ISSN 1745-8587

Birkbeck Centre for Applied Macroeconomics

BCAM 1405

Higher moments of MSVARs and the business cycle

Alexander Karalis Isaac
Birkbeck, University of London
University of Warwick

October 2014

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Higher moments of MSVARs and the business cycle

Alexander Karalis Isaac*

October 2014

Abstract

I derive the first four moments of the Markov-switching VAR and use the results to reconsider the conflict between the Great Moderation and Financial Crisis literatures. In contrast to the linear model, a three-regime Markov-switching model captures the skewness and kurtosis of US GDP growth 1954-2011. However, a specification with four regimes splits the sample in 1984, a result familiar from the Great Moderation literature. The higher moments of the MSVAR, not previously studied in the literature, reveal the Great Moderation to be a trade off between variance and kurtosis. U.S. GDP growth shifts from an almost Gaussian structure 1954-84 into a pattern with low variance, negative skewness and high kurtosis. The Markov-switching model which splits the sample accurately captures the new moment structure.

1 Introduction

Since their introduction in Hamilton [1989] Markov-switching models have proved a popular method for studying parameter instability in macroeconomics. Hamilton considered the classical business cycle, Sims and Zha [2006] used a multivariate Bayesian extension to measure the contribution of policy regimes and variance regimes to uncertainty in the US macroeconomy. Recently, Hubrich and Tetlow [2012] have used Markov-switching methods to consider the transmission of financial crises while Svensson and Williams [2007] and Farmer et al. [2011] have pioneered the extension of Markov switching techniques to forward looking and micro-founded-DSGE models. The work most closely related to this is Bianchi [2013]. Bianchi derives the first two moments of the MS-VAR and systematically analyses the evolution of expectations and uncertainty in an economy with regime changes. This can

*PhD candidate, Birkbeck College and Teaching Fellow, University of Warwick. Email: a.karalis-isaac@warwick.ac.uk

have important consequences for forecasts, variance decompositions and welfare calculations. Once we accept a non-Gaussian world, variance does not completely characterise uncertainty and I show below that the Markov-switching approach can also have important consequences for our understanding of the higher moments of macroeconomic data.

One attractive feature of Markov-switching models in the light of the financial crisis is that they allow the researcher to model non-Gaussian features of the data in a simple way. Despite this, little attention has been paid to the higher moments of the MS-VAR. This paper fills a gap in the literature by deriving explicit solutions for the third and fourth moments of the MS-VAR. This will allow researchers to consider information in the higher moments when estimating and evaluating competing macroeconomic models. The results should be of use to those working in both the applied and theoretical literatures.[1] The remainder of this introduction considers a simple application: can consideration of the higher moments, through a Markov switching approach, improve our understanding of the change in the behaviour of U.S. GDP growth suggested by the 'Great Moderation' literature but called into question by the financial crisis?

The large empirical and theoretical literature on the Great Moderation finds the period since the mid-1980s to be one of increased stability, as measured by the lower variance of macroeconomic aggregates such as GDP, consumption and inflation, see for example McConnell and Perez-Quiros [2000], Kim and Nelson [1999] and Stock and Watson [2003]. Models with financial frictions, however, can suggest that more intense financial management will lead to more extreme uncertainty. For example, in the theoretical framework proposed by Brunnermeier and Sannikov [2011] the financial sector provides diversification for firms and households which, facing lower variation in returns, increase their leverage. This results in an economy with more extreme behaviour - more time is spent in the very good and very bad parts of the state space, with little time in transition between the two.

Despite the growing criticism of the Great Moderation literature in the wake of the financial crisis, a reduction in the variance of many macroeconomic variables in the mid 1980s remains a distinct possibility. The gap between the moderation literature and the financial crisis literature may be due to the limited focus of the moderation literature on the first two moments of the variables considered. This focus on two moments is evident also in the recent literature concerned with the 'end' of the Great Moderation, for example Broer and Kero [2011].

This paper considers the first four moments of U.S. GDP growth. Measured on the whole sample the empirical standardised moments reveal a mild non-Gaussianity with slight negative skew and excess kurtosis. Linear autoregressions necessarily ignore information in the higher moments. However, a

[1] In theoretical models the MSV solution to the DSGE is a Markov-switching VAR, Farmer et al. [2011].

