\left[y_{t+i}-\lambda y_{t+i-1}\right] \tag{25}$$

Thus $\eta_t$ is a linear combination of current and future values of $y_t$.

The bounds in (21) and (22) can be used to illustrate limiting cases.

If $\theta$ is close to $\lambda$, so that $y_t$ is close to being white noise, $R^2_{\min}$is close to zero. If $\theta$ is close to zero, $R^2_{\max}$ is close to one. But only if $\theta$ and $\lambda$ are *both* sufficiently close to zero (implying that both $y_t$ and the single predictor $x_t$ 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 3 entirely devoid of content.

In marked contrast, as $|\theta|$ tends to unity the range of possible values of $R^2$ collapses to a single point (which is $\frac{1-sgn(\theta)\lambda}{2}$). Thus any ARMA$(1,1)$ process with high $|\theta|$ implies that there is very little scope for the underlying predictive model to outperform the ARMA.

[20] Note that, while in general, as discussed in Section 3.2 there will be multiple $(2^q-1)$ nonfundamental representations of the same order, in this particular case, with $q=1$, there is only one.

## 4.5 How similar will multivariate forecasts be to univariate forecasts?

Figure 1 in Section 2 showed that multivariate and univariate forecasts of inflation were fairly strongly correlated. A corollary to Proposition 3 provides a simple explanation.

**Corollary 4** ***(The Prediction Correlation)*** *Letting* $\rho = corr\left(E\left(y_{t+1}|\mathbf{x}_t\right), E\left(y_{t+1}|y^t\right)\right)$ *be the correlation coefficient between the predictions from the predictive regression (3) and from the fundamental ARMA representation, (14), then*

$$\rho = \sqrt{\frac{R^2_{\min}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R^2}} \geq \rho_{\min} = \sqrt{\frac{R^2_{\min}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R^2_{\max}\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}} \geq 0 \qquad (26)$$

*where both inequalities are strong for* $q > 0$*.*

**Proof.** See Appendix E. ■

Corollary 4 shows that the prediction correlation $\rho$ is monotonically decreasing in $R^2$. Thus the closer is $R^2$ to its lower bound, then the more its predictions must resemble those from the ARMA model. But, since Proposition 3 places an upper bound on $R^2$, it also follows that the narrower is the gap between the $R^2$ bounds, the closer to unity is $\rho$; i.e., the univariate and multivariate forecasts must be strongly correlated.

## 4.6 How much would observing structural shocks help prediction?

As discussed in Section 3.4, much applied macroeconomic analysis has focused on identifying structural shocks. In light of the key role that the minimum variance nonfundamental representation plays in our results, it is also notable that there has been a growing literature on nonfundamentalness in the context of structural models. These papers focus on the possibility that, even when the number of structural shocks equals the number of observables, structural shocks may still not be recoverable from the data - whether in the case of a scalar observable, as here, or a vector of observables (Fernández-Villaverde *et al.*, 2007; Leeper *et al.*, 2013). Most commonly this nonfundamentalness is assumed to reflect differences in information sets between the econometrician and the agents in the economy, who are typically assumed to observe the structural shocks.[21]

[21] All nonfundamentalness problems are due to some difference in information sets. They may also arise in models where agents themselves have incomplete information (e.g., Graham and Wright, 2010).

An important implication of our results is that - whatever the implicit interest in attempting to identify structural shocks, or in exploring models in which they may or may not be identified - in simple predictive terms, it may not actually *matter* very much that shocks are hidden. Proposition 3 says that, in cases where the gap between $R^2_{\min}$ and $R^2_{\max}$ is narrow, even if it were possible to use the predictive regression for $y_t$ that conditions on the history of the structural shocks (as captured in the vector of predictors, $\mathbf{x}_{t-1}$) this would not do much better than the (fundamental) ARMA. As a direct implication, from Corollary 4, the predictions from the ARMA would look very similar to the predictions from the structural model.

In terms of the discussion of the relationship between structural shocks and the ARMA innovation (see (13) in Section 3.4), this similarity arises because, even in cases where there are multiple shocks, the true relationship $\varepsilon_t = \mathbf{f}(L)\,\mathbf{s}_t$ may be quite well approximated by $\varepsilon_t \approx \mathbf{f}_0\mathbf{s}_t$, the impact effect in period $t$ of a particular linear combination of structural shocks, even when this linear combination is not fundamental for $y^t$.

It is striking that, in at least some cases, this feature arises as a direct consequence of the structural model. Thus Fernández-Villaverde *et al.* (2007) and Leeper *et al.* (2013) both provide examples of simple structural models which are special cases of our ARMA(1,1) example with a single structural shock, $s_t$, which is (up to a scaling factor) equal to $\eta_t$, the shock to the nonfundamental representation (20). Thus $s_t$ is not observable to the econometrician; but is assumed to be observable to the agents in the economy. In both cases, it is striking that the structure of the model means that the resulting macroeconomist's ARMA is an MA(1) process with a value of $\psi\,(=\theta)$ very close to unity.

Both sets of authors focus on the differences in impulse responses to structural shocks, depending on whether the econometrician treats them as fundamental or nonfundamental (in the time series sense). These differences can be nontrivial.[22]. But in both cases, Proposition 3 tells us that, even if we could directly observe the structural shock, $s_t$, differences in predictive power that would result from actually observing the structural shocks in both models would be minimal. In the case analysed by Fernández-Villaverde *et al.* (2007), for example, $y_t$ is the savings of a forward-looking agent responding optimally to $s_t$, a structural income shock, and the MA parameter of the fundamental representation is simply given by $\psi = \theta = \frac{1}{1+r}$, where $r$ is the real interest rate, so on any plausible calibration can never be far below unity. In this particularly simple case, with a single nonfundamental structural shock, the predictive regression associated with the structural model would simply *be* the nonfundamental representation; and thus its $R^2$

[22]Particularly if, for example, $y_t = \Delta Y_t$ for some unit root process $Y_t$; in which case even small differences above and below unity change the signs of long run responses to structural shocks.

would attain the upper bound in Proposition 3. But, by application of the formulae for the bounds in the ARMA(1,1) case, (21) and (22), setting $\lambda = 0$, we have $R^2_{\min} \approx \frac{1}{2}\left(\frac{1}{1+r}\right)$, $R^2_{\max} \approx \frac{1}{2}\left(\frac{1+2r}{1+r}\right)$: i.e., both upper and lower bounds would be very close to $\frac{1}{2}$. Thus even without observing the structural shock, we would be able to predict $y_t$ quite well. But, more crucially, observing the true structural shock would barely improve our predictions. Furthermore, from Corollary 4 in this simple model the correlation between the unobservable structural shock $s_t$ and the observable univariate innovation would simply be equal to $\rho$, the correlation between the predictions, and both would be equal to $\theta$, here given by $\frac{1}{1+r}$, and hence very close to unity: thus $s_t$ might not be observable, but a very close proxy would be observable.

