R^2 bounds for predictive models: what univariate properties tell us about multivariate predictability[*]

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

November 19, 2017

Abstract

A longstanding puzzle in macroeconomic forecasting has been that a wide variety of multivariate models have struggled to out-predict univariate models consistently. We seek an explanation for this puzzle in terms of population properties. We derive bounds for the predictive R^2 of the true, but unknown, multivariate model from univariate ARMA parameters alone. These bounds can be quite tight, implying little forecasting gain even if we knew the true multivariate model. We illustrate using CPI inflation data.

Keywords: Forecasting; Macroeconomic Models; Autoregressive Moving Average Representations; Predictive Regressions; Nonfundamental Representations; Time-Varying ARMA; Inflation Forecasts

JEL codes: C22, C32, C53, E37

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

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

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

1 Introduction

A long-standing and, on the face of it, puzzling feature of macroeconomic forecasting (that goes back at least as far as Nelson, 1972) has been that a wide variety of multivariate models have struggled to out-predict univariate models, particularly in terms of a consistent performance both over time and over a range of variables.[1] Indirect evidence of the power of univariate models can also be inferred from the relative forecasting success of Bayesian VARs that utilise Minnesota type priors (e.g., see Banbura *et al.*, 2010; Canova, 2007, p. 378), since these effectively give greater weight in estimation to finite order univariate autoregressive representations.

In this paper we seek insights into this puzzle in terms of population properties. We analyse a stationary univariate time series process, y_t, data for which are assumed to be generated by a multivariate macroeconomic model. We then take a backwards look at the relationship between multivariate and univariate properties, by asking what the univariate ARMA representation can tell us about the properties of the true multivariate model that generated the data.

We first ask: how much better could we predict y_t if we could condition on the true state variables of the underlying multivariate model, rather than just use the ARMA? We show that the resulting one-step-ahead predictive R^2 must lie between bounds, $R^2_{\min}$ and $R^2_{\max}$, that can be derived from ARMA parameters alone. The R^2 bounds will usually lie strictly within $[0, 1]$. We first derive these bounds for a time-invariant framework, and then show how they can be generalised to models with time-varying parameters. Hence our core results do not rely on the assumption of structural stability.

The lower bound, $R^2_{\min}$, is simply the one-step-ahead R^2 of the fundamental ARMA representation. We show that $R^2_{\max}$ is the (strictly notional) R^2 of a particular "nonfundamental" (Lippi and Reichlin, 1994) representation. While such nonfundamental representations are nonviable as predictive models their *properties*, and hence $R^2_{\max}$, can be derived from the ARMA parameters.

For some time series, ARMA properties imply that the gap between $R^2_{\min}$ and $R^2_{\max}$ is quite narrow. In such cases our results show that little improvement in predictive performance would be possible, even if we had the true state variables for y_t. We show

---

[1]On the problems of providing consistent forecasting performance over time, for a range of macro time series, see e.g., D'Agostino and Surico (2012); Chauvet and Potter (2013); Rossi (2013a); Estrella and Stock (2015); Stock and Watson (2007, 2009, 2010, 2016). In contrast, Banbura *et al.* (2010), Koop (2013) and Carriero *et al.* (2016), for example, find that large Bayesian VAR models can (but do not always) outpredict smaller models, including univariate (AR) models; and Stock and Watson (2002) find that forecasts from factor models can outperform univariate (AR) benchmarks, but typically less so for nominal than real variables.

that this case is particularly likely to occur when y_t is the first difference of an I(1) process, with a Beveridge and Nelson (1981) unit root permanent component with relatively low volatility.

The R^2 bounds are a population property. Clearly in a finite sample the true ARMA representation, and hence the true R^2 bounds, are not known. However, we *can* calculate the bounds for commonly used univariate representations, and we show that this provides important insights. Furthermore, we show that, even if these representations are mis-specified, because the true ARMA is higher order, but close to cancellation, univariate properties can still provide important information about the nature of multivariate predictability. In particular, we note the implications for the time series properties of one-step ahead predictions and the covariance structure of the underlying system.

We illustrate our analysis using data on CPI inflation in eight countries. (For space reasons we focus on results for the US in the main paper, with those for seven other countries discussed in the online appendix.) We calculate the (time-varying) R^2 bounds implied by two commonly used univariate unobserved components (UC) representations, both of which are nested within a time-varying parameter ARMA(1,1) model.[2] These two models are Stock and Watson's (2007) UC-stochastic volatility model and Chan, Koop and Potter's (2013) UC model with an autoregression in the transitory component.

The rest of the paper is structured as follows. Section 2 sets out the links between the ARMA representation and the multivariate model; and describes the R^2 bounds and their implications. In Section 3 we illustrate our results for the special case of an ARMA(1,1). Section 4 shows that our core results can be generalised to accommodate time variation in parameters. Section 5 considers the implications of cancellation, or near-cancellation, of AR and MA polynomials in the true ARMA for inference in finite samples. Section 6 presents the empirical application. Section 7 concludes. Online appendices provide proofs and derivations, estimation results for the seven other countries and background detail for our empirical application.

---

[2]We derive moment conditions for the ARMA(1,1) models implied by these UC representations. As far as we are aware, these derivations are also new.

2 What the population ARMA representation tells us about the true multivariate system: the time-invariant case

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

Consider a univariate time series y_t that is generated by a linear (or linearised) multivariate macroeconomic model:[3]

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

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

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

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

We make the following assumptions:

Assumptions

A1 $\mathbf{A}$ *can be diagonalised as* $\mathbf{A} = \mathbf{T}^{-1}\mathbf{M}\mathbf{T}$ *where* $\mathbf{M}$ *is an* $n \times n$ *diagonal matrix.*

A2 $eig(\mathbf{M}) = \{\mu_i\}$, *with* $|\mu_i| < 1 \ \forall i$.

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

Assumption A1, that the ABCD system can be diagonalised, is in most cases innocuous.[4] Assumption A2, that the system is stationary, is also simply convenient: some or all of the elements of $\mathbf{y}_t$ and $\mathbf{z}_t$ may in principle be stationary transformations of underlying nonstationary series. Assumption A3 follows Fernández-Villaverde *et al.* (2007);

---

[3]We use the notation of the generic ABCD representation of Fernández-Villaverde *et al.* (2007). They assume that this system represents the rational expectations solution of a DSGE model (in which cases the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ are usually functions of a lower dimensional vector of deep parameters, $\boldsymbol{\delta}$). But the representation is sufficiently general to capture the key properties of a wide range of multivariate models, including VAR and factor models. Note that the state vector $\mathbf{z}_t$ may contain information from the history of $\mathbf{y}_t$ itself. In the benchmark structural DSGE model of Smets and Wouters (2007), for example, $\mathbf{z}_t$ contains levels of 6 out of the 7 observables in $\mathbf{y}_t$. The system can also represent the companion form of a VAR.

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

it is convenient (but not essential) to assume normality to equate expectations to linear projections; while the normalisation of the structural shocks to be orthogonal, with unit variances, is simply an identifying assumption, with the matrices $\mathbf{B}$ and $\mathbf{D}$ accounting for scale factors and mutual correlation. The assumption that the structural disturbances $\mathbf{s}_t$ are serially uncorrelated, while standard is, however, crucial - as we discuss below in Lemma 1.

Note that the time-invariant nature of the model is *not* crucial; it merely simplifies the exposition. In Section 4 we consider generalisations to cases where the parameters of the structural model may vary over time.

These assumptions allow us to derive a particularly simple specification for the true predictive regression for y_t, a single element of $\mathbf{y}_t$. This conditions on a minimal set of AR(1) predictors that are linear combinations of the state variables in the system in (1) and (2):[5]

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

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

where $\mathbf{x}_t = (x_{1t}, ..., x_{rt})'$ is an $r \times 1$ vector of predictors with law of motion

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

with $\boldsymbol{\Lambda} = diag\left(\lambda_1, \ldots, \lambda_r\right)$, $i = 1, \ldots, r$, where the λ_i are elements of $\{\mu_i\} = eig\left(\mathbf{M}\right)$ such that $\beta_i \neq 0$, and $\lambda_i \neq \lambda_j$, $\forall i$, and hence $r \leq n$.

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

Remark: *Elements of the predictor vector $\mathbf{x}_t$ in the true predictive regression may be aggregates of the elements of the underlying true state vector $\mathbf{z}_t$ if $\mathbf{A}$, the autoregressive matrix of the states, has repeated eigenvalues. Additionally, if the ABCD representation has a block-recursive structure, there may be state variables with no predictive role for y_t. Thus r, the dimension of the predictor vector, may be substantially less than n, the dimension of the true underlying states. At most one element may have $\lambda_i = 0$ in which case x_{it} is NIID.*

---

[5]All proofs are in the online appendix.

2.2 The Macroeconomist's ARMA

Exploiting standard results (e.g., applying Corollary 11.1.2 in Lütkepohl (2007)), it is straightforward to derive the true univariate reduced form for y_t:[6]

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

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

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

The θ_i are solutions to a set of r moment conditions that match the autocorrelations of y_t, as set out in Appendix B. The condition $|\theta_i| \leq 1$, $\forall i$, gives the unique fundamental solution (Hamilton, 1994, pp. 64-67; Lippi and Reichlin, 1994) since it ensures that $\varepsilon_t = \theta(L)^{-1} \lambda(L) y_t$ is recoverable as a non-divergent sum of current and lagged values of y_t.[7]

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

2.3 Bounds for the predictive R^2

We have derived the ARMA representation from the underlying structural model. We now look at this process backwards, and ask: what do the population univariate properties of y_t, as captured by $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, tell us about the properties of the structural multivariate system that generated the data for y_t?

We first show that the degree of predictability measured by the R^2 of the true predictive regression (3) lies between bounds that can be defined solely in terms of population ARMA parameters. Denote $\sigma_u^2 = Var(u_t)$, $\sigma_y^2 = Var(y_t)$ and $\sigma_\varepsilon^2 = Var(\varepsilon_t)$.

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

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

---

[6]This draws on the seminal work of Zellner and Palm (1974) and Wallis (1977).

[7]The limiting case $|\theta_i| = 1$, for some i, which is not invertible but is still fundamental, may in principle arise if y_t has been over-differenced. But since this case essentially arises from a mis-specification of the structural (multivariate) model we do not consider it further.

be the one-step-ahead predictive R^2 for the true predictive regression for y_t (3), that is derived from the ABCD representation (1) and (2) of the true multivariate model. Under A1 to A3, R^2 satisfies

$$0 \leq R^2_{\min}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \leq R^2 \leq R^2_{\max}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \leq 1 \tag{7}$$

where

$$R^2_{\min}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = 1 - \sigma^2_\varepsilon / \sigma^2_y \tag{8}$$

is the predictive R^2 from the ARMA representation (5) and

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

Corollary 1 (R^2 **bounds for a minimal ARMA**). *If the macroeconomist's ARMA (5) is a minimal representation (i.e., $\theta_i \neq \lambda_j, \theta_i \neq 0, \forall i, \forall j$) then the R^2 bounds lie strictly within $[0, 1]$.*

2.3.1 The lower bound for R^2

The intuitive basis for the lower bound, $R^2_{\min}$, is straightforward and follows from known results (e.g., see Lütkepohl (2007), Proposition 11.2). Predictions generated by the fundamental ARMA representation condition only on the history of y_t; so they cannot be worsened by conditioning on the true state variables.[8] Indeed, the true R^2 must be strictly greater than $R^2_{\min}$ except in the limiting case that $u_t = \varepsilon_t$. Furthermore, for any y_t process that is not IID (which would imply a *non*-minimal ARMA in (5)) this lower bound is itself strictly positive.

2.3.2 The upper bound for R^2

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

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

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

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

---

[8]Which may in principle, as noted above, contain information from the history of y_t itself.

Recall that, in deriving the ARMA from the structural model, we noted that the MA parameters, θ must satisfy r moment conditions to match the autocorrelations of y_t, subject to the constraint that all the θ_i live within $(-1, 1)$. However, there are a further $(2^r - 1)$ *non*fundamental ARMA representations, in which one or more of the θ_i is replaced by its reciprocal, (Lippi and Reichlin, 1994),[9] each of which also satisfies the moment conditions, and thus generates identical autocorrelations to (5). In the particular nonfundamental representation, (10), relevant to Proposition 1 all the θ_i in (5) are replaced by their reciprocals.[10]

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

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

In Section 2.4 below we provide further intuition for the existence of an upper bound; in Section 3.4 we illustrate in a simple analytical example.

2.3.3 The R^2 bounds and observable predictors

Our R^2 bounds apply to predictions that condition on the true state variables that generated the data for y_t. In practice, of course, we must make do with predictors we can actually observe. Suppose, for some observable predictor vector, $\mathbf{q}_t$, we simply run a predictive regression that is just a least squares projection of the form $y_t = \boldsymbol{\gamma}' \mathbf{q}_{t-1} + \xi_t$. If $\mathbf{q}_t \neq \mathbf{x}_t$, but contains elements that are at least somewhat correlated with elements of $\mathbf{x}_t$, any such regression may have predictive power, but we would *not* necessarily expect the resulting predictive R^2 to exceed our lower bound, $R^2_{\min}$.[11]

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

Corollary 2 (R^2 *Bounds for observable predictors with efficient filtering*) *Consider some set of estimates $\widehat{\mathbf{x}}_t = E\left(\mathbf{x}_t | \mathbf{q}^t, y^t\right)$ derived by the Kalman Filter, that condition*

---

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

[10]Note that if $\theta_i = 0$ for some i (hence the ARMA is not a minimal representation) the nonfundamental representation is undefined but we can still use (9) to calculate $R^2_{\max} = 1$.

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

on the joint history of a vector of observable predictors, $\mathbf{q}^t$ and y^t. The predictive R^2 for a predictive regression of the same form as (3), but replacing $\mathbf{x}_{t-1}$ with $\widehat{\mathbf{x}}_{t-1}$, also satisfies $R^2_{\widehat{\mathbf{x}}} \in [R^2_{\min}, R^2_{\max}]$, as in Proposition 1.

If the observable predictor vector $\mathbf{q}_t$ has any informational content about the true state variables that is independent of the history y^t, then $R^2_{\widehat{\mathbf{x}}}$ *must* be strictly greater than $R^2_{\min}$, since this comes from a predictive model that conditions only on the history y^{t-1}. Clearly the more information $\mathbf{q}_t$ reveals about the true states, the closer $R^2_{\widehat{\mathbf{x}}}$ can get to R^2. If, in contrast, $\mathbf{q}_t$ reveals no information about $\mathbf{x}_t$ that cannot be recovered from y^t, it is predictively redundant, in which case $E(\mathbf{x}_t | \mathbf{q}_t, y^t) = E(\mathbf{x}_t | y^t)$, implying $R^2_{\widehat{\mathbf{x}}} = R^2_{\min}$.[12]

2.4 The Predictive Space

While the focus of this paper is on the R^2 bounds, this is not the only information that the population ARMA representation provides about the predictive system. Nor indeed is it necessary to know the full set of ARMA parameters; even a restricted set of univariate characteristics can also provide information.

Following Mitchell, Robertson and Wright (2017), let $\mathcal{P}_r$ be the parameter space of all possible predictive models with r predictors. The parameters of the predictive model map to some set of univariate properties, $\mathbf{u}$. Such properties might, for example, be the full set of ARMA parameters (i.e., $\mathbf{u} = (\boldsymbol{\lambda}, \boldsymbol{\theta})$) but they might simply the be property that y_t is, for example a near-IID process (i.e., $R^2_{\min}$ is less than some particular value) or has a Beveridge and Nelson (1981) permanent component $c_y(1)$, where $y_t = c_y(L)\varepsilon_t$, that is less than unity.

Suppose that we observe - or possibly simply wish to assume - some set of univariate properties $\mathbf{u}$. Then there is an inverse mapping that describes the parameter space of all possible predictive models that could have generated the univariate property $\mathbf{u}$. Mitchell, Robertson and Wright (2017) denote this the "Predictive Space", $\mathbb{P}_{\mathbf{u}}$, a strict (and often quite restricted) subset of the full parameter space, $\mathcal{P}_r$.[13]

In this paper we let $\mathbf{u} = (\boldsymbol{\lambda}, \boldsymbol{\theta})$, and the R^2 bounds derived above are a key defining characteristic of $\mathbb{P}_{\boldsymbol{\lambda}, \boldsymbol{\theta}}$, but by no means the only one.

Mitchell, Robertson and Wright (2017) provide some intuition for this broader class of restrictions on the parameter space of the underlying predictive model, and how they

---

[12]This is indeed the null hypothesis of no Granger Causality from $\mathbf{q}_t$, as originally formulated by Granger (1969) (although in practice in most econometric testing y^{t-1} is typically only included via a finite set of autoregressive terms).

[13]$\mathbb{P}_{\mathbf{u}}$ is the pre-image of $\mathbf{u}$ in $\mathcal{P}_r$.

relate to the R^2 bounds.[14] Consider the case where we observe (or assume) that y_t has some particular univariate property, or set of properties, **u.** It must immediately follow that, for any predictive model with predictions $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_{t-1}$, then in the limit as $R^2 \to 1$, $\widehat{y}_t \to y_t$, and hence $\widehat{y}_t$ must *also* have **u.** Conversely, for any predictive model for which $\widehat{y}_t$ does *not* have **u,** i.e., for which $\mathbf{u}_{\widehat{y}} \neq \mathbf{u},$ this must imply (at least) an upper bound on R^2. Thus even when we do not know the true ARMA representation, knowing **u** and at least some general properties of $\mathbf{u}_{\widehat{y}}$ must in general imply R^2 bounds. In Proposition 4 we provide an example.

But a second important aspect of the predictive space $\mathbb{P}_\mathbf{u}$ is that since the sum of the predictions and the prediction errors must by construction match the univariate properties of y_t, then if the predictions themselves do not display some univariate property **u** then this must imply restrictions on the covariance matrix of innovations in (3) and (4).

We focus here on on two particular features of the predictive space $\mathbb{P}_{\boldsymbol{\lambda},\boldsymbol{\theta}}$ that contains the parameters of the true predictive model, both of which arise as corollaries of Proposition 1.

2.4.1   Time series properties of the predictions

Corollary 3 *(**Time series properties of the predictions**) Whereas y_t has an $ARMA(r,r)$ representation in population, the predictions $\widehat{y}_t = \boldsymbol{\beta}'\mathbf{x}_t$ have an $ARMA(r, r-1)$ representation.*

The key insight here is that the time series properties of the predictions $\widehat{y}_t$ are inherently different from those of y_t itself (in terms of the analysis of the previous section, $\mathbf{u}_{\widehat{y}} \neq \mathbf{u}$). Indeed this inherent difference in time series properties is an essential part of the explanation of *why* there must be an upper bound for R^2, as in Proposition 1. We discuss this issue further in Section 3.5.2 below. We shall also see that this difference in time series properties provides important insights into the empirical application discussed in Section 6.

2.4.2   Covariance properties

Since both R^2 bounds are associated with ARMA representations, Proposition 1 also provides an example of the implications for the covariance properties of the underlying

---

[14]One of the referees objected to our use of the term "restrictions" on the predictive system. Clearly in causal terms the properties of the predictive system determine univariate properties, and not vice-versa. However, in strict mathematical terms, if we observe (or assume) a population univariate property, this does indeed restrict the parameter space of predictive systems that could have generated that property.

structural model and its associated predictive regression:

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

Thus not only do the R^2 bounds reveal limits to the degree of multivariate predictability, they also shed light on the necessary properties of innovations to any predictive system within the predictive space $\mathbb{P}_{\boldsymbol{\lambda},\boldsymbol{\theta}}$. In the neighbourhood of either bound, prediction errors for y_t and innovations to the predictor vector $\mathbf{x}_t$ must be close to perfectly correlated, so they must be close to being generated by a single structural shock. Thus, even in cases where the gap between R^2 bounds is wide, the closer a predictive is to *attaining* the upper bound, the tighter is the parameter space it can inhabit.

In Section 3 we also show that, for an important special case, Corollary 4 has stronger implications: the correlation between innovations to $\widehat{y}_t$ and those to y_t may be bounded for *any* predictive system within the predictive space.

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

As an illustrative example we explore an important special case, in which y_t admits an ARMA(1,1) representation, which arises from a single predictor model, but is also consistent with a widely used univariate unobserved components model.

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

Consider the case in which data for y_t are generated by an ABCD model with a single state variable and a 2×1 vector of structural shocks. In this case the structural and predictive models coincide:

$$y_t = \beta x_{t-1} + u_t \tag{11}$$

$$x_t = \mu x_{t-1} + v_t \tag{12}$$

where in terms of the ABCD representation we have $\mathbf{x}_t = x_t = z_t$, $\mathbf{A} = \mu$, $\mathbf{C} = \beta$, $v_t = \mathbf{B}s_t$ and $u_t = \mathbf{D}s_t$, with $\mathbf{B}$ and $\mathbf{D}$ both 1×2 row vectors that generate a covariance structure

for u_t and v_t; let $\sigma_{uv} = cov(u_t, v_t)$. This simple system has been widely employed.[15] A predictive system of this form can also easily subsume the case of an underlying structural ABCD representation in which the state vector $\mathbf{z}_t$ has $n > 1$ elements, with multiple eigenvalues, but where the subset of state variables that predict y_t can be reduced to a single predictor, with AR parameter μ. We note below that this framework also nests a very commonly used unobserved components representation.

3.2 The moment condition for θ

By substitution from (11) into (12) we have

$$(1 - \lambda L) y_t = \beta v_{t-1} + (1 - \lambda L) u_t \tag{13}$$

with $\lambda = \mu = eig(A)$. The right-hand-side of this expression is an MA(1) so y_t admits a fundamental ARMA(1,1) representation

$$(1 - \lambda L) y_t = (1 - \theta L) \varepsilon_t \tag{14}$$

with $|\theta| < 1$. The first order autocorrelation of the MA(1) process on the right-hand-side of (14) matches that of the right-hand-side of (13): i.e., the single MA parameter θ is the the solution in $(-1, 1)$ to the moment condition

$$\frac{-\theta}{1 + \theta^2} = \frac{\beta \sigma_{uv} - \lambda \sigma_u^2}{\left(1 + \lambda^2\right) \sigma_u^2 + \beta^2 \sigma_v^2 - 2\lambda\beta\sigma_{uv}} \tag{15}$$

Since the autocorrelation on the right-hand-side of (15) is derived from the parameters of the ABCD representation, we have $\theta = \theta\left(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\right)$.

3.3 An unobserved components decomposition

An alternative way to derive the univariate ARMA(1,1) is to consider the unobserved components (UC) decomposition of an $I(1)$ process, Y_t, into a random walk trend component,

---

[15]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.

τ_t, and a stationary AR(1) component, c_t:[16]

$$
\begin{aligned}
Y_t &= c_t + \tau_t \\
c_t &= \mu c_{t-1} + s_{c,t} \\
\tau_t &= \tau_{t-1} + g + s_{\tau,t}
\end{aligned}
\tag{16}
$$

where $s_{c,t} \sim i.i.d.(0, \sigma_c^2)$, $s_{\tau,t} \sim i.i.d.(0, \sigma_\tau^2)$ and $\sigma_{c\tau} = cov(s_{c,t}, s_{\tau,t})$.

In this UC representation the trend may have a deterministic element (if $g > 0$) and a unit root stochastic component (when $\sigma_\tau > 0$).

The representation, (16), can be viewed in two distinct ways.

It is straightforward to show that, without imposing any restrictions on the structure of the model, it can be reparameterised as a predictive system with the same structure as (11) and (12), in which the stationary AR(1) component c_t is the single predictor variable for $y_t = \Delta Y_t$.[17]

More commonly, this representation is used as alternative (implicit) derivation of the univariate ARMA(1,1) for $y_t = \Delta Y_t$ by imposing the identifying assumption that the (innovations to the) trend and stationary components are orthogonal ($\sigma_{c\tau} = 0$) since then (14) and (16) contain the same number of parameters. In this case, the stationary component c_t can be interpreted in filtering terms as an *estimate* (up to a scale factor) of the true state variable x_t, conditional only upon data for y_t. But we note that it also imposes a nontrivial restriction on the parameter space of the ARMA:

Lemma 3 *In the UC representation in (16), if $\mu \geq 0$, $\sigma_\tau > 0$ and $\sigma_{c\tau} = 0$, then $y_t = \Delta Y_t$ admits a restricted ARMA(1,1) representation as in (14) with $0 < \lambda < \theta < 1$. Hence the Beveridge-Nelson decomposition of $Y_t = \frac{c_y(L)}{(1-L)}\varepsilon_t$ has $c_y(1) = (1-\theta)/(1-\lambda) < 1$.*

The restricted nature of this ARMA representation makes it particularly suitable for "near-stationary" Y_t processes: a prime example being inflation, which we analyse later in this paper. In terms of our analysis of R^2 bounds we shall see that one key feature of this representation is that if c_t, the AR(1) component, is strongly persistent, i.e., $\mu\,(=\lambda)$ is close to unity, then the MA parameter θ must be even closer to unity. We shall see that this feature of the state space model has strong implications for the nature of the R^2 bounds for the first difference of any series that can be represented as in (16).

---

[16]We focus here on the time-invariant case; but in Section 6 below we extend the analysis to the case where μ, σ_τ^2 and σ_c^2 are all potentially time-varying.

[17]See Appendix I.1 for the reparameterisation in the time-varying case, which nests the time-invariant case here.