Markov-switching model with one lag and three regimes matches the first four standardised moments well. This is discussed in Section 3.1. The one low-growth regime in this estimation recurs throughout the sample period. Due to this recurrence, the ergodic probability is shared across all regimes and the unconditional moments reflect the empirical moments estimated on the whole sample.

Closer inspection of the smoothed regime probabilities reveals that a high-growth, high-variance regime is present in the period 1954-1984 and is then replaced by a high-growth, low-variance regime. This suggests a one off change in the behaviour of U.S. GDP growth which the three-regime model is not fully able to capture. A more complex model, with four regimes, is presented in Section 3.2. With four regimes the model breaks the sample into two periods: a pair of regimes alternate on the 1954-1984 period and a further pair alternate on the 1984-2011 period. As a result the ergodic probability is assigned only to the regimes holding in the second half of the sample. The implied unconditional moments are significantly less Gaussian, with a skewness of -1.4 and a kurtosis of 6.9. When compared to the sample moments estimated on the same period, skewness -1.5 and kurtosis 8.2, it is clear the four-regime Markov-switching model more accurately reflects the recent U.S. business cycle than either the linear model or the three-regime estimation.

Modelling the higher moments closes the gap between the Great Moderation literature and the Financial Crisis literature. Since the financial liberalisation of the early 1980s variance has declined but the risk of tail events has increased. Rather than pronouncing the end of the Great Moderation, I find that low-variance and high-kurtosis characterise the post 1984 business cycle; the estimations do not try to return to the business cycle of the pre-1984 era. This is an important difference. We should not try to characterise the financial crisis as announcing the return of the high-variance, near Gaussian structure which characterised the 1954-84 economy. Rather policy makers, households and firms should try to better understand how to make decisions in a low-variance, high-kurtosis world. One way to read these results is therefore as supporting evidence for the wider Markov-switching research agenda which is able to accommodate such a moment structure.

The Markov-switching approach in the paper is similar to that pioneered by Hamilton [1989], though the autoregressions here are written in terms of intercepts rather than means. This is in line with much of the recent literature on Markov-switching in macroeconomics. As suggested in the title of Hamilton's classic paper, such a modelling approach is concerned with business cycles, in the traditional sense of expansions and contractions in economic activity. Hamilton's paper considered U.S. GNP growth and its association with the NBER definition of the business cycle. Similarly the regimes in GDP growth extracted in this paper often have an association with the NBER cycle dates. This is different from, but no less important than, the levels concept of the business cycle as a deviation from a long-run trend which dominates the DSGE literature. The negative skewness and high kurtosis of the post-1984 period suggest the modern business cycle may be better characterised by the asymmetry of its large deviations than by alternation between periods of positive and negative growth.

Before presenting the estimation results summarised in this introduction, it is necessary to derive expressions for the first four unconditional moments of the Markov-switching VAR, in intercept switching form. This is the object of Section 2 which gives results for the third and fourth moments of the MSI(N)-VAR(p) in r variables, assuming Mean Square Stability. The first and second moments are similar to the Markov Jump Linear System considered by Costa et al. [2005] and are treated in Appendix A. The unconditional first and second moments are also available in Bianchi [2013], where the main focus is on the conditional moments. Although not the focus of this paper, the conditional higher moments follow straightforwardly from the recursions employed here to find expressions for the limiting, unconditional, moments. Appendix B studies the special case of the MSI(M)-AR(1) and derives results under the stationarity assumption employed in Francq and Zakoïan [2001] who consider the first two moments of a general MS-VARMA system. This is quite a costly restriction in the context of Markov-switching models and would rule out the kind of degenerate distribution over regimes reported by the four-regime estimation.

2 Moments of the MSI(M)-VAR(p)

This section derives the unconditional moments, up to order four, of the Markov-switching VAR in which the intercept, autoregressive parameters and variance matrix are all regime dependent. There may be r variables, p lags and N regimes in the model. The stability concept employed is 'mean square stability', which requires the system to be ergodic, but not necessarily stationary.

