Online Appendix to Accompany "R2 bounds for predictive models: what univariate properties tell us about multivariate predictability" by Mitchell, Robertson and Wright

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

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

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

$$u_t = \begin{bmatrix} 1 & \mathbf{0}_{1\times(n-1)} \end{bmatrix} \mathbf{D}\mathbf{s}_t \tag{A.8}$$

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

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

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

$$y_t = \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t \tag{A.10}$$

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

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

which could be substantially less than n.■

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

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

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

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

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

wherein $\widetilde{y}_t$ is then an MA(r), $\mathbf{w}_t = \begin{pmatrix} u_t & \mathbf{v}_t' \end{pmatrix}'$ and the final equality implicitly defines a set of 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

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

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

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

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

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

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

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

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

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 $\theta^N(L)$ in (10) in which θ_i is replaced with θ_i^{-1} for all i, has innovations η_t with the minimum variance, with*

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

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

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

so

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

and

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

Since ε_t is IID we have

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

implying

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

Hence we have

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

thus

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

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

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

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

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

where

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

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

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

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

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

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

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

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, $\Omega^F = \sigma_\varepsilon^2 \mathbf{1}\mathbf{1}'$, and $\Omega^N = \sigma_\eta^2 \mathbf{1}\mathbf{1}'$, 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 $R^2 = R_N^2 (\lambda, \theta)$.

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

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

Lemma 5 *In the population regression*

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

where the true process for y_t is as in (3), and $\mathbf{x}^F_t$ is the vector of fundamental univariate predictors defined in (C.20), all elements of the coefficient vector $\boldsymbol{\nu}_F$ are zero.

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

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

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

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

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

Lemma 6 *In the population regression*

$$y_t = \boldsymbol{\phi}'_{\mathbf{x}}\mathbf{x}_{t-1} + \boldsymbol{\phi}'_N\mathbf{x}^N_{t-1} + \zeta_t \tag{C.27}$$

where $\mathbf{x}^N_t$ is the vector of nonfundamental univariate predictors defined in (C.21), all elements of the coefficient vector $\boldsymbol{\phi}_{\mathbf{x}}$ are zero.

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

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

Using

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

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

$$\eta_t = F^r \prod_{i=1}^{r}(-\theta_i)\left(\frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right)\left(\beta_1\frac{v_{1t-1}}{1-\lambda_1 L} + \beta_2\frac{v_{2t-1}}{1-\lambda_2 L} + \ldots + \beta_r\frac{v_{rt-1}}{1-\lambda_r L} + u_t\right) \tag{C.29}$$

Now

$$\begin{aligned} F^r \frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\frac{v_{kt-1}}{(1-\lambda_k L)} &= F^r\left(\frac{\prod_{i\neq k}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right)v_{kt-1} \\ &= v_{kt} + c_1 v_{kt+1} + c_2 v_{kt+2} + \ldots \end{aligned} \tag{C.30}$$

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

$$F^r\left(\frac{\prod_{i=1}^{r}(1-\lambda_i L)}{\prod_{i=1}^{r}(1-\theta_i F)}\right)u_t = u_t + b_1 u_{t+1} + b_2 u_{t+2} + \ldots \tag{C.31}$$

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

$$R_{min}^2(\lambda, \theta) \leq R^2 \leq R_{max}^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) \tag{C.32}$$

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

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 ARMA(r, r). The minimal ARMA(p, q) representation will only be of lower order if we have *either* cancellation of some MA and AR roots, *or* an MA or AR coefficient precisely equal to zero. Thus we have

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

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

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

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

$$c_t = \mu c_{t-1}+s_{c,t} \tag{G.2}$$

$$\tau_t = \tau_{t-1}+s_{\tau,t} \tag{G.3}$$

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

$$y_t=\triangle Y_t=\triangle c_t+\triangle\tau_t \tag{G.4}$$

$$=\left(\mu-1\right)c_{t-1}+s_{c,t}+s_{\tau,t} \tag{G.5}$$

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

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

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

$$\begin{aligned} -\frac{\theta}{1+\theta^2} &= \frac{Cov\left(s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}, s_{c,t-1} - s_{c,t-2} + s_{\tau,t-1} - \lambda s_{\tau,t-2}\right)}{Var\left(s_{c,t} - s_{c,t-1} + s_{\tau,t} - \lambda s_{\tau,t-1}\right)} && \text{(G.8)} \\ &= \frac{-\sigma_c^2 - \lambda\sigma_\tau^2}{2\sigma_c^2 + \left(1+\lambda^2\right)\sigma_\tau^2} && \text{(G.9)} \end{aligned}$$

So

$$\begin{aligned} -\frac{\theta}{1+\theta^2} &= \frac{-\sigma_c^2 - \lambda\sigma_\tau^2}{2\sigma_c^2 + \left(1+\lambda^2\right)\sigma_\tau^2} && \text{(G.10)} \\ &= \frac{-\sigma_c^2 + \lambda\sigma_c^2 - \lambda\left(\sigma_\tau^2 + \sigma_c^2\right)}{2\sigma_c^2 - \left(1+\lambda^2\right)\sigma_c^2 + \left(1+\lambda^2\right)\left(\sigma_\tau^2+\sigma_c^2\right)} && \text{(G.11)} \end{aligned}$$

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

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

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

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

$$\begin{aligned}\left(1 + \lambda^2\right)(\lambda + (1-\lambda) q) - \lambda\left(1 + \lambda^2 + \left(1 - \lambda^2\right) q\right) &= \lambda + (1-\lambda) q \\ &\quad + \lambda^3 + \lambda^2 (1-\lambda) q \\ &\quad - \lambda - \lambda^3 - \lambda\left(1 - \lambda^2\right) q \\ &= \left((1-\lambda) + \lambda^2 (1-\lambda) - \lambda\left(1 - \lambda^2\right)\right) q \\ &= \left(1 - 2\lambda + \lambda^2\right) q \\ &= (1-\lambda)^2 q > 0\end{aligned} \tag{G.16}$$

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

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

will have $\theta > \lambda$ (see Figure G.1).■

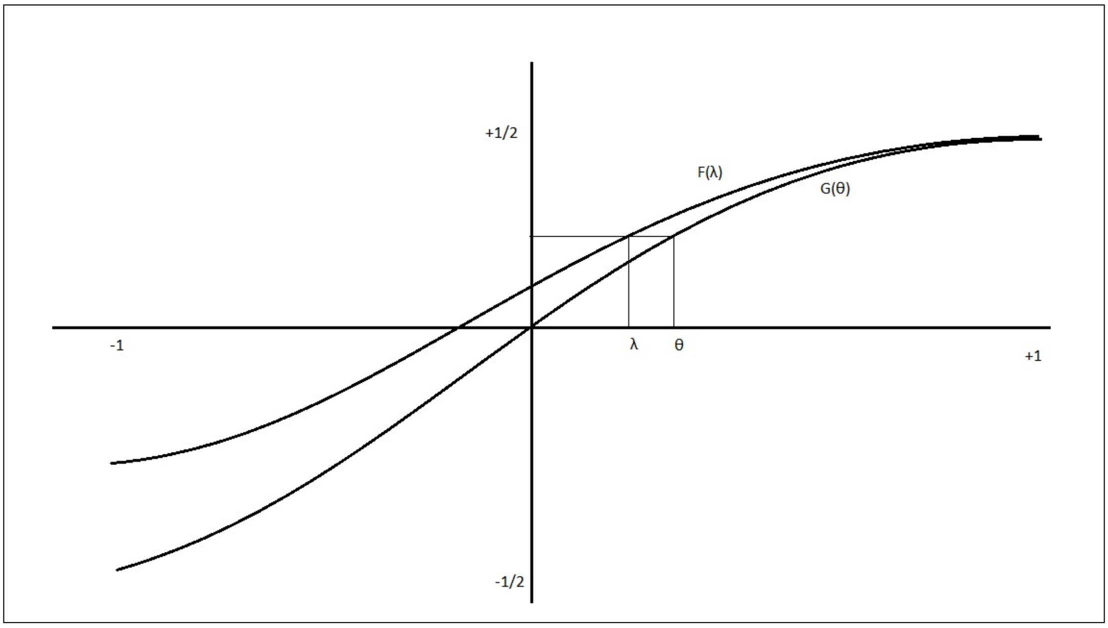

Figure G.1: Proof of Lemma 3 (Beveridge-Nelson decomposition)

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

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

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 $\rho_{uv}\left(\theta, \lambda, R^2\right)$ as R^2 covers that range.

The first order condition yields a possible stationary point where:

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

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,

$$\begin{aligned} y_t &= \beta_t x_{t-1} + u_t & \text{(I.1)} \\ x_t &= \mu_t x_{t-1} + v_t & \text{(I.2)} \end{aligned}$$

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

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

$$\begin{aligned} y_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}} y_{t-1} &= \beta_t x_{t-1} - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}}\beta_{t-1} x_{t-2} + u_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}} u_{t-1} \\ &= \beta_t \left(x_{t-1} - \mu_{t-1} x_{t-2}\right) + u_t - \mu_{t-1}\frac{\beta_t}{\beta_{t-1}} u_{t-1} \qquad \text{(I.3)} \end{aligned}$$

thus

$$y_t - \lambda_t y_{t-1} = \beta_t v_{t-1} + u_t - \lambda_t u_{t-1} \qquad \text{(I.4)}$$

wherein

$$\lambda_t = \mu_{t-1}\frac{\beta_t}{\beta_{t-1}} \qquad \text{(I.5)}$$

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

$$
\begin{aligned}
y_t - \lambda_t y_{t-1} &= \varepsilon_t - \theta_t \varepsilon_{t-1} && \text{(I.6)} \\
y_t - \lambda_t y_{t-1} &= \eta_t - \gamma_t \eta_{t-1} && \text{(I.7)}
\end{aligned}
$$

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:

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

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

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

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

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

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

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

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 the sequences

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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,

$$\begin{aligned} Y_t &= \tau_t + c_t && \text{(I.1)} \\ \tau_t &= \tau_{t-1} + s_{\tau,t} && \text{(I.2)} \\ c_t &= \mu_t c_{t-1} + s_{c,t} && \text{(I.3)} \end{aligned}$$

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

$$\begin{aligned} y_t &= \Delta Y_t && \text{(I.4)} \\ x_t &= c_t && \text{(I.5)} \\ \beta_t &= \mu_t - 1 && \text{(I.6)} \\ u_t &= s_{c,t} + s_{\tau,t} && \text{(I.7)} \\ v_t &= s_{c,t} && \text{(I.8)} \end{aligned}$$

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)

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

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*

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

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);* $\rho_{BN}(\mathbf{\Psi}) = corr\left(u_t, \boldsymbol{\delta}'\mathbf{v}_t\right)$, *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)

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

this gives

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

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)); $V_{\hat{y}}(1,1) = \frac{1+\lambda}{1-\lambda}$ 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 $\frac{dV_{\hat{y}}}{dR^2} < 0$ as stated in the proposition.

Since

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

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

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

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

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

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

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

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

But given the assumptions in the proposition we have

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

hence

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

as required.