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

The R^2 of the true predictive regression (11) that conditions on the single true state variable $x_t = z_t$ has a lower bound given by

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

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

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

which would be the notional R^2 of the nonfundamental representation associated with (14)

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

which is a special case of (10).[18]

3.4.1 The upper bound for R^2 and the nonfundamental representation

To provide intuition for the upper bound, note that using straightforward manipulations we can reparameterise (19) as a special case of the predictive system in (11) and (12) as

$$y_t = \beta_N x^N_{t-1} + \eta_t \tag{20}$$
$$x^N_t = \lambda x^N_{t-1} + \eta_t$$

with $\beta_N = \lambda - \theta^{-1}$, where the maximal R^2 would be attained by the state variable $x^N_t = (1 - \lambda L)^{-1} \eta_t$. Since the resulting predictive system is a reparameterisation of a nonfundamental representation x^N_t cannot be derived as a convergent sum of past y_t. However we can write, using (19),

$$x^N_t = \left(1 - \theta^{-1}L\right)^{-1} y_t = -\frac{\theta L^{-1}}{(1 - \theta L^{-1})} y_t = -\sum_{i=0}^{\infty} \theta^i y_{t+i+1} \tag{21}$$

so x^N_t is a convergent sum of *future* values of y_t. Thus predictive power comes about because x^N_t acts as a window into the future: the lower is θ, the more it will reveal.[19]

---

[18]Note that the moment condition (15) is satisfied by θ and also by θ^{-1}. While in general, as discussed in Section 2.2, there will be multiple nonfundamental representations of the same order, in this particular case, with $r = q = 1$, there is only one.

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

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

3.4.2 The R^2 bounds in some special cases of the ARMA(1,1)

The bounds in (17) and (18) can be used to illustrate some important special cases.

As a benchmark case, we start by considering the single limiting case in which the R^2 bounds are *not* interesting. If θ is close to λ, so that y_t is close to being white noise, $R^2_{\min}$ is close to zero. If θ is also close to zero, $R^2_{\max}$ is close to one. But only if θ and λ are *both* sufficiently close to zero does the inequality for R^2 open up to include the entire range from zero to unity. Thus only in this doubly limiting case is Proposition 1 entirely devoid of content. Note also that in this case both y_t and the single predictor x_t are close to white noise.

In marked contrast, as $|\theta|$ approaches unity the value of R^2 tends to a single point $(\frac{1-sign(\theta)\lambda}{2})$. This has the important implication that for any ARMA$(1,1)$ process with high $|\theta|$ there is very little scope for the true predictive regression to outperform the ARMA.

The unobserved components model in (16), in which $y_t = \Delta Y_t$ is the first difference of a unit root process, is an important, and commonly applied, special case in which there are strong *a priori* grounds to expect this to be the case. As noted above, from Lemma 3, we must have $\theta > \lambda > 0$. If the transitory component c_t is strongly persistent (λ close to unity) then θ must be even closer to unity, implying that $R^2_{\max}$ will be close to $R^2_{\min}$.

3.5 The Predictive Space for an ARMA(1,1)

In Section 2.4 we noted that univariate properties do not just provide us with R^2 bounds. The parameters of the true predictive system must live within the "Predictive Space" of predictive systems that generate these univariate properties. We noted two corollaries of Proposition 1 that illustrate the features of this parameter space. The ARMA(1,1) case provides a useful illustration since it can be shown that in this case the predictive space has a particularly simple representation.

3.5.1 Corollary 4 in the ARMA(1,1) case: bounds on the innovation correlation

Corollary 4 showed that, for the general case, if the true predictive system has an R^2 in the neighbourhood of either of the R^2 bounds, its innovation covariance matrix must be

close to being rank 1: i.e., innovations to (all) predictors would be strongly correlated with u_t, the prediction error. In the context of one particular case of the ARMA(1,1) this can apply for *any* value of R^2:

Proposition 2 (Bounds for ρ_{uv} for an ARMA(1,1)/Unobserved Components Representation) *Consider the fundamental $ARMA(1, 1)$ univariate representation (14) which is the reduced form of a predictive system with $r = 1$, with predictive eror u_t and a single $AR(1)$ predictor with innovations v_t. For $0 < \lambda < \theta \leq 1$ (as implied by the univariate unobserved components model (16)) the absolute value of the innovation correlation $\rho_{uv} = corr\,(u_t, v_t)$ satisfies*

$$|\rho_{uv}| = |corr\,(u_t, v_t)| \geq \rho_{\min} = \frac{2\sqrt{(\theta - \lambda)\,(1 - \lambda\theta)\,\theta}}{1 - \lambda^2 + (\theta - \lambda)^2} > 0 \qquad (22)$$

The proof of the proposition (see Appendix H) exploits the particularly simple form of the predictive space in this special case. A single predictor model, as in (11) and (12), can be parameterised by the triplet $(\lambda, \rho_{uv}, R^2)$. These map to the two ARMA parameters (λ, θ). The inverse mapping describes a parameter space which, for given λ, can be represented by a curve in (R^2, ρ_{uv}) space with a unique stationary point at $|\rho_{uv}| = \rho_{\min}$.

We show later for our empirical application in Section 6 that the lower bound for $|\rho_{uv}|$, $\rho_{\min}$, can be quite close to unity even where the gap between the upper and lower bound for R^2 may be quite wide. In these circumstances, although the predictive regression may offer an improvement relative to the fundamental ARMA, this can only be the case if the true predictor variable has innovations closely resembling those of the predicted variable.[20]

3.5.2 Corollary 3 in the ARMA(1,1) case: $R^2_{\max}$ and the time series properties of the predictions

Corollary 3 noted a key general feature of any predictive system, that the predictions it generates must of necessity be of a lower MA order than the predicted series y_t itself. This provides additional intuition for the the upper bound for R^2 in Proposition 1. For the true state variable to predict y_t well must ultimately require the the predictions it generates to mimic the time series properties of y_t itself. But if the time series properties of y_t and x_t are inherently different, this must imply a limit on how well x_t can predict y_{t+1}.

---

[20]Mitchell, Robertson and Wright (2017) prove a generalisation of this result for $r \geq 1$, for any predicted series with $c_y(1) < 1$.

The UC representation analysed in Lemma 3 and Proposition 2 provides a powerful illustration. As noted there the Beveridge Nelson decomposition of such a y_t has the property that $c_y(1) < 1$. The predictor, however, is an AR(1). Writing (12), as $x_t = C_x(L)v_t$ then $C_x(1) = \frac{1}{1-\lambda} > 1$. For strongly persistent predictors $C_x(1)$ can be well above unity. The Beveridge-Nelson decomposition of the process for x_t (and hence for the predictions $\widehat{y}_t = \beta x_{t-1}$) is thus distinctly different from that of y_t itself. As such, its ability to predict y_t can be severely constrained.

We also noted above, in Section 3.4.2, that only in the unique limiting case where $\lambda = \theta = 0$ do our bounds cease to bind. In this case (and only in this case) the time series properties of x_t and y_t are identical: a white noise can predict another white noise arbitrarily well (or arbitrarily badly). In all other cases the difference in time series properties must imply R^2 bounds.

4   Time-varying parameters

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

Before considering an extension of our analysis to time-varying parameters, it is 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 ABCD and ARMA representations for y_t but with non-Gaussian shocks.[21] Second, even forms of structural

---

[21]Any ARMA model has a state space representation (Hamilton, 1994, chapter 13, pp. 375-6). Per-

instability that cannot be captured in this way should arguably still imply a time-invariant representation in *some* form. Thus, for example, each of the two unobserved components models of inflation analysed in Section 6 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, for the special case of a time-varying ARMA(1,1)[22], which nests commonly used unobserved components models which we exploit in the empirical example in the next section:

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

$$y_t = \beta_t x_{t-1} + u_t \tag{23}$$

$$x_t = \mu_t x_{t-1} + v_t \tag{24}$$

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

$$(1 - \lambda_t L) y_t = (1 - \theta_t L) \varepsilon_t \tag{25}$$

with

$$\lambda_t = \mu_{t-1} \frac{\beta_t}{\beta_{t-1}} \tag{26}$$

(thus if $\beta_t = \beta_{t-1}, \lambda_t = \mu_t$) and ε_t is a serially uncorrelated error orthogonal to y^{t-1}, with $E(\varepsilon_t^2) = \sigma_{\varepsilon,t}^2$. Fundamentalness here requires

$$\lim_{i \to \infty} \prod_{j=0}^{i} \theta_{t-j} = 0, \ \forall t \tag{27}$$

---

manent mean shifts induce a unit root that can be differenced out to derive a stationary ABCD representation.

[22]The methodology could be generalised to higher order ARMA representations.

17

implying that ε_t can be recovered from y^t. Letting

$$R_t^2 = 1 - \sigma_{u,t}^2/\sigma_{y,t}^2 \tag{28}$$

be the time-varying R^2 for the predictive regression that conditions on the true state variable x_t (23), then

$$0 < R_{\min,t}^2 \leq R_t^2 \leq R_{\max,t}^2 < 1 \tag{29}$$

where $R_{\min,t}^2 = 1 - \sigma_{\varepsilon,t}^2/\sigma_{y,t}^2$ is the time-varying R^2 of (25), and $R_{\max,t}^2 = 1 - \sigma_{\eta,t}^2/\sigma_{y,t}^2$ is the time-varying R^2 of the associated unique time-varying nonfundamental representation

$$(1 - \lambda_t)\, y_t = (1 - \gamma_t L)\, \eta_t \tag{30}$$

where γ_t satisfies

$$\lim_{i \to \infty} \prod_{j=0}^{i} \gamma_{t+j}^{-1} = 0 \ \forall t \tag{31}$$

implying that η_t can only be recovered from current and future values of y_t.

Remark: *Corollaries 1 to 4 in the time-invariant case of Proposition 1 also apply in the time-varying case of Proposition 3.*

Time-varying parameters introduce simultaneity into the moment conditions for θ_t and $\sigma_{\varepsilon,t}^2$ (whereas in the time-invariant case these can be solved independently). As far as we are aware the exact derivation of the processes for θ_t and $\sigma_{\varepsilon,t}^2$, and of the associated nonfundamental representation, has not been carried out before.[23] While solution of the moment conditions is as a result distinctly more complicated for the time-varying case, once this problem has been solved the proof of the (time-varying) R^2 bounds follows quite straightforwardly, and analogously to the proof of Proposition 1.[24] All the associated formulae nest the time-invariant results for the ARMA(1,1) model as a special case.

The Stock and Watson (2007) unobserved components stochastic volatility model, discussed in Section 6, is a special case with $\lambda_t = 0$, $\forall t$ (i.e., a time-varying MA(1)). In this case the properties of their structural state space model (see below), constrain θ_t to be strictly positive and less than unity.

In contrast, a striking feature of the more general time-varying ARMA(1,1) case is

---

[23]Stock and Watson (2007), for example, note that their unobserved components stochastic volatility model (as employed in the next section) implies a time-varying MA(1) representation, but the estimates of θ_t that they present are derived using a time-invariant formula.

[24]As such the methodology applied here could in principle be extended to higher order predictive systems.

that θ_t is not bounded above by 1. The fundamentalness condition (27) only requires that the product of the sequence converges to zero, which can be satisfied with some individual values of θ_t greater than unity. Furthermore, even if μ_t, the AR(1) parameter of the predictor, is bounded to lie within $(-1, 1)$, from (26), the same bounds do not apply to λ_t. Indeed we show in Section 6 that estimates of both λ_t and θ_t exceed unity, at some points in time, in our empirical applications.[25]

5 The R^2 bounds and the predictive space when $r > q$

All our analysis has thus far been in terms of population properties. We have assumed that r, the true order of the predictive system for y_t is known, and hence that the true "macroeconomist's ARMA" (5), from which we calculate the R^2 bounds, is an ARMA(r, r). Corollary 1 noted that if this is also a minimal representation (with no cancellation of AR and MA polynomials, or zero MA terms) then the R^2 bounds will lie strictly within $[0, 1]$.

In a finite sample, clearly we cannot know r. To what extent does this limit the value of our population-based analysis?

Even in population, we cannot rule out the logical possibility that the macroeconomist's ARMA may be *non*-minimal. However, suppose that in population y_t admits an ARMA(p, q) representation that *is* minimal. Then it is straightforward to show that the MA order q can only be less than r if the $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ parameters of the structural model satisfy $r - q$ restrictions that ensure exact cancellation of AR and MA polynomials in the univariate reduced form.[26] Lippi and Reichlin (1994) note that there are no obvious theoretical properties of structural macroeconomic models that would imply such restrictions. Hence as a population property we would usually expect $r = q$.[27]

Consider now the finite sample case in which standard model selection criteria point to a particular ARMA(p, q) representation of y_t. To be specific, assume that the data admit a time-invariant ARMA(1,1) representation,[28] and that the estimated parameters

---

[25]Nor is the nonfundamental MA parameter, γ_t, equal to θ_t^{-1}, except in the limiting time-invariant case.

[26]In Lippi and Reichlin's (1994) terms, this would imply that the minimal ARMA(p, q) is the fundamental representation, which provides the lower R^2 bound, while there would exist a nonfundamental "nonbasic" ARMA(r, r) representation, with $r > q$, in which all the θ_i in the macroeconomist's ARMA (5) are replaced with their reciprocals, which provides the true upper bound. But the nonbasic nature of this representation would mean that the true upper bound would be unknowable.

[27]Note that this would also rule out $q = 0$, i.e., a pure AR(p). While such representations are widely used in empirical applications, the derivation from a structural model shows that, absent restrictions on the ABCD parameters, such representations can only be rationalised as approximations for the true ARMA(r, r).

[28]Most of the arguments presented here also apply in the time-varying case, to which we revert below

$\left(\widehat{\lambda},\widehat{\theta}\right)$ and calculated R^2 bounds (using (17) and (18), as in the example analysed above in Section 3) are reasonably well-estimated, conditional upon that order.

Irrespective of the true value of r, it follows immediately that the finite history $\{y_t\}_0^T$ is at least *consistent* with a predictive system with a single predictor with AR parameter $\widehat{\lambda}$, with an R^2 bounded between the estimated values $R^2_{\min}\left(\widehat{\lambda},\widehat{\theta}\right)$ and $R^2_{\max}\left(\widehat{\lambda},\widehat{\theta}\right)$. We have seen that if $\widehat{\lambda}$ and $\widehat{\theta}$ are both close to unity, which is likely to occur in particular if y_t is the first difference of a near-stationary $I(1)$ process, then the gap between these estimated bounds may be quite narrow.

Of course, in a finite sample, it is perfectly possible, in principle, that the true value of r may be greater than 1. The higher order elements of the AR and MA polynomials in the true population ARMA could be sufficiently *close* to cancellation that it may be impossible to infer the true value of r in any finite sample. What if we get r wrong? There are three key implications:

1. The true *lower* bound $R^2_{\min}$ is likely to be quite similar to the estimated value derived from the parameters of a low order ARMA, simply because of the feature of near-cancellation: if the additional parameters are close to cancelling they will barely impact on goodness-of-fit.

2. In contrast, simply by inspection of the general (time-invariant) formula (9) for the upper bound, $R^2_{\max}$ in Proposition 1, it is evident that a higher true value of r, almost certainly means that the true value of $R^2_{\max}$ is likely to be higher than that implied by the estimated ARMA$(1,1)$ representation, and all the more so if the true value of θ_i is close to zero for some i.

3. However, for the true model to *attain* a higher value of R^2 does *not* simply require it to have more than one predictor. The key insight of Section 2.4 is that whatever is the true value of r, the parameters of the true predictive system must still live within the "Predictive Space" consistent with the univariate properties of y_t: in this particular example, that y_t can be *represented* (to some arbitrary degree of precision) by an ARMA(1,1).

While the problem above arises in finite samples, population properties still allow us to gain some insights into the nature of the predictive space in such cases by considering the special case that the minimal population ARMA representation is *exactly* an

---

after discussing our empirical application.

ARMA$(1, 1)$ with the particular form implied by the univariate unobserved components model of Lemma 3, but that the true value of r is greater than 1:

Proposition 4 (Escaping the ARMA$(1, 1)$ bounds). *Let y_t admit a minimal fundamental ARMA(1,1) representation with $0 \leq \lambda < \theta < 1$. Hence, from Lemma 3 y_t has a Beveridge-Nelson decomposition with $c_y(1) = \frac{1-\theta}{1-\lambda} < 1$. Let the true data-generating process for y_t be a structural ABCD model that, from Lemma 1, reduces to a predictive system with $r > 1$ predictors. (The structural model must therefore satisfy $r - 1$ restrictions such that $p = 1, q = 1$.) For any predictive model of this form,*

$$R^2 > R_{\max}^2(\lambda, \theta) \Leftrightarrow c_{\widehat{y}}(1) < \frac{1}{1 - \lambda} \tag{32}$$

where $R_{\max}^2(\lambda, \theta)$ is the calculated upper bound for a single predictor model from Proposition 1, using (18), and the predictions from the true structural model, $\widehat{y}_t = \boldsymbol{\beta}' \mathbf{x}_{t-1}$ have the Beveridge-Nelson decomposition $\widehat{y}_t = c_{\widehat{y}}(L) \boldsymbol{\beta}' \mathbf{v}_t$.

Thus a higher order predictive model may in principle exceed the R^2 bounds calculated from the parameters of the ARMA(1,1) representation, but it can only do this if the predictions it generates have lower persistence than those of an AR(1) predictor consistent with the ARMA(1,1) formulation (which, as discussed in Section 3.5.2, would have $c_x(1) = \frac{1}{1-\lambda}$). So the ARMA(1,1) representation still provides us with important information about the nature of the predictive space that contains the parameters of the true predictive system.

In summary, Proposition 4 illustrates the more general property that whatever the true value of r, univariate properties can still provide us with information on whether a predictive system has the "right kind of predictions" - in time series terms. We explore this issue further, in the context of our empirical example, in Section 6.8.

6   An empirical application: R^2 bounds for inflation

6.1   Key points

We use the framework above to: a) analyse the univariate properties of US inflation, with results for an additional seven OECD countries presented and discussed in (the online) Appendix L; and b) make inference about both the potential predictive performance and nature of the true multivariate models that generated the data. We use two time-varying univariate unobserved components models: Stock and Watson (2007) and Chan, Koop

and Potter (2013). Both models have been used previously to model and forecast inflation, in particular in the US, and have been found to forecast well relative to competitors. We show that both are nested within the time-varying parameter ARMA(1,1) representation of Proposition 3, and hence can be used to derive time-varying R^2 bounds. As expected, the two representations imply similar values, and time paths, of $R^2_{\min,t}$. However at times they imply very different values of $R^2_{\max,t}$, with Chan *et al.*'s model typically implying a much narrower gap between the two bounds. We note that this reflects the distinctly different multivariate frameworks implicit in the two univariate representations.

Since both representations are implicitly single predictor models, we also consider whether a higher order predictive system might in principle have an R^2 outside our estimated bounds. We show, using simulation evidence, that even if the benchmark Smets and Wouters (2007) DSGE model were the true data generating process for inflation, it would fall foul of Proposition 4, because it generates "the wrong kind of predictions".

6.2 Data

We analyse quarterly headline CPI inflation data for the US (seasonally adjusted), downloaded from FRED (the underlying data are from the OECD's MEI database) over the sample 1961Q1 to 2017Q1.

6.3 Unobserved components models through the lens of the time-varying ARMA(1,1)

Consider the following general unobserved components model with stochastic volatility that nests both Stock and Watson (2007) and Chan *et al.* (2013):

$$
\begin{aligned}
Y_t &= \tau_t + c_t \tag{33}\\
\tau_t &= \tau_{t-1} + s_{\tau,t}, \text{ where } s_{\tau,t} = \sigma_{\tau,t}\zeta_{\tau,t}\\
c_t &= \mu_t c_{t-1} + s_{c,t}, \text{ where } s_{c,t} = \sigma_{c,t}\zeta_{c,t}\\
\ln \sigma^2_{\tau,t} &= \ln \sigma^2_{\tau,t-1} + \upsilon_{\tau,t}\\
\ln \sigma^2_{c,t} &= \ln \sigma^2_{c,t-1} + \upsilon_{c,t}
\end{aligned}
$$

where inflation, Y_t, is decomposed as the sum of a a random walk (permanent) component, τ_t, and a (transitory) AR(1) component, c_t. $\zeta_t = (\zeta_{\tau,t}, \zeta_{c,t})$ is assumed $NIID(0, I_2)$, $\upsilon_t = (\upsilon_{\tau,t}, \upsilon_{c,t})$ is $NIID(0, diag(\sigma^2_{\upsilon,\tau}, \sigma^2_{\upsilon,c}))$. Allowing for stochastic volatility has been found to improve both in-sample and out-of-sample fit (e.g., see Stock and Watson (2007)).

The Stock and Watson (2007) model assumes $\mu_t = 0$. We follow Chan's (2017) generalisation of Stock-Watson by estimating $\sigma_{v,\tau}^2$ and $\sigma_{v,c}^2$ rather than setting both equal to 0.2, as in Stock and Watson (2007). This flexibility is helpful, as shown by Chan (2017), certainly in considering applications beyond the US (see appendix L). Henceforth, we denote our first UC model, "SWC".

The SWC model implies that $y_t = \Delta Y_t$ has a time-varying MA(1) representation.[29] In our framework, for $r = q = 1$, the SWC representation implies inflation is generated by an underlying multivariate system with time-varying parameters as in (23) and (24), in which, in the absence of restrictions across the underlying structural model, the single predictor x_t must be IID.[30] The predictor x_t could itself be some aggregate of state variables for inflation, but this would usually require that these are all themselves IID (i.e., may in principle be some aggregate of "news" about a range of series).

The Chan, Koop and Potter (2013) ("CKP") model allows both the volatility of the transitory component, $\sigma_{c,t}$, and its AR parameter, μ_t, to be time-varying; but in contrast to SWC it assumes the innovation to the permanent component has constant volatility ($\sigma_{\tau,t} = \sigma_\tau$, $\forall t$). This implies a time-varying ARMA(1,1) representation, in which the AR parameter, λ_t, is a recursive function of μ_t, as in Proposition 3.[31] If again we assume $r = q = 1$, the underlying predictor for inflation, x_t, consistent with this ARMA representation is itself a time-varying AR(1) process with AR parameter μ_t.[32]

The time invariant versions of both SWC and CKP impose certain restrictions on their ARMA representations (as discussed in Section 3.3 and summarised in Lemma 3): both require $\theta > \lambda \geq 0$. In the time-varying case we are not aware of any results showing that this must necessarily still hold, although in practice it does hold in all the applications we have examined. Note that (as discussed in relation to Proposition 3) the CKP representation does *not* require either θ_t or λ_t to be bounded above by unity.

One feature of both representations is also quite clear-cut, whether in time-varying or time-invariant form: both exclude the possibility that either y_t or Y_t admit a pure AR

---

[29]Note that Stock and Watson use a time-invariant formula to derive an estimate of the implied time-varying MA parameter; however we show below that in this context this generates very similar answers to the exact recursive formula.

[30]In the SWC framework, with no AR component, stochastic volatility in the implicit single predictor can be captured by time variation in β_t.

[31]See Appendix I.1. Note that CKP also utilise restrictions that bound both τ_t and μ_t. We impose bounds on μ_t, as in CKP, but not on τ_t, since this would change the order of the ARMA representation. However, we find that our estimated unobserved components are affected only minimally by whether we impose the bound on τ_t.

[32]Note that, as in the time-invariant case analysed in Section 3.3, c_t can be viewed in filtering terms as an estimate of the true predictor, conditional upon the history of y_t, and the identifying assumption $E(\zeta_{\tau,t}, \zeta_{c,t}) = 0$, $\forall t$.

representation.

Both the SWC and CKP models are estimated, with the same priors and starting values, using Bayesian methods as in Chan (2017) and Chan *et al.* (2013), respectively; we refer the reader to these papers for background specification and estimation details.[33]

6.4 Unobserved Components representations of quarterly CPI inflation

Figure 1 summarises our estimation results and the properties of the derived ARMA representations.