The MS-VAR is given by

$$y(t+1) = \Gamma_{\theta(t+1)} y(t) + \psi_{\theta(t+1)} + G_{\theta(t+1)} w(t+1) \tag{1}$$

where $y(0)$ and $\theta(0)$ are given. $\theta(t)$ and $w(l)$ are independent for all t, l and $w(0)$ is independent of $y(0)$ and $\theta(0)$, so that $y(t)$, $w(l)$ are independent for $l > t$. The innovation w is assumed multivariate standard normal. y is an $n.1$ vector; if the original model has p lags and r variables, then $n = rp$ and (1) is the 'companion form' of the VAR.

The timing convention in econometrics changes the first and second moment operators slightly, relative to the benchmark model in Costa et al. [2005]. Costa et. al.'s matrix $\mathcal{B}$ for the first moment is replaced by the matrix $\mathcal{M}_1$, defined in Appendix A. The second moment operator is thier $\mathcal{V}$ rather than their $\mathcal{T}$.

2.1 The Fourth Moment of the MSVAR

In this section the fourth moment of the Markov-Switching model (1) is studied. The first sub-section considers the homogeneous version of the system, and establishes an operator $\mathcal{L}(.)$ and a matrix $\mathcal{M}_4$ that play a role equivalent to $\mathcal{V}(.)$ and $\mathcal{M}_2$ in the case of the second moment. Dealing with the non-homogenous version of the system in Section 2.1.2 introduces further complication as various third-moment terms are required.

2.1.1 The Homogenous System

The fourth moment of the vector process y can be defined as $\mathbb{E}[yy' \otimes yy']$, see for example, Schott [2005] or Magnus and Neudecker [1988]. First, consider the system

$$y(t+1) = \Gamma_{\theta(t+1)} y(t) \tag{2}$$

Define the fourth moment of the joint process $(y(t), \theta(t))$, and the fourth moment of $y(t)$ itself

$$\begin{aligned} H_i(t) &= \mathbb{E}[y(t)y(t)' \otimes y(t)y(t)' \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\ H(t) &= \begin{bmatrix} H_1(t) & \dots & H_N(t) \end{bmatrix} \\ \mathbb{H}(t) &= \mathbb{E}[y(t)y(t)' \otimes y(t)y(t)'] \end{aligned}$$

Now we find a recursion for the fourth moment of the joint process $(y(t), \theta(t))$.

$$
\begin{aligned}
H_j(t+1) &= \mathbb{E}[y(t+1)y(t+1)' \otimes y(t+1)y(t+1)'\, 1_{\{\theta(t+1)=j\}}] \\
&= \sum_{i=1}^{N} \mathbb{E}[\Big(\Gamma_{\theta(t+1)}y(t)y(t)'\Gamma'_{\theta(t+1)}\Big) \otimes \Big(\Gamma_{\theta(t+1)}y(t)y(t)'\Gamma'_{\theta(t+1)}\Big)\, 1_{\{\theta(t+1)=j\}}\, 1_{\{\theta(t)=i\}}] \\
&= \sum_{i=1}^{N} \mathbb{E}[\Big(\Gamma_j y(t)y(t)'\Gamma'_j\Big) \otimes \Big(\Gamma_j y(t)y(t)'\Gamma'_j\Big)\, 1_{\{\theta(t)=i\}}]p_{ij} \\
&= \sum_{i=1}^{N} \Big(\Gamma_j \otimes \Gamma_j\Big)\mathbb{E}[\Big(y(t)y(t)' \otimes y(t)y(t)'\Big)\, 1_{\{\theta(t)=i\}}]\Big(\Gamma'_j \otimes \Gamma'_j\Big)p_{ij} \\
&= \sum_{i=1}^{N} p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)H_i(t)\big(\Gamma_j \otimes \Gamma_j\big)' \qquad (3) \\
&= \mathcal{L}_j(H(t))
\end{aligned}
$$

The third line of (3) used the fact that $\Gamma_{\theta(t+1)}$ is non-random given $\theta(t+1) = j$; the fourth line used the property of the Kronecker product that

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

while the final line defines the fourth moment operator $\mathcal{L}_j(.)$. As usual, let $\mathcal{L}(.) = [\mathcal{L}_1(.), \dots, \mathcal{L}_N(.)]$. In order to study the spectral radius of $\mathcal{L}(.)$, consider the vectorised recursion below, and recall $\varphi(ABC) = (C' \otimes A)\varphi(B)$.