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

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

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

Thus we only need to show that $H\left(V_{\hat{y}}, R^2\right)$ is positive for $R^2_{\max}(\lambda,\theta) < \frac{1}{2}$. Given that $R^2_{\max}(\lambda,\theta)$ 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

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

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

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

Hence

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

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

The first column of Table 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 $V_y = \left(1 - R^2_{\min}\right) c(1)^2$, 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 $V_{\widehat{y}} = \frac{1+\lambda}{1-\lambda} \Rightarrow \lambda = \frac{V_{\widehat{y}}-1}{V_{\widehat{y}}+1}$, thus a median value of $V_{\widehat{y}} \approx 4$ would arise from an AR(1) predictor with $\lambda = 0.6$, thus a value of $V_{\widehat{y}}$ 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 $V_{\widehat{y}} = 1$.

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 $Y_t = 400 \log (CPI_t / CPI_{t-1})$, 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.

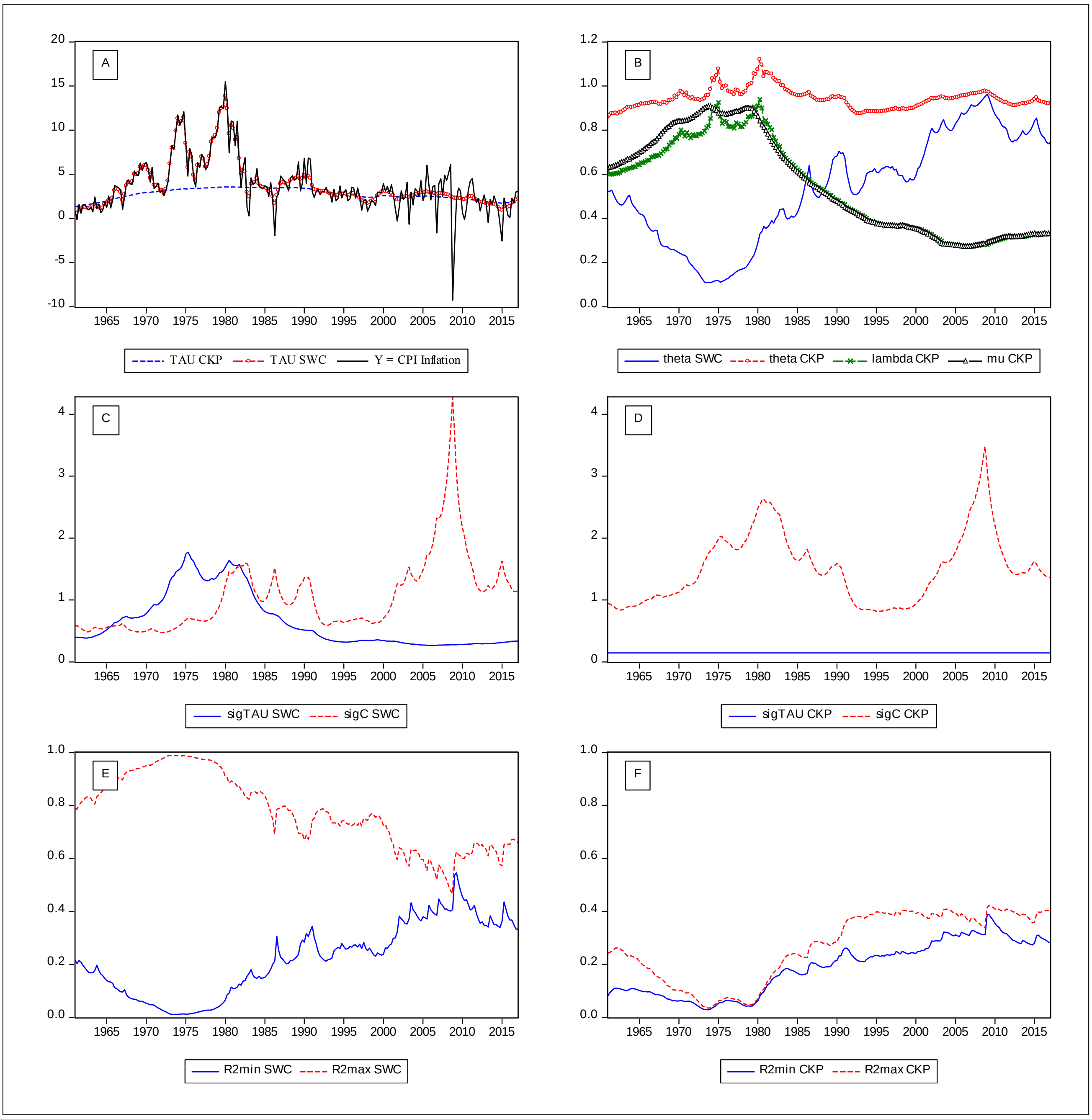

Figure L.1: US. Panel A plots posterior median estimates of the permanent component, τ_t, of inflation from the SWC and CKP models alongside CPI inflation. Panel B plots posterior median estimates of θ_t, λ_t and μ_t from the SWC and CKP models (where $\lambda_t = \mu_t = 0$ for SWC). Panels C and D 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.