---

[33]We gratefully acknowledge use of Joshua Chan's Matlab code for both the SWC and CKP models, available at http://joshuachan.org/code.html. As detailed in the discussion below, with associated Figures in Appendices L and M, we do investigate the robustness of results to some of these specification choices.
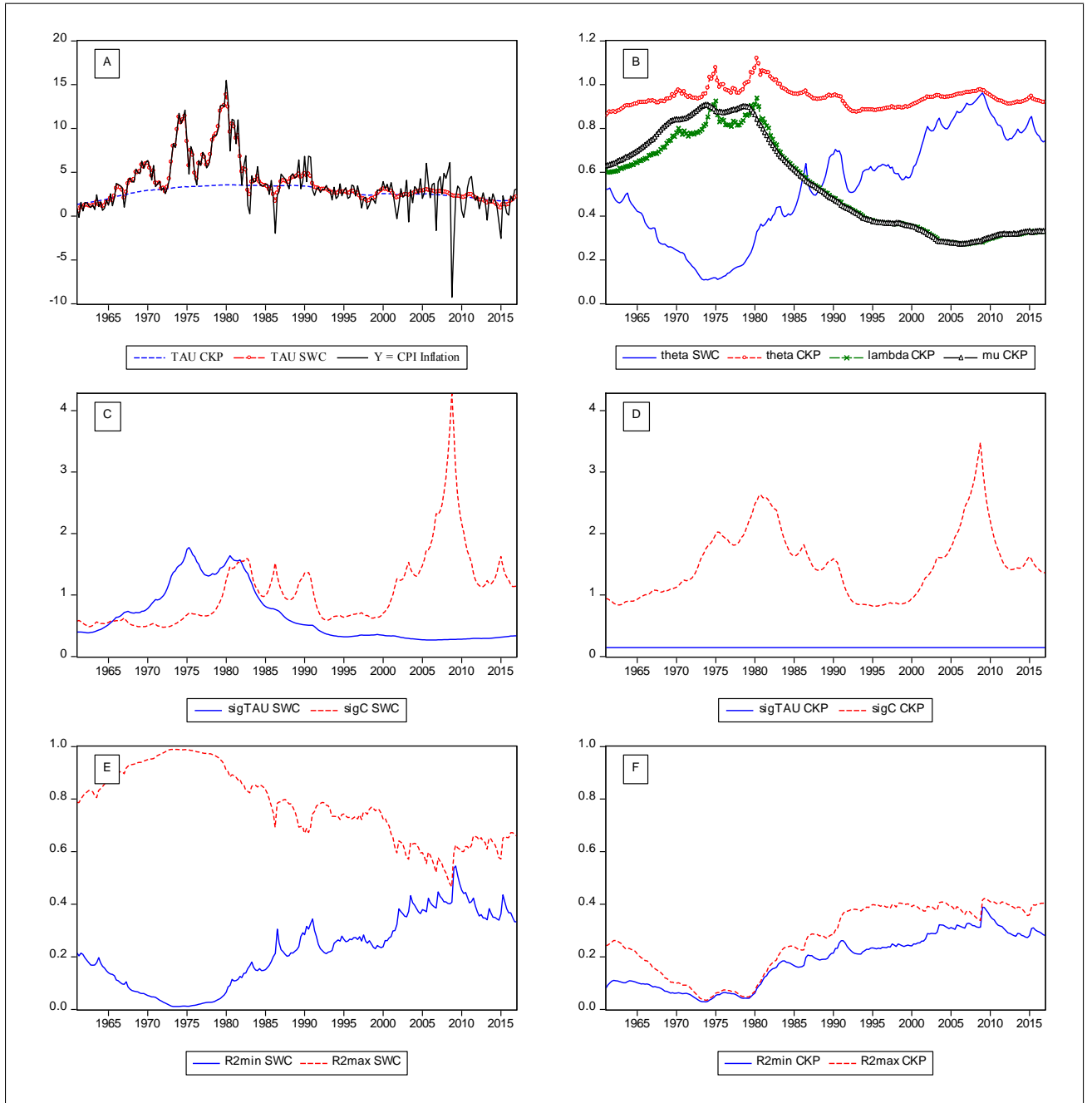
Figure 1: US inflation. Panel A plots posterior median estimates of the permanent component, τ_t, of inflation from the SWC and CKP models alongside CPI inflation. Panel B plots posterior median estimates of θ_t, λ_t and μ_t from the SWC and CKP models (where $\lambda_t = \mu_t = 0$ for SWC). Panels C and D plot posterior median of estimates of $\sigma_{\tau,t}$ and $\sigma_{c,t}$ from the SWC and CKP models. Panels E and F plot posterior median estimates of $R^2_{\min,t}$ and $R^2_{\max,t}$ from the SWC and CKP models as defined in Proposition 3.

Panel A of Figure 1 plots annualised quarterly inflation, Y_t, alongside the estimated permanent components, τ_t, in the SWC and CKP representations.[34] For the first half of the sample especially during the periods of high inflation in the 1970s and early 1980s, Panel A shows that the CKP estimates of τ_t are much smoother than those from SWC. Panels C and D show that during this period shocks to inflation in the US are largely interpreted as permanent in SWC (hence at these times the path for τ_t is very similar to that for inflation itself), but must be allocated to the transitory component in CKP. However, from the late 1980s onwards (i.e., post Great Moderation), the SWC and CKP estimates of τ_t (and hence the implied cycles, c_t) are much more similar, with the SWC estimates of $\sigma^2_{\tau,t}$ falling and then stabilising at similar values to CKP. Both representations therefore imply that transitory shocks have dominated in more recent data.

6.5   The lower bound, $R^2_{\min,t}$ and the ARMA representations

Comparison of Panels E and F, of Figure 1, shows that, as we would expect (see Section 5) both SWC and CKP generate fairly similar estimates of $R^2_{\min,t}$ (for $y_t = \Delta Y_t$).[35] Estimates of $R^2_{\min,t}$ fell to near-zero during the high inflation of the mid-1970s, but have recovered in more recent years.

How can both SWC and CKP show such similar patterns of time variation in univariate predictability, while having such distinctly different patterns of $\sigma^2_{\tau,t}$ and $\sigma^2_{c,\tau}$? The reconciliation comes from an examination of the implied ARMA structure of both representations.

In the SWC representation, for example, Figure 1 shows that the fall in the estimated value of $R^2_{\min,t}$ (Panel E) to near-zero in the mid-1970s was, of necessity, matched by a fall in $\widehat{\theta}_t$, (Panel B).[36] But in the CKP representation, $R^2_{\min,t}$ is driven primarily by the *difference* between $\widehat{\theta}_t$ and $\widehat{\lambda}_t$.[37] Panel B of Figure 1 shows that during this inflationary period $\widehat{\mu}_t$, the estimated AR(1) parameter of CKP's transitory component of US inflation,

---

[34]Panels C and D of Figure M.1 (see online appendix) also show that results are robust to consideration of a more diffuse prior for σ^2_τ in CKP. Such a diffuse prior is in line with the similarly diffuse prior employed in SWC.

[35]Chan *et al.* (2013)'s out-of-sample predictability tests (their Table 5) also show that differences between the CKP and Stock-Watson's UC model are relatively modest, certainly for 1-step ahead forecasts which are our focus in this paper.

[36]The time-invariant formula in (17), for the SWC/MA(1) case is simply $R^2_{\min} = \theta^2/(1+\theta^2)$. In Panels A and B of Figure M.1 we show that applying the time-invariant formulae from Section 3 to the time-varying UC and ARMA estimates usually gives good, or (in the case of the SWC representation very good) approximations to the true, recursive values we derive from our moment conditions. Exceptions to this general rule arise when estimates of θ_t exceed unity.

[37]From (17) the time-invariant formula is $R^2_{\min} = (\lambda - \theta)^2 / (1 - \lambda^2 + (\lambda - \theta)^2))$.

26

rose to a peak of around 0.9. There was a similar, if somewhat more volatile implied pattern in $\widehat{\lambda}_t$ (which is driven not just by $\widehat{\mu}_t$ but also by its rate of change). This narrowing of the gap between $\widehat{\theta}_t$ and $\widehat{\lambda}_t$, despite a few periods when $\widehat{\theta}_t$ exceeded unity,[38] implied falls in estimates of $R^2_{\min,t}$ to close to zero during this period of higher inflation.

While SWC and CKP, despite their different interpretations, both capture the very low degree of univariate predictability, especially during the mid-1970s, there are significant differences in what the two representations imply about multivariate predictability, to which we now turn.

6.6   The upper bound, $R^2_{\max,t}$ and the ARMA representations

Panels E and F of Figure 1 show that while the time paths for the estimates of $R^2_{\min,t}$ are similar for both SWC and CKP, at times their estimates of $R^2_{\max,t}$ differ very markedly, particularly in the period when inflation was high and $R^2_{\min,t}$ was low. Comparison of Panels E and F shows that the estimated paths for $R^2_{\max,t}$ from CKP are typically much lower than those implied by SWC. The gap between $R^2_{\min,t}$ and $R^2_{\max,t}$ from the CPK model has widened in recent years, but remains distinctly narrower than for SWC.[39]

To help understand how these contrasting estimates for $R^2_{\max,t}$ arise from the two ARMA models implied by SWC and CKP, we again exploit the formula, (18), for $R^2_{\max}$ in the time-invariant ARMA(1,1) given in Section 3: $R^2_{\max} = R^2_{\min} + \left(1 - \theta^2\right)\left(1 - R^2_{\min}\right)$.[40] For SWC, a low estimate of $R^2_{\min,t}$ requires $\widehat{\theta}_t$ to be close to zero, which must imply that $\widehat{R}^2_{\max,t}$ is close to unity; in contrast in the CKP model a similarly low $\widehat{R}^2_{\min,t}$ reflects a high value of $\widehat{\lambda}_t$ that is close to an even higher value of $\widehat{\theta}_t$ (which in some periods exceeds unity). This results in the implied $\widehat{R}^2_{\max,t}$ being very close to the lower bound.

---

[38] As noted in the discussion of Proposition 3, fundamentalness does not impose an upper bound of unity in every period. Note also that the proof also shows that the nonfundamental MA parameter γ_t is only equal to θ_t^{-1} on average, so when Figure 1 shows $\widehat{\theta}_t > 1$ this does not imply that $\widehat{\gamma}_t < 1$; indeed it is always higher than $\widehat{\theta}_t$.

[39] In Panels E and F of Figure M.1 we show 16.5%, 50% and 83.5% quantiles of the posterior distribution of $(R^2_{\max,t} - R^2_{\min,t})$ for SWC and CKP. The range of values of the gap between the upper and lower bounds is more revealing of the impact of parameter uncertainty than for either in isolation, since $R^2_{\min,t}$ and $R^2_{\max,t}$ are strongly correlated across replications. The posterior intervals are much narrower for CKP than SWC.

[40] As noted above (see footnote 36), Figure M.1 shows that the time-invariant formulae mostly provide a good approximation to the true values.

6.7 Implications for multivariate models with $r = 1$

The differences between the SWC and CKP estimates of $R^2_{\max,t}$ reflect the very different implicit assumptions about the nature of the underlying multivariate predictive systems for inflation that generated the ARMA reduced form. At this stage we focus on the implications under the maintained assumption that $r = 1$; we consider the impact of relaxing this assumption in the next sub-section.

The SWC representation constrains the AR parameter $\lambda_t = 0$; this implies there must be a single white noise predictor in the underlying multivariate model.[41] But the CKP representation both allows the single predictor to be persistent and produces estimates of $\widehat{\mu}_t$ (and hence $\widehat{\lambda}_t$) that are, at times, quite close to unity. Note that while CKP estimate this parameter as the AR(1) of the transitory component, when viewed through the lens of our analytical framework it is an estimate of the AR(1) parameter of the true (but unobserved) predictor.

The contrast between estimates of $R^2_{\max,t}$ can therefore be seen as arising from different implicit assumptions about the underlying macroeconomic drivers of inflation. The (at times) strongly persistent predictor implied by the CKP representation is what we might expect in a traditional Phillips Curve framework, if some indicator of demand pressure from the real economy was both persistent (for which there is much evidence) and had predictive power for changes in inflation (for which evidence is of course much more mixed; cf. Stock and Watson, 2007). However, the mid-1970s, when inflation was most hard to predict from its own history, was also the period in which the predictor implicit in the CKP representation was estimated to be most strongly persistent. From Corollary 3, and the analysis of Section 3.5.2, the time series properties of such a strongly persistent predictor would have been radically different from the time series properties of inflation, which was at the time near white-noise. Thus the associated estimate of $R^2_{\max,t}$ tells us that no such predictor could have done much better during this period than the ARMA (which itself had minimal predictive power): an AR(1) with high μ_t would have been just too different from a near-white noise to have more than marginal predictive power.

In contrast, the (at times, much) wider gap between the R^2 bounds from the SWC representation (Panel E of Figure 1[42]) leaves open the possibility of nontrivial improvements in predictability of inflation relative to the ARMA, but *only* if the true predictor is white noise. This could, at least in principle, be consistent with a "news" model of informational shocks driving forward-looking price setting. Unlike the case with a persis-

---

[41] Time variation in β_t allows us to make the single predictor pure white noise.

[42] Results for other OECD countries are similar; see Panel E of Figures L2-L8 in the online appendix.

tent predictor, during the period of higher inflation the estimated $R^2_{\max,t}$ tells us that a white noise predictor could in principle have predicted inflation extremely well. Corollary 3 again provides a rationale. Since we know that the change in inflation was near-white noise during this period, the time series properties of the predictor and the predicted variable would have been very close. An upper bound close to unity then simply tells us that there could in principle have been a white noise predictor in period t that could have predicted the outturn for inflation in period $t+1$.

At face value this suggests a resolution of the Predictive Puzzle for inflation: that macroeconomists looking at persistent predictors of inflation have been looking in the wrong place, given the much greater scope for predictability from a white noise predictor. However a distinct note of caution is needed. As discussed in Section 2.4, the R^2 bounds are not the only information provided by the univariate representation. Proposition 2 showed that, at least in the time-invariant ARMA(1,1) case, there is also a lower bound on $|\rho_{uv}|$, the absolute *correlation* between any such predictor and innovations to inflation itself.[43] In the time-invariant MA(1) case this lower bound is given by $\rho_{\min} = \frac{2\theta}{1+\theta^2}$. If we plug the SWC time-varying value $\widehat{\theta}_t$ into this time-invariant formula, the implied estimate $\widehat{\rho}_{\min,t}$ has been close to unity in recent years.[44] Thus, while the SWC representation implies that a white noise predictor could, *in principle*, predict better than the time-varying MA(1), the "predictive space" of parameters of the true predictive system consistent with this MA(1) representation is very tightly defined by this bound on ρ_{uv}.

Thus an alternative, and distinctly more pessimistic, resolution of the Predictive Puzzle in relation to inflation is that in recent data the properties required for even a white noise predictor to out-predict the univariate representation are so tightly defined that there may be little or no scope for such a predictor to exist.[45] This conclusion holds *a fortiori* for a persistent predictor consistent with the CKP representation, given the combination of tight R^2 bounds *and* the lower bound for $|\rho_{uv}|$.

---

[43] For the general case this result relates to the correlation between u_t and v_t, the *innovation* to the predictor, but in the MA(1) case $x_t = v_t$.

[44] We have not found a way to generalise Proposition 2 to the time-varying case; however we would defend the approach used here on the basis both of the (usually) fairly good approximations provided by time-invariant formulae for the R^2 bounds, and the logic of Corollaries 3 and 4, together with the concept of the predictive space, all of which must apply even in a time-varying context.

[45] This conclusion also holds, to varying extents, in most other countries (see online Appendix L). An exception is Italian inflation, where the SWC estimates of $\widehat{\theta}_t$ (see Figure L.6, Panel B) are much lower, implying lower values for $\widehat{\rho}_{\min,t}$ (using the time-invariant formula) than in the US.

6.8 Implications for higher-order predictive models of inflation

The potential resolution of the Predictive Puzzle for inflation offered in the previous section is of course dependent on the crucial assumption that the predictive regression has just a single (possibly composite) predictor. Yet, as discussed above, in Section 5, while the data for CPI inflation may be well described by a low-order ARMA in a finite sample, it is always possible that r, the true number of predictors derived from the underlying ABCD model, is greater than q, the MA order found in the data, because the the true reduced form (the "macroeconomist's ARMA", (5)) has near-cancellation, such that increasing the ARMA order would give no significant improvement in fit.[46]

Since most multivariate predictive models that generate inflation forecasts have larger numbers of state variables this would appear, on the face of it at least, to re-open the Predictive Puzzle, since, as discussed in Section 5, a higher true value of r would almost certainly mean that the true value of $R^2_{\max}$ would be higher than that implied by an estimated low-order ARMA representation. Yet, as documented by Stock and Watson (2007) in practice additional predictors still struggle to predict inflation better than a univariate representation.

We can get some insights into why this might be the case by analysing some of the properties of the predictive system for inflation implied by the benchmark DSGE model of Smets and Wouters (2007).

The Smets-Wouters model has $n = 16$ linearly independent state variables with distinct eigenvalues. Even if we allow for the possibility (discussed in relation to Lemma 1 above) of a near-block-recursive structure, such that r might in principle be very much smaller, this still in principle opens up at least the logical possibility that the true value of $R^2_{\max}$ might be distinctly higher than implied by the R^2 bounds we calculated in the previous section.

Proposition 4 provides one crucial insight into why, in practice, a higher true value of r need not imply that the true predictive system actually *attains* a value of R^2 beyond the upper bounds implied by our estimated representations, which implicitly assume $r = 1$. It showed that, for the time-invariant case at least, this will only be the case if the true structural model, with $r > 1$ generates predictions with lower persistence (a lower value

---

[46]The discussion in online Appendix L shows that in most countries it is quite hard even to distinguish conclusively between time-varying MA(1) and ARMA(1,1) representations of CPI inflation. So it seems highly unlikely that higher order representations could be estimated. In practice we are not aware of readily available estimation routines that would allow us to estimate a higher order ARMA model with time-varying parameters (or equivalently a UC model for the level of inflation, Y_t with multiple stationary components). However, applying standard time-invariant ARMA estimation techniques to CPI inflation suggests minimal gains from increasing ARMA order in any of the 8 countries we examine.

of $c_{\widehat{y}}(1)$) than the single predictor model consistent with the ARMA(1,1) representation.

But in Appendix K we demonstrate, via simulation, that the Smets and Wouters DSGE model as fitted to US macroeconomic data, generates the *wrong* kind of predictions: it consistently delivers implied values of $c_{\widehat{y}}(1)$ well above unity and at least as large as those implied by the CKP time-varying ARMA(1,1) representation in recent years (hence, by implication, very much higher than that implied by the SWC representation, which always implies $c_{\widehat{y}}(1) = 1$).

Nor indeed should we be very surprised by this result. Inflation in the Smets-Wouters model, as in most structural models (at least those with a New Keynesian core) is generated by a hybrid Phillips Curve with both forward- and backward-looking components. Inflation predictions are thus driven by a combination of strongly persistent shocks to marginal costs from the real economy and (less persistent) margin shocks (which to some extent capture some of the univariate characteristics of inflation itself). A structure like this simply cannot generate predictions with time series properties that allow it to escape the R^2 bounds implied by the low-order ARMA representations we analysed in the previous section.

Of course, the Smets-Wouters model is only one model, albeit an important benchmark. We cannot rule out the possibility that the true model that generated the data for inflation - in the US and elsewhere - may escape the R^2 bounds calculated in Section 6.7. But the insights of the "predictive space" still apply. While the theoretical R^2 bounds may widen in higher dimensional systems (possibly even to include values of $R^2_{\max}$ close to unity) the existence of the Predictive Puzzle suggests that the parameter space of any predictive system that could actually get anywhere near this notionally possible higher upper bound is likely to be sparse.[47]

7   Conclusions

We motivated the analysis in this paper with reference to the "Predictive Puzzle" in empirical macroeconomics: that multivariate time series models often struggle to outpredict univariate representations.

---

[47]For example Mitchell, Robertson and Wright (2017) work through the implications of case of a true ARMA(2,2) that generates data for inflation consistent (within the range of sampling variation) with Stock and Watson's MA(1) representation. They show that the bounds from the ARMA(1,1) provide a very good approximation, in the sense that the predictive space" either has very little mass outside these bounds, or only contains predictive systems with properties that we would rule out on a priori grounds (eg, cases in which both predictors have $\lambda_i < 0$, or are perfectly negatively correlated.

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

This paper uses population-based analysis to analyse the Predictive Puzzle. If a multivariate macroeconomic model yields an $ARMA(r, r)$ reduced form for one of its endogenous variables, we show that we can use the ARMA parameters to derive bounds for the R^2 for the predictive regression that conditions upon the true states. If r is low then these bounds can be tight, so that little improvement beyond univariate modelling is possible.

Many contemporary macro models imply that r may be quite high, in which case the R^2 bounds could in principle be (substantially) looser. At the same time, univariate analysis typically finds little role for higher order AR or MA terms (Diebold and Rudebusch, 1989; Cubbada *et al.*, 2009). So either multivariate macro models are being specified with too many state variables, or the structure of these models must generate exact (or near-) cancellation of AR and MA terms.

In the former case, the R^2 bounds found from estimating the low order ARMA may therefore provide a good characterisation of the (often limited) degree of improvement that multivariate prediction may give over univariate analysis.

But even in the latter case, where the true R^2 bounds may *potentially* be much wider than those implied by low order estimation, we show that improvements in predictability can be obtained only if the predictor variables introduced have characteristics that are consistent with properties of the univariate representation. In our application to CPI inflation we argue that improvements in predictability will typically require low (or even zero) persistence in the predictor variables, and that this is not a feature displayed by, for example, the benchmark Smets and Wouters (2007) DSGE model.

Our analysis suggests new avenues that future researchers might pursue. In the case of our empirical example, it suggests that future researchers looking for predictors of inflation should be focussed on low persistence or IID ("news") predictors. More generally researchers should be looking for multivariate predictability consistent with the univariate properties of the series they seek to predict.

References

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

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

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

[4] Carriero, A., T.E. Clark and M. Marcellino (2016), "Common Drifting Volatility in Large Bayesian VARs", *Journal of Business and Economic Statistics*, 34, 375-390.

[5] Chan, J.C.C. (2017), "Specification Tests for Time-Varying Parameter Models with Stochastic Volatility", *Econometric Reviews.* Forthcoming.

[6] Chan, J.C.C., G. Koop and S.M. Potter (2013), "A New Model of Trend Inflation", *Journal of Business and Economic Statistics*, 31, 94-106.

[7] 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.

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

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

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

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

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

[13] Estrella, A. and J.H. Stock (2015), "A State-Dependent Model for Inflation Forecasting", in *Unobserved Components and Time Series Econometrics* (edited by S.J. Koopman and N. Shephard), Oxford University Press, Ch 3, 14-29.

[14] 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.

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

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

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

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

[19] Koop, G. (2013), "Forecasting with Medium and Large Bayesian VARs", *Journal of Applied Econometrics*, 28, 177-203.

[20] Mitchell, J., Robertson, D. and S. Wright (2017), "The Predictive Space: If $\mathbf{x}$ predicts y, what does y tell us about $\mathbf{x}$?", Birkbeck College Working Paper.

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

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

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

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

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

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

[27] 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.

[28] 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.

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

[30] Stock, J.H. and M.W. Watson (2016), "Core Inflation and Trend Inflation", *Review of Economics and Statistics*, 98(4), 770-784.

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

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

A Proof of Lemma 1 (the predictive system for y_t)

We can write the state equation (1) as

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

$$\mathbf{T}\mathbf{z}_t = \mathbf{M}\mathbf{T}\mathbf{z}_{t-1} + \mathbf{T}\mathbf{B}\mathbf{s}_t \tag{A.2}$$

$$\mathbf{x}_t^* = \mathbf{M}\mathbf{x}_{t-1}^* + \mathbf{v}_t^* \tag{A.3}$$

where $\mathbf{x}_t^* = \mathbf{T}\mathbf{z}_t$ is $n \times 1$ and $\mathbf{v}_t^* = \mathbf{T}\mathbf{B}\mathbf{s}_t$. The observables equation (2) is then

$$\mathbf{y}_t = \mathbf{C}\mathbf{T}^{-1}\mathbf{T}\mathbf{z}_{t-1} + \mathbf{D}\mathbf{s}_t \tag{A.4}$$

$$= \mathbf{C}\mathbf{T}^{-1}\mathbf{x}_{t-1}^* + \mathbf{D}\mathbf{s}_t \tag{A.5}$$

Let

$$\boldsymbol{\beta}^{*\prime} = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times(n-1)} \end{array} \right] \mathbf{C}\mathbf{T}^{-1} \tag{A.6}$$

and we can write a predictive equation for the first element of $\mathbf{y}_t$ as

$$y_t = \boldsymbol{\beta}^{*\prime}\mathbf{x}_{t-1}^* + u_t \tag{A.7}$$

where

$$u_t = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times(n-1)} \end{array} \right] \mathbf{D}\mathbf{s}_t \tag{A.8}$$

This representation may in principle have state variables with identical eigenvalues (for example multiple IID states) or state variables with zero $\boldsymbol{\beta}^*$ entries (states that do not directly affect $\mathbf{y}_t$). To derive a minimal representation we first eliminate from $\mathbf{x}_t^*$ those elements with zero $\boldsymbol{\beta}^*$ entries (and rewrite $\boldsymbol{\beta}^*$ appropriately). Then if state variables x_i^* and x_j^* correspond to identical eigenvalues $\mu_i = \mu_j$ (and so have the same autoregressive parameter in the transition equation) we combine these into a new state variable $x_i = x_i^* + \frac{\beta_j^*}{\beta_i^*} x_j^*$ (and note that x_i will also be an AR(1) with parameter μ_i) and we can then rewrite the prediction equation in terms of x_i with parameter $\beta_i = \beta_i^*$. We are then left with an $r \times 1$ vector $\mathbf{x}_t$ obeying

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{A.9}$$

where $\mathbf{v}_t$ contains the appropiate elements of $\mathbf{v}_t^*$ and a prediction equation

$$y_t = \boldsymbol{\beta}^{\prime}\mathbf{x}_{t-1} + u_t \tag{A.10}$$

where $\mathbf{\Lambda}$ then contains the distinct eigenvalues of $\mathbf{M}$ corresponding to the r variables in $\mathbf{x}_t$ (which are either original states in $\mathbf{x}_t^*$ or combinations thereof that are relevant for y_t); $\boldsymbol{\beta}$ contains the matching elements from $\boldsymbol{\beta}^*$, and $\mathbf{v}_t$ those from $\mathbf{v}_t^*$.