$$
\begin{aligned}
\varphi\big(H(t+1)\big) &= \begin{bmatrix} \varphi\Big(\sum_i p_{i1}\big(\Gamma_1 \otimes \Gamma_1\big)H_i(t)\big(\Gamma_1 \otimes \Gamma_1\big)'\Big) \\ \vdots \\ \varphi\Big(\sum_i p_{iN}\big(\Gamma_N \otimes \Gamma_N\big)H_i(t)\big(\Gamma_N \otimes \Gamma_N\big)'\Big) \end{bmatrix} \\
&= \begin{bmatrix} \sum_i p_{i1}\Big(\big(\Gamma_1 \otimes \Gamma_1\big) \otimes \big(\Gamma_1 \otimes \Gamma_1\big)\Big)\varphi\big(H_i(t)\big) \\ \vdots \\ \sum_i p_{iN}\Big(\big(\Gamma_N \otimes \Gamma_N\big) \otimes \big(\Gamma_N \otimes \Gamma_N\big)\Big)\varphi\big(H_i(t)\big) \end{bmatrix} \qquad (4)
\end{aligned}
$$

Notice that each row in (4) can be written as an inner product

$$
\Big[p_{1j}(\Gamma_j \otimes \Gamma_j) \otimes (\Gamma_j \otimes \Gamma_j) \quad \dots \quad p_{Nj}(\Gamma_j \otimes \Gamma_j) \otimes (\Gamma_j \otimes \Gamma_j)\Big] \begin{bmatrix} \varphi\big(H_1(t)\big) \\ \vdots \\ \varphi\big(H_N(t)\big) \end{bmatrix}
$$

Let $\bar{\Gamma}_i = (\Gamma_i \otimes \Gamma_i)$, then we can write (4) as $\varphi(H(t+1)) = \mathcal{M}_4 \varphi(H(t))$, where $\mathcal{M}_4$ is given by

$$\begin{aligned}\mathcal{M}_4 &= \begin{bmatrix} p_{11}\left(\bar{\Gamma}_1 \otimes \bar{\Gamma}_1\right) & \dots & p_{N1}\left(\bar{\Gamma}_1 \otimes \bar{\Gamma}_1\right) \\ \vdots & \ddots & \vdots \\ p_{1N}\left(\bar{\Gamma}_N \otimes \bar{\Gamma}_N\right) & \dots & p_{NN}\left(\bar{\Gamma}_N \otimes \bar{\Gamma}_N\right) \end{bmatrix} \\ &= \operatorname{diag}\left(\bar{\Gamma}_i \otimes \bar{\Gamma}_i\right)\left(P' \otimes I_{n^4}\right) \end{aligned} \tag{5}$$

If the largest eigenvalue of $\mathcal{M}_4 < 1$ then the fourth moment of (2) will converge to zero.

2.1.2 The Non-Homogeneous System

We now return to (1) as the system equation. In much of the following, notation concerning the specific regime and the time period is supressed but the meaning should be clear from the previous derivations. As with the second moment, the fourth can be written as the sum of the homogeneous term and the non-homogeneous terms, say $S_j(t)$.

$$H_j(t+1) = \mathcal{L}_j(H(t)) + S_j(t) \tag{6}$$

If (1) is Mean Square Stable, that is if the vector of ergodic probabilities exists, $\pi = P'\pi$, and $\mathcal{M}_2$ is a stable matrix, first and second moment terms in the non-homogenous part, $S_j(t)$, converge to long run values, Q_i, q_i, see Appendix A for definitions. If we further have that all third moment matrices $\mathcal{M}_{i3}$ discussed in Section 2.1.3 are stable, then $S_j(t)$ will itself converge to a long-run value S_j. In such a case we will be able to solve a recursion based on (6) for the unconditional fourth moment of (1).