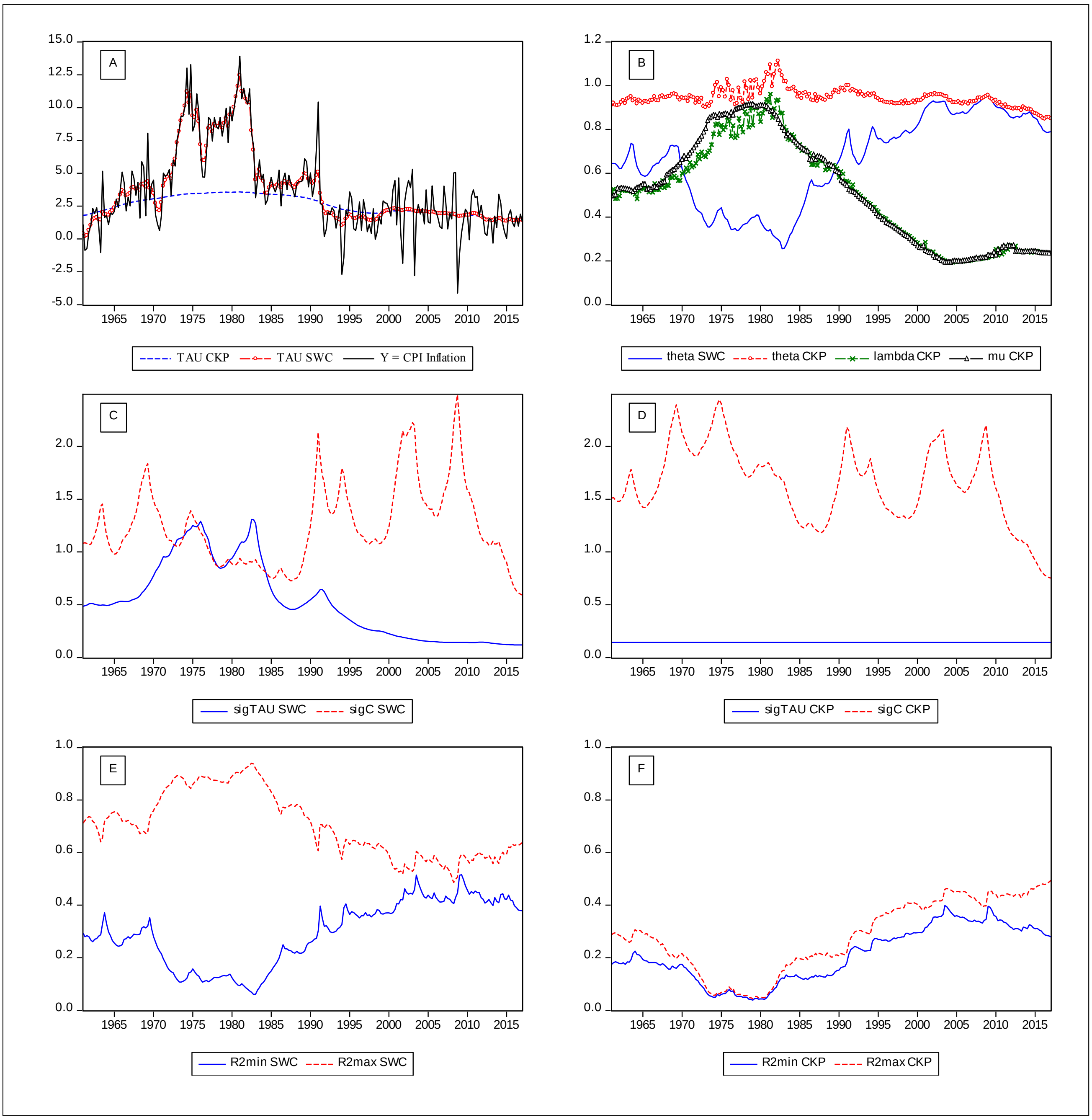

Figure L.2: Canada. See note to Figure L.1

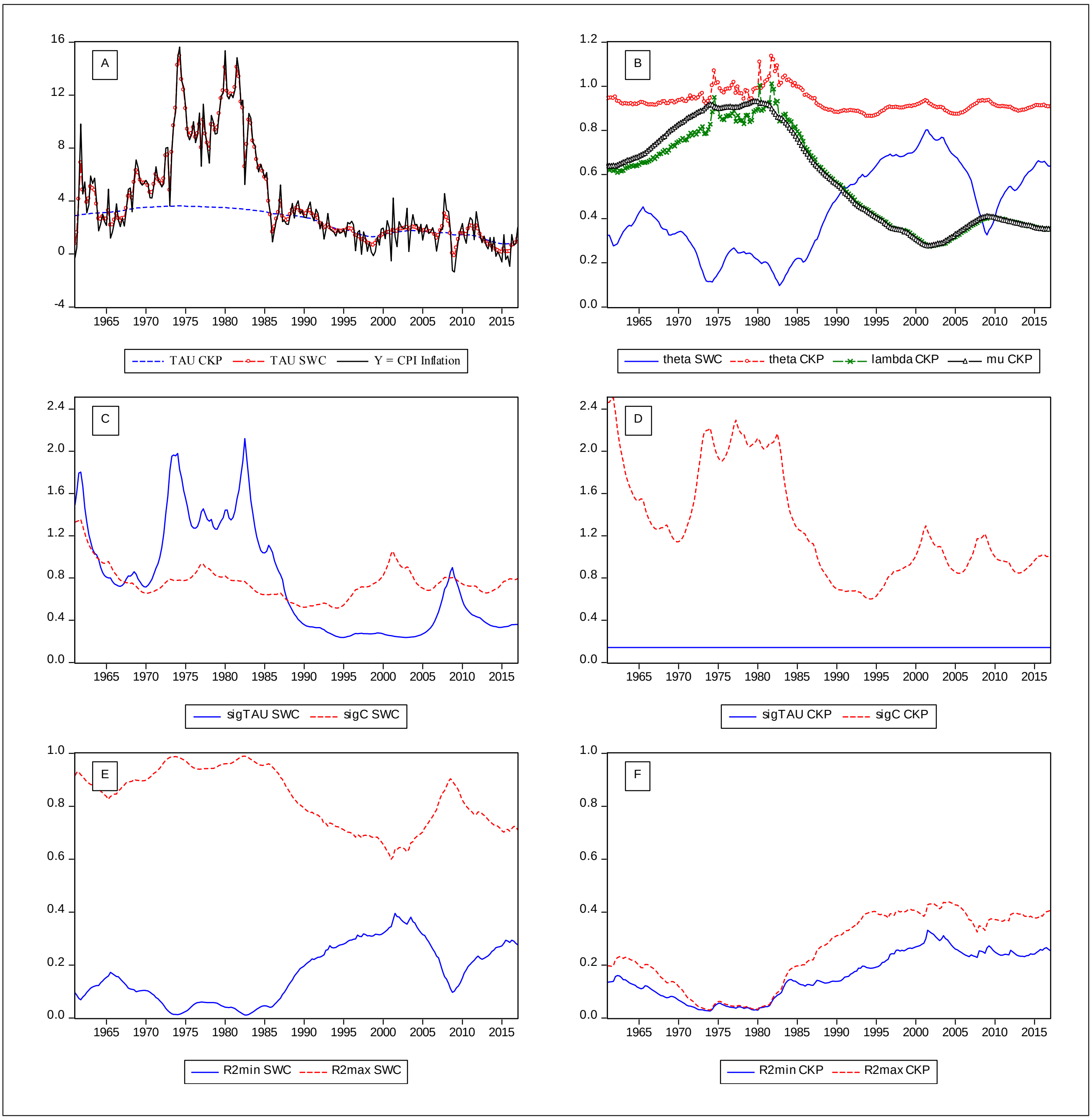

Figure L.3: France. See note to Figure L.1

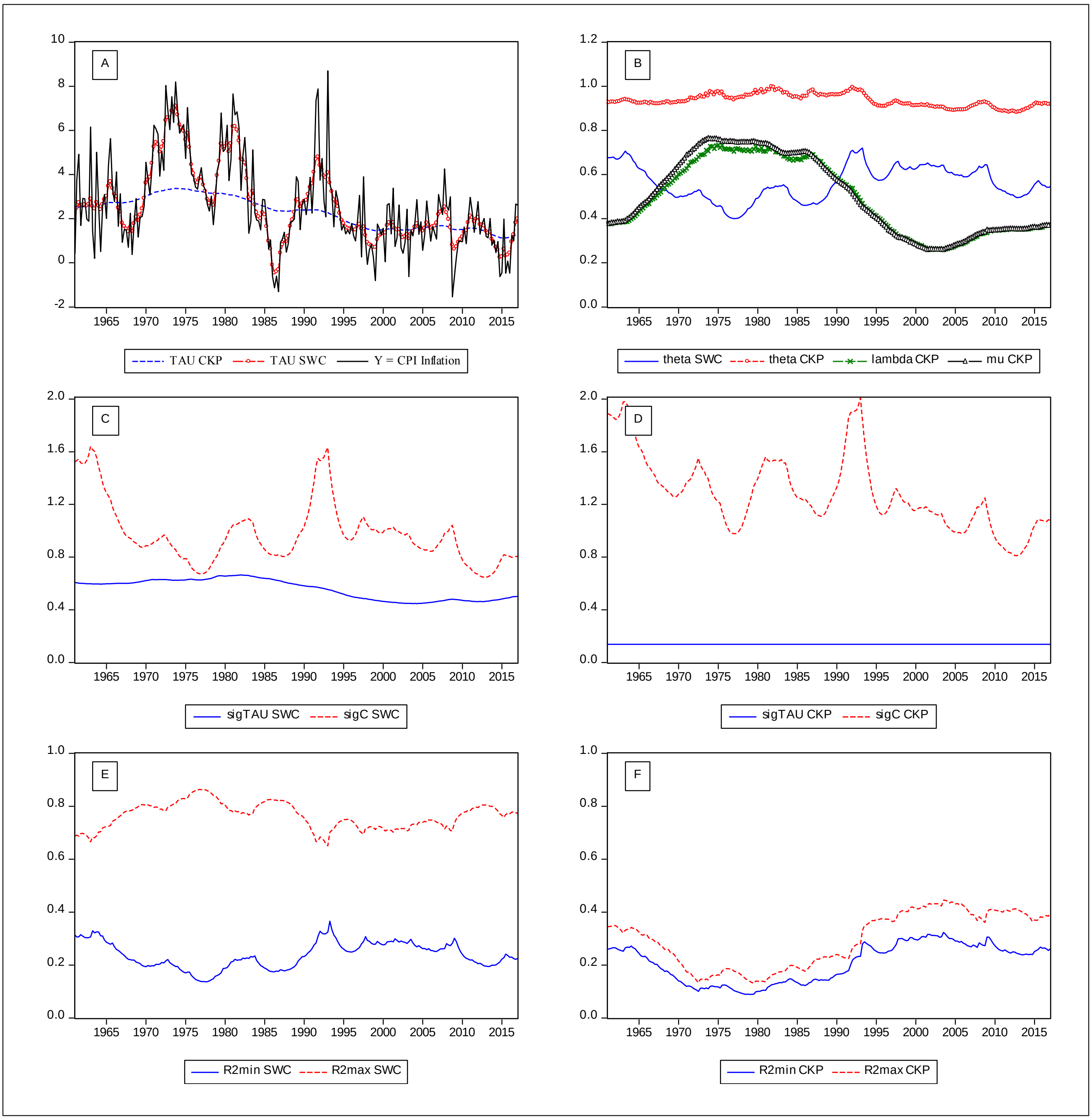

Figure L.4: Germany. See note to Figure L.1

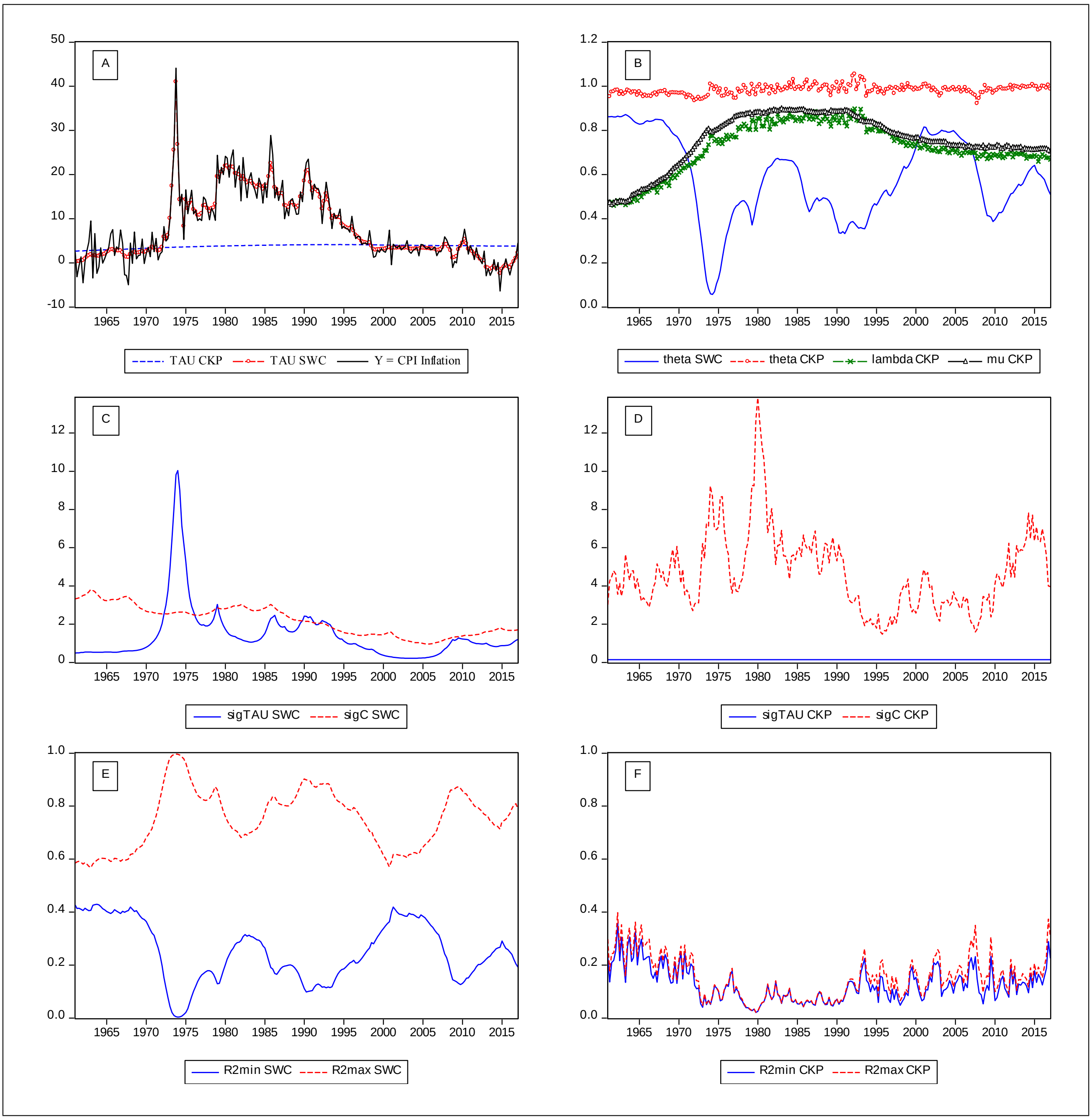

Figure L.5: Greece. See note to Figure L.1