Note that

$$r = n - \#\left\{\{\text{states that do not predict } y_t\} \cup \{\text{repeated eigenvalues of } \mathbf{M}\}\right\} \quad \text{(A.11)}$$

which could be substantially less than n.$\blacksquare$

B   Proof of Lemma 2 (The Macroeconomist's ARMA)

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

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

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

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

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

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

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

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

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

$$\prod_{i=1}^{r}\left(1 - \theta_i L\right) = 1 + c_1 L + c_2 L^2 + \ldots + c_r L^r \quad \text{(B.4)}$$

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

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

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

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

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

C   Proof of Proposition 1 (Bounds for the Predictive R^2)

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

Lemma 4 *In the set of all possible nonfundamental $ARMA(r,r)$ representations consistent with (5) in which $\theta_i > 0$, $\forall i$, and θ_i is replaced with θ_i^{-1} for at least some i, the moving average polynomial $\theta^N\left(L\right)$ in (10) in which θ_i is replaced with θ_i^{-1} for all i, has innovations η_t with the minimum variance, with*

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

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

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

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

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

Now define

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

so

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

and

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

Since ε_t is IID we have

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

implying

$$\sum_{j=0}^{\infty} c_j c_{j+k} = 0 \ \forall k > 0 \qquad (C.7)$$

Hence we have

$$\sum_{j \neq k} c_j c_k = 2 \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+k} = 0 \qquad (C.8)$$

thus

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

Thus using (C.9) and (C.3) we have (C.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, θ_i is replaced with θ_i^{-1}, $i = 1, \ldots, s(k)$, for $s \leq r$. For any such representation, with innovations $\eta_{k,t}$, we have

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

This is minimised for $s(k) = r$, which is only the case for the single representation in which θ_i is replaced with θ_i^{-1} for all i, and thus this will give the minimum variance nonfundamental representation. Note that 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

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

and

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

and note that immediately from the above we have

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

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

$$R^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_{\hat{y}}^2 + \sigma_u^2} \tag{C.14}$$

where

$$\sigma_{\hat{y}}^2 = \boldsymbol{\beta}' E \left(\mathbf{x}_t \mathbf{x}_t'\right) \boldsymbol{\beta} \tag{C.15}$$

We now show that we can recast the macroeconomist's ARMA and its minimal variance nonfundamental counterpart as special cases of the predictive system in Lemma 1.

For these two ARMA representations

$$\prod_{i=1}^r (1 - \lambda_i L) \, y_t = \prod_{i=1}^r (1 - \theta_i L) \, \varepsilon_t \tag{C.16}$$

$$\prod_{i=1}^r (1 - \lambda_i L) \, y_t = \prod_{i=1}^r \left(1 - \theta_i^{-1} L\right) \eta_t \tag{C.17}$$

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

$$1 + \sum_{i=1}^r \frac{\beta_{F,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^r (1 - \theta_i L)}{\prod_{i=1}^r (1 - \lambda_i L)} \tag{C.18}$$

$$1 + \sum_{i=1}^r \frac{\beta_{N,i} L}{1 - \lambda_i L} = \frac{\prod_{i=1}^r \left(1 - \theta_i^{-1} L\right)}{\prod_{i=1}^r (1 - \lambda_i L)} \tag{C.19}$$

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

$$\mathbf{x}_t^F = \mathbf{\Lambda}\mathbf{x}_{t-1}^F + \mathbf{1}\varepsilon_t \tag{C.20}$$

$$\mathbf{x}_t^N = \mathbf{\Lambda}\mathbf{x}_{t-1}^N + \mathbf{1}\eta_t \tag{C.21}$$

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

$$y_t = \boldsymbol{\beta}_F' \mathbf{x}_{t-1}^F + \varepsilon_t \tag{C.22}$$

$$y_t = \boldsymbol{\beta}_N' \mathbf{x}_{t-1}^N + \eta_t \tag{C.23}$$

The predictive regressions in (C.22) and (C.23), together with the processes for the two univariate predictor vectors in (C.20) and (C.21), are both special cases of the general predictive system of Lemma 1, but with rank 1 covariance matrices, $\Omega^F = \sigma_\varepsilon^2 \mathbf{11}'$, and $\Omega^N = \sigma_\eta^2 \mathbf{11}'$, thus proving Corollary 4.[48] 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 5 *In the population regression*

$$y_t = \boldsymbol{\nu}_{\mathbf{x}}' \mathbf{x}_{t-1} + \boldsymbol{\nu}_F' \mathbf{x}_{t-1}^F + \xi_t \tag{C.24}$$

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 (C.20), all elements of the coefficient vector $\boldsymbol{\nu}_F$ are zero.

---

[48]Note that we could also write (C.22) 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 $\widehat{x}_t = \mathbf{\Lambda}\widehat{x}_{t-1} + \mathbf{k}\varepsilon_t$, where $\mathbf{k}$ is a vector of steady-state Kalman gain coefficients (using the Kalman gain definition as in Harvey, 1989). The implied reduced form process for y_t must be identical to the fundamental ARMA representation (Hamilton, 1994) hence we have $\beta_{F,i} = \beta_i k_i$.

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

$$y_t = \frac{\prod_{i=1}^r (1 - \theta_i L)}{\prod_{i=1}^r (1 - \lambda_i L)} \varepsilon_t = \frac{\beta_1 v_{1t-1}}{1 - \lambda_1 L} + \frac{\beta_2 v_{2t-1}}{1 - \lambda_2 L} + \ldots + \frac{\beta_r v_{rt-1}}{1 - \lambda_r L} + u_t \qquad (\text{C.25})$$

So we may write, using (C.20),

$$
\begin{aligned}
x_{jt-1}^F &= \frac{\varepsilon_{t-1}}{1 - \lambda_j L} \\
&= \left(\frac{L}{1 - \lambda_j L} \right) \frac{\prod_{i=1}^r (1 - \lambda_i L)}{\prod_{i=1}^r (1 - \theta_i L)} \times \\
&\quad \left(\frac{\beta_1 L v_{1t-1}}{1 - \lambda_1 L} + \frac{\beta_2 L v_{2t-1}}{1 - \lambda_2 L} + \ldots + \frac{\beta_r L v_{rt-1}}{1 - \lambda_r L} + u_t \right)
\end{aligned}
\qquad (\text{C.26})
$$

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 6 *In the population regression*

$$y_t = \phi_{\mathbf{x}}' \mathbf{x}_{t-1} + \phi_N' \mathbf{x}_{t-1}^N + \zeta_t \qquad (\text{C.27})$$

where $\mathbf{x}_t^N$ is the vector of nonfundamental univariate predictors defined in (C.21), all elements of the coefficient vector $\phi_{\mathbf{x}}$ 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 (10) and (3), and substituting from (4), we have

$$y_t = \frac{\prod_{i=1}^r (1 - \theta_i^{-1} L)}{\prod_{i=1}^r (1 - \lambda_i L)} \eta_t = \beta_1 \frac{v_{1t-1}}{1 - \lambda_1 L} + \beta_2 \frac{v_{2t-1}}{1 - \lambda_2 L} + \ldots + \beta_r \frac{v_{rt-1}}{1 - \lambda_r L} + u_t \qquad (\text{C.28})$$

Using

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

42

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

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

(C.29)

Now

$$F^r \frac{\prod_{i=1}^{r}(1 - \lambda_i L)}{\prod_{i=1}^{r}(1 - \theta_i F)} \frac{v_{kt-1}}{(1 - \lambda_k L)} = F^r \left(\frac{\prod_{i \neq k}(1 - \lambda_i L)}{\prod_{i=1}^{r}(1 - \theta_i F)} \right) v_{kt-1}$$

$$= v_{kt} + c_1 v_{kt+1} + c_2 v_{kt+2} + \ldots$$

(C.30)

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

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

(C.31)

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

Now let $R_1^2 = 1 - \sigma_\xi^2/\sigma_y^2$ be the predictive R^2 of the regression (C.24) analysed in Lemma 5. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^F$ in (C.22) are both nested within (C.24) we must have $R_1^2 \geq R^2$ and $R_1^2 \geq R_F^2$. But Lemma 5 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 (C.27) analysed in Lemma 6. Since the predictive regressions in terms of $\mathbf{x}_t$ in (3) and in terms of $\mathbf{x}_t^N$ in (C.23) are both nested in (C.27) we must have $R_2^2 \geq R^2$ and $R_2^2 \geq R_N^2$. But Lemma 6 implies that, given $\boldsymbol{\phi}_{\mathbf{x}} = 0$ we must have $R_2^2 = R_N^2$, hence $R_N^2 \geq R^2$. From above we have 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

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

(C.32)

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 $rank \begin{bmatrix} B \\ D \end{bmatrix} = 1$.

This completes the proof of the Proposition. ∎

D Proof of Corollary 1 (R^2 bounds for a minimal ARMA)

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

$$q = r - \#\{\theta_i = 0\} - \#\{\theta_i = \lambda_i \neq 0\} \tag{D.1}$$

$$p = r - \#\{\lambda_i = 0\} - \#\{\theta_i = \lambda_i \neq 0\} \tag{D.2}$$

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

E Proof of Corollary 2 (R^2 Bounds for observable predictors with efficient filtering)

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

F Proof of Corollary 3 (Time series properties of the predictions)

Using (B.1), restated here

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

implies

$$\det\left(\mathbf{I}-\mathbf{\Lambda}L\right)\widehat{y}_t = \boldsymbol{\beta}'\mathrm{adj}\left(\mathbf{I}-\mathbf{\Lambda}L\right)\mathbf{v}_{t-1} \tag{F.2}$$

where the right-hand side of (F.2) is an MA$(r-1)$, since each element of adj$(\mathbf{I}-\mathbf{\Lambda}L)$ is a polynomial of order $\leq r-1$. Hence $\widehat{y}_t$ is an ARMA$(r, r-1)$.∎

G Proof of Lemma 3 (Beveridge-Nelson decomposition)

The UC model of equation (16) is, setting the deterministic component $g = 0$ as this does not affect this proof

$$\begin{align}
Y_t &= c_t + \tau_t \tag{G.1}\\
c_t &= \mu c_{t-1} + s_{c,t} \tag{G.2}\\
\tau_t &= \tau_{t-1} + s_{\tau,t} \tag{G.3}
\end{align}$$

Assume $s_{\tau,t} \sim (0, \sigma_\tau^2)$, $s_{c,t} \sim (0, \sigma_c^2)$ and assume $\sigma_{c\tau} = Cov\left(s_{c,t}, s_{\tau,t}\right) = 0$, i.e., the innovations to the random walk and to the cyclical components are orthogonal.

We have

$$\begin{align}
y_t = \triangle Y_t &= \triangle c_t + \triangle \tau_t \tag{G.4}\\
&= \left(\mu - 1\right)c_{t-1} + s_{c,t} + s_{\tau,t} \tag{G.5}
\end{align}$$

Now we can write $c_{t-1} = \left(1 - \mu L\right)^{-1} s_{c,t-1}$

$$\begin{align}
y_t &= \left(\mu - 1\right)\left(1 - \mu L\right)^{-1} s_{c,t-1} + s_{c,t} + s_{\tau,t} \tag{G.6}\\
y_t &= \mu y_{t-1} + \left(\mu - 1\right)s_{c,t-1} + s_{c,t} - \mu s_{c,t-1} + s_{\tau,t} - \mu s_{\tau,t-1}\\
y_t &= \mu y_{t-1} + s_{c,t} - s_{c,t-1} + s_{\tau,t} - \mu s_{\tau,t-1}
\end{align}$$

or, since $\mu = \lambda$,

$$y_t = \lambda y_{t-1} + s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1} \tag{G.7}$$

which is an ARMA(1,1), as the second order autocorrelation of $s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}$ is zero.

The first order autocorrelation of $\varepsilon_t - \theta\varepsilon_{t-1}$, cf. (14), is $-\frac{\theta}{1+\theta^2}$ so this has to match the first order autocorrelation of $s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}$. This implies

$$-\frac{\theta}{1+\theta^2} = \frac{Cov\left(s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}, s_{c,t-1} - s_{c,t-2} + s_{\tau,t-1} - \lambda s_{\tau,t-2}\right)}{Var\left(s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}\right)} \tag{G.8}$$

$$= \frac{-\sigma_c^2 - \lambda\sigma_\tau^2}{2\sigma_c^2 + \left(1+\lambda^2\right)\sigma_\tau^2} \tag{G.9}$$

So

$$-\frac{\theta}{1+\theta^2} = \frac{-\sigma_c^2 - \lambda\sigma_\tau^2}{2\sigma_c^2 + \left(1+\lambda^2\right)\sigma_\tau^2} \tag{G.10}$$

$$= \frac{-\sigma_c^2 + \lambda\sigma_c^2 - \lambda\left(\sigma_\tau^2 + \sigma_c^2\right)}{2\sigma_c^2 - \left(1+\lambda^2\right)\sigma_c^2 + \left(1+\lambda^2\right)\left(\sigma_\tau^2 + \sigma_c^2\right)} \tag{G.11}$$

and

$$-\frac{\theta}{1+\theta^2} = \frac{-\lambda - (1-\lambda)q}{1+\lambda^2 + \left(1-\lambda^2\right)q} \tag{G.12}$$

where $q = \sigma_c^2 / \left(\sigma_\tau^2 + \sigma_c^2\right)$.

Thus

$$\frac{\theta}{1+\theta^2} = \frac{\lambda + (1-\lambda)q}{1+\lambda^2 + \left(1-\lambda^2\right)q}. \tag{G.13}$$

Now consider the curves $G\left(\theta\right) = \frac{\theta}{1+\theta^2}$ and $F\left(\lambda\right) = \frac{\lambda+(1-\lambda)q}{1+\lambda^2+\left(1-\lambda^2\right)q}$ for $-1 \le \theta, \lambda \le 1$. Note that G is monotonic with $G\left(-1\right) = -\frac{1}{2}$ and $G\left(1\right) = \frac{1}{2}$. We show that $F\left(\lambda\right)$ lies everywhere above $G\left(\lambda\right)$

$$F\left(\lambda\right) - G\left(\lambda\right) = \frac{\lambda + (1-\lambda)q}{1+\lambda^2 + \left(1-\lambda^2\right)q} - \frac{\lambda}{1+\lambda^2} \tag{G.14}$$

$$= \frac{\left(1+\lambda^2\right)\left(\lambda + (1-\lambda)q\right) - \lambda\left(1+\lambda^2 + \left(1-\lambda^2\right)q\right)}{\left(1+\lambda^2 + \left(1-\lambda^2\right)q\right)\left(1+\lambda^2\right)} \tag{G.15}$$

Now the denominator is positive so we need only consider the numerator

$$\left(1+\lambda^2\right)\left(\lambda+(1-\lambda)q\right)-\lambda\left(1+\lambda^2+\left(1-\lambda^2\right)q\right)=\lambda+(1-\lambda)q \tag{G.16}$$
$$+\lambda^3+\lambda^2\left(1-\lambda\right)q$$
$$-\lambda-\lambda^3-\lambda\left(1-\lambda^2\right)q$$
$$=\left((1-\lambda)+\lambda^2\left(1-\lambda\right)-\lambda\left(1-\lambda^2\right)\right)q$$
$$=\left(1-2\lambda+\lambda^2\right)q$$
$$=(1-\lambda)^2\,q>0$$

So the curve F lies above the curve G and hence for any λ the solution to

$$G\left(\theta\right)=F\left(\lambda\right) \tag{G.17}$$

will have $\theta>\lambda$ (see Figure G.1).∎
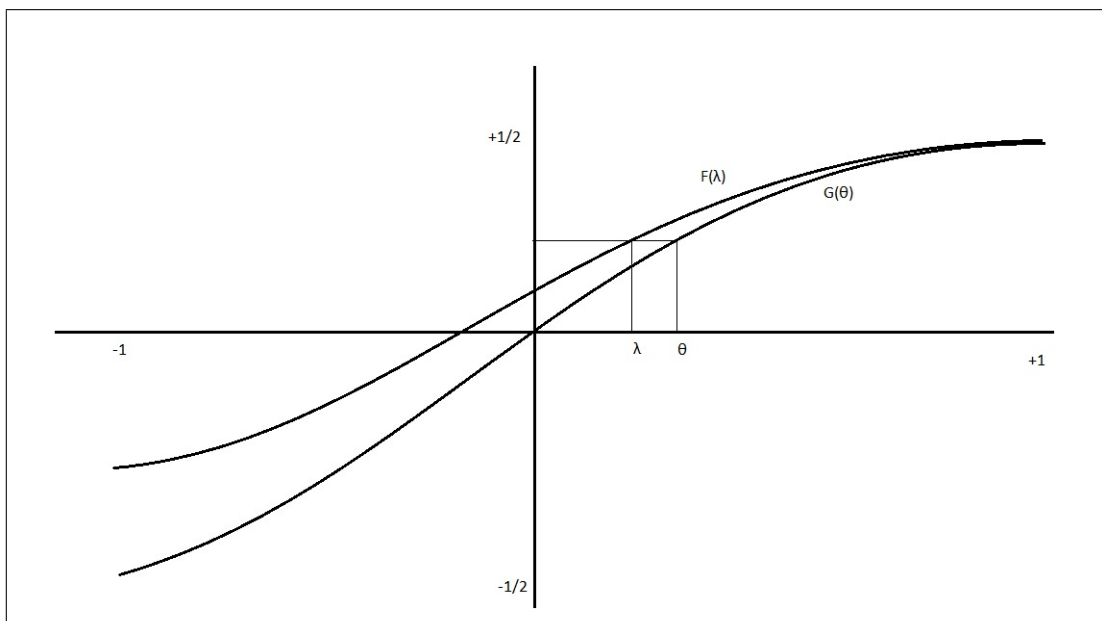
Figure G.1: Proof of Lemma 3 (Beveridge-Nelson decomposition)

H   Proof of Proposition 2 (Bounds for ρ_{uv})

We can re-write (15), the moment condition for the ARMA(1,1), as

$$\frac{-\theta}{1+\theta} = \frac{-\lambda + \beta\rho_{uv}s}{\left(1 - \lambda^2\right) + \beta^2 s^2 - 2\lambda\beta\rho_{uv}s} \tag{H.1}$$

where $\rho_{uv} = corr\left((u_t, v_t)\right)$ and $s = \frac{\sigma_u}{\sigma_v}$, and we note that the predictive equation here has

$$R^2 = \frac{Var\left(\beta x_{t-1}\right)}{Var\left(y_t\right)} = \frac{\frac{\beta^2\sigma_v^2}{1-\lambda^2}}{\frac{\beta^2\sigma_v^2}{1-\lambda^2} + \sigma_u^2} = \frac{\beta^2 s^2}{\left(1 - \lambda^2\right) + \beta^2 s^2}$$

Without loss of generality, assume $\beta > 0$, implying

$$\beta s = \sqrt{\frac{\left(1 - \lambda^2\right) R^2}{1 - R^2}}. \tag{H.2}$$

Subsituting into (H.1) we can (with some tedious but straightforward manipulations) invert to obtain an expression for ρ_{uv} in terms of λ, θ and R^2, giving

$$\rho_{uv}(\theta, \lambda, R^2) = -\left(\frac{(\theta - \lambda)(1 - \theta\lambda) + \frac{(1-\lambda^2)R^2}{1-R^2}\theta}{\left(1 - \lambda^2 + (\theta - \lambda)^2\right)\sqrt{\frac{(1-\lambda^2)R^2}{1-R^2}}}\right) \tag{H.3}$$

This equation describes the predictive space $\mathbb{P}_{\lambda,\theta}$: a necessary relation between parameters that describe the predictive system that generates the ARMA(1,1), and has powerful consequences. For example if for a given triplet (θ, λ, R^2) the solved value for ρ_{uv} lies outside the unit interval then there can be no possible predictive model described by that particular combination of (θ, λ, R^2).

We have already seen in Corollary 4 that the maximum and minimum values of R^2 correspond to $|\rho_{uv}| = 1$. Values of R^2 between these limits will correspond to different values of ρ_{uv}. If the limits are both attained at $\rho_{uv} = +1$ (or at $\rho_{uv} = -1$) then there must be a turning point in the function $\rho_{uv}(\theta, \lambda, R^2)$ as R^2 covers that range.

The first order condition yields a possible stationary point where:

$$\frac{\partial\rho_{uv}(\theta, \lambda, R^2)}{\partial R^2} = 0 \Rightarrow R^2 = \frac{(\theta - \lambda)(1 - \theta\lambda)}{\theta - \lambda + \theta(1 - \theta\lambda)}$$

which after substituting into (H.3) yields a solution as long as

$$(\theta - \lambda)\theta > 0$$

which is satisfied for $\theta > \lambda$, given $\lambda > 0$. Given the definition above the second-order condition confirms a maximum for ρ_{uv}, hence a minimum for $|\rho_{uv}|$ at the value

$$\rho_{\min} = \frac{2\sqrt{(\theta - \lambda)(1 - \theta\lambda)\theta}}{\left(1 - \lambda^2 + (\theta - \lambda)^2\right)} > 0. \blacksquare$$

I   Proof of Proposition 3 (the time-varying ARMA(1,1))

Restating the predictive model (23) and (24) from the proposition,

$$
\begin{align}
y_t &= \beta_t x_{t-1} + u_t \tag{I.1}\\
x_t &= \mu_t x_{t-1} + v_t \tag{I.2}
\end{align}
$$

it can be characterised by the sequence $\left\{\beta_t, \mu_t, \sigma_{v,t}^2, \sigma_{u,t}^2, \sigma_{uv,t}\right\}$.

Assumption: $\beta_t \neq 0$

This assumption is the time-varying equivalent of that in the time-invariant case in Lemma 1. (In this context we are simply ruling out a measure zero case in any model that generates β_t as a random sequence from a continuous error distribution.)

This then implies a time varying ARMA(1,1) since

$$y_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}y_{t-1} = \beta_t x_{t-1} - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}\beta_{t-1}x_{t-2} + u_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}u_{t-1}$$

$$= \beta_t\left(x_{t-1} - \mu_{t-1}x_{t-2}\right) + u_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}u_{t-1} \tag{I.3}$$

thus

$$y_t - \lambda_t y_{t-1} = \beta_t v_{t-1} + u_t - \lambda_t u_{t-1} \tag{I.4}$$

wherein

$$\lambda_t = \mu_{t-1}\frac{\beta_t}{\beta_{t-1}} \tag{I.5}$$

and the right-hand side is a time-varying MA(1) process (its 2nd order autocorrelation is zero).

As in the time-invariant case we define two time-varying ARMA(1,1) representations

$$y_t - \lambda_t y_{t-1} \;=\; \varepsilon_t - \theta_t \varepsilon_{t-1} \qquad (I.6)$$

$$y_t - \lambda_t y_{t-1} \;=\; \eta_t - \gamma_t \eta_{t-1} \qquad (I.7)$$

Note that the equality of the AR parameter of the predictor to the AR parameter in the ARMA representation that occurs in the time-invariant case no longer holds; but there is still a direct recursive mapping in terms of μ_t, β_t and β_{t-1} (with equality as a special case if the β_t are constant).

The representation (I.6) is fundamental if we can derive ε_t as a convergent sum of current and lagged values of y_t:

$$\varepsilon_t = \widetilde{y}_t + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} \theta_{t-j} \right) \widetilde{y}_{t-i} \qquad (I.8)$$

where $\widetilde{y}_t = y_t - \lambda_t y_{t-1}$, thus for fundamentalness we require

$$\lim_{i \to \infty} \prod_{j=0}^{i} \theta_{t-j} = 0 \;\; \forall t \qquad (I.9)$$

In the time-invariant case, with $\theta_t = \theta \;\forall t$, a necessary and sufficient condition is $|\theta| < 1$. For the time-varying case a sufficient condition is $|\theta_t| < 1$ for all t, however this is not a necessary condition (indeed we find in our application that the fundamental representation can have $|\theta_t| > 1$ for some t).

As in the time-invariant case, for the nonfundamental representation (I.7) we have

$$\eta_t = - \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \gamma_{t+j}^{-1} \right) \widetilde{y}_{t+i} \qquad (I.10)$$

which gives a convergent sum in terms of current and future values of y_t if

$$\lim_{i \to \infty} \prod_{j=1}^{i} \gamma_{t+j}^{-1} = 0 \;\; \forall t \qquad (I.11)$$

Note also that we now no longer have $\gamma_t = \theta_t^{-1}$, except in the time-invariant case. Conditional upon the sequences $\left\{ \beta_t, \mu_t, \sigma_{v,t}^2, \sigma_{u,t}^2, \sigma_{uv,t} \right\}_{t=0}^{T}$ the predictive model implies

50

the sequences

$$W_{0t} \quad : \quad = var\left(\widetilde{y}_t\middle| \{\beta_t, \mu_t, \sigma^2_{v,t}, \sigma^2_{u,t}, \sigma_{uv,t}\}^T_{t=0}\right) \tag{I.12}$$

$$W_{1t} \quad : \quad = cov\left(\widetilde{y}_t, \widetilde{y}_{t-1}\middle| \{\beta_t, \mu_t, \sigma^2_{v,t}, \sigma^2_{u,t}, \sigma_{uv,t}\}^T_{t=0}\right) \tag{I.13}$$

where as before $\widetilde{y}_t = y_t - \lambda_t y_{t-1}$. These autocovariances, conditional upon the parameter sequence, are given by

$$W_{0t} \quad = \quad \beta^2_t \sigma^2_{v,t-1} + \sigma^2_{u,t} + \left(\mu_{t-1}\frac{\beta_t}{\beta_{t-1}}\right)^2 \sigma^2_{u,t-1} - 2\mu_{t-1}\frac{\beta^2_t}{\beta_{t-1}}\sigma_{uv,t-1} \tag{I.14}$$

$$W_{1t} \quad = \quad \beta_t\sigma_{uv,t-1} - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}\sigma^2_{u,t-1} \tag{I.15}$$

We now have a recursive moment matching problem: for a given sequence $\{\lambda_t\}$ (from (I.5)) we require sequences $\{\theta_t, \sigma^2_{\varepsilon,t}\}$ such that the time-varying moments implied by the fundamental ARMA representation, conditional upon $\{\theta_t, \sigma^2_{\varepsilon,t}, \lambda_t\}$, match those of the structural model given in (I.14) and (I.15), i.e.

$$cov\left(\widetilde{y}_t, \widetilde{y}_{t-1}\middle| \{\theta_t, \sigma^2_{\varepsilon,t}, \lambda_t\}^T_{t=0}\right) \quad = \quad -\theta_t\sigma^2_{\varepsilon,t-1} = W_{1t} \tag{I.16}$$

$$var\left(\widetilde{y}_t\middle| \{\theta_t, \sigma^2_{\varepsilon,t}, \lambda_t\}^T_{t=0}\right) \quad = \quad \sigma^2_{\varepsilon,t} + \theta^2_t\sigma^2_{\varepsilon,t-1} = W_{0t} \tag{I.17}$$

and analogously for the sequences $\{\gamma_t, \sigma^2_{\eta,t}\}$ from the nonfundamental representation:

$$cov\left(\widetilde{y}_t, \widetilde{y}_{t-1}\middle| \{\gamma_t, \sigma^2_{\eta,t}, \lambda_t\}^T_{t=0}\right) \quad = \quad -\gamma_t\sigma^2_{\eta,t-1} = W_{1t} \tag{I.18}$$

$$var\left(\widetilde{y}_t\middle| \{\gamma_t, \sigma^2_{\eta,t}, \lambda_t\}^T_{t=0}\right) \quad = \quad \sigma^2_{\eta,t} + \gamma^2_t\sigma^2_{\eta,t-1} = W_{0t}. \tag{I.19}$$

Re-writing (I.17) and (I.19) as

$$\sigma^2_{\varepsilon,t} \quad = \quad W_{0t} - \theta^2_t\sigma^2_{\varepsilon,t-1} \tag{I.20}$$

$$\sigma^2_{\eta,t} \quad = \quad W_{0t} - \gamma^2_t\sigma^2_{\eta,t-1} \tag{I.21}$$

then by recursive substitution, for given θ_t, the solution for $\sigma^2_{\varepsilon,t}$ becomes invariant to starting values as $t \to \infty$ if

$$\lim_{t\to\infty}\prod_{j=0}^{t}\theta^2_{t-j} = 0 \tag{I.22}$$

which is clearly satisfied by (I.9), the property of fundamentalness.