Substituting (1) into the recursion for the fourth moment we find (7)

$$\begin{aligned} H_j(t+1) =& \mathbb{E}[\Big(\left(\Gamma y + \psi + Gw\right)\left(\Gamma y + \psi + Gw\right)'\Big) \otimes \Big(\left(\Gamma y + \psi + Gw\right)\left(\Gamma y + \psi + Gw\right)'\Big) 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\ =& \mathbb{E}[\Big(\Gamma yy'\Gamma' + \Gamma y\psi' + \Gamma yw'G' + \psi y'\Gamma' + \psi\psi' + \psi w'G' + Gwy'\Gamma' + Gw\psi' + GwwG'\Big) \\ & \otimes \Big(\Gamma yy'\Gamma' + \Gamma y\psi' + \Gamma yw'G' + \psi y'\Gamma' + \psi\psi' + \psi w'G' + Gwy'\Gamma' + Gw\psi' + Gww'G'\Big) 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \end{aligned} \tag{7}$$

The first term $\mathbb{E}[\Gamma yy'\Gamma \otimes \Gamma yy'\Gamma\, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}]$ is the homogeneous term $\mathcal{L}_j(H(t))$, and the remaining terms define $S_j(t)$. Considering each term in this product individually, the problem of finding the fourth moment of the vector $y(t)$ is reduced to finding second moment terms for the random matrices in (7). The theorems in Section 4 of Ghazal and Neudecker [2000] apply; theorems (4.3) and (4.4) are particularly useful and are stated below without the proof, which appears in the original article.

Theorems on second moments of random matrices**,** Ghazal and Neudecker

$$(4.3) \qquad \mathbb{E}[X \otimes X] = \sum_{ij}(E'_{ij} \otimes I_n)\Omega_{XX}(I_n \otimes E'_{ij}) + M \otimes M$$

$$(4.4) \qquad \mathbb{E}[X' \otimes X] = \sum_{ij}(E'_{ij} \otimes I_n)\Omega_{XX}(E'_{ij} \otimes I_n)K_{nn} + M' \otimes M$$

where $\mathbb{E}[X] = M$, E_{ij} are the $n \times n$ basis matrices and K_{nn} is the $n^2 \times n^2$ commutation matrix. To apply the theorems note that $\Omega_{XX} = var(x)$ where $x = \varphi(X)$, see Magnus and Neudecker [1979].

Conditionally on the regime $\theta(t+1) = j$, most non-zero terms in (7) contain only one random element, y or w. For such terms it is quite easy to derive the expectation without specific reference to basis matrices and the commutation matrix, as we are not dealing with *products* of random matrices; rather there will be a random matrix and various fixed multiplying factors. In cases where $y(t)$ and $w(t+1)$ enter only on opposite sides of the Kronecker product, independence of $w(t+1)$ and $y(t)$ allows us to take expectations on each side of the product so these terms are again straightforward. However, terms where $y(t)$ and $w(t+1)$ both enter twice, once on each side of a Kronecker product, require more care. Examples are given by $\mathbb{E}[(\Gamma yw'G') \otimes (\Gamma yw'G')]$ and $\mathbb{E}[(\Gamma yw'G') \otimes (Gwy'\Gamma')]$.

Consider the term $\mathbb{E}[(\Gamma yw'G') \otimes (\Gamma yw'G')]$. Put $X = yw'$, giving $\mathbb{E}[X] = \mathbb{E}[y] \cdot \mathbb{E}[w'] = \mathbf{0}_{n.n}$ by the independence of $y(t)$ and $w(t+1)$. Therefore $\mathrm{var}(x) = \mathbb{E}[xx'] = \mathbb{E}[(w \otimes y)(w \otimes y)'] = \mathbb{E}[w \otimes w' \otimes y \otimes y'] = I_{n^2} \otimes Q_i(t)$. This used the property of the Kronecker product that, for a, b both vecotors

$$ab' = a \otimes b' = b' \otimes a, \qquad \varphi(ab') = b \otimes a$$

Then we have

$$\begin{aligned}\mathbb{E}[\big(\Gamma yw'G'\big) \otimes \big(\Gamma yw'G'\big)] &= \big(\Gamma \otimes \Gamma\big)\mathbb{E}[X \otimes X]\big(G' \otimes G'\big) \\ &= \big(\Gamma \otimes \Gamma\big)\sum_{kl}(E'_{kl} \otimes I_n)\Omega_{XX}(I_n \otimes E'_{kl})\big(G' \otimes G'\big)\end{aligned}$$

with $\Omega_{XX} = I_{n^2} \otimes Q_i(t)$. For the second term, $\mathbb{E}[(\Gamma yw'G') \otimes (Gwy'\Gamma')]$, put $X = wy'$ so that