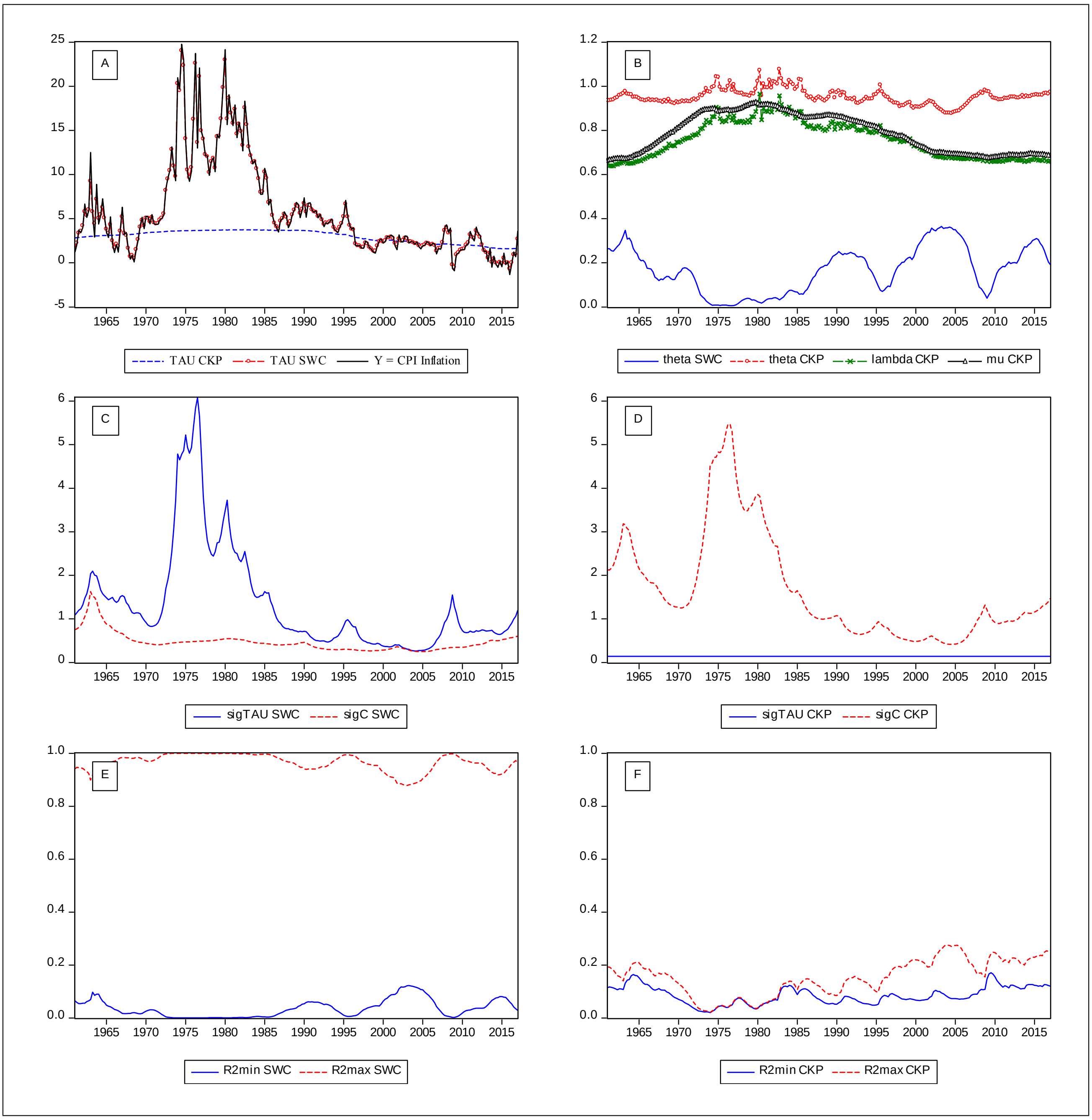

Figure L.6: Italy. See note to Figure L.1

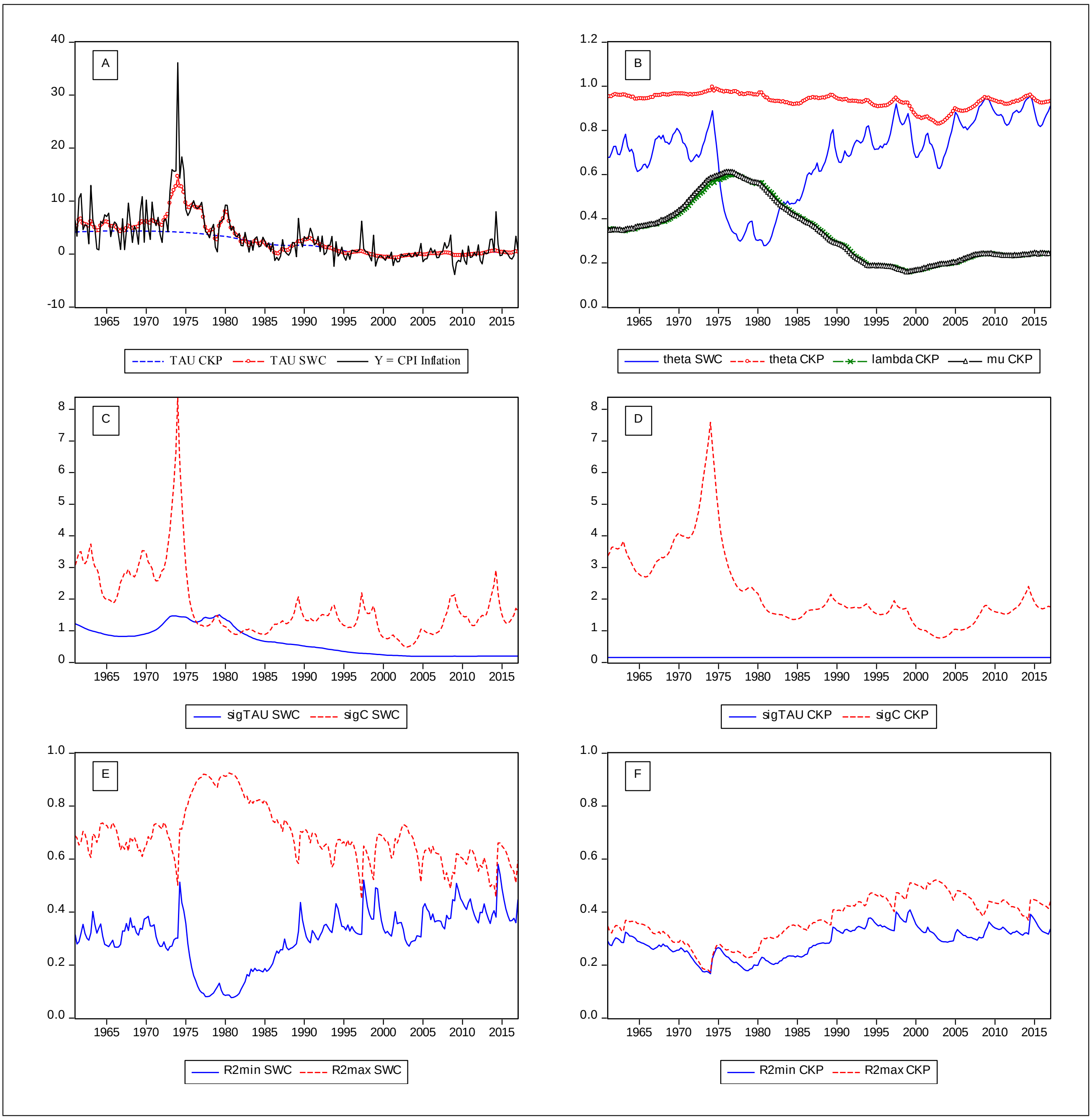

Figure L.7: Japan. See note to Figure L.1

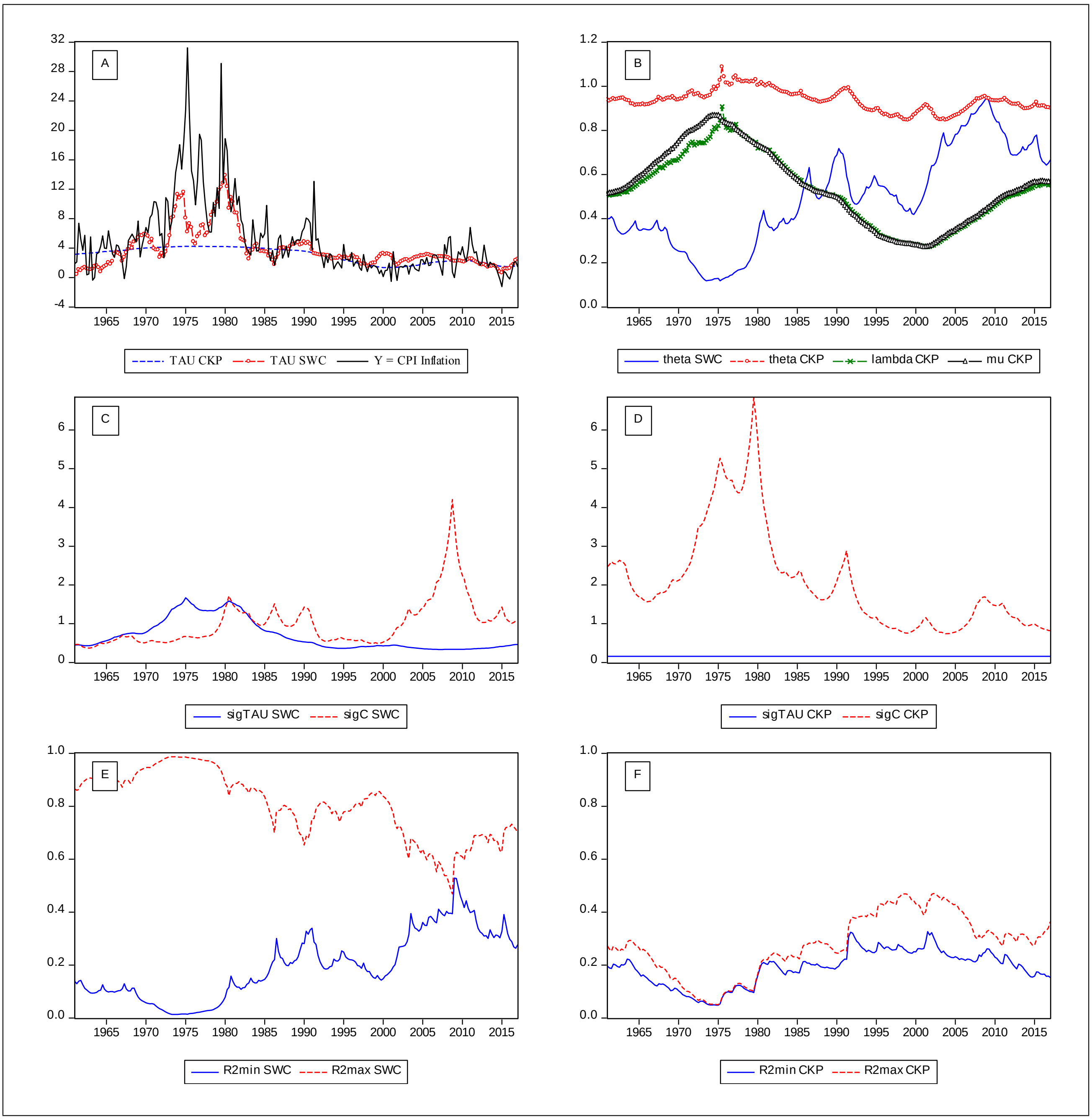

Figure L.8: UK. See note to Figure L.1

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 $\widehat{\theta}_t$ (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.