To solve, substituting from (I.16) into (I.17) we have

$$\sigma_{\varepsilon,t}^2 = W_{0t} - \frac{W_{1t}^2}{\sigma_{\varepsilon,t-1}^2} \tag{I.23}$$

which we can solve recursively forward, and then use (I.16) to find θ_t. By inspection in (I.20) the impact of the initial value $\sigma_{\varepsilon,0}^2$ tends to zero as $t \to +\infty$, thus we have a unique fundamental representation in population.

In the time-invariant case, once we know θ we know $\gamma = \theta^{-1}$, but here it is not so simple. Substituting for γ_t using (I.18) the equivalent recursion for the nonfundamental representation is

$$\sigma_{\eta,t}^2 = W_{0t} - \frac{W_{1t}^2}{\sigma_{\eta,t-1}^2} \tag{I.24}$$

However if we solve forward, by inspection of (I.21), the impact of the initial value diverges. But, if we rewrite as the backward recursion

$$\sigma_{\eta,t-1}^2 = \frac{W_{1t}^2}{\left(W_{0t} - \sigma_{\eta,t}^2\right)} \tag{I.25}$$

we can then solve for γ_t using (I.18). As $t \to -\infty$, the impact of starting values tends to zero, thus the representation is again unique in population.

The proof of the inequality then follows analogously to the proof of Proposition 1, since this only requires serial independence, it does not require that w_t is drawn from a time-invariant distribution. To see this, from (I.8) ε_t is a combination of current and lagged $\widetilde{y}_t$, whereas from (I.10) η_t is a combination of strictly future values of $\widetilde{y}_t$. Thus η_t must have predictive power for all possible predictors (except itself), but not vice versa. ∎

I.1 Application of Proposition 3 to the unobserved components model

It is straightforward to show that the unobserved components model of Section 6.3 can also be put into the form of the predictive model in the proposition.

Restating (33), the model for inflation Y_t,

$$Y_t \;=\; \tau_t + c_t \tag{I.1}$$

$$\tau_t \;=\; \tau_{t-1} + s_{\tau,t} \tag{I.2}$$

$$c_t \;=\; \mu_t c_{t-1} + s_{c,t} \tag{I.3}$$

Then we can restate as the predictive model in (I.1) and (I.2), by defining

$$y_t \;=\; \Delta Y_t \tag{I.4}$$

$$x_t \;=\; c_t \tag{I.5}$$

$$\beta_t \;=\; \mu_t - 1 \tag{I.6}$$

$$u_t \;=\; s_{c,t} + s_{\tau,t} \tag{I.7}$$

$$v_t \;=\; s_{c,t} \tag{I.8}$$

where our assumption in the proof above that $\beta_t \neq 0$ clearly translates to the assumption $\mu_t \neq 1$. We can then apply the formulae in the proof of the proposition.

J   Proof of Proposition 4 (Escaping the ARMA(1,1) bounds)

To prove the proposition, first define the limiting variance ratio (Cochrane, 1988) of the predicted series, y_t, as $V_y = \sigma_P^2/\sigma_y^2$ where $\sigma_P^2 = c_y(1)^2 \sigma_\varepsilon^2$ is the variance of the Beveridge-Nelson (1981) permanent component (see Lemma 3). It is straightforward to show (see Robertson and Wright, 2009, Appendix C1) that in the case of an ARMA(1,1)

$$V_y < 1 \iff \theta > \lambda > 0 \iff c(1) < 1 \tag{J.1}$$

We now exploit a necessary linkage between V_y and three summary features of any multivariate system, proved in Mitchell, Robertson and Wright (2017), Proposition 2, reproduced below as Proposition 5 for convenience:

Proposition 5 *Let V_y be the limit of the variance ratio (Cochrane,1988) of the predicted process y_t, defined by*

$$V_y = \frac{\sigma_P^2}{\sigma_Y^2} = 1 + 2\sum_{i}^{\infty} corr\left(y_t, y_{t-i}\right) \tag{J.2}$$

The parameters $\mathbf{\Psi} = (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ of the predictive system must satisfy

$$g(\mathbf{\Psi}) = V_y \tag{J.3}$$

$$\text{where } G\left(R^2, V_{\widehat{y}}, \rho_{BN}\right) = 1 + R^2\left(V_{\widehat{y}} - 1\right) + 2\rho_{BN}\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)}$$

where $R^2(\mathbf{\Psi})$ is the predictive R^2 from (3); $\rho_{BN}(\mathbf{\Psi}) = corr(u_t, \boldsymbol{\delta}'\mathbf{v}_t)$, with $\boldsymbol{\delta}' = \boldsymbol{\beta}'\left[I - \Lambda\right]^{-1}$, is the correlation between innovations to 1-step ahead and long-run (Beveridge-Nelson) forecasts; and $V_{\widehat{y}}(\mathbf{\Psi})$ is the variance ratio of the predicted value $\widehat{y}_t \equiv \boldsymbol{\beta}'\mathbf{x}_{t-1}$, calculated by replacing y_t with $\widehat{y}_t$ in (J.2).

Proof. See Mitchell, Robertson and Wright (2017). ∎

To show that Proposition 5 leads directly to Proposition 4, if we totally differentiate (J.3)

$$0 = G_1 d\rho_{BN} + G_2 dR^2 + G_3 dV_{\widehat{y}} \tag{J.4}$$

this gives

$$\frac{dV_{\widehat{y}}}{dR^2} = -\frac{G_2}{G_3} - \frac{G_1}{G_3}\frac{d\rho_{BN}}{dR^2} \tag{J.5}$$

We evaluate this expression at the calculated upper bound for an ARMA(1,1), where $R^2_{\max}(\lambda, \theta) = \frac{(1 - \lambda\theta)^2}{1 - \lambda^2 + (\theta - \lambda)^2}$ (using (18)); $V_{\widehat{y}}(1,1) = \frac{1+\lambda}{1-\lambda}$ and $\rho_{BN} = -1$ since, exploiting (20) the reparameterisation of the nonfundamental ARMA(1,1) in Section 3.4, at the upper bound $\rho_{BN} = corr\left(\varepsilon_t, \left(\frac{\lambda - \theta^{-1}}{1 - \lambda}\right)\varepsilon_t\right) = -1$, given $0 \leq \lambda < \theta \leq 1$ as assumed in the proposition.

We now establish that at this point $G_1 > 0, G_2 > 0$ and $G_3 > 0$, implying $\frac{dV_{\widehat{y}}}{dR^2} < 0$ as stated in the proposition.

Since

$$G_1 = 2\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)} > 0 \tag{J.6}$$

for all possible values of $V_{\widehat{y}}$ and R^2, we thus need to establish the signs of G_2 and G_3 at this point, using

$$G_2 = V_{\widehat{y}} - 1 + \frac{\rho_{BN}V_{\widehat{y}}\left(1 - 2R^2\right)}{\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)}} \tag{J.7}$$

$$G_3 = R^2 + \frac{\rho_{BN}R^2\left(1 - R^2\right)}{\sqrt{V_{\widehat{y}}R^2\left(1 - R^2\right)}} \tag{J.8}$$

If we first evaluate (J.8) at $\rho = -1$ then

$$G_3\left(-1, R^2, V_{\widehat{y}}\right) > 0 \iff \sqrt{V_{\widehat{y}}} > \sqrt{\frac{1 - R^2}{R^2}} \iff V_{\widehat{y}} + 1 > \frac{1}{R^2} \tag{J.9}$$

54

Now $V_{\hat{y}} + 1 = \frac{1+\lambda}{1-\lambda} + 1 = \frac{2}{1-\lambda}$ hence

$$G_3\left(-1, R^2, V_{\hat{y}}\left(1,1\right)\right) > 0 \iff R^2 > \frac{1-\lambda}{2} \tag{J.10}$$

But given the assumptions in the proposition we have

$$R^2_{\max}\left(\lambda, \theta\right) \geq \frac{1-\lambda}{2} \tag{J.11}$$

hence

$$G_3\left(-1, R^2_{\max}\left(\lambda, \theta\right), V_{\hat{y}}\left(1,1\right)\right) > 0 \tag{J.12}$$

as required.

Now evaluate G_2 at $\rho_{BN} = -1$

$$
\begin{aligned}
G_2\left(-1, R^2, V_{\hat{y}}\right) &= \left(V_{\hat{y}} - 1\right) - \left(V_{\hat{y}} R^2 \left(1 - R^2\right)\right)^{-1/2} V_{\hat{y}}\left(1 - 2R^2\right)\\
&= \frac{\left(V_{\hat{y}} - 1\right)\sqrt{R^2\left(1 - R^2\right)} - \sqrt{V_{\hat{y}}}\left(1 - 2R^2\right)}{\sqrt{R^2\left(1 - R^2\right)}} = \frac{H\left(V_{\hat{y}}, R^2\right)}{\sqrt{R^2\left(1 - R^2\right)}} \quad \text{(J.13)}
\end{aligned}
$$

Now given $V_{\hat{y}}\left(1,1\right) > 1$ the numerator $H\left(V_{\hat{y}}, R^2\right)$ is certainly positive if $\left(1 - 2R^2_{\max}\left(\lambda, \theta\right)\right) < 0$ i.e. if $R^2_{\max}\left(\lambda, \theta\right) > \frac{1}{2}$.

Thus we only need to show that $H\left(V_{\hat{y}}, R^2\right)$ is positive for $R^2_{\max}\left(\lambda, \theta\right) < \frac{1}{2}$. Given that $R^2_{\max}\left(\lambda, \theta\right)$ always satisfies the inequality (J.11), if we evaluate $H\left(V_{\hat{y}}, R^2\right)$ at $R^2 = \frac{1-\lambda}{2}$ and $V_{\hat{y}} = \frac{1+\lambda}{1-\lambda}$ we have

$$
\begin{aligned}
H\left(V_{\hat{y}}\left(1,1\right), \frac{1-\lambda}{2}\right) &= \left(\frac{1+\lambda}{1-\lambda} - 1\right)\sqrt{\frac{1-\lambda}{2}\left(1 - \frac{1-\lambda}{2}\right)} - \sqrt{\frac{1+\lambda}{1-\lambda}}\left(1 - 2\frac{1-\lambda}{2}\right)\\
&= \frac{2\lambda}{1-\lambda}\sqrt{\left(\frac{1-\lambda}{2}\right)\left(\frac{1+\lambda}{2}\right)} - \lambda\sqrt{\frac{1+\lambda}{1-\lambda}} = 0 \quad \text{(J.14)}
\end{aligned}
$$

and since

$$H_1 = \frac{1}{2}\frac{\left(V_{\hat{y}} - 1\right)\left(1 - 2R^2\right)}{\sqrt{R^2\left(1 - R^2\right)}} + 2\sqrt{V_{\hat{y}}} > 0 \tag{J.15}$$

we must have

$$R^2_{\max}\left(\lambda, \theta\right) < \frac{1}{2} \Rightarrow H\left(V_{\hat{y}}\left(1,1\right), R^2_{\max}\left(\lambda, \theta\right)\right) > 0 \tag{J.16}$$

Hence

$$G_2\left(-1, R^2_{\max}\left(\lambda, \theta\right), V_{\hat{y}}\left(1,1\right)\right) > 0 \tag{J.17}$$

as required.

Hence at $R^2 = R^2_{\max}(\lambda, \theta)$, $\frac{dV_{\hat{y}}}{dR^2} < 0$ so higher values of R^2 require lower values of $V_{\hat{y}}$.

∎

K   Time series properties of the inflation predictions from the Smets and Wouters (2007) DSGE model

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

Table K1 summarises the results.

Table K1: Time Series Properties of Simulated Inflation Predictions, $\widehat{y}_t \equiv \Delta\widehat{\pi}_t$,
in the Smets-Wouters (2007) model at various forecast horizons

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

The first column of Table K1 shows the first-order autocorrelation coefficient of the simulated predictions; the remaining columns show estimates of $V_{\hat{y}}$ using sample variance ratios (using the small sample correction proposed by Cochrane, 1988) at a range of finite horizons.[49] Table K1 makes it clear that the Smets-Wouters model generates predictions with strong positive persistence - as would be expected given that predicted changes in inflation in the model are driven by strongly persistent processes in the real economy.

---

[49]Note that for the the general case, for y_t $V_y = (1 - R^2_{\min}) c(1)^2$, hence $c(1) < 1$ implies $V_y < 1$, and analogously for $V_{\hat{y}}$. For the ARMA(1,1) case the reverse also applies.

As a benchmark for comparison, in the ARMA(1,1) case $V_{\hat{y}} = \frac{1+\lambda}{1-\lambda} \Rightarrow \lambda = \frac{V_{\hat{y}}-1}{V_{\hat{y}}+1}$, thus a median value of $V_{\hat{y}} \approx 4$ would arise from an AR(1) predictor with $\lambda = 0.6$, thus a value of $V_{\hat{y}}$ well *above* the value implied by the CKP representation in recent decades, and shown in Panel B of Figure L.1 for US CPI inflation and in Figure M.11 for US GDP deflator inflation (strictly speaking the relevant comparator for the Smets-Wouters model). As such the Smets-Wouters model is even further from generating IID predictions, consistent with the SWC representation, since this would imply $V_{\hat{y}} = 1$.

Thus, using Proposition 4, if the Smets-Wouters model were the true DGP it would generate the "wrong kind of predictions" to have an R^2 exceeding the the calculated upper bound derived for recent year from a single predictor model.

L   CPI inflation in 8 OECD countries

This appendix complements the results for the US, in Section 6 of the main paper, by both analysing the univariate properties of inflation in a further seven OECD countries (Canada, France, Germany, Greece, Italy, Japan and the UK) and by making inference about both the potential predictive performance and nature of the true multivariate models that generated the data.

The quarterly headline CPI inflation data are downloaded from FRED (the underlying data are from the OECD's MEI database) over the sample 1961Q1 to 2017Q1. With the exception of the US the published CPI series are not seasonally adjusted; but in most countries there is significant evidence of quarterly seasonality. For all countries except the US we therefore seasonally adjust the annualised quarterly inflation series, defined as $Y_t = 400 \log \left(CPI_t / CPI_{t-1} \right)$, using X12.[50]

To ensure this online appendix and its discussion of the eight OECD countries is self-contained, and to facilitate cross-country comparisons, we include the US results, also discussed in Section 6. Thus, Figure L.1 reproduces Figure 1 in the main paper.

Figures L.1 to L.8 summarise our estimation results and the properties of the derived ARMA representations.