$\Omega_{XX} = Q_i(t) \otimes I_{n^2}$. Then

$$\begin{aligned}\mathbb{E}[\big(\Gamma y w' G'\big) \otimes \big(G w y' \Gamma'\big)] &= \big(\Gamma \otimes G\big)\mathbb{E}[X' \otimes X]\big(G' \otimes \Gamma'\big) \\ &= \big(\Gamma \otimes G\big) \sum_{kl} (E'_{kl} \otimes I_n)\Omega_{XX}(E'_{kl} \otimes I_n) K_{nn}\big(G' \otimes \Gamma'\big)\end{aligned}$$

The terms in S will depend on all lower moments, including the third. To see consider the second term in (7)

$$\begin{aligned}&\mathbb{E}[(\Gamma y y' \Gamma') \otimes (\Gamma y \psi')\, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\ &\qquad = \sum_i \mathbb{E}[(\Gamma y y' \Gamma') \otimes (\Gamma y \psi')\, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\ &\qquad = \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\mathbb{E}[y(t)y(t)' \otimes y(t)\psi'_j\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]\big(\Gamma'_j \otimes I_n\big) \\ &\qquad = \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)F_i(t, \psi_j, 4)\big(\Gamma'_j \otimes I_n\big)\end{aligned}$$

The central term is a third moment term $\mathbb{E}[yy' \otimes yc']$. The notation $F_i(t, c, l)$ is adopted to describe third moment terms, with c standing for the appropriate constant and $l \in \{1, 2, 3, 4\}$ describing the position of the constant in terms of the four vectors. In this notation $\mathbb{E}[y(t)y(t)' \otimes y(t)\psi'_j\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] = F_i(t, \psi_j, 4)$, giving the final line above. This third-moment term is discussed in detail in Section 2.1.3.

For the final term in $S_j(t)$, $\mathbb{E}[Gww'G' \otimes Gww'G']$ define $\Omega_{WW} = \mathbb{E}[ww' \otimes ww']$. Recall, $w(t) = [\epsilon(t)', 0_{1,n(p-1)}]'$ and $\epsilon \sim N(0_{r.1}, I_r)$, are the shocks to the r variables in the VAR. This gives $\mathbb{E}[\epsilon\epsilon' \otimes \epsilon\epsilon'] = I_{r^2} + K_{rr} + \varphi(I_r)\{\varphi(I_r)\}' := \Omega_{\epsilon\epsilon}$, see for example Schott, Theorem 10.19. Then

$$\Omega_{WW} = \begin{bmatrix} \Omega_{\epsilon\epsilon} & 0_{r,r^2(p^2-1)} \\ 0_{r^2(p^2-1),r} & 0_{r^2(p^2-1),r^2(p^2-1)} \end{bmatrix}$$

Provided all lower moments converge, $S_j(t)$ will converge to a matrix S_j, in which all probabilities $\pi_i(t)$ are replaced by their long-run values π_i and all time-dependent lower moments $q_i(t)$ etc. are replaced by their ergodic values, q_i. Definitions of these long-run values are given in Appendix A. Further provided that $\mathcal{M}_4$ is a stable matrix, the fourth moment of the intercept-switching MSVAR will be well defined. Section 2.1.4 lists the non-zero terms in (7), while the section below gives the intermediate third moment terms that this will require.

2.1.3 Calculating the third moment terms, $F_i(t,c,l)$

Consider a third moment matrix

$$\mathbb{E}[y(t)y(t)' \otimes y(t)c' \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] := F_i(t,c',4) \in \mathbb{R}^{n^2,n^2}$$