Of course the two examples cited above are both very stylised, and have the feature that the structural model results in a particular MA representation with $\psi = \theta$ very close to unity. Other structural models might result in values much further from unity, in which case the differences in predictive power would be much more significant. Nonetheless, this analysis is a reminder that lack of observability *per se* may not have major implications in predictive terms.

# 5 Time-varying parameters

In general, if any of the parameters in the structural model (1) and (2) are non-constant over time, this must translate into time variation in the parameters of the predictive system (3) and (4), i.e., the coefficient vector $\boldsymbol{\beta}$, the vector of AR parameters $\boldsymbol{\mu}$ and the error covariance matrix $\boldsymbol{\Omega} = E\left(\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix}\begin{bmatrix} u_t \\ \mathbf{v}_t \end{bmatrix}'\right)$, all of which can be derived as functions of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. This will translate into time variation in the parameters of the univariate representation for $y_t$ (we shall see an example of this in our empirical example, in Section 6). However, this does not of itself detract from the key insights that our analysis provides; it merely complicates the algebra. The proof of our core result, the $R^2$ bounds in Proposition 3, 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 $\boldsymbol{\mu}$, $\boldsymbol{\beta}$ or $\boldsymbol{\Omega}$.

Before considering how to extend our analysis to cases with 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.[23] 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 nests our empirical application as a special case.

**Proposition 5** ***(Bounds for the Predictive $R^2$ of a Time-Varying ARMA(1,1))*** *Consider the time-varying parameter predictive system, with $r = 1$,*

$$y_t = \beta_t x_{t-1} + u_t \tag{27}$$

$$x_t = \lambda_t x_{t-1} + v_t \tag{28}$$

*where $x_t$ is a scalar predictor with 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 $y_t$ admits the time-varying fundamental ARMA(1,1) representation*

$$(1 - \lambda_t L)\, y_t = (1 - \theta_t L)\, \varepsilon_t \tag{29}$$

*with $0 < |\theta_t| < 1$, $0 < |\lambda_t| < 1$, $\lambda_t \neq \theta_t$ (the macroeconomist's and econometrician's ARMA are identical) and $\varepsilon_t$ is a serially uncorrelated error orthogonal to $y^t$, with $E(\varepsilon_t^2) = \sigma_{\varepsilon,t}^2$. Letting*

$$R_t^2 = 1 - \sigma_{u,t}^2 / \sigma_{y,t}^2 \tag{30}$$

*be the time-varying $R^2$ for the predictive regression (27) (with $\sigma_{y,t}^2 = \sigma_{\varepsilon,t}^2 + \theta_t^2 \sigma_{\varepsilon,t-1}^2$) then*

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

[23] 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.

*where* $R^2_{\min,t}$ *is the time-varying* $R^2$ *of (29), and* $R^2_{\max,t}$ *is the time-varying* $R^2$ *of the associated time-varying nonfundamental representation*

$$(1-\lambda_t)\, y_t = (1-\gamma_t L)\, \eta_t \tag{31}$$

*where*

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

**Proof.** See Appendix F ■

The proof of this proposition shows that time-varying parameters introduce simultaneity into the moment conditions for $\theta_t$ and $\sigma^2_{\varepsilon,t}$, which, for the time-invariant case, can be solved independently. While this makes solution of the moment conditions distinctly more complicated for the time-varying case[24] 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 3. All the associated formulae nest the time-invariant results for the ARMA(1,1) model, as given above, as a special case.

While the moment conditions become 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 3.

Note that the case analysed in Proposition 5 nests the Stock Watson (2007) unobserved components stochastic volatility model as a limiting case as $\lambda_t \to 0$, $\forall t$. In that limiting case the inequality constraints on the sequence $\{\theta_t\}$ are imposed by the (time-invariant) structural state space model, which implies that $\theta_t$ is non-negative and less than unity by construction.[25]

We conjecture that the result can be generalised to higher order time-varying ARMA representations[26]; although in practice estimated versions of such models on macroeconomic data (as in our example below) have thus far been of low order.

[24] As far as we are aware the exact derivation of the processes for $\theta_t$ and $\sigma^2_{\varepsilon,t}$, and of the associated nonfundamental representation, has not been carried out before. For example, in Stock and Watson's (2007) unobserved components model of inflation, which reduces to a time-varying MA(0,1) in differences, Stock and Watson derive $\theta_t$ using the time-invariant formula. In practice this provides a good approximation to the formula given above, since the ratio $\frac{\sigma^2_{\varepsilon,t}}{\sigma^2_{\varepsilon,t-1}}$ mostly remains close to unity.

[25] The sequence for $\theta_t$ is driven by the ratio of the variances of the "permanent" and "transitory" innovations, both of which are random walks in logarithms.

[26] It is trivial to extend to higher order AR polynomials, via quasi-differencing.

# 6 An empirical application: U.S. inflation

In this section we look further at the univariate properties of U.S. inflation previously examined in Section 2; and use the results above to consider what they tell us about multivariate macroeconomic models of inflation. In Section 6.1 we examine the implications of Proposition 2 for the dimensions of the predictive model; in Section 6.2 we examine the implications of Propositions 3 and 5 for its predictive power.

Following Stock and Watson (2007) our focus is on inflation as measured by the GDP deflator, although their analysis shows considerable commonality across a range of price indices.