---

[50]As implemented in EViews 9.5. In Appendix M (Figures M.9-M.10) we show that when we apply X12 to the unadjusted CPI inflation series for the US (which publishes both adjusted and unadjusted series) and compare the results with the adjusted series they are extremely similar. We also report results (Figures M.11-M.12) for the US using GDP deflator inflation (as analysed by Stock and Watson (2007)), and show that the results are qualitatively similar. However, while the R^2 bounds (in Panel E of Figure M.11) still narrow in recent data they do not do so to the same extent as for CPI inflation (Panel E of Figure L.1), implying that there is more scope to forecast changes in GDP deflator inflation with a multivariate model than CPI inflation.
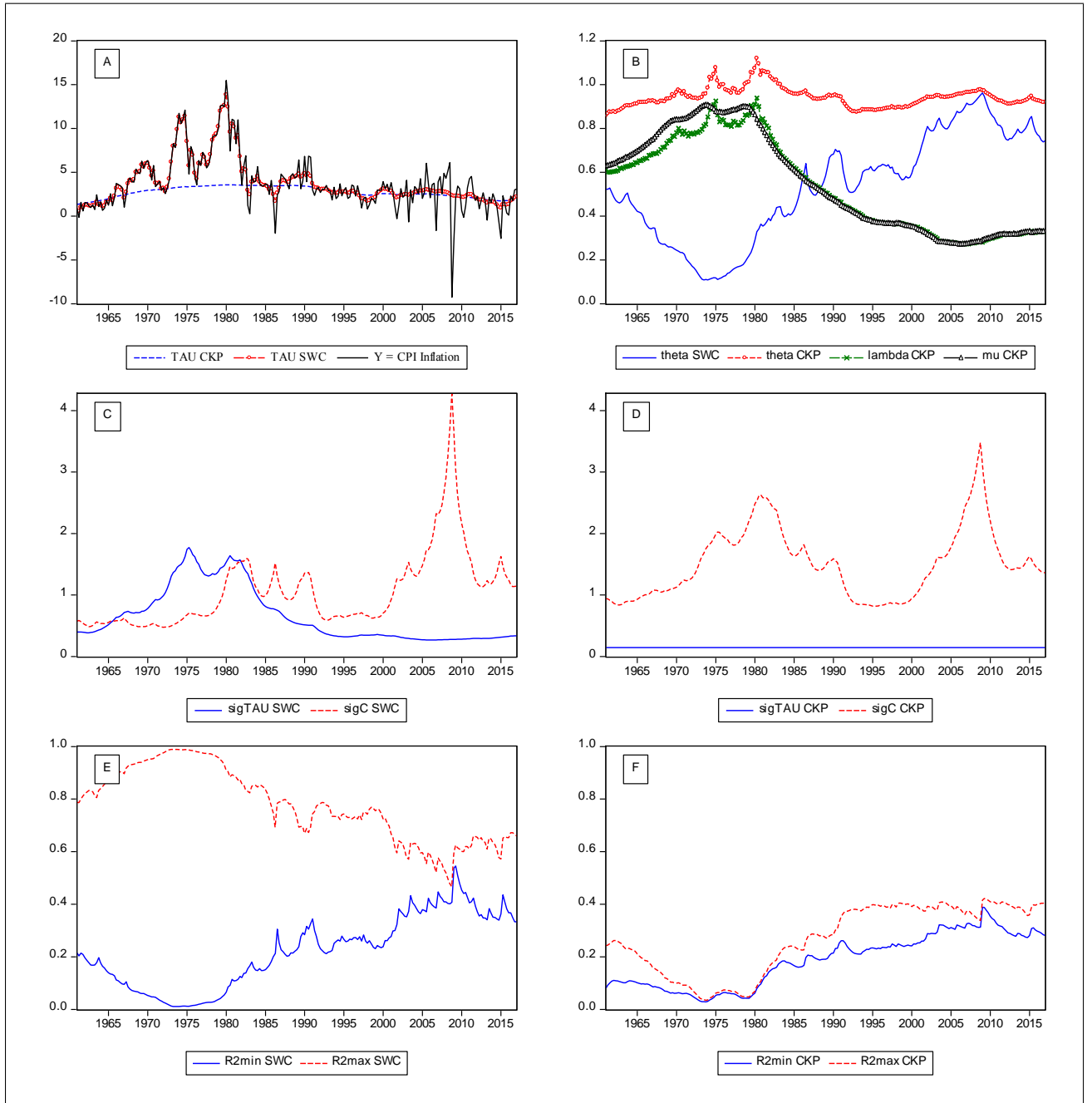
Figure L.1: US. Panel A plots posterior median estimates of the permanent component, τ_t, of inflation from the SWC and CKP models alongside CPI inflation. Panel B plots posterior median estimates of θ_t, λ_t and μ_t from the SWC and CKP models (where $\lambda_t = \mu_t = 0$ for SWC). Panels C and D plot posterior median of estimates of $\sigma_{\tau,t}$ and $\sigma_{c,t}$ from the SWC and CKP models. Panels E and F plot posterior median estimates of $R^2_{\min,t}$ and $R^2_{\max,t}$ from the SWC and CKP models as defined in Proposition 3.
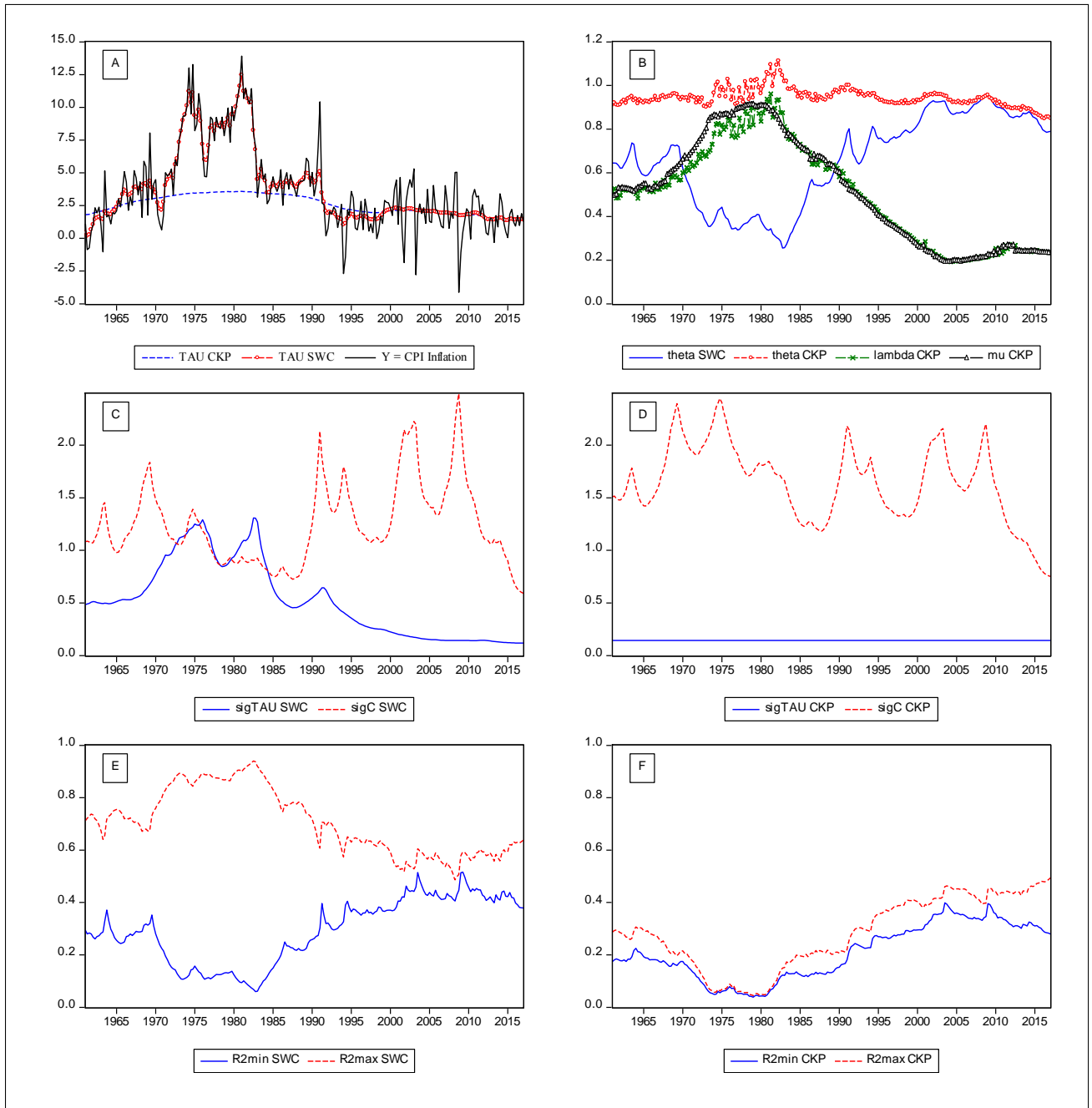
Figure L.2: Canada. See note to Figure L.1
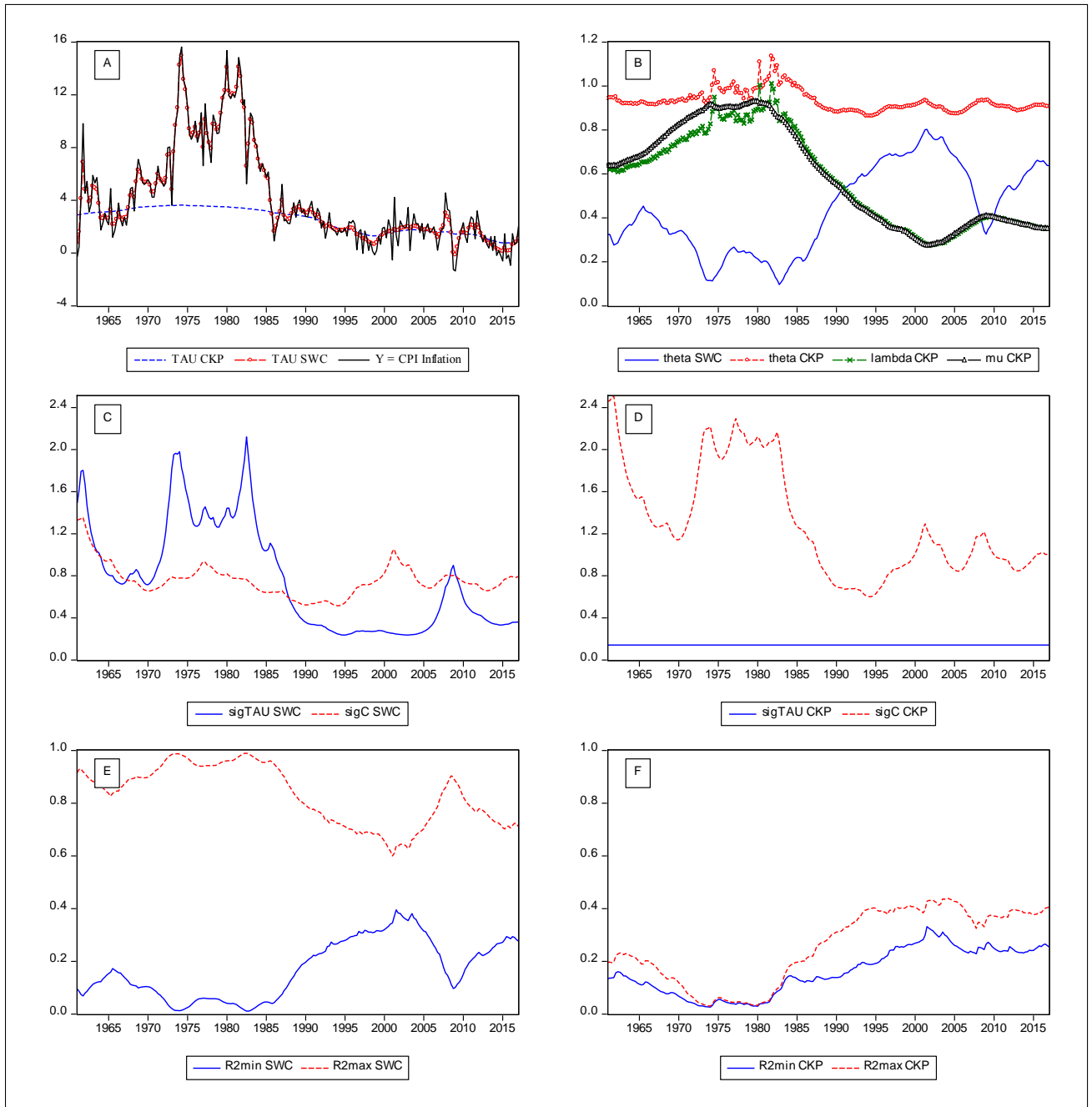
Figure L.3: France. See note to Figure L.1
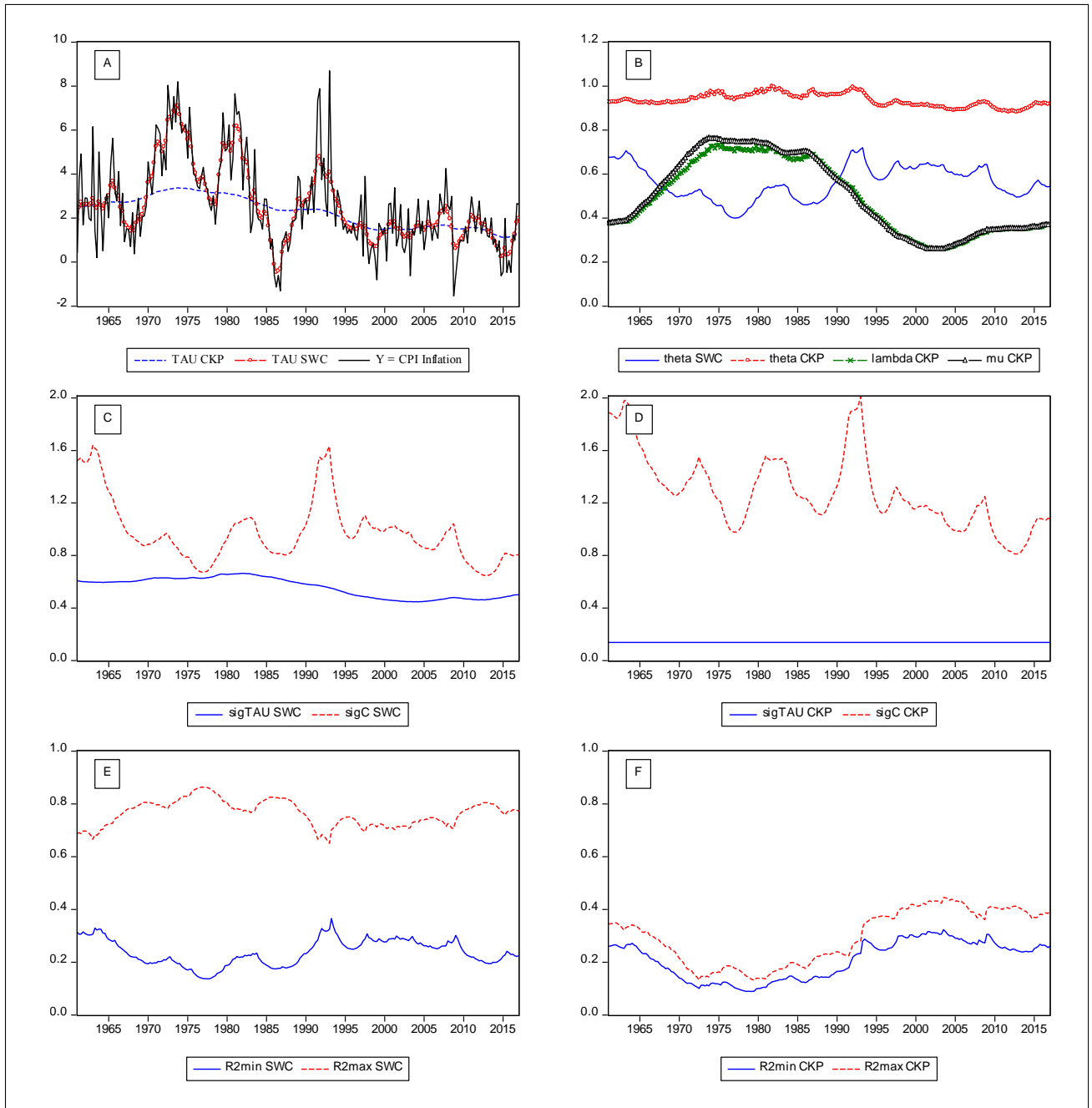
Figure L.4: Germany. See note to Figure L.1
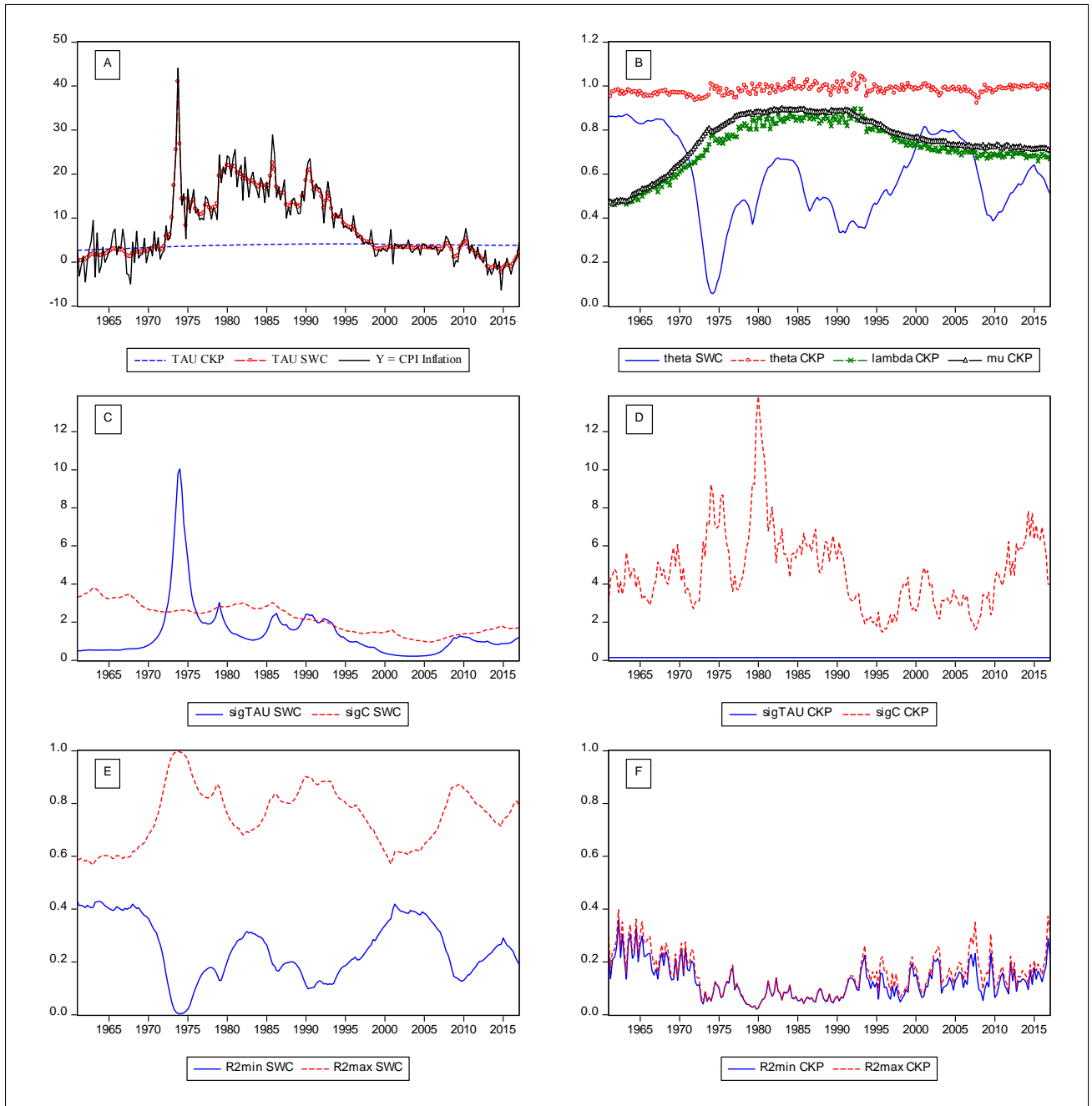
Figure L.5: Greece. See note to Figure L.1
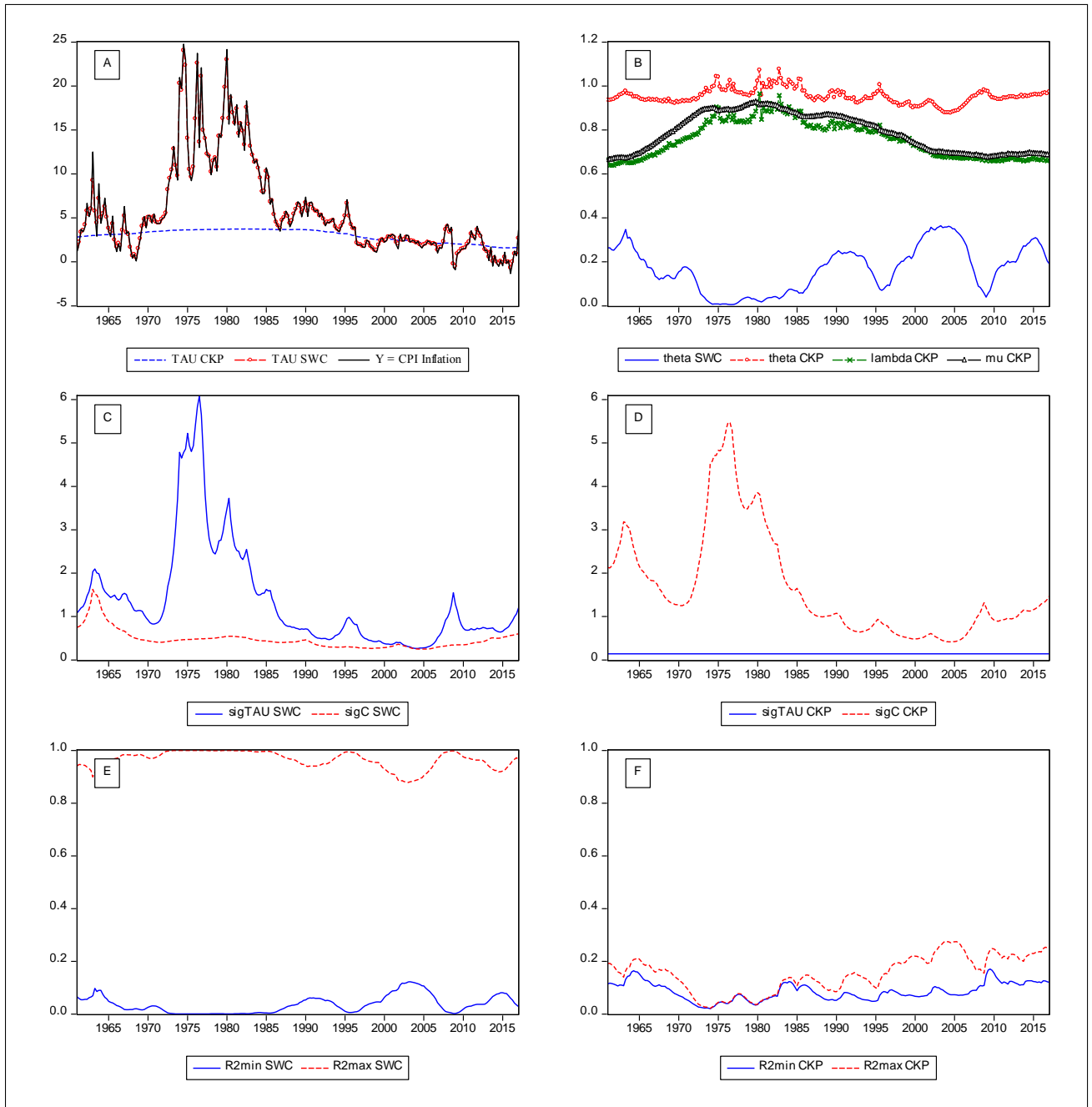
Figure L.6: Italy. See note to Figure L.1
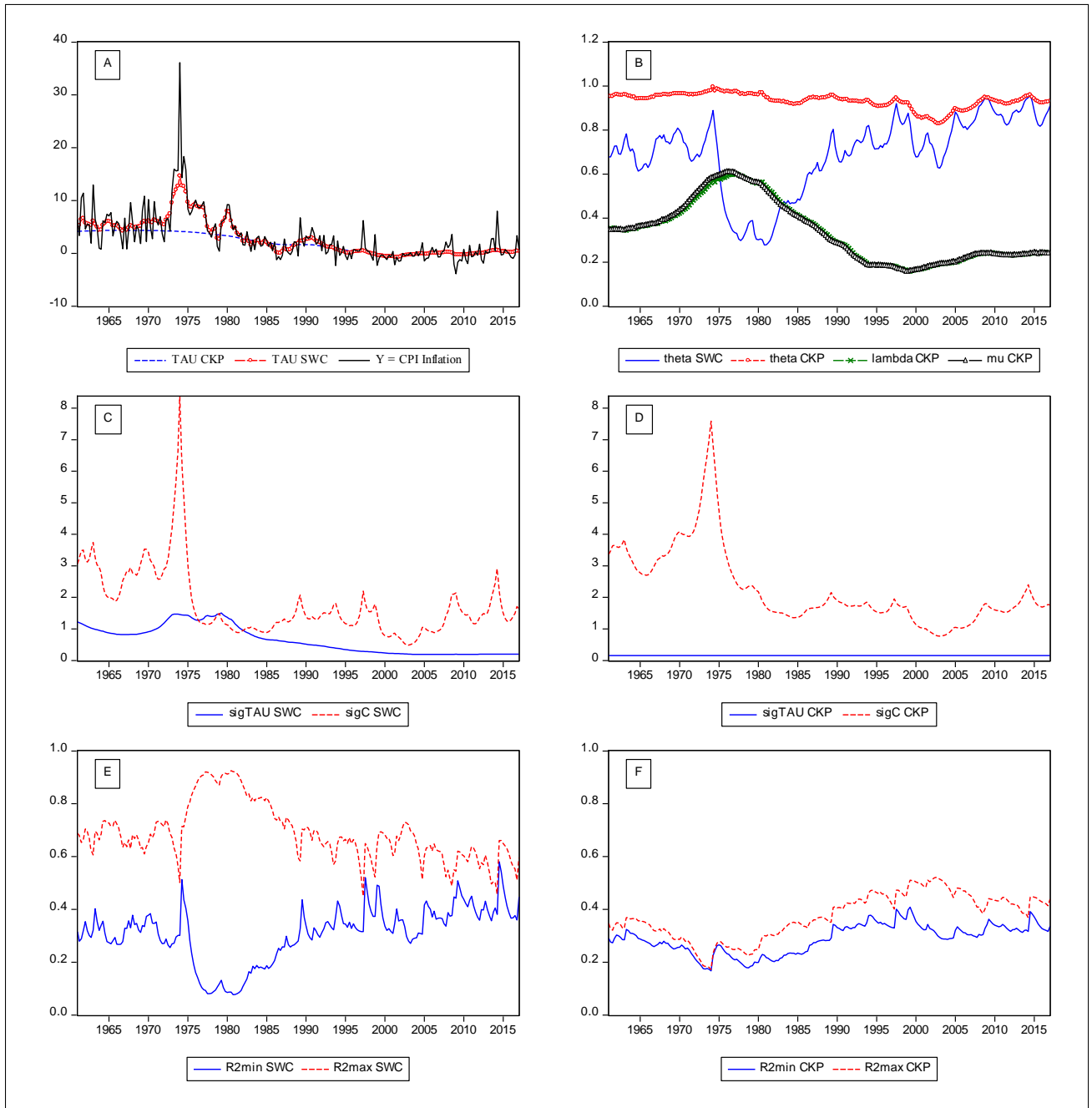
Figure L.7: Japan. See note to Figure L.1
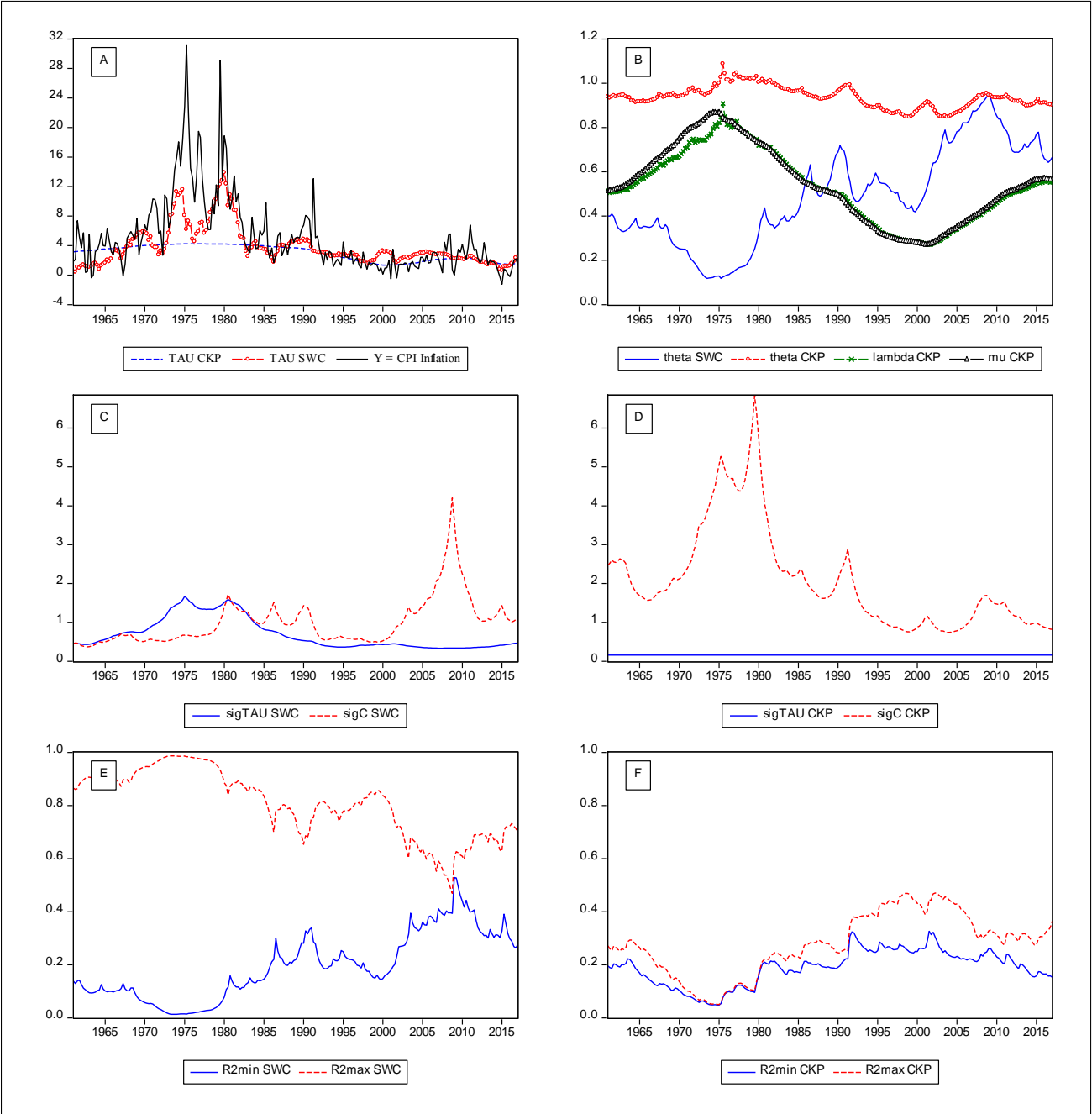
Figure L.8: UK. See note to Figure L.1

Panels A of Figures L.1 to L.8 plot, for each country, annualised quarterly inflation, Y_t, alongside the estimated permanent components, τ_t, in the SWC and CKP representations.[51] The CKP estimates of τ_t are seen, from Panel A, to be much smoother than those from SWC, including during the periods of higher inflation through the 1970s and early 1980s. This is explained by Panels C and D; these Panels reveal that during this period shocks to inflation in the US - and Canada, France, Greece, Italy and to a lesser degree the UK too - are largely interpreted as permanent in SWC (hence at these times the path for τ_t is very similar to that for inflation itself), but allocated to the transitory component in CKP. However, in more recent (post 1990s) data, the SWC and CKP estimates of τ_t (and hence the implied cycles, c_t) are more similar, with the SWC estimates of the variance of the permanent component falling and then stabilising at similar values to CKP. Transitory shocks have tended to dominate in more recent data. A striking contrast is found in Germany (Figure L.4) and Japan (Figure L.7) where transitory shocks play a greater role in both the SWC and CKP representations. While both SWC and CKP estimates of trend inflation in Germany stay within a very narrow range (as might be expected, given the putative stabilising role of the Bundesbank for most of the sample), the two estimates also do *not* converge in later data, with the SWC trend still affected quite strongly by current inflation.[52]

Comparison of Panels E and F, of Figures L.1 to L.8, shows that, as we would expect (see Section 5) both SWC and CKP generate similar estimates of $R^2_{\min,t}$ (for $y_t = \Delta Y_t$).[53] In the US, Canada, France and the UK (and to a lesser degree in Japan) estimates of $R^2_{\min,t}$ fell to near-zero during the high inflation of the mid-1970s but then rose thereafter. In Germany there is a less pronounced dip in estimates of $R^2_{\min,t}$; but then inflation did not, unlike in the other countries, rise to double-digit levels in the 1970s. In Italy estimates of $R^2_{\min,t}$ have remained consistently low throughout the sample period; while in Greece

---

[51]Panels C and D of Figures M.1-M.8 also show that results are robust, in all countries except Greece and Italy, to consideration of a more diffuse prior for σ_τ in CKP. Such a diffuse prior is in line with the similarly diffuse prior employed in SWC. In Greece and Italy the relatively tight priors used by CKP imply time-varying ARMA parameters that make us sceptical of the results. In both countries, the implied paths for θ_t (in Panel B of Figures L.5 and L.6) are very close to unity for most of the sample. This appears to suggest over-differencing, reflecting very low (time-invariant) estimates of σ_τ. However, we think it unlikely that inflation in these countries was so close to being stationary. Furthermore, in the case of Greece, a more diffuse prior results in more plausible time paths for $\widehat{\tau}_t$ (see Panel C of Figure M.5).

[52]Note that this feature also differs strikingly from that in Chan (2017) who finds that the transitory component dominates German inflation. However the difference here appears to reflect his use of unadjusted CPI data: the seasonal component derived from X12 is very volatile.