To start the recursion, as usual put

$$\begin{aligned}
F_j(t+1,c',4) =& \mathbb{E}[y(t+1)y(t+1)' \otimes y(t+1)c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\
=& \sum_{i=1}^{N} \mathbb{E}[y(t+1)y(t+1)' \otimes y(t+1)c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}} \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
=& \sum_{i=1}^{N} \mathbb{E}[\Big(\big(\Gamma_{\theta(t+1)}y(t) + \psi_{\theta(t+1)} + G_{\theta(t+1)}w(t+1)\big)\big(\Gamma_{\theta(t+1)}y(t) + \psi_{\theta(t+1)} + G_{\theta(t+1)}w(t+1)\big)'\Big) \\
& \otimes \Big(\big(\Gamma_{\theta(t+1)}y(t) + \psi_{\theta(t+1)} + G_{\theta(t+1)}w(t+1)\big)c'\Big) \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}} \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
=& \sum_{i=1}^{N} \mathbb{E}[\big(\Gamma yy'\Gamma' + \Gamma y\psi' + \Gamma yw'G' + \psi y'\Gamma' + \psi\psi' + \psi w'G' + Gwy'\Gamma' + Gw\psi' + Gww'G'\big) \\
& \otimes \big(\Gamma yc' + \psi c' + Gwc'\big) \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}} \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \qquad (8)
\end{aligned}$$

There are 27 terms in (8). The following discusses a couple of the archetypal terms, then all non-zero terms in (8) are listed. Consider the first term in (8)

$$\begin{aligned}
\sum_i \mathbb{E}[\Gamma yy'\Gamma' \otimes \big(\Gamma yc'\big) \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}} \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] & \\
&= \sum_i \mathbb{E}[\big(\Gamma_j y(t)y(t)'\Gamma_j'\big) \otimes \big(\Gamma_j y(t)c'\big) \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]p_{ij} \\
&= \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\mathbb{E}[y(t)y(t)' \otimes y(t)c' \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]\big(\Gamma_j' \otimes I_n\big) \\
&= \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)F_i(t,c',4)\big(\Gamma_j' \otimes I_n\big)
\end{aligned}$$

This first term is the term associated with the homogenous version of the system, (2). For the homogenous case, define a 'third moment operator matrix' by noticing

$$
\begin{aligned}
\varphi\big(F_i(t+1)\big) &= \varphi\Big(\sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)F_i(t)\big(\Gamma_j' \otimes I_n\big)\Big) \\
&= \sum_i p_{ij}\Big(\big(\Gamma_j \otimes I_n\big) \otimes \big(\Gamma_j \otimes \Gamma_j\big)\Big)\varphi\big(F_i(t)\big)
\end{aligned}
$$

We can write the final line as a dot product:

$$
\Big[p_{1j}(\Gamma_j \otimes I_n) \otimes (\Gamma_j \otimes \Gamma_j) \quad \dots \quad p_{Nj}(\Gamma_j \otimes I_n) \otimes (\Gamma_j \otimes \Gamma_j)\Big]\begin{bmatrix}\varphi\big(F_1(t)\big) \\ \vdots \\ \varphi\big(F_N(t)\big)\end{bmatrix}
$$

Giving

$$
\begin{aligned}
\hat{\varphi}(F(t+1)) &= \Big\{\operatorname{diag}[\big(\Gamma_j \otimes I_n\big) \otimes \bar{\Gamma}_j][P' \otimes I_{n^4}]\Big\}\hat{\varphi}\big(F(t)\big) \\
&= \mathcal{M}_{31}\hat{\varphi}\big(F(t)\big)
\end{aligned}
$$

where we re-use the notation $\bar{\Gamma}_j = (\Gamma_j \otimes \Gamma_j)$. Taking the second term of (8) we have

$$
\begin{aligned}
&\sum_i \mathbb{E}[\big(\Gamma y y' \Gamma'\big) \otimes \big(\psi c'\big)\, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
&\qquad = \sum_i \mathbb{E}[\Gamma_j y(t) y(t)' \Gamma_j' \otimes \psi_j c'\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] p_{ij} \\
&\qquad = \sum_i \big(\Gamma_j \mathbb{E}[y(t)y(t)'\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]\Gamma_j' \otimes \psi_j c'\big) p_{ij} \\
&\qquad = \sum_i p_{ij}\big(\Gamma_j Q_i(t) \Gamma_j'\big) \otimes \psi_j c' \qquad (9)
\end{aligned}
$$