## 6.1 What do the univariate properties of inflation tell us about the nature of a multivariate predictive system for inflation?

Stock and Watson's (2007) preferred model for U.S. inflation, $\pi_t$, as discussed previously in Section 2 is in fact a univariate model of the simplest kind: an MA(1) in $y_t \equiv \Delta\pi_t = (1-\theta L)\,\varepsilon_t$. They capture the changing nature of U.S. inflation in their more than forty year sample either by discrete regime change in 1984 (with the onset of the Great Moderation) or continual parameter drift. Discrete change is modelled by estimating the preferred MA(1) model separately on pre- and post-1984 data. Continual change is captured by the time-varying MA(1) model.

In both cases, the model is a special case of the ARMA(1,1) example, with $\lambda = 0$, as analysed in its time-invariant form at various earlier points in the paper; and in Section 5 for the time-varying case. The fact that $\lambda = 0$ in all versions of this univariate representation immediately provides a crucial piece of information about any multivariate model or predictive system for inflation: that, at least in a single predictor model, any such predictor must be (or be indistinguishable from) an IID process.[27] This puts very strong restrictions on candidate predictors.

In light of this crucial feature of inflation, it is therefore no surprise that Stock and Watson (2007) find that the range of output gap estimates they investigate, that are commonly used to try and forecast inflation, which are all highly persistent time-series, have at best transient predictive power for inflation.

A very similar consideration arises for Smets and Wouters' (2007) DSGE model. The

[27]Note that this statement is consistent with the time-varying framework of Section 5, with possibe time-varying variances in the univariate representation, since we can still let the predictor be IID, if $\beta_t$ is time-varying.

real marginal cost process that they use as the key linkage between the real economy and inflation is strongly persistent. Their implied predictive regression also includes a very important role for an additional markup process (assumed exogenous), which in practice captures at least some of the univariate predictability of inflation.[28] But nonetheless the resulting predictions for inflation, shown in Panel (b) of Figure 1, inherit at least some of the positive persistence of the real economy, and thus are well away from being IID.

The predicted values for inflation from the unrestricted VAR, shown in the bottom panel of Figure 1, provide an interesting contrast. The chart shows that the VAR does predict inflation somewhat better (in sample) than either the univariate model or the Smets-Wouters model. But, strikingly, the predictions that the VAR generates *are* very close to being IID - and thus could be represented, in our framework, as coming from a single composite predictor with $\lambda = 0$. But this is, on the face of it, a surprising result. The predictions are generated by a VAR that includes lagged levels, and thus contains many highly persistent series. Yet the data select a particular combination of series, and lags thereof, to predict inflation, that ends up being extremely close to being IID (or at least appearing to be, in sample).

By extending our analysis to the predictions, $\widehat{y}_t^{VAR}$, generated by the VAR (rather than $y_t$) it is evident that this can only be the case if the univariate reduced form for $\widehat{y}_t^{VAR}$ which must have multiple AR roots (given the large number of persistent series in the VAR) also has MA roots that nearly precisely cancel the AR roots. Nor, in a large high order VAR, with large numbers of unconstrained coefficients, is this particularly difficult to achieve, if the resulting predictive values end up providing the best fit, at least in sample. But this invites the questions: a) is $\widehat{y}_t^{VAR}$, a particular linear combination of lagged elements of the VAR, *actually* an IID process?; and b) would it arise from any structural model? The fact that the predictive power of the VAR for inflation is very significantly reduced out of sample suggests strongly that, while $\widehat{y}_t^{VAR}$ may appear to be IID in sample, this property is in fact largely or wholly an artefact of the particular sample period.[29] Furthermore, if this is the case, and the predicted values are simply an arbitrary linear combination of lags of a large number of persistent and/or nonstationary series, that spuriously appear to be IID in sample, then in reality this linear combination

[28] For more detailed analysis see Mitchell and Wright (2014).

[29] A very similar result can be found in relation to the Smets-Wouters model. If the parameter restrictions implied by the theoretical model are not imposed, and inflation is simply regressed on all 16 (linearly independent) lagged states with unrestricted coefficients, the resulting predicted values for inflation are also extremely close to univariate white noise, and (in sample at least) predict inflation distinctly better, because the model generates the "right kind" of predictions. But of course, as in the case of the VAR, the resulting predictions lose any structural interpretation

is likely to be highly persistent, and may well have a unit root, so that instability in recursive estimation is exactly what we would *expect* to find.[30]

Two further implications of our analysis for inflation are worth noting.

First, it is not *necessarily* the case that the fitted values for inflation from a correctly specified predictive model must be IID. By extension of the argument in Section 3.8, it is trivially the case that if we could see into the future perfectly our predictions for changes in inflation must have exactly the same time series properties as inflation itself - i.e., they must be MA(1). By extension, suppose that we had two underlying state variables for inflation in the ABCD representation, (1) and (2), one of which, $z_{1t}$, was IID, but the other, $z_{2t}$, was itself an MA(1). Then, for the general case the predictions for inflation, $\widehat{y}_t$, a linear combination of $z_{1t}$ and $z_{2t}$ will also be MA(1); and hence could in principle predict arbitrarily well. But except in the case of complete perfect foresight (i.e., $\sigma_u = 0$) a structural model of this type would generate a process for inflation that would in general be MA(2) rather than MA(1), so could only be consistent with the observed properties of inflation if one of the $\theta_i$ was equal to zero - thus again restricting the parameters of the predictive model.[31]

We cannot, of course, rule out the possibility that this might be the case, or at least that the implied restriction on the predictive system might be hard to reject on the available dataset. But if this were the case, and we then estimated, say, a VAR, which contained both $z_{1t}$ and $z_{2t}$, the resulting predicted values $\widehat{y}_t$ would *not* be IID, but would instead be a MA(1). But since, as noted, our unrestricted VAR generates predictions that are indistinguishable from IID, there appears to be no evidence in the data that conflicts with our assumption that there is a single (if possibly composite) IID predictor of changes in inflation.