[53]Chan *et al.* (2013)'s out-of-sample predictability tests (their Table 5) also show that differences between the CKP and Stock-Watson's UC model are relatively modest, certainly for 1-step ahead forecasts which are our focus in this paper.

they have bounced around more, but were also low in the inflationary 1970s.

These movements in $R^2_{\min,t}$, as discussed in the main paper, can be understood and decomposed by inspecting the estimates for $\widehat{\theta}_t$ and $\widehat{\lambda}_t$. For the SWC representations these falls in the estimated value of $R^2_{\min,t}$ (Panel E) to near-zero in the mid-1970s are, of necessity, matched by a fall in $\widehat{\theta}_t$, (Panel B). For the CKP representations these falls in $R^2_{\min,t}$ during this inflationary period are driven by both $\widehat{\mu}_t$, the estimated AR(1) parameter of the transitory component of inflation, and $\widehat{\lambda}_t$ rising to peaks (Panel B). These peaks are around 0.8 to 0.9 in the US, Canada, France, Greece and the UK. In Germany, and in particular in Japan, these peaks are lower; while in Italy, the average values of $\widehat{\mu}_t$ and $\widehat{\lambda}_t$ are higher although these estimates still peak at around 0.9 in the late 1970s (see Figure L.6, Panel B).

Panels E and F of Figures L.1 to L.8 also show that while the time paths of estimates of $R^2_{\min,t}$ are similar for both SWC and CKP, their estimates of $R^2_{\max,t}$ can differ very markedly, particularly in the period when inflation was high and $R^2_{\min,t}$ was low.[54] For all eight countries we observe only a small gap between $R^2_{\min,t}$ and $R^2_{\max,t}$ from the CPK model, especially during the inflationary 1970s; in contrast the SWC model suggests much larger gaps, even during the 1970s. In the SWC model estimates of $R^2_{\max,t}$ are highest and close to unity during the 1970s in the US, Canada, France, Greece and the UK; in Japan there is a lower peak at around 0.9. These estimates of $R^2_{\max,t}$ then declined as inflation fell from the 1980s onwards. However, for Germany (and Italy) the estimates of $R^2_{\max,t}$ from SWC exhibit less variation: estimates are consistently higher, averaging around 0.7 (and 0.95), across the 1961Q1 to 2017Q1 sample. Comparison of Panels E and F, across all eight countries, shows that the estimated paths for $R^2_{\max,t}$ from CKP are much lower than those implied by SWC. Again, as discussed in the main paper, we can understand and decompose these movements in $R^2_{\max,t}$ by relating them to the observed movements of $\widehat{\theta}_t$ and $\widehat{\lambda}_t$.

M   Supplementary empirical results: CPI inflation in 8 OECD countries

Here we present additional Figures referred to both in the main body of the paper and in Appendix L to provide background information on the estimation results.

---

[54]In Panels E and F of Figures M.1-M.8 we show, by country, 16.5%, 50% and 83.5% quantiles of the posterior distribution of $(R^2_{\max,t} - R^2_{\min,t})$ for SWC and CKP. For all countries, the posterior intervals are much narrower for CKP than SWC.
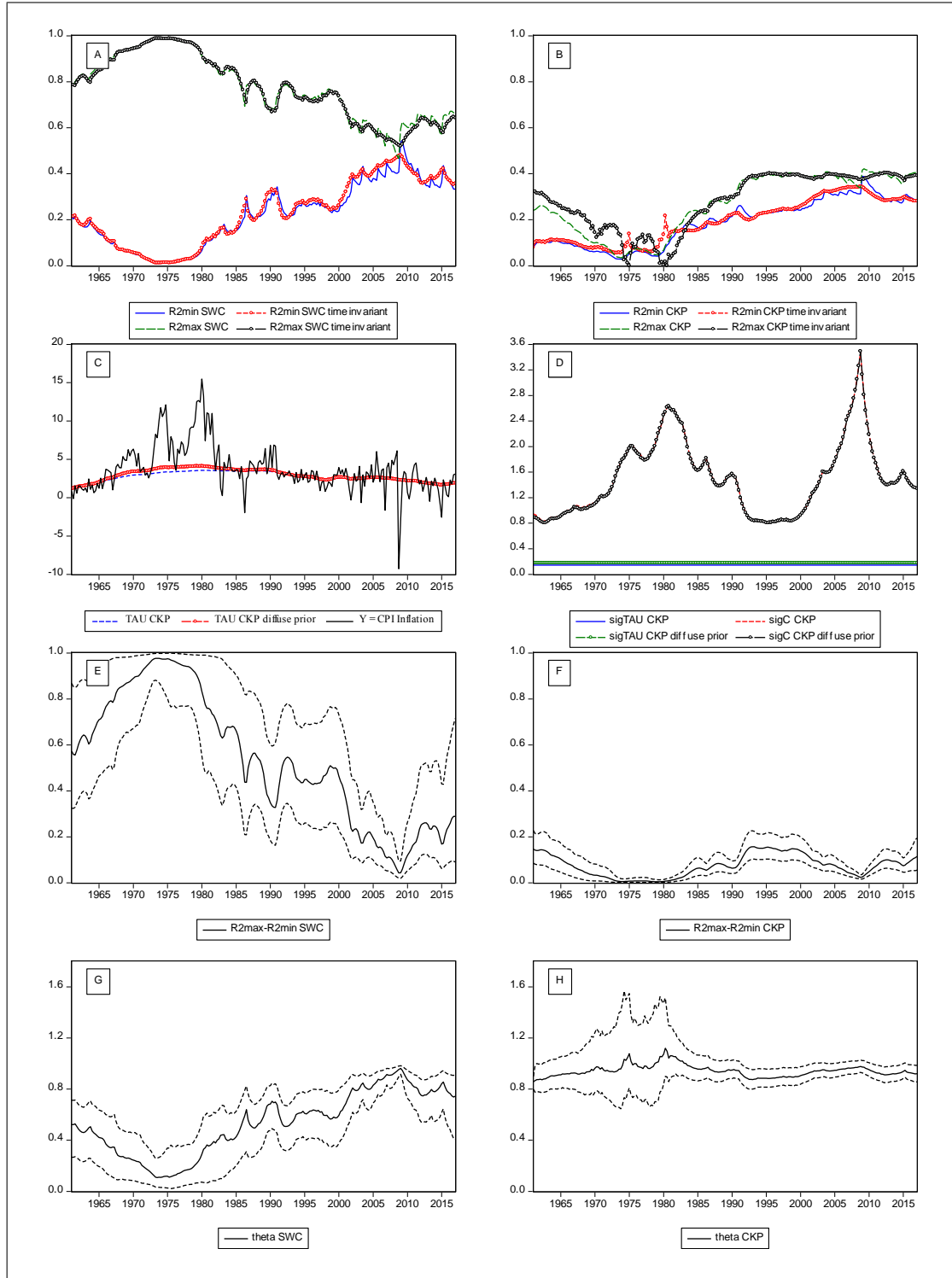
Figure M.1: US. Panels A and B plot posterior median estimates of $R^2_{\min,t}$ and $R^2_{\max,t}$ from Proposition 3 and the time-invariant approximations from Section 3 for the SWC and CKP models, respectively. Panel C plots posterior median estimates of the permanent component, τ_t, of inflation from the CKP model both where the priors are as in CKP (calibrated for US inflation data) and when $\sqrt{E(\sigma^2_\tau)} \neq 0.141$, as in CKP, but the priors are chosen so that this is 100 times bigger. This "diffuse" prior imposes less smoothness on the permanent component. Panel D plots posterior median estimates for $\sigma_{\tau,t}$ and $\sigma_{c,t}$ for both variants of the CKP model. Panels E and F plot 16.5%, 50% and 83.5% quantiles of the posterior distributions of $(R^2_{\max,t} - R^2_{\min,t})$ for the SWC and CKP models (using CKP's prior). Panels G and H plot 16.5%, 50% and 83.5% quantiles of the posterior distributions of θ_t for the SWC and CKP models (using CKP's prior).
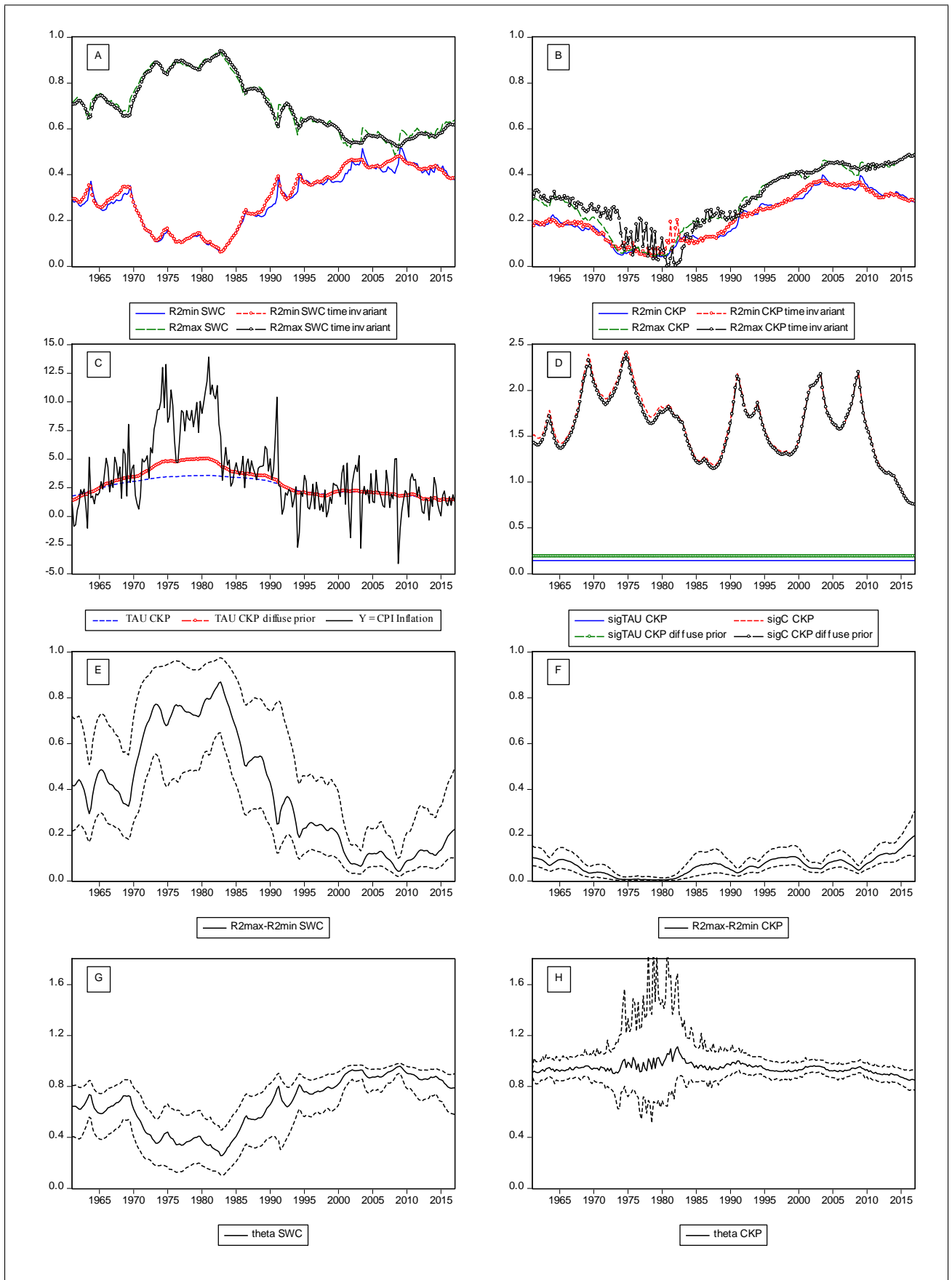
Figure M.2: Canada. See notes to Figure M.1
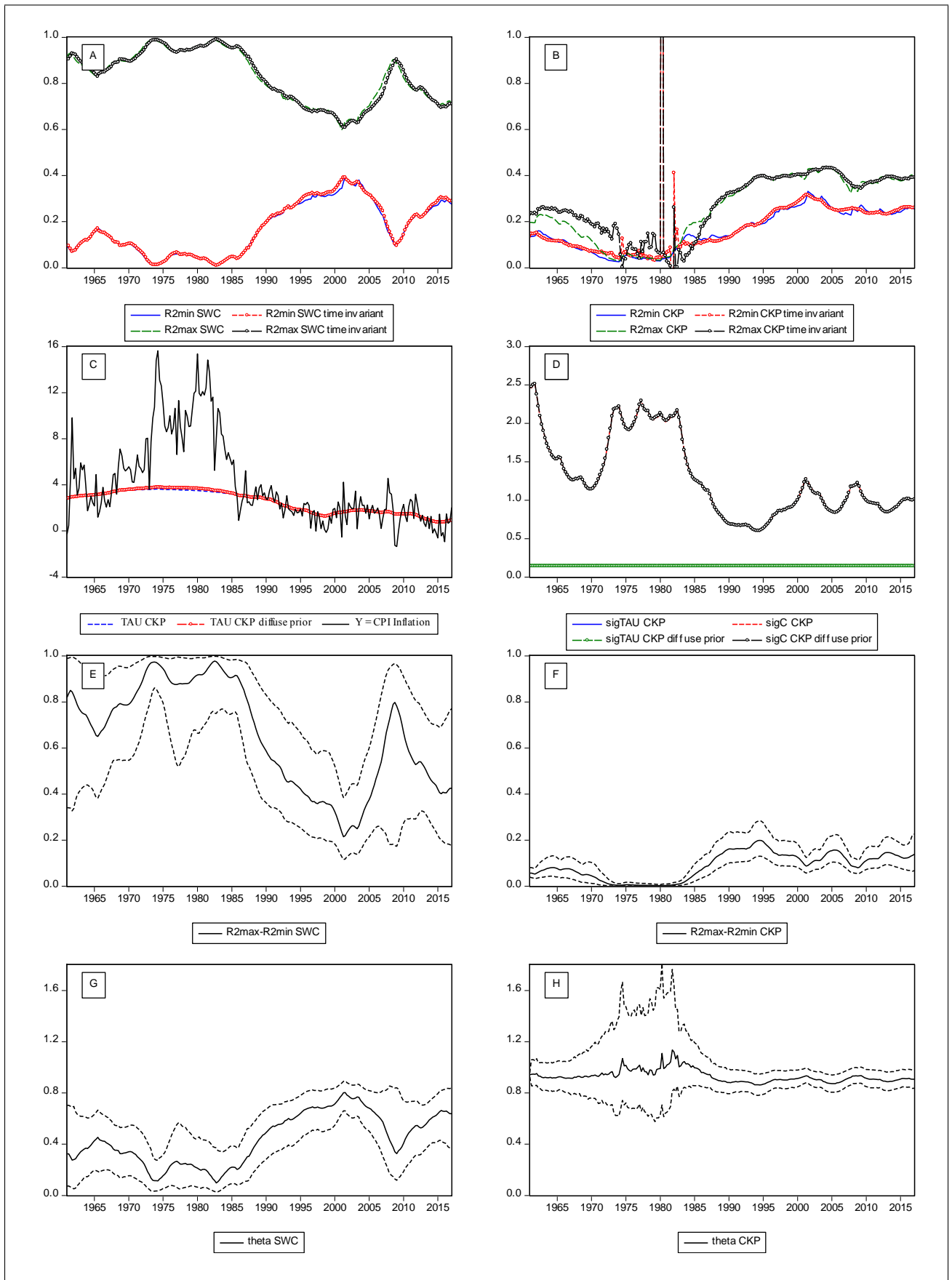
Figure M.3: France. See note to Figure M.1

71
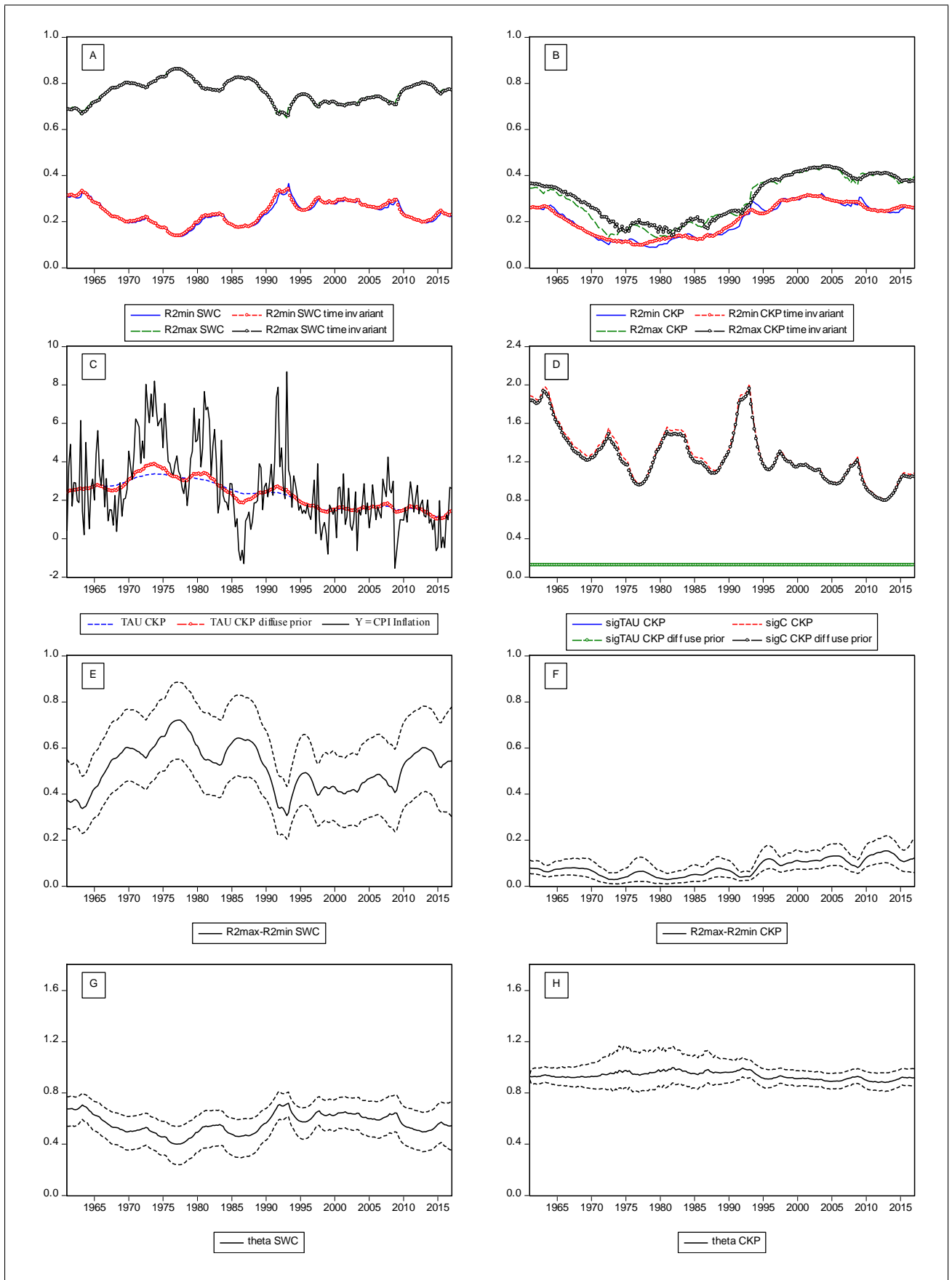
Figure M.4: Germany. See note to Figure M.1

72
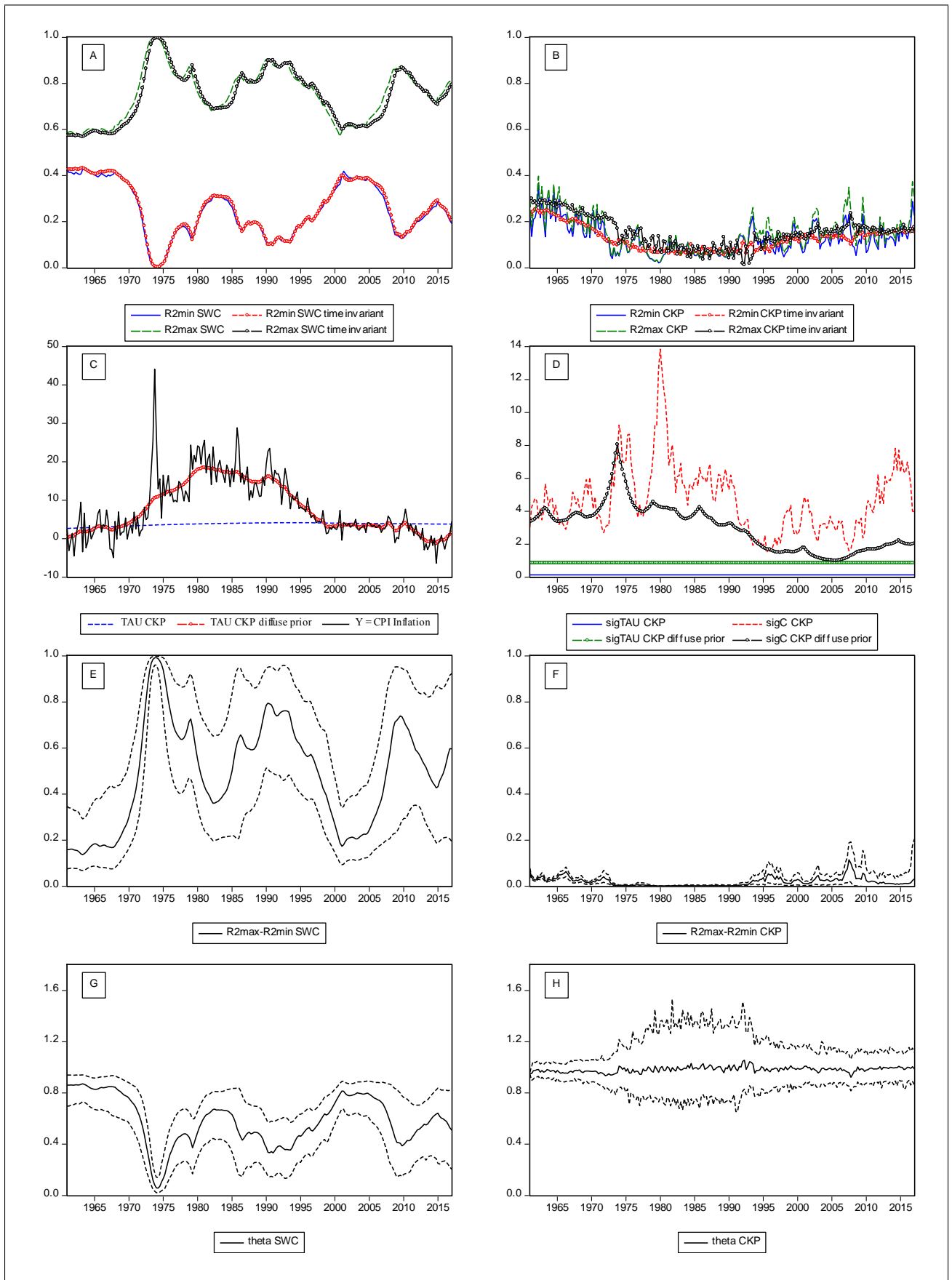
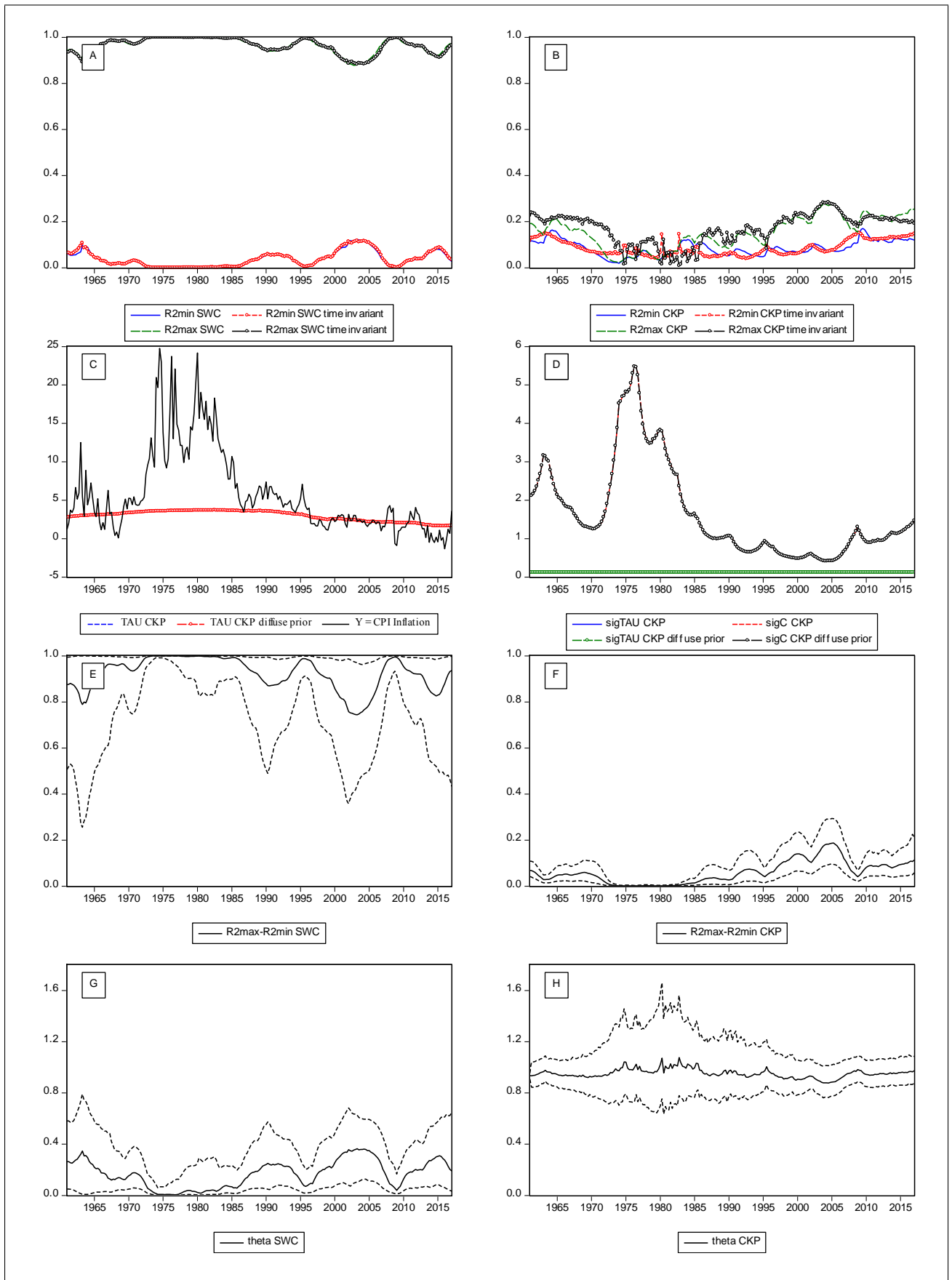Figure M.5: Greece. See note to Figure M.1
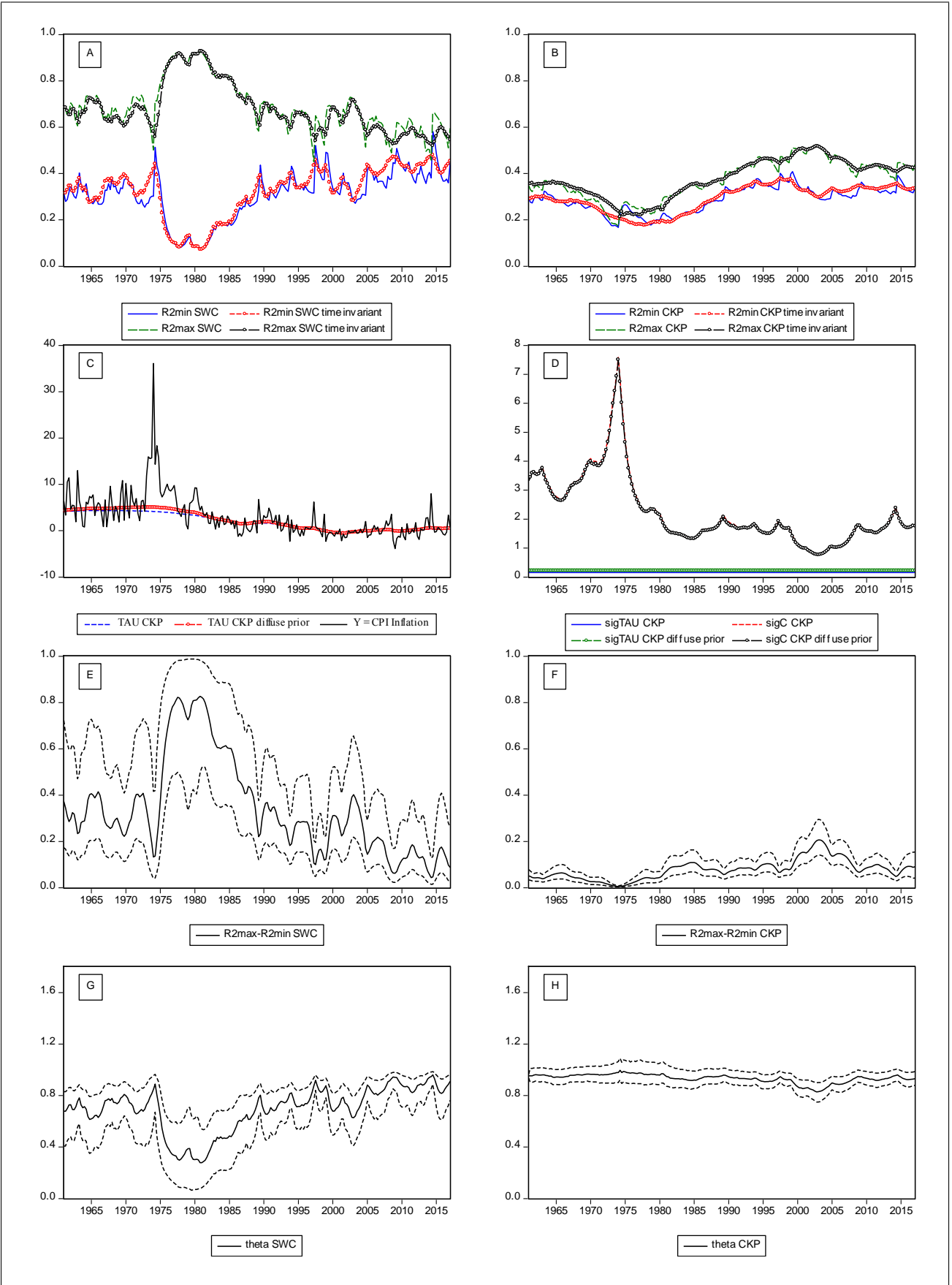
Figure M.6: Italy. See note to Figure M.1