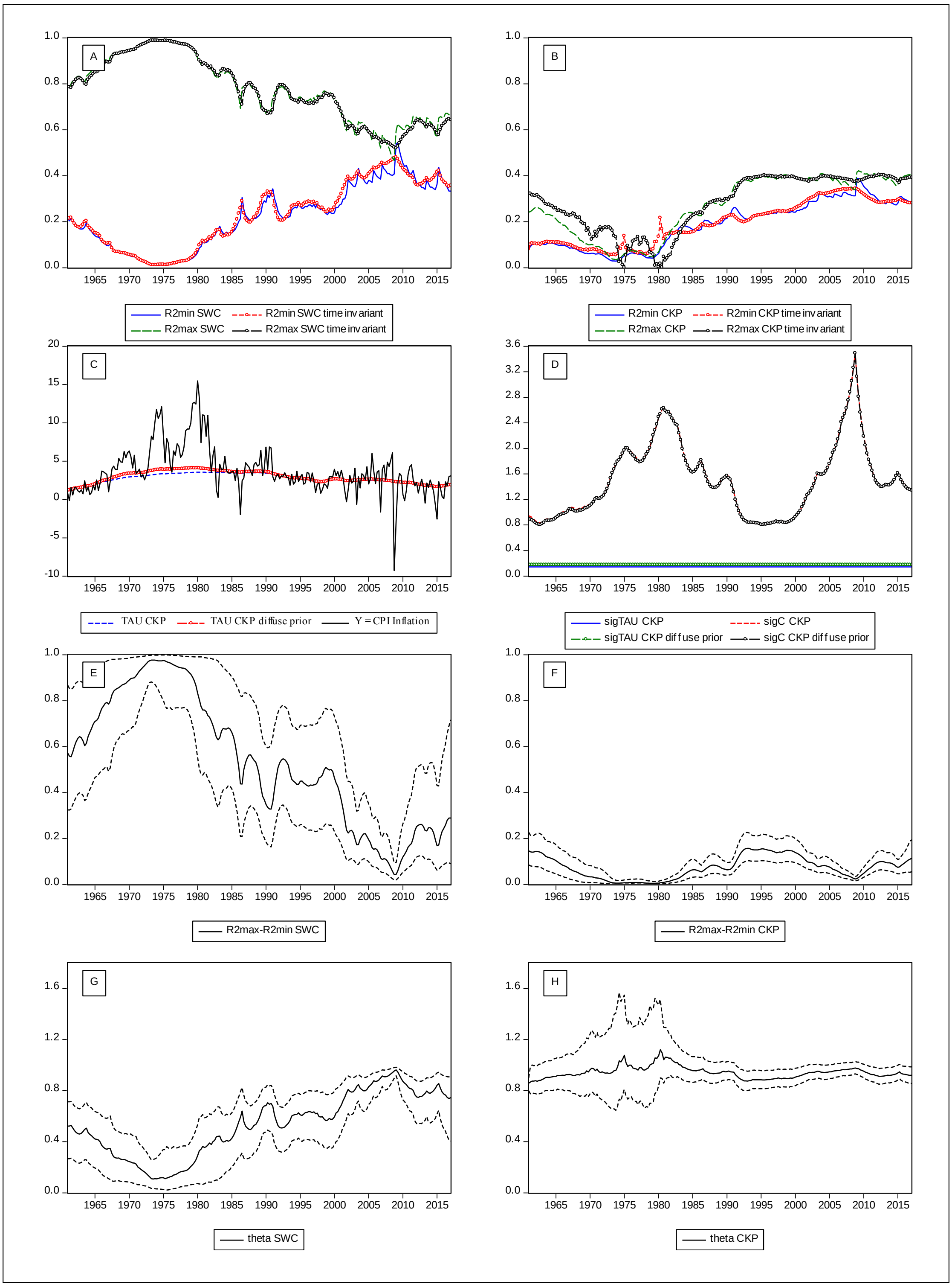

Figure M.1: US. Panels A and B plot posterior median estimates of $R^2_{\min,t}$ and $R^2_{\max,t}$ from Proposition 3 and the time-invariant approximations from Section 3 for the SWC and CKP models, respectively. Panel C plots posterior median estimates of the permanent component, τ_t, of inflation from the CKP model both where the priors are as in CKP (calibrated for US inflation data) and when $\sqrt{E(\sigma_\tau^2)} \neq 0.141$, 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).