This in turn leads on to a second implication that should be of particular interest to macroeconomists. If we are looking for IID predictors of inflation that are consistent with the univariate properties of inflation then, in standard macro models, at least, we must conclude that we are looking in the wrong place. As these models pretty much all generate strongly persistent state variables for inflation. One obvious source of IID predictions is a model where the prediction is itself an innovation to some information set: i.e., is "news" (as in, for example, Leeper *et al.*, 2013). But our analysis implies that,

[30]By contrast the predicted value in the univariate model, which is simply $\theta_t \varepsilon_{t-1}$, is by construction a linear combination of lags of $\Delta\pi_t$, and hence must at least be stationary. This almost certainly helps to explain its greater stability in recursive estimation.

[31]A structural model of this type cannot be diagonalised as in Assumption A1; however the model is nested in a more general specification in which $\mathbf{M}$ takes the Jordan Form, which, as noted above in footnote 10, admits ARMA models with $q > p$.

given the univariate properties of inflation, *only* news - in the sense of an IID process - can predict inflation.

We now go on to show that, at least in recent data, the *extent* of this predictive power must also be quite limited; and any predictions are likely to look quite similar to those from a univariate model.

## 6.2 How the univariate properties of inflation constrain the multivariate predictive system for inflation

A key, and widely cited, insight from Stock and Watson (2007) is that for a range of widely used predictors of inflation there is little evidence that it is possible to out-forecast a univariate model, particularly in recent data. Our results shed light on this by showing that in recent data the gap between the $R^2$ bounds implied by the univariate model has narrowed significantly. We demonstrate this result in two ways.

We first take Stock and Watson's (2007) own MA(1) model for changes in inflation, estimated separately on both pre- and post- 1984 data. From their estimated and published values of $\theta$ (reported in their Table 3) we report the implied values of $R^2_{\min}$ and $R^2_{\max}$ using (21) and (22), setting $\lambda = 0$.[32]

Table 1: Implied values of $R^2_{\min}$ and $R^2_{\max}$ from Stock and Watson's (2007) estimated MA(1) model for inflation

| Estimates | MA(1) coefficient | |
|---|---|---|
| | 1960:I-1983:IV | 1984:1-2004:IV |
| $\widehat{\theta}\,(=\widehat{\rho_{\min}})$ | 0.275 (0.085) | 0.656 (0.088) |
| $R^2_{\min}(\widehat{\theta})$ | 0.070 (0.040) | 0.301 (0.042) |
| $R^2_{\max}(\widehat{\theta})$ | 0.930 (0.040) | 0.699 (0.042) |

Notes to Table. Estimates in the first row are taken from Stock and Watson (2007, Table 3); estimated standard errors are in parentheses. Remaining rows use formulae in equations (21) and (22) and (26), setting $\lambda = 0$ with $\theta$ equal to the estimated value in the relevant column of the first row. Standard errors for $R^2$ are approximated using the delta method.

[32] Note that here we follow Stock and Watson (2007) in treating it as known that $\lambda = 0$, and hence only consider the impact of sampling variation in $\widehat{\theta}$. Robertson and Wright (2012, Appendix B.2) consider the case where the true value of $\lambda$ may differ from zero; but this does little to change the conclusions presented here.

The two sub-samples (columns) in Table 1 show a distinct contrast. In the earlier pre-1984 sample, $\widehat{\theta}$ is closer to zero, with the result that the implied range of values of $R^2$ for any predictor is barely constrained, with $R^2_{\min}$ close to zero and $R^2_{\max}$ close to unity. In the second of the two samples $\widehat{\theta}$ is distinctly closer to unity. We are closer to the limiting case discussed in Section 4.4, in which the range of possible values of $R^2$ is considerably narrower.

This narrowing becomes even more striking if, secondly, we follow Stock and Watson (2007) and fit the time-varying unobserved components trend-cycle model with stochastic volatility, seen in Figure 1. This model, as the authors note, is equivalent to a time-varying MA(1) representation (and is thus a special case of Proposition 5). Figure 2 plots the implied but now time-varying values of $R^2_{\min}$ and $R^2_{\max}$. Figure 2, Panel (a) shows that the contraction of the range of feasible values of $R^2$, seen in Table 1 is even more marked when we explicitly allow for time variation in $\theta$. Panel (b) shows that this narrowing of the gap between the bounds is driven by a steady rise in $\theta_t$ in recent decades, which simultaneously raises the lower bound (as univariate predictability rises) and lowers the upper bound.

In the light of these calculations, Stock and Watson (2007)'s conclusion that inflation has become much harder to forecast in recent data becomes readily interpretable in terms of their own univariate representation. Essentially in recent inflation data there is quite limited scope for even the best possible predictor of inflation consistent with the properties of inflation to out-predict the fundamental MA representation.

As a direct implication, wherever $R^2$ lies in this narrow range, from Corollary 4 the predictions from the correctly specified predictive regression must also closely resemble those from the univariate representation. For the time-invariant case this is straightforward to show, since using (21), (22) and (26), setting $\lambda = 0$, the prediction correlation $\rho$ has a lower bound $\rho_{\min}$ which is simply equal to $\theta$. And it should of course be borne in mind that, for any predictive model with an $R^2$ reasonably close to the lower bound, Corollary 4 showed that $\rho$ will be very much closer to unity.