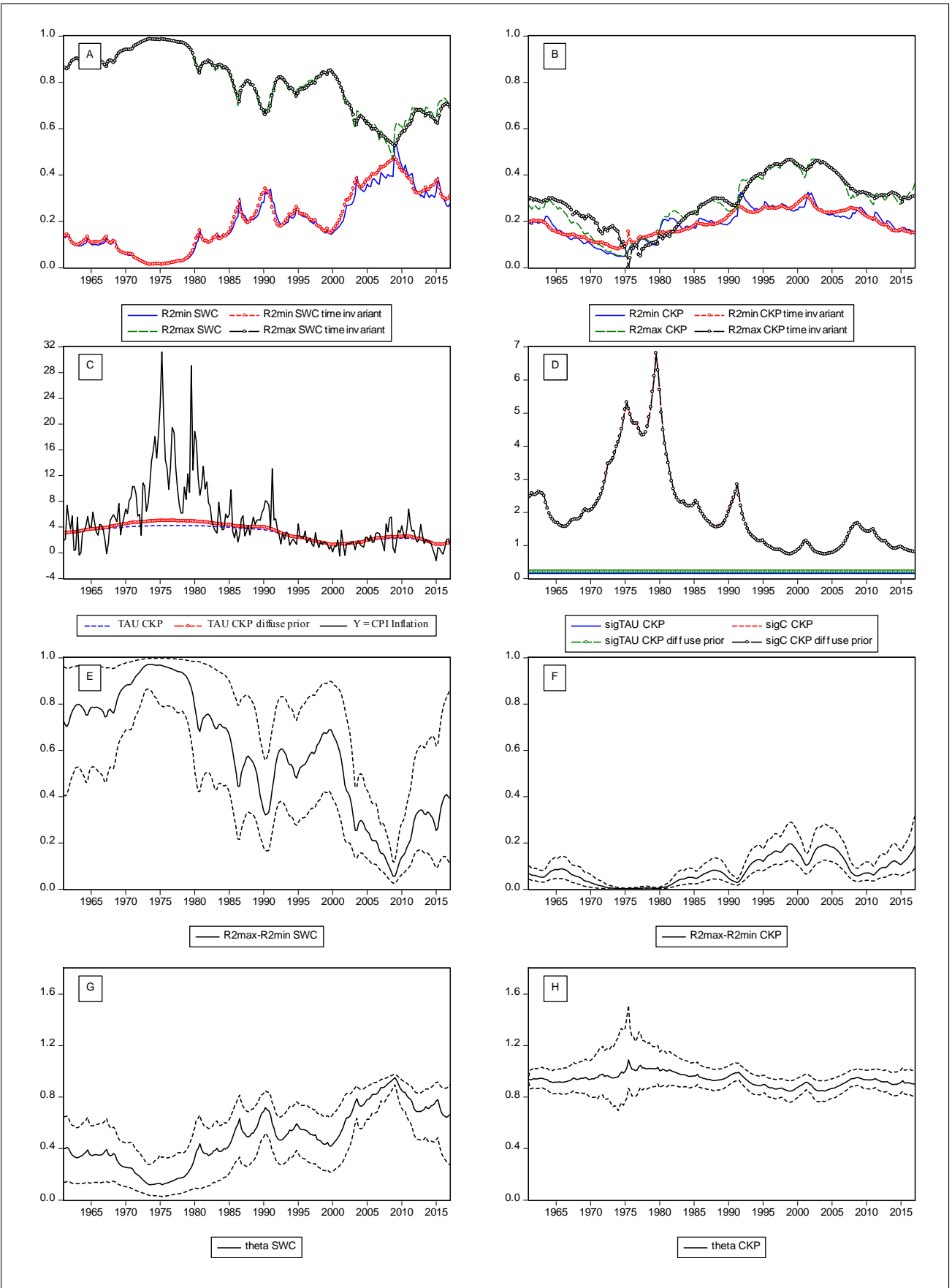Figure M.7: Japan. See note to Figure M.1
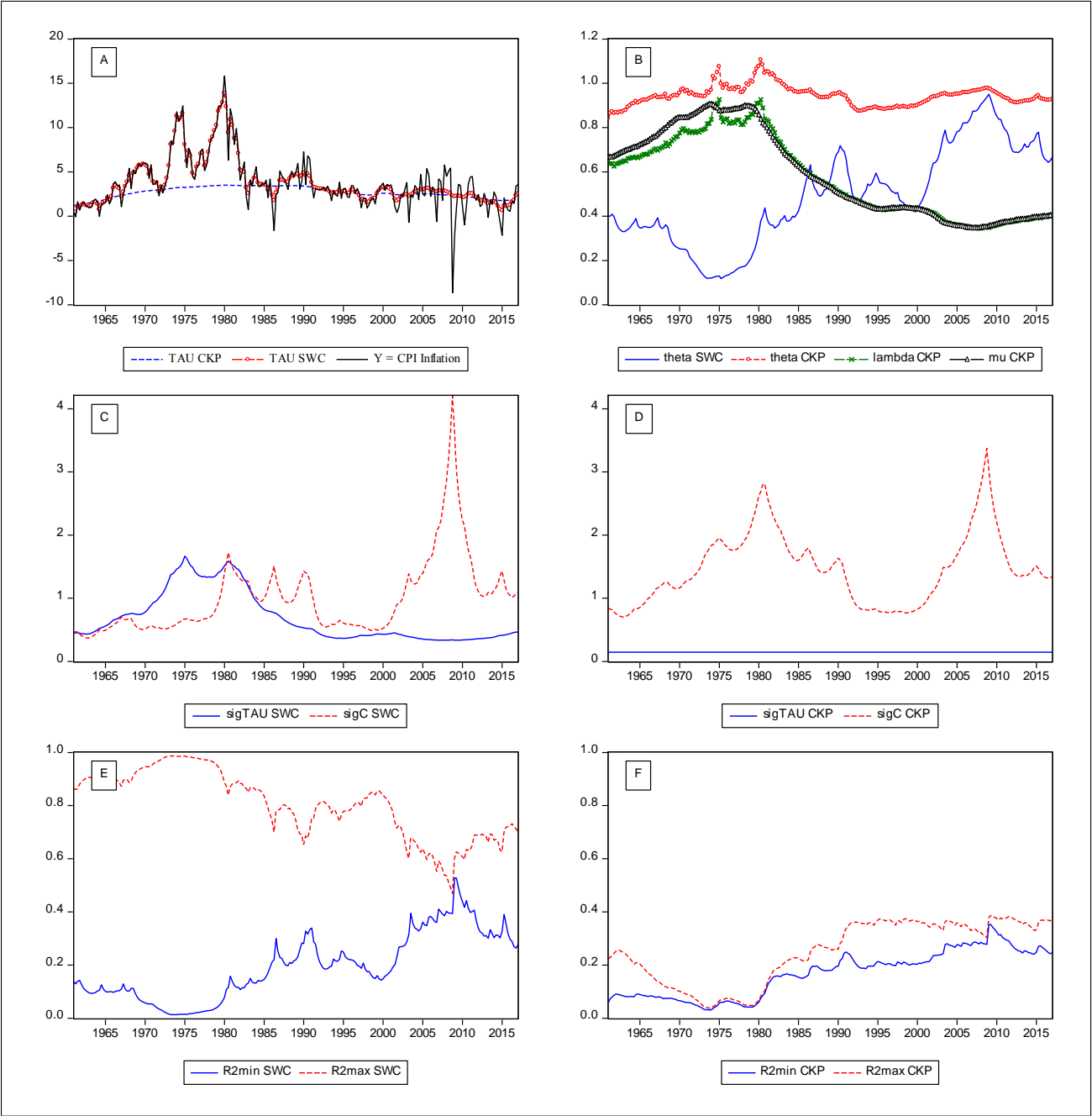
Figure M.8: UK. See note to Figure M.1

Figure M.9: US: using X12 to seasonally adjust CPI inflation. See note to Figure L.1
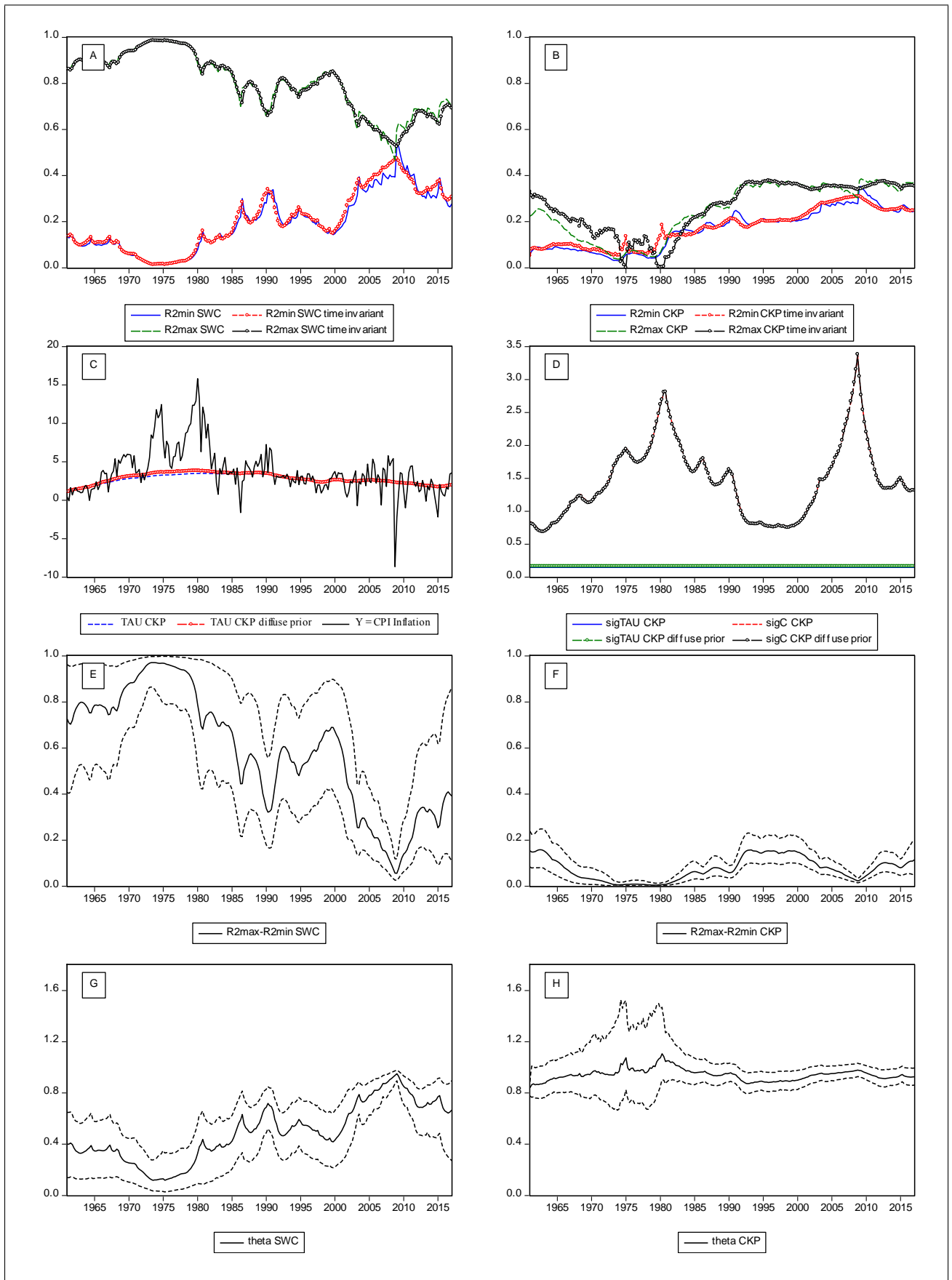
Figure M.10: US: using X12 to seasonally adjust CPI inflation (cont.). See note to Figure M.1
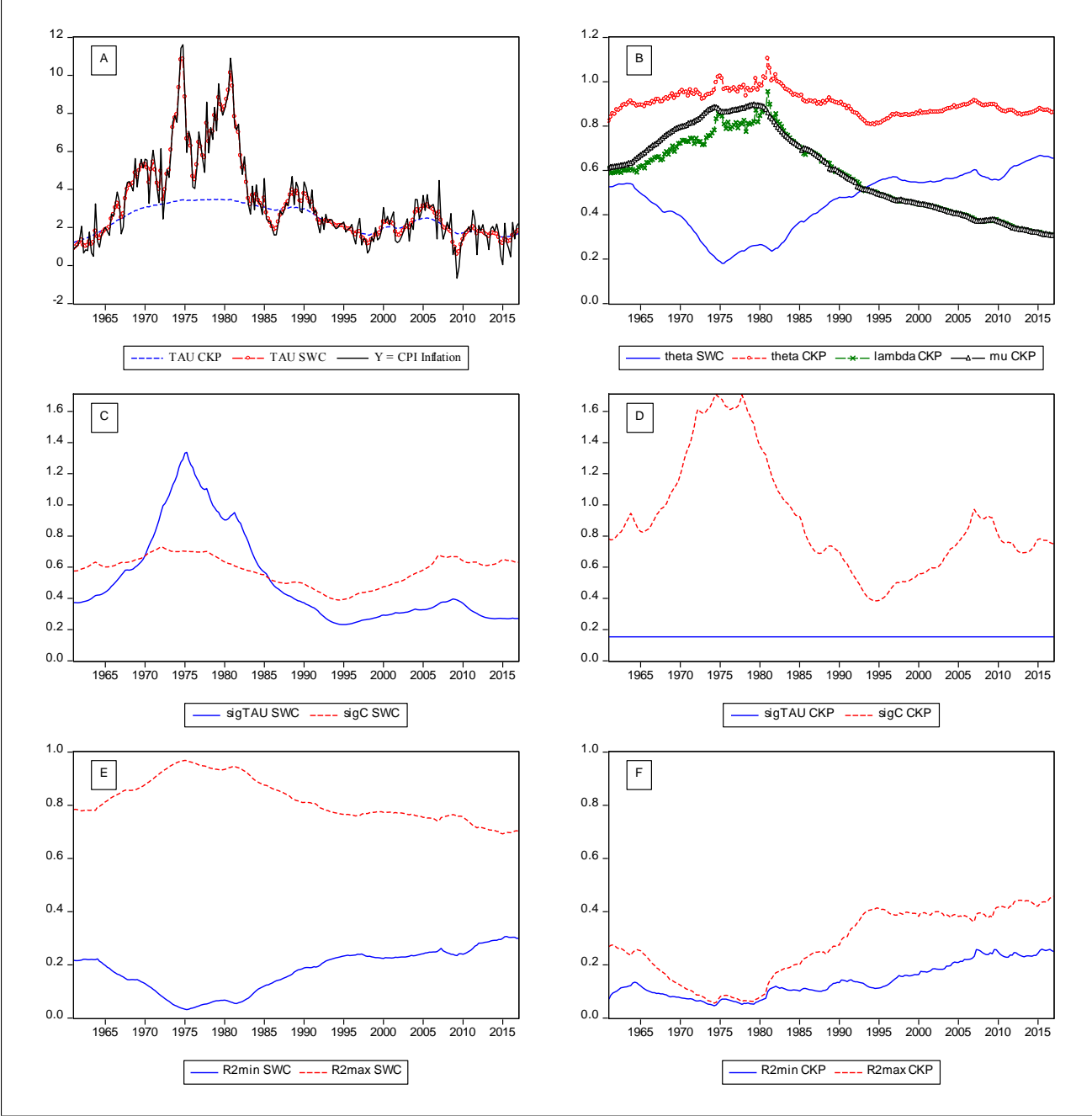
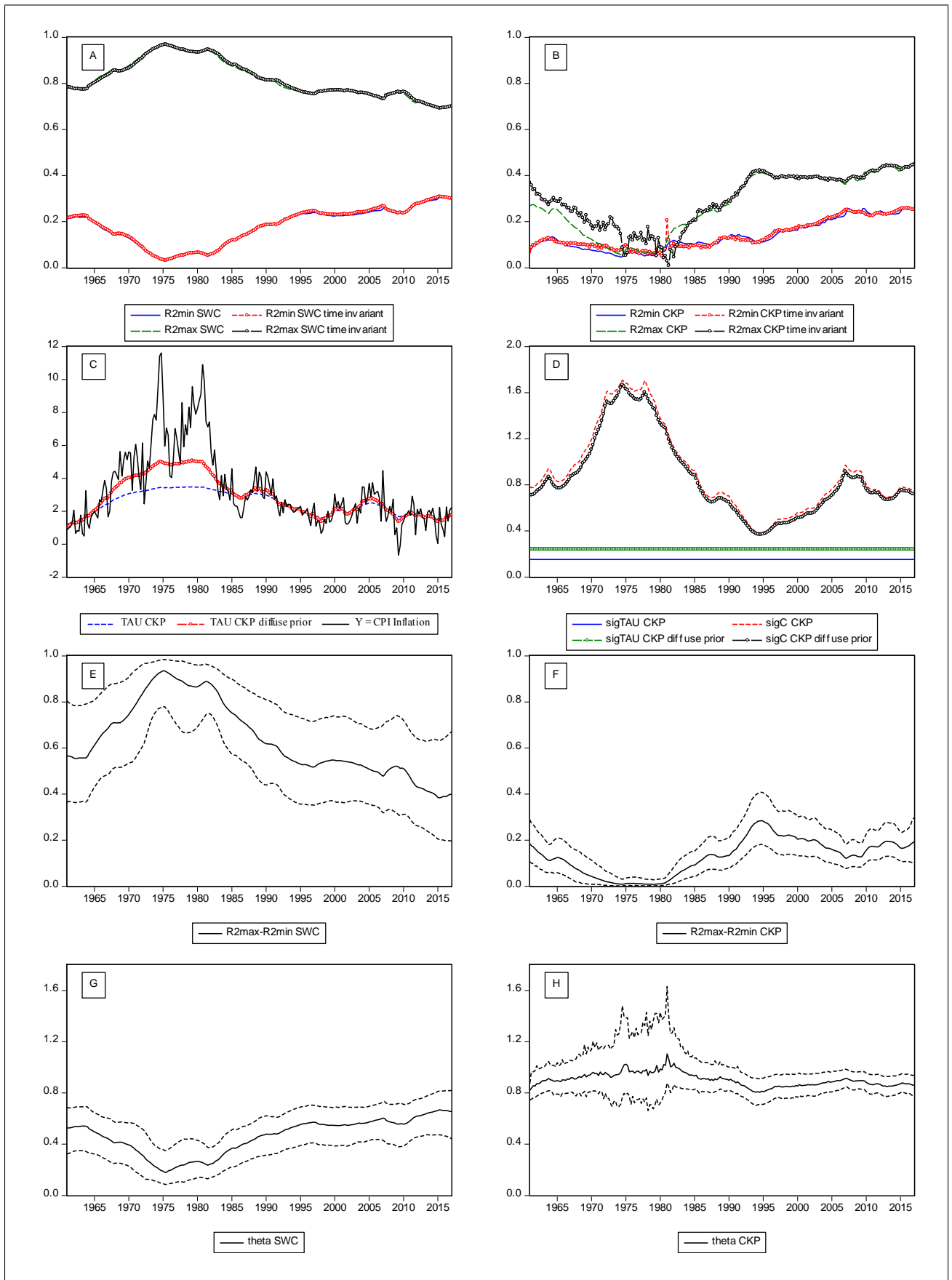Figure M.11: US: GDP deflator inflation. See note to Figure L.1

Figure M.12: US: GDP deflator inflation (cont.). See notes to Figure M.1

References

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

[2] Chan, J.C.C. (2017), "Specification Tests for Time-Varying Parameter Models with Stochastic Volatility", *Econometric Reviews*. Forthcoming.

[3] Chan, J.C.C., G. Koop and S.M. Potter (2013), "A New Model of Trend Inflation", *Journal of Business and Economic Statistics*, 31, 94-106.

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

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

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

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

[8] Mitchell, J., Robertson, D. and S. Wright (2017), "The Predictive Space: If $\mathbf{x}$ predicts y, what does y tell us about $\mathbf{x}$?", Birkbeck College Working Paper.

[9] Robertson, D. and S. Wright (2009), "The limits to stock return predictability", Birkbeck College Working Paper.

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

[11] 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.