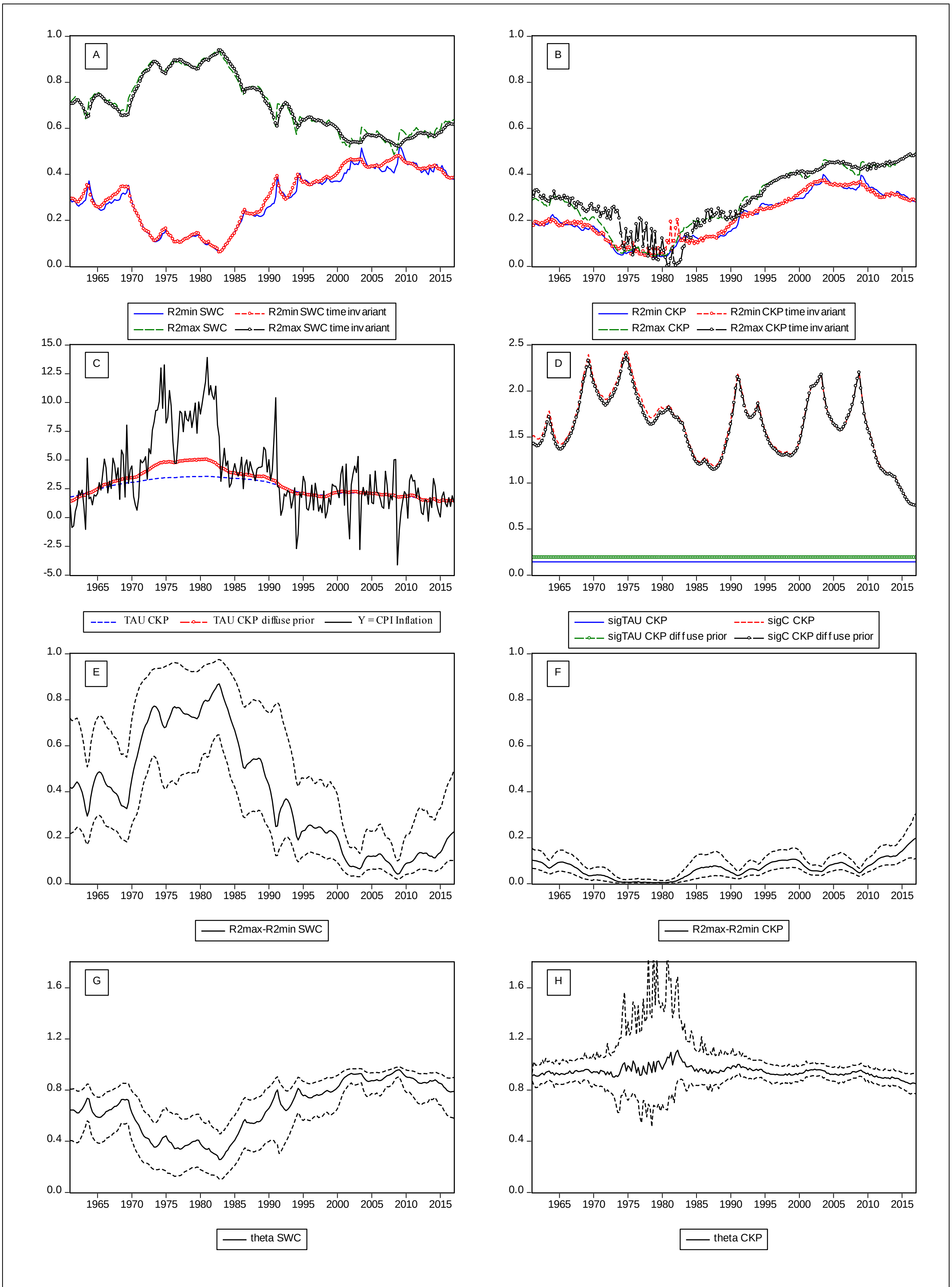

Figure M.2: Canada. See notes to Figure M.1

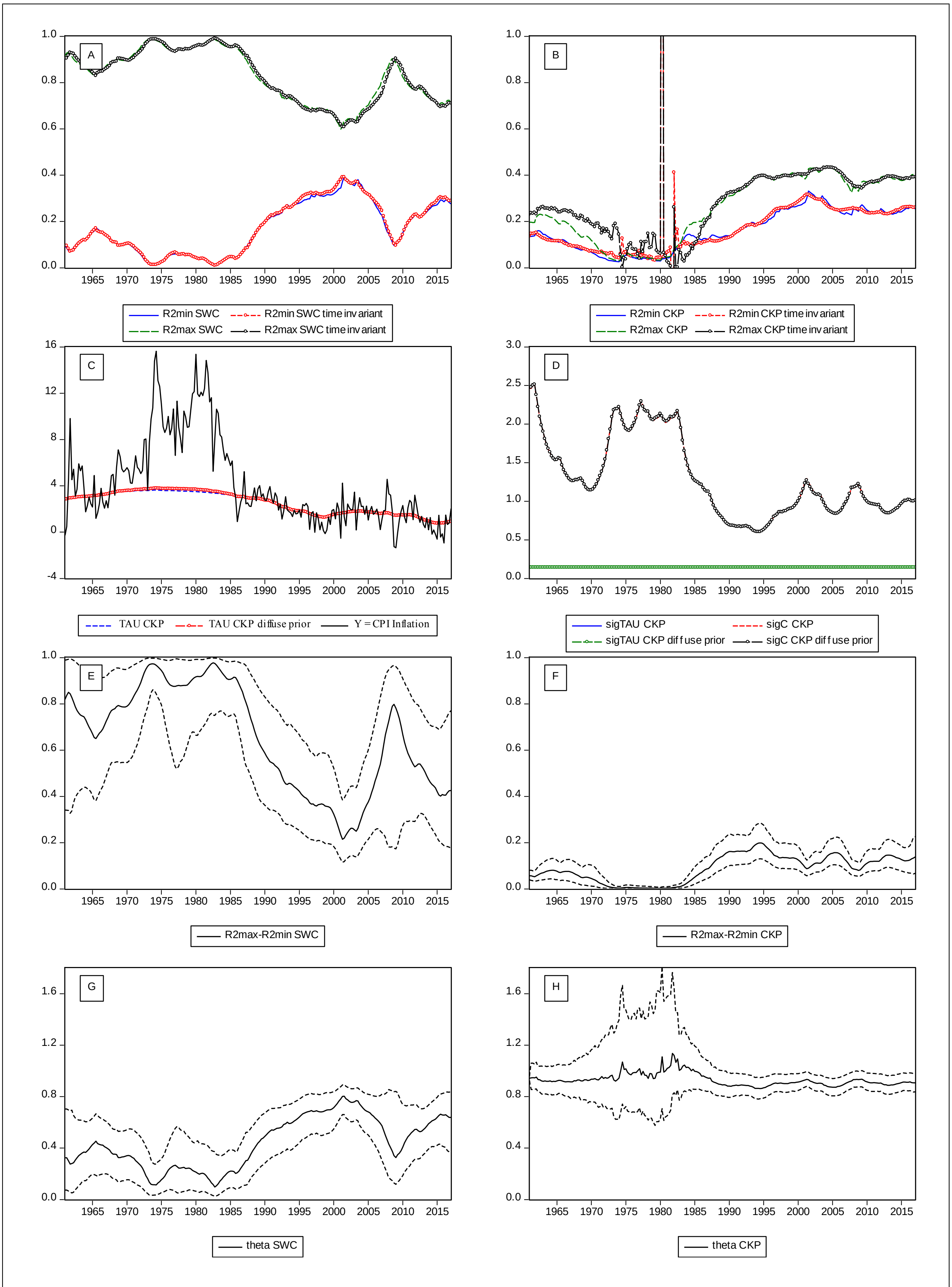

Figure M.3: France. See note to Figure M.1

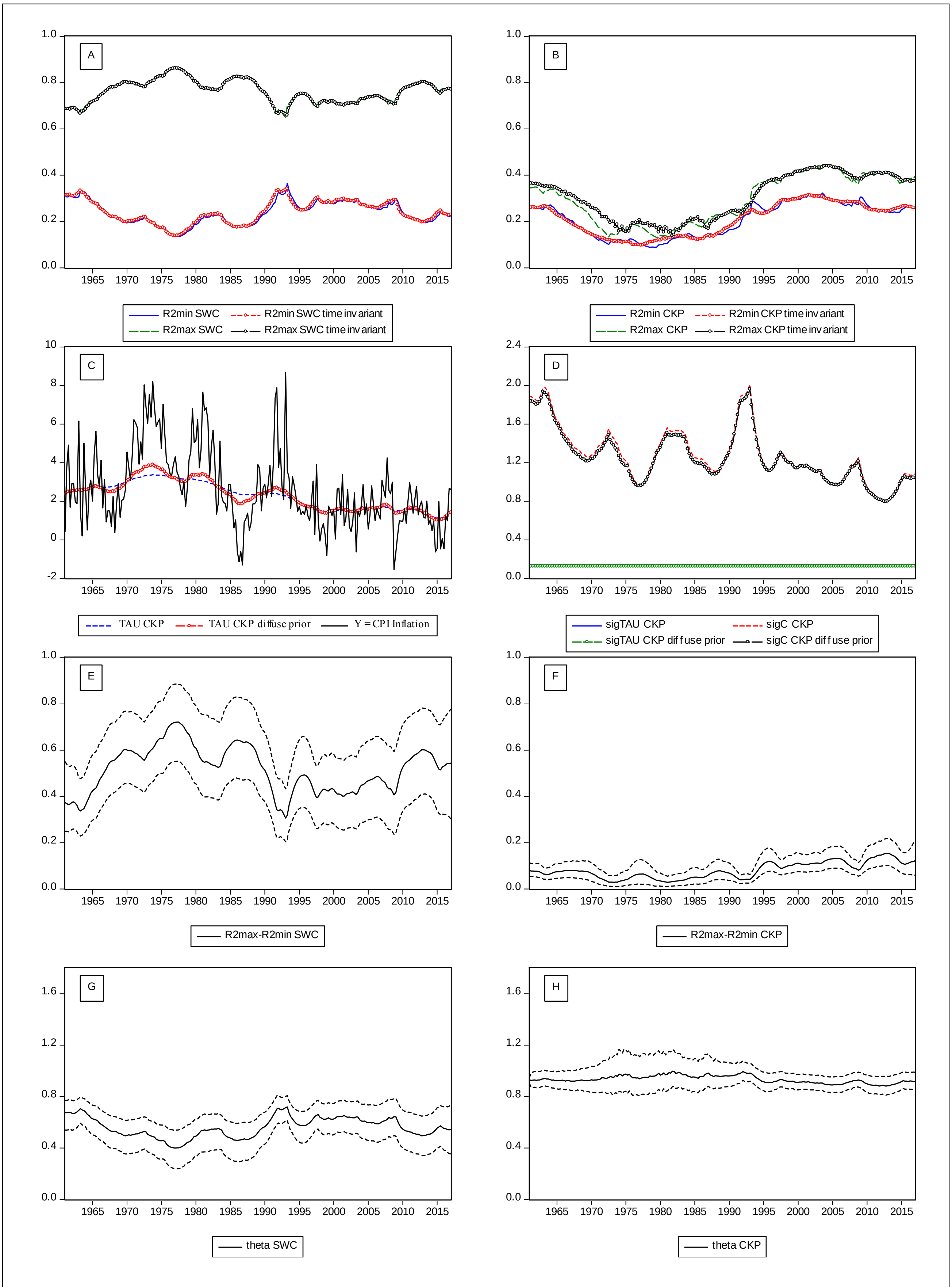

Figure M.4: Germany. See note to Figure M.1

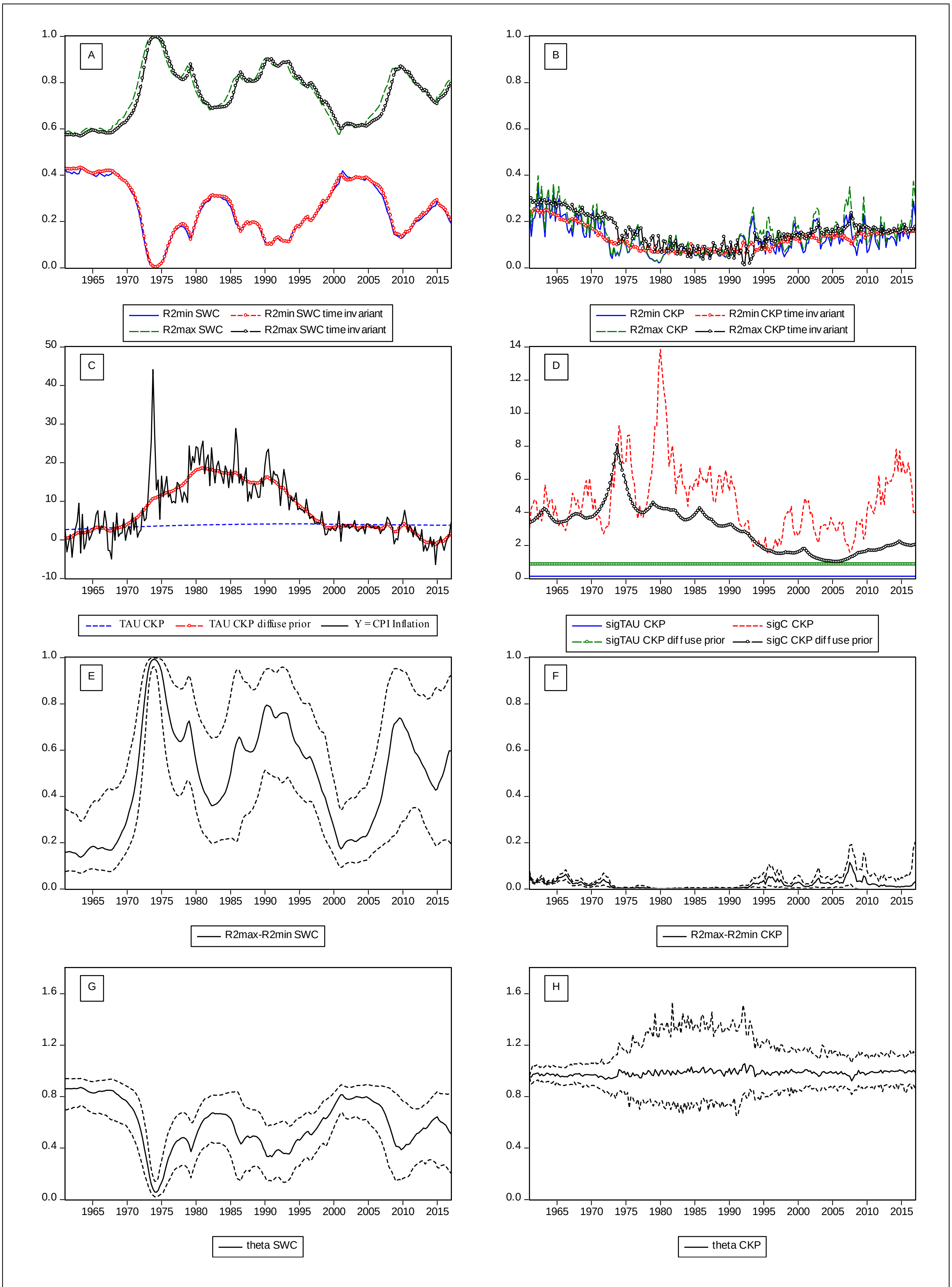

Figure M.5: Greece. See note to Figure M.1