Figure 1 showed that predictions from both the Smets and Wouters DSGE model and the VAR were indeed reasonably strongly correlated with those from the appropriate univariate benchmark, the time-invariant MA(1); both had correlation coefficients of around one half. However, we should be cautious in interpreting this feature of the data since arguably both of these multivariate predictive systems should be regarded as imperfect predictors - i.e., as proxies for the true predictors, as discussed in Section 4.2.1. As noted there, only if predictive information from such predictors is used *efficiently*, such that

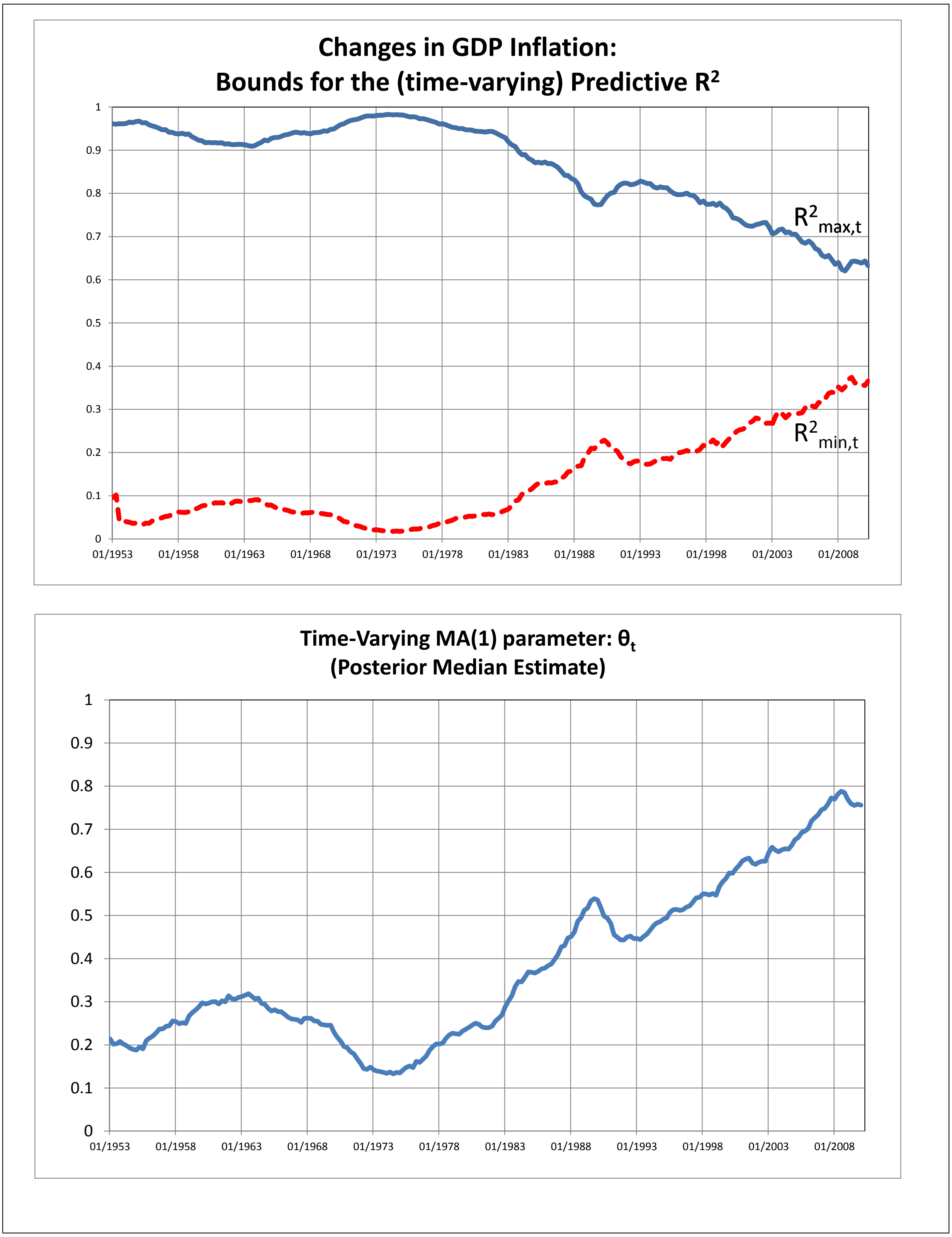


Figure 2: Changes in U.S. inflation: Bounds for the time-varying predictive $R^2_{\min}$ and posterior estimates of $\theta_t$

the resulting predictive errors are truly IID, conditional upon the joint history of the predictors and inflation, can we apply Proposition 3 and Corollary 4.

Thus, for example, in the case of the Smets and Wouters DSGE model, we can ask whether the predictions are efficient, in the sense that they capture all available information, including the history of inflation itself. A simple way of assessing this in-sample is to regress actual inflation on both the Smets and Wouters DSGE predictions for inflation and the univariate predictions.[33] If the Smets and Wouters model were efficient, the univariate predictions would be predictively redundant. Instead the data select what is close to an unweighted average of the two sets of predictions. Unsurprisingly, the correlation between these (approximately) efficient predictions and the univariate predictions is very much higher (at around 0.84) well above the full sample estimate of $\widehat{\rho_{\min}} = \widehat{\theta} = 0.4$. This is indeed to be expected, given that the associated $R^2$ is only around 0.19, that is, very much closer to $R^2_{\min}$ than to $R^2_{\max}$.

Such high values of the prediction correlation $\rho$ bring out the feature that - especially in the recent past - all predictive models for inflation must to a greater or lesser extent be "ARMA-like" in the sense that they generate fairly similar predictions. But a further implication of our analysis, explored in more detail in Robertson and Wright (2012) brings out another respect in which this must be the case. They show that the "Stambaugh Correlation" (Stambaugh, 1999), $corr\,(u_t, \boldsymbol{\beta}'\mathbf{v}_t))$, between predictive errors in $y_t$ and innovations to the predicted values (i.e., in the context of our simple one predictor model) must be very close to unity in absolute value, for *all* predictive models for inflation, wherever $R^2$ lies within its bounds. This means that any predictive system for inflation must also be "ARMA-like" in the sense that it must suffer from acute "Stambaugh Bias". The problems of estimation and inference of ARMA models are well known; but this suggests that, since all (efficient) predictive regressions for inflation must be "ARMA-like", in both the senses above, these problems will apply almost equally to *any* efficient predictive regression for inflation.

# 7 Conclusions

Economists do not forecast in an informational vacuum. Our analysis has shown that what we know about the univariate time-series properties of a process can tell us a lot about the properties of the multivariate model that generated the data, even before we

[33] On the assumption that the true predictor must be IID this is not only the simplest but also the only way to carry out such a test; but we do not wish to set too much store by this.

look for data on predictors or run a single predictive regression. For U.S. inflation, and if we follow Stock and Watson (2007) and assume a univariate MA(1) model for inflation, our results imply: a) that it will be hard in recent data to find multivariate models or predictive regressions that predict much better than the univariate model; b) that the predictions for inflation from any (efficient) predictive regression must be very similar to those from the univariate model; and c) good predictor variables must have little time-series persistence.

# References


[1] Ang, A., G. Bekaert and M. Wei (2007), "Do macro variables, asset markets, or surveys forecast inflation better?", *Journal of Monetary Economics*, 54(4), 1163-1212.

[2] 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.

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

[4] Campbell, J.Y., A.W. Lo and A.C. MacKinlay (2007), *The Econometrics of Financial Market*s. Princeton University Press.

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

[6] 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.

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

[8] 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.

[9] Giannone, D., L. Reichlin and D. Small (2008), "Nowcasting: The real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55(4), 665-676.

[10] Graham, L. and S. Wright (2010), "Information, heterogeneity and market incompleteness", *Journal of Monetary Economics*, 57(2), 164-174.

[11] Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica,* 37(3), 424-438.

[12] Hall, A.R., A. Inoue, J.M. Nason and B. Rossi (2012), "Information criteria for impulse response function matching estimation of DSGE models", *Journal of Econometrics*, 170(2), 499-518.

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

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

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

[16] Komunjer, I. and Ng, S. (2011), "Dynamic Identification of Dynamic Stochastic General Equilibrium Models", *Econometrica*, 79, 1995–2032.

[17] Lahiri, K., G. Monokroussos and Y. Zhao (2013), "Forecasting Consumption: the role of consumer confidence in real time with many predictors", University ot Albany SUNY working paper.

[18] Leeper, E.M. T.B. Walker, S.S. Yang (2013), "Fiscal Foresight and Information Flows", *Econometrica*, 81(3), 1115-1145.

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

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

[21] Mitchell, J. and S. Wright (2014), "The predictable and the unpredictable economy in the Smets-Wouters Model", mimeo.

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

[24] Rapach, D. and G. Zhou (2013), "Forecasting Stock Returns". In *Handbook of Economic Forecasting, Volume 2A*, (eds. Elliott, G. and A. Timmermann), Amsterdam: Elsevier, pp. 328–383.

[25] 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.

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

[27] Samuelson, P. (1965), "Proof that Properly Anticipated Prices Fluctuate Randomly", *Industrial Management Review*, 6, 41-49.

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

[29] Stambaugh, R.F. (1999), "Predictive Regressions", *Journal of Financial Economics,* 54, 375-421.

[30] Stock, J.H and M.W. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20(2), 147-162.

[31] 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.

[32] 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.

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

# Appendix

## 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

$$\boldsymbol{\beta}^{*\prime} = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times(n-1)} \end{array}\right]\mathbf{C}\mathbf{T}^{-1}$$

we can write

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

where

$$u_t = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times(n-1)} \end{array}\right]\mathbf{D}\mathbf{s}_t$$

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 $\beta$. Then we can write

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

as in Lemma 1 with

$$\boldsymbol{\beta} = \left[ \overset{r\times r}{\mathbf{I}} \quad \overset{r\times(n-r)}{\mathbf{0}} \right] \boldsymbol{\beta}^*$$

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

$$\mathbf{x}_t = \mathbf{M}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{A.2}$$

where $\mathbf{M} = diag\,(\mu_1, ..., \mu_r)$ and $\mathbf{v}_t = \mathbf{K}\mathbf{v}_t^*$.

To derive (A.2), partition $\mathbf{K}$, $\mathbf{x}_t^*$ and $\mathbf{M}^*$ conformably as

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

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

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

and

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

so

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

Thus

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

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 (A.2) 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 (A.2) above.■

# B Moment Conditions for the Macroeconomist's ARMA

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

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

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

$$\widetilde{y}_t \equiv \prod_{i=1}^{r}\left(1 - \mu_i L\right) y_t = \sum_{i=1}^{r}\beta_i \prod_{j\neq i}\left(1 - \mu_j L\right) L v_{it} + \prod_{i=1}^{r}\left(1 - \mu_i L\right) u_t \equiv \sum_{i=0}^{r}\boldsymbol{\gamma}'_i L^i \mathbf{w}_t \tag{B.2}$$

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, $\boldsymbol{\gamma}_i\left(\boldsymbol{\beta}, \mu\right)$, for $i = 0, \ldots, r$ each of which is $(r+1) \times 1$.

Let $acf_i$ be the $i$th order autocorrelation of $\widetilde{y}_t$ implied by the predictive system. Write

$\mathbf{\Gamma} = E\left(\mathbf{w}_t \mathbf{w}_t'\right)$ and we have straightforwardly

$$acf_i\left(\boldsymbol{\beta}, \mu, \boldsymbol{\Omega}\right) = \frac{\sum_{j=0}^{r-i} \boldsymbol{\gamma}_j' \boldsymbol{\Omega} \boldsymbol{\gamma}_{j+i}}{\sum_{j=0}^{r} \boldsymbol{\gamma}_j' \boldsymbol{\Omega} \boldsymbol{\gamma}_j} \tag{B.3}$$

To obtain explicitly the coefficients of the MA($r$) representation write the right hand side of (B.2) as an $MA(r)$ process $\sum_{i=0}^{r} \boldsymbol{\gamma}_i' L^i \mathbf{w}_t = \prod_{i=1}^{r} (1 - \psi_i L)\, \varepsilon_t = \psi\left(L\right) \varepsilon_t$ for some white noise process $\varepsilon_t$ and $r^{\text{th}}$ order lag polynomial $\psi\left(L\right)$.