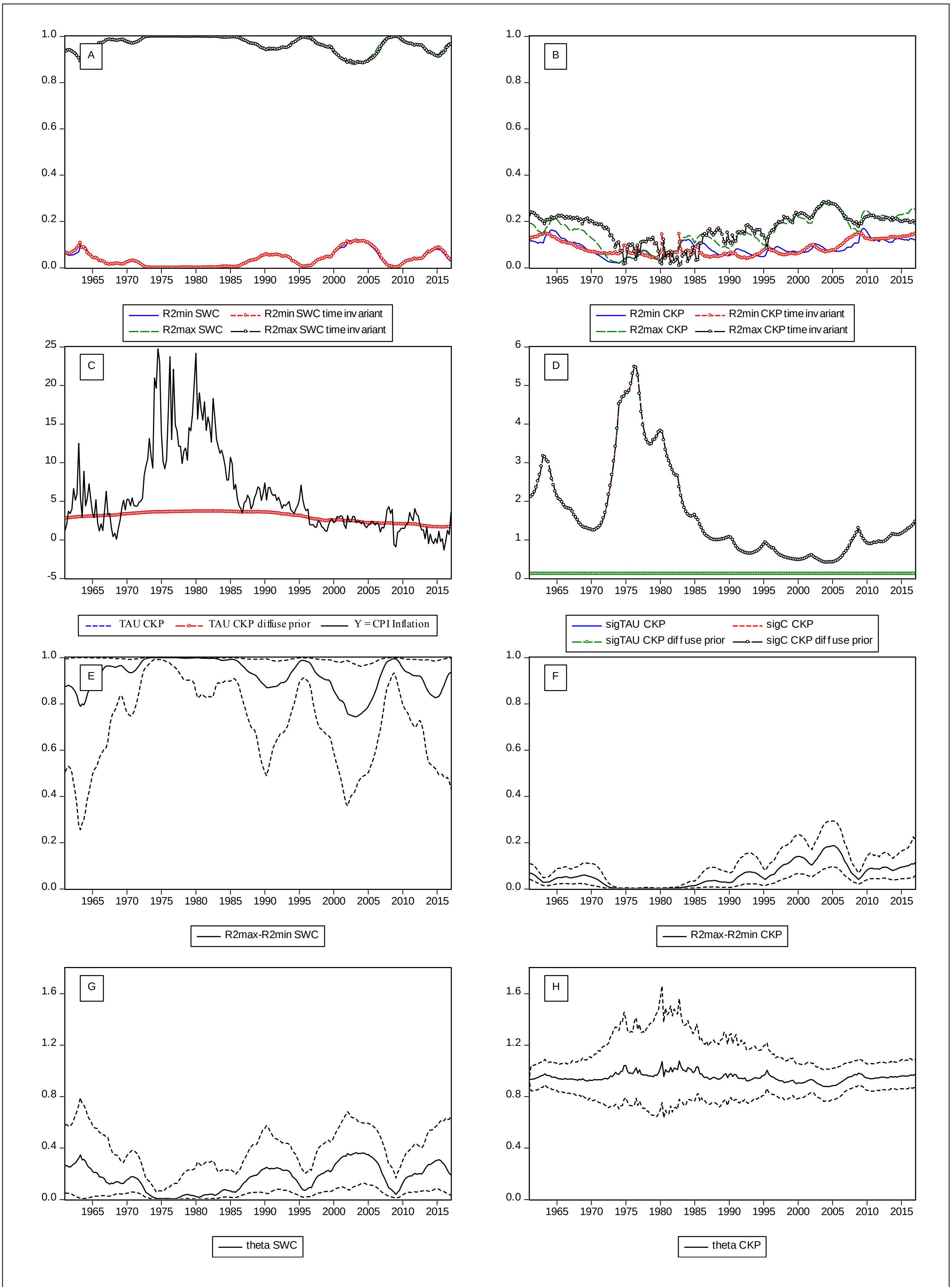

Figure M.6: Italy. See note to Figure M.1

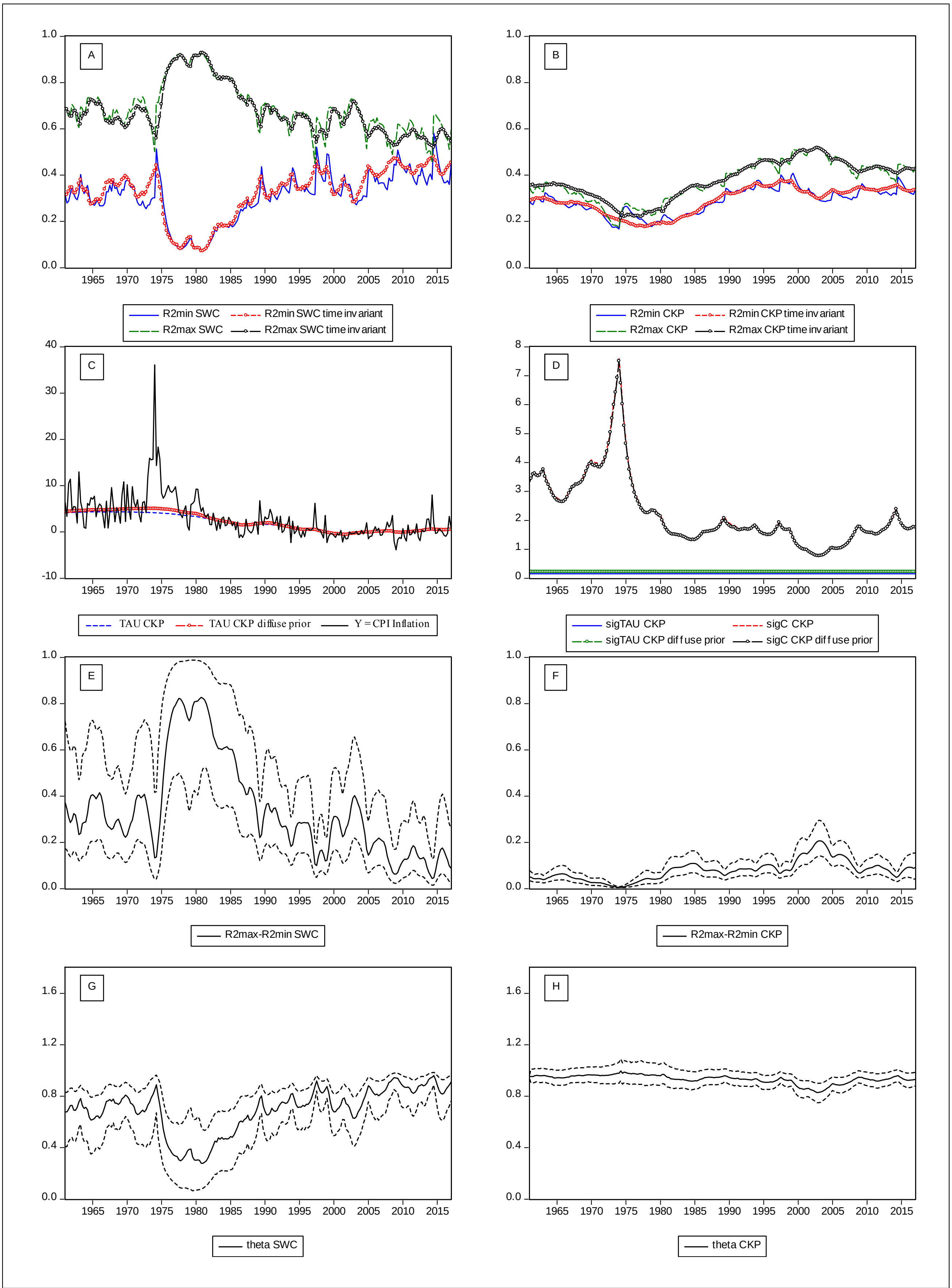

Figure M.7: Japan. See note to Figure M.1

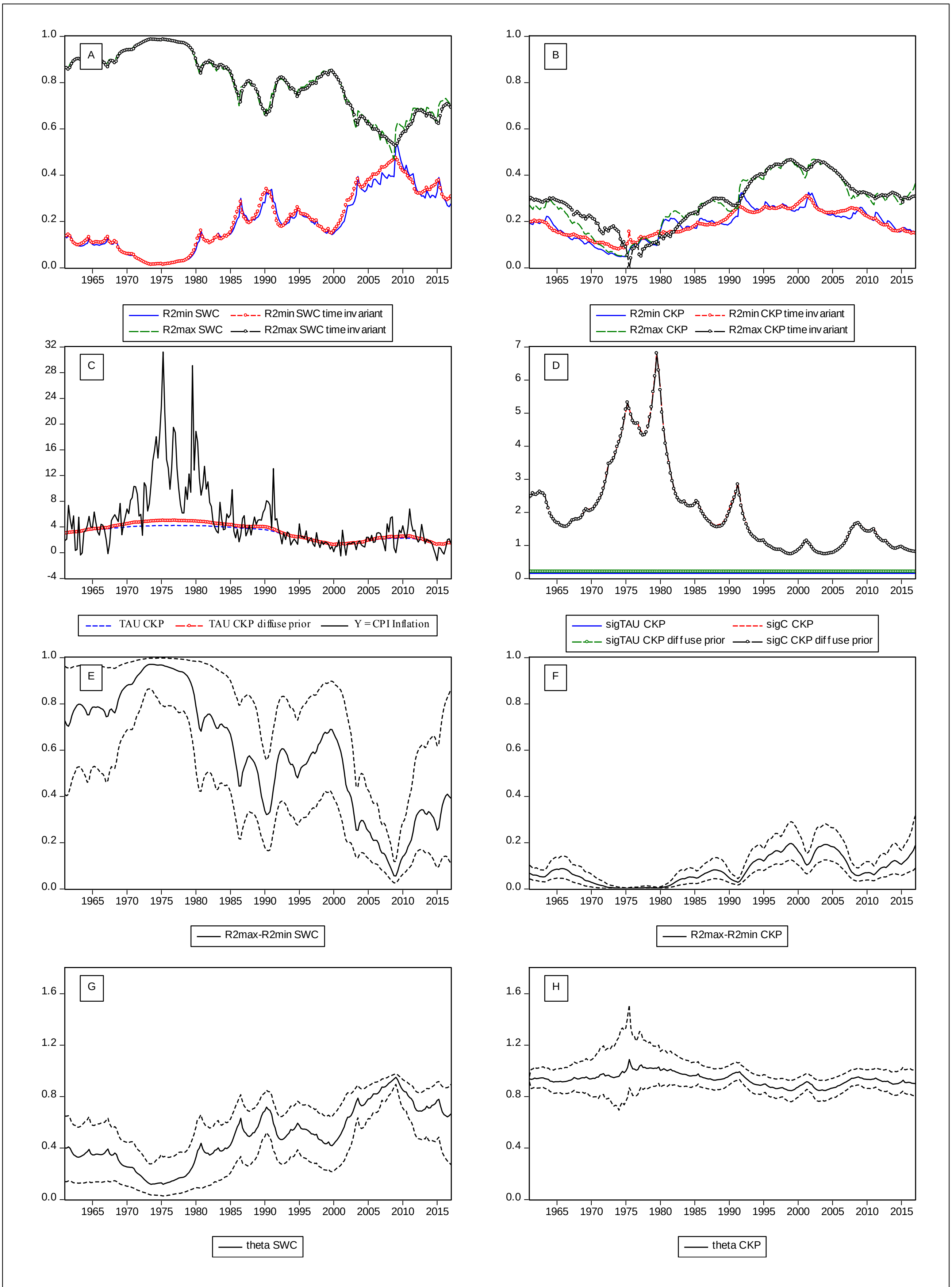

Figure M.8: UK. See note to Figure M.1

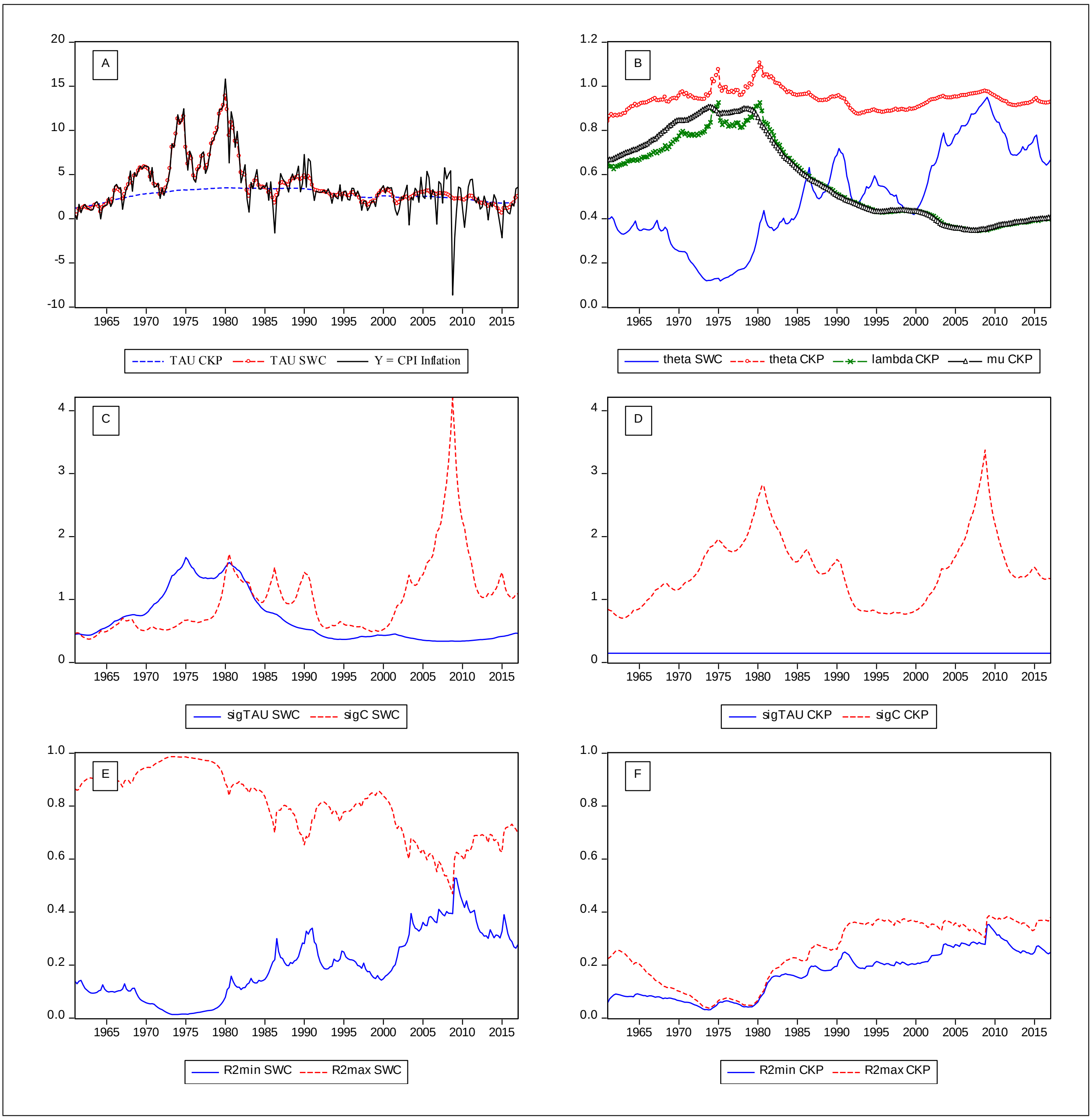

Figure M.9: US: using X12 to seasonally adjust CPI inflation. See note to Figure L.1