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

$$\prod_{i=1}^{r} (1 - \psi_i L) = 1 + c_1 L + c_2 L^2 + ... + c_r L^r \tag{B.4}$$

Then the $i$th order autocorrelation of $\psi\left(L\right) \varepsilon_t$is given by (Hamilton, 1994, p.51)

$$\frac{c_i + c_{i+1} c_1 + c_{i+2} c_2 + .... + c_r c_{r-i}}{1 + c_1^2 + c_2^2 + ... + c_r^2} \quad i = 1, \ldots, r \tag{B.5}$$

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

$$\frac{c_i + c_{i+1} c_1 + c_{i+2} c_2 + .... + c_r c_{r-i}}{1 + c_1^2 + c_2^2 + ... + c_r^2} = acf_i\left(\boldsymbol{\beta}, \mu, \boldsymbol{\Omega}\right), \quad i = 1, \ldots, r \tag{B.6}$$

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

# C Proof of Proposition 2

Consistency with the macroeconomist's ARMA($r, r$) representation in (6) requires that the lag polynomials $\lambda\left(L\right)$ and $\theta\left(L\right)$ in the econometrician's ARMA($p, q$) representation in (14) must satisfy

$$\frac{\theta\left(L\right)}{\lambda\left(L\right)} = \frac{\varphi\left(L\right)}{\mu\left(L\right)} \tag{C.1}$$

so we have

$$q = r - \#\{\varphi_i = 0\} - \#\{\varphi_i = \mu_i \neq 0\}$$
$$p = r - \#\{\mu_i = 0\} - \#\{\varphi_i = \mu_i \neq 0\}$$

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 the coefficients must match, i.e., $\mu_i = \lambda_i$ and $\theta_i = \psi_i$.■

# D Proof of Proposition 3

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

**Lemma 6** *In the set of all possible nonfundamental ARMA$(p,q)$ representations consistent with (14) in which, for $q > 0$, $\theta_i$ is replaced with $\theta_i^{-1}$ for at least some $i$, the moving average polynomial $\theta^N(L)$ in (20) in which $\theta_i$ is replaced with $\theta_i^{-1}$ for all $i$, has innovations $\eta_t$ with the minimum variance, with*

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

**Proof.** Equating (14) to (20) the non-fundamental and fundamental innovations are related by

$$\varepsilon_t = \prod_{i=1}^{r} \left( \frac{1 - \theta_i^{-1} L}{1 - \theta_i L} \right) \eta_t = \sum_{j=0}^{\infty} c_j \eta_{t-j} \tag{D.2}$$

for some square summable $c_j$. Therefore, since $\eta_t$ is itself IID,

$$\sigma_\varepsilon^2 = \sigma_\eta^2 \sum_{j=0}^{\infty} c_j^2 \tag{D.3}$$

Now define

$$c(L) = \sum_{j=0}^{\infty} c_j L^j = \prod_{i=1}^{r} \left( \frac{1 - \theta_i^{-1} L}{1 - \theta_i L} \right) \tag{D.4}$$

so

$$c(1) = \prod_{i=1}^{r} \left( \frac{1 - \theta_i^{-1}}{1 - \theta_i} \right) = \prod_{i=1}^{r} \left( \frac{-1}{\theta_i} \right) \tag{D.5}$$

and

$$c(1)^2 = \prod_{i=1}^{r} \frac{1}{\theta_i^2} = \left( \sum_{j=0}^{\infty} c_j \right)^2 = \sum_{j=0}^{\infty} c_j^2 + \sum_{k \neq j} c_j c_k \tag{D.6}$$

Since $\varepsilon_t$ is IID we have

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

implying

$$\sum_{j=0}^{\infty} c_j c_{j+k} = 0 \; \forall k > 0 \tag{D.7}$$

Hence we have

$$\sum_{j \neq k} c_j c_k = 2 \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+k} = 0 \tag{D.8}$$

thus

$$\sum_{j=0}^{\infty} c_j^2 = c(1)^2 = \prod_{i=1}^{r} \frac{1}{\theta_i^2} \tag{D.9}$$

Thus using (D.9) and (D.3) we have (D.1).

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, $\theta_i$ is replaced with $\theta_i^{-1}$, $i = 1, \ldots, s(k)$, for $s \leq r$. For any such representation, with innovations $\eta_{k,t}$, we have

$$\sigma_{\eta,k}^2 = \sigma_\varepsilon^2 \prod_{i=1}^{s(k)} \theta_i^2 \tag{D.10}$$

This is minimised for $s(k) = r$, which is only the case for the single representation in which $\theta_i$ is replaced with $\theta_i^{-1}$ 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

$$R_F^2 = R_F^2(\lambda, \theta) = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2} \tag{D.11}$$

and

$$R_N^2 = R_N^2(\lambda, \theta) = 1 - \frac{\sigma_\eta^2}{\sigma_y^2} \tag{D.12}$$

and note that immediately from the above we have

$$R_N^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) = 1 - \left(1 - R_F^2(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \prod_{i=1}^{r} \theta_i^2 \tag{D.13}$$

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

$$R^2 = \frac{\sigma^2_{\widehat{y}}}{\sigma^2_{\widehat{y}} + \sigma^2_u} \tag{D.14}$$

where

$$\sigma^2_{\widehat{y}} = \boldsymbol{\beta}' E\left(\mathbf{x}_t \mathbf{x}_t'\right) \boldsymbol{\beta} \tag{D.15}$$

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

Start from these two ARMA representations

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

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

$$1 + \sum_{i=1}^{r} \frac{\beta_{F,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^{r} \left(1 - \theta_i L\right)}{\prod_{i=1}^{r} \left(1 - \lambda_i L\right)} \tag{D.16}$$
$$1 + \sum_{i=1}^{r} \frac{\beta_{N,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^{r} \left(1 - \theta_i^{-1} L\right)}{\prod_{i=1}^{r} \left(1 - \lambda_i L\right)} \tag{D.17}$$

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

$$\mathbf{x}_t^F = \mathbf{M}\mathbf{x}_{t-1}^F + \mathbf{1}\varepsilon_t \tag{D.18}$$
$$\mathbf{x}_t^N = \mathbf{M}\mathbf{x}_{t-1}^N + \mathbf{1}\eta_t \tag{D.19}$$

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

$$y_t = \boldsymbol{\beta}_F' \mathbf{x}_{t-1}^F + \varepsilon_t \tag{D.20}$$
$$y_t = \boldsymbol{\beta}_N' \mathbf{x}_{t-1}^N + \eta_t \tag{D.21}$$

The predictive regressions in (D.20) and (D.21), together with the processes for the two univariate predictor vectors in (D.18) and (D.19), are both special cases of the general predictive system in (3) and (4), but with rank 1 covariance matrices, $\Omega^F = \sigma_\varepsilon^2 \mathbf{1}\mathbf{1}'$, and $\Omega^N = \sigma_\eta^2 \mathbf{1}\mathbf{1}'$.[34] 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 7** *In the population regression*

$$y_t = \boldsymbol{\nu}_{\mathbf{x}}' \mathbf{x}_{t-1} + \boldsymbol{\nu}_F' \mathbf{x}_{t-1}^F + \xi_t \tag{D.22}$$

*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 (D.18), all elements of the coefficient vector $\boldsymbol{\nu}_F$ are zero.*

**Proof.** The result will follow automatically if we can show that the $x_{it-1}^F$ are all orthogonal to $u_t \equiv y_t - \beta' \mathbf{x}_{t-1}$. Equalising (14) and (3), and substituting from (4), we have (noting that $p = q = r$)

$$y_t = \frac{\prod_{i=1}^r (1 - \theta_i L)}{\prod_{i=1}^r (1 - \lambda_i L)} \varepsilon_t = \frac{\beta_1 v_{1t-1}}{1 - \lambda_1 L} + \frac{\beta_2 v_{2t-1}}{1 - \lambda_2 L} + \ldots + \frac{\beta_r v_{rt-1}}{1 - \lambda_r L} + u_t \tag{D.23}$$

[34] Note that we could also write (D.20) 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, 1981). The implied reduced form process for $y_t$ must be identical to the fundamental ARMA representation (Hamilton, 1994) hence we have $\beta_{F,i} = \beta_i k_i$.

So we may write, using (D.18),

$$\begin{aligned} x_{jt-1}^F &= \frac{\varepsilon_{t-1}}{1-\lambda_j L} \\ &= \left(\frac{L}{1-\lambda_j L}\right) \frac{\prod_{i=1}^r (1-\lambda_i L)}{\prod_{i=1}^r (1-\theta_i L)} \times \\ &\left(\frac{\beta_1 L v_{1t-1}}{1-\lambda_1 L} + \frac{\beta_2 L v_{2t-1}}{1-\lambda_2 L} + \ldots + \frac{\beta_r L v_{rt-1}}{1-\lambda_r L} + u_t\right) \end{aligned} \tag{D.24}$$

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

**Lemma 8** *In the population regression*

$$y_t = \boldsymbol{\phi}_{\mathbf{x}}' \mathbf{x}_{t-1} + \boldsymbol{\phi}_N' \mathbf{x}_{t-1}^N + \zeta_t \tag{D.25}$$

*where $\mathbf{x}_t^N$ is the vector of nonfundamental univariate predictors defined in (D.19), 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 $\eta_t \equiv y_t - \beta_N' \mathbf{x}_{t-1}^N$. Equating (20) and (3), and substituting from (4), we have

$$y_t = \frac{\prod_{i=1}^r (1-\theta_i^{-1} L)}{\prod_{i=1}^r (1-\lambda_i L)} \eta_t = \beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_r \frac{v_{rt-1}}{1-\lambda_r L} + u_t \tag{D.26}$$

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

$$\eta_t = F^r \prod_{i=1}^r (-\theta_i) \left(\frac{\prod_{i=1}^r (1-\lambda_i L)}{\prod_{i=1}^r (1-\theta_i F)}\right) \left(\beta_1 \frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_r \frac{v_{rt-1}}{1-\lambda_r L} + u_t\right) \tag{D.27}$$

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 $\eta_t$ 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 (D.22) analysed in Lemma 7. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^F$ in (D.20) are both nested within (D.22) we must have $R_1^2 \geq R^2$ and $R_1^2 \geq R_F^2$. But Lemma 7 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 (D.25) analysed in Lemma 8. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^N$ in (D.21) are both nested in (D.25) we must have $R_2^2 \geq R^2$ and $R_2^2 \geq R_N^2$. But Lemma 8 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 $rank \begin{bmatrix} \mathrm{B} \\ \mathrm{C} \end{bmatrix} = 1$ . Furthermore for $q > 0$ we have $R_F^2 > 0$ and $R_N^2 < 1$.

This completes the proof of the Proposition. ■

# E Proof of Corollary 4

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

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

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

$$= \left[corr\left(\boldsymbol{\beta}'_F\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right)\right]^2 R^2 \tag{E.3}$$

where we use as derived in the Proof of Proposition 3 above that $Cov\left(\boldsymbol{\beta}'_F\mathbf{x}_t^F, u_t\right) = 0$.

This then gives, exploiting the inequality in the Proposition,

$$\rho = corr\left(\boldsymbol{\beta}'_F\mathbf{x}_t^F, \boldsymbol{\beta}'\mathbf{x}_t\right) = \sqrt{\frac{R_F^2}{R^2}} \geq \sqrt{\frac{R_{\min}^2\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}{R_{\max}^2\left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)}} \geq 0 \tag{E.4}$$

completing the proof of the Corollary.■

# F Proof of Proposition 5

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 (27) as

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

with

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

This nests both (29) and (31) as special cases, with

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

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

where $\theta_t$ and $\sigma^2_{\varepsilon,t}$ jointly satisfy the moment conditions

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

where $\rho_t = corr\left(u_t, v_t\right) \equiv \sigma_{uv,t}/\left(\sigma_{u,t}\sigma_{v,t}\right)$ which can be solved recursively for some initial values $\sigma^2_{\varepsilon,0}$, $\theta_0$ (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

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

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 (F.2), setting $\rho_{t-1} = 1$ throughout, and solving recursively backwards. The proof of the inequality follows analogously to the proof of Proposition 3, since this only requires serial independence, it does not require that $\mathrm{w}_t$ 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 $\varepsilon_t$ is a combination of current and lagged $y_t$, whereas $\eta_t$ is a combination of strictly future values of $y_t$. Thus $\eta_t$ must have predictive power for all possible predictors (except itself), but not vice versa.

To extend to the ARMA(1,1) case, substitute from (28) into (27), 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 (F.1) we can then apply the same arguments as for the MA(1) case.■