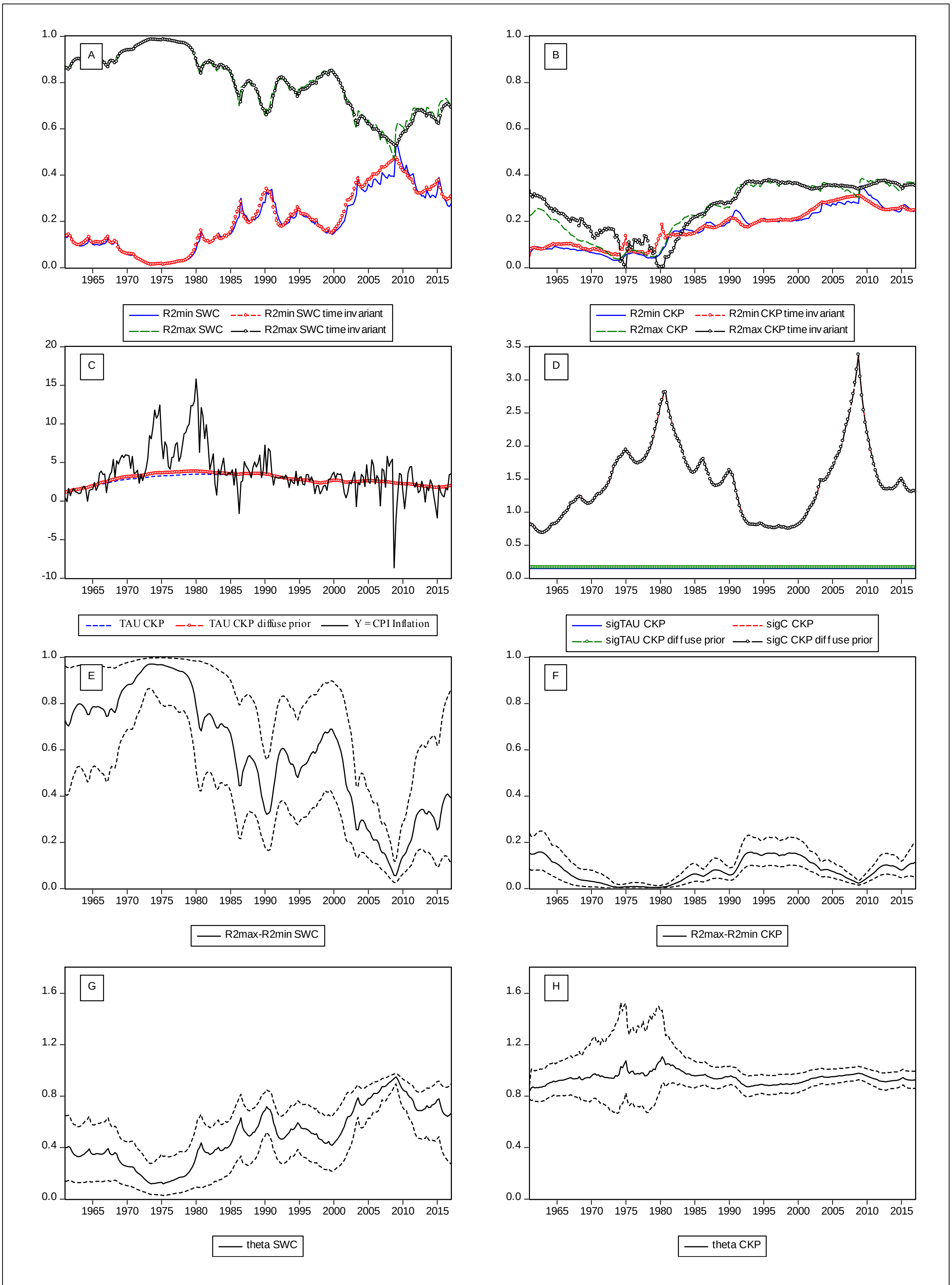

Figure M.10: US: using X12 to seasonally adjust CPI inflation (cont.). See note to Figure M.1

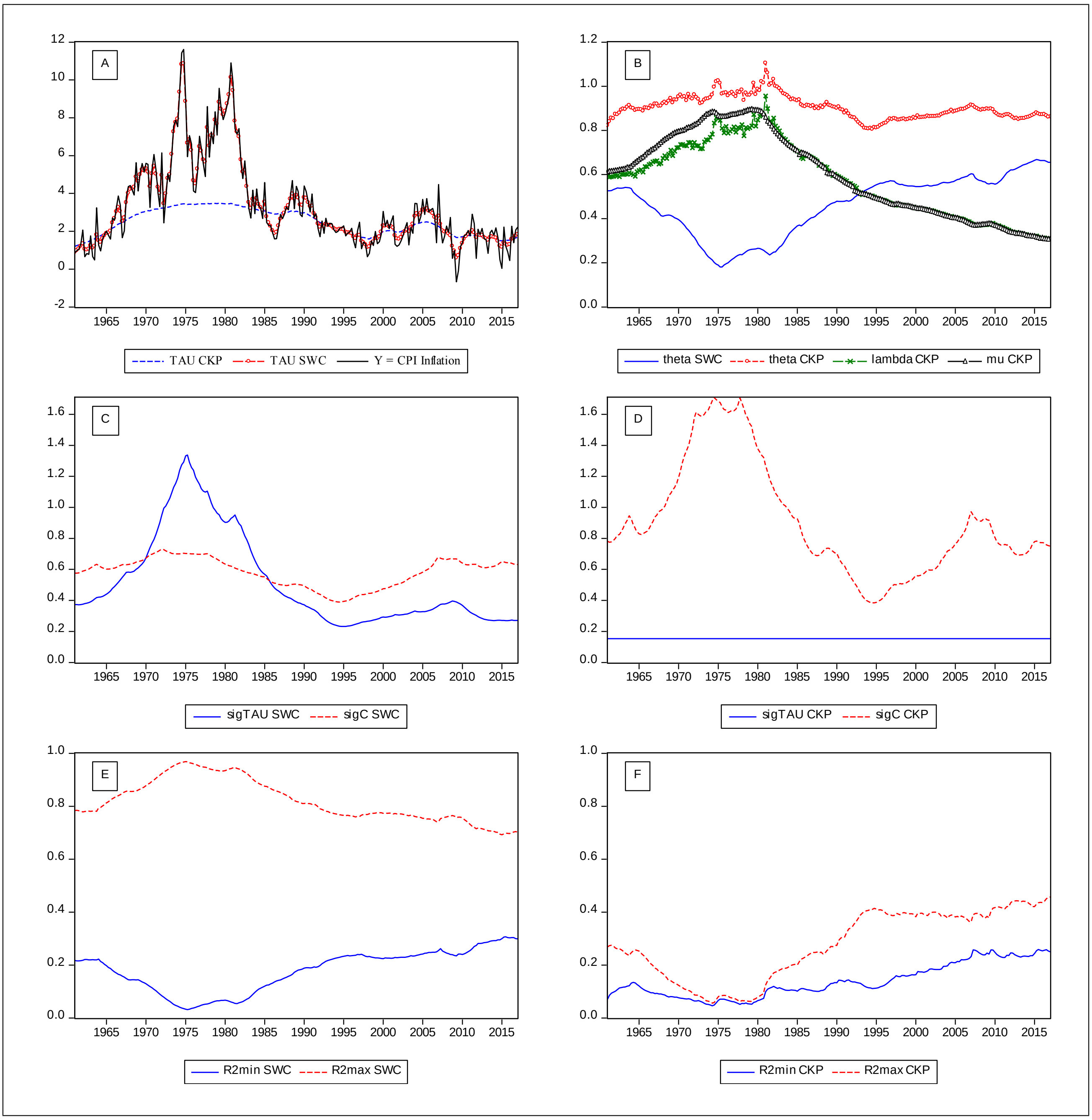

Figure M.11: US: GDP deflator inflation. See note to Figure L.1

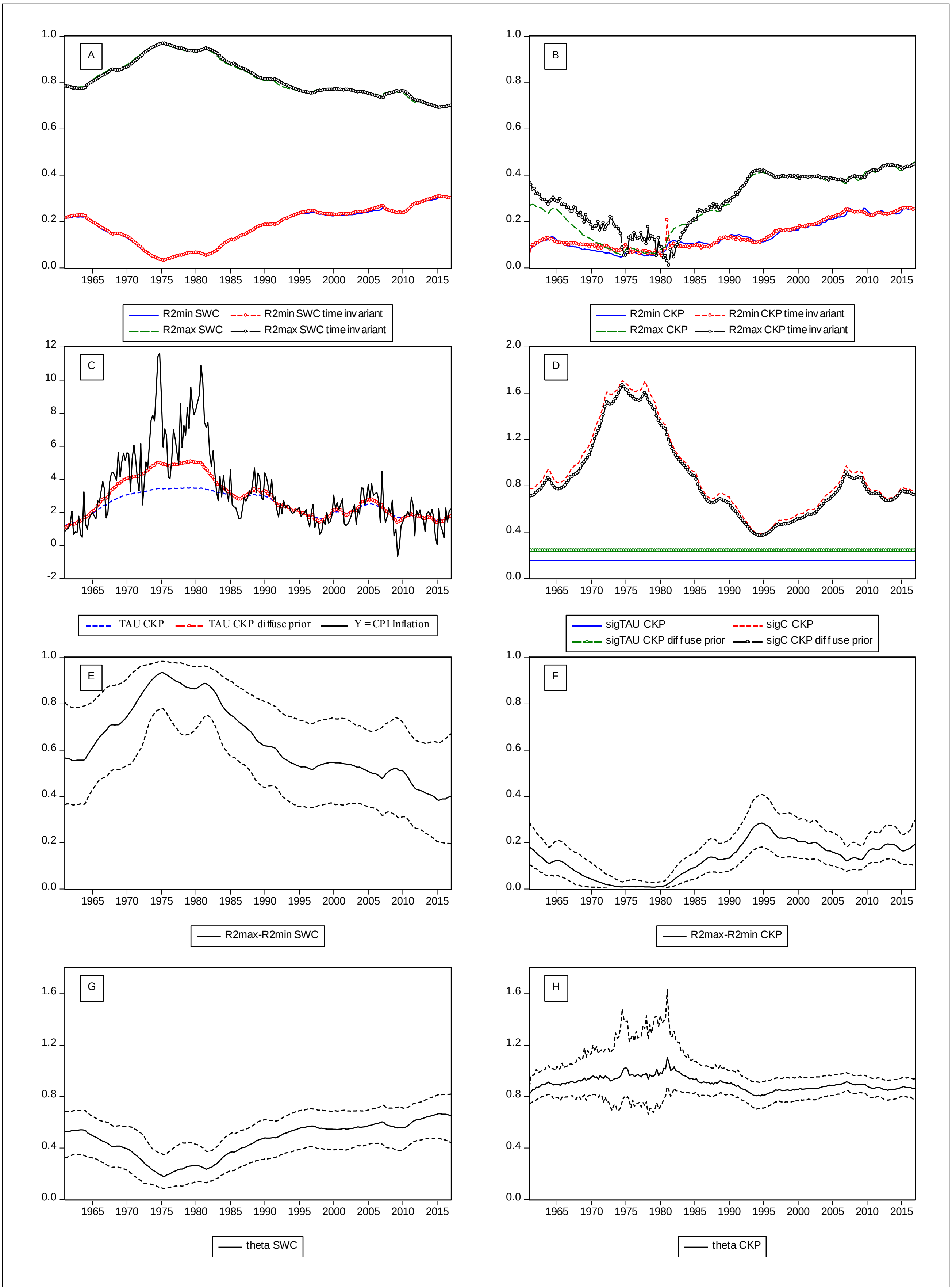

Figure M.12: US: GDP deflator inflation (cont.). See notes to Figure M.1

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.