which demonstrates the dependence of the third moment terms on the second moment, $Q(t)$. Provided that the second moment converges, and the summation in (9) is also stable, this term will converge. The fourth term is also useful to study. Again the properties of the Kronecker product for a, b both vectors are employed. This gives

$$
\begin{aligned}
&\mathbb{E}[\Gamma y\psi' \otimes \Gamma yc' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\
&\qquad = \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\mathbb{E}[y(t)\psi_j' \otimes y(t)c' \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
&\qquad = \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\big(\psi_j' \otimes \mathbb{E}[y(t) \otimes y(t) \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \otimes c'\big) \\
&\qquad = \sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\Big(\psi_j' \otimes \varphi\big(Q_i(t)\big) \otimes c'\Big)
\end{aligned}
$$

Proceeding similarly, we find the following non-zero terms in (8):

$$
\begin{aligned}
\mathbb{E}[\Gamma y\psi \otimes \psi c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(\Gamma_j q_i(t)\psi_j') \otimes (\psi_j c') \\
\mathbb{E}[\Gamma y w' G' \otimes G w c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(\Gamma_j \otimes G_j)\big(q_i(t) \otimes I_n \otimes c'\big)(G_j' \otimes I_n) \\
\mathbb{E}[\psi y \Gamma' \otimes \Gamma y c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(I_n \otimes \Gamma_j)\big(\psi_j \otimes Q_i(t) \otimes c'\big)(\Gamma_j' \otimes I_n) \\
\mathbb{E}[\psi y' \Gamma' \otimes \psi c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(\psi_j q_i(t)'\Gamma_j') \otimes (\psi_j c') \\
\mathbb{E}[\psi \psi' \otimes \Gamma y c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(\psi_j \psi_j') \otimes (\Gamma_j q_i(t) c') \\
\mathbb{E}[\psi \psi' \otimes \psi c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}\pi_i(t)(\psi_j \psi_j' \otimes \psi_j c') \\
\mathbb{E}[\psi w' G' \otimes G w c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}\pi_i(t)(I_n \otimes G_j)\big(\psi_j \otimes I_n \otimes c'\big)(G_j' \otimes I_n) \\
\mathbb{E}[G w y' \Gamma' \otimes G w c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(G_j \otimes G_j)\big(q_i(t) \otimes \varphi(I_n) \otimes c'\big)(\Gamma_j' \otimes I_n) \\
\mathbb{E}[G w \psi' \otimes G w c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}\pi_i(t)(G_j \otimes G_j)\big(\psi_j \otimes \varphi(I_n)' \otimes c'\big) \\
\mathbb{E}[G w w' G' \otimes \Gamma y c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}(G_j \otimes \Gamma_j)\big(I_n \otimes q_i(t) c'\big)(G_j' \otimes I_n) \\
\mathbb{E}[G w w' G' \otimes \psi c' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] &= \sum_i p_{ij}\pi_i(t)\big(G_j G_j' \otimes \psi_j c'\big)
\end{aligned}
$$

Finally, note that other third moment terms, $\mathbb{E}[cy' \otimes yy']$ etc., can be defined analogously.

We are now ready to complete the definition of the fourth moment of the MSI(M)-VAR(p). The non-zero terms in the fourth moment recursion are listed in the following tables.

2.1.4 Terms in the 4^{th} moment $H_j(t+1)$

The 'Term' column refers to the number of the term in the expansion of (7), the column 'Number' refers to the page of the array `NHT` in the Matlab code which implements these expressions.

Table 1: First twenty terms

Term	Value	Number
1	$\mathcal{L}_j(H(t))$	NA
2	$\sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)F_i(k,\psi_j,4)\big(\Gamma_j' \otimes I_n\big)$	1
4	$\sum_i p_{ij}\big(\Gamma_j \otimes I_n\big)F_i(k,\psi_j,3)\big(\Gamma_j \otimes \Gamma_j\big)'$	2
5	$\sum_i p_{ij}\big(\Gamma_j Q_i(t)\Gamma_j'\big) \otimes \big(\psi_j\psi_j'\big)$	3
9	$\sum_i p_{ij}\big(\Gamma_j Q_i(t)\Gamma_j'\big) \otimes \big(G_j G_j'\big)$	4
10	$\sum_i \big(\Gamma_j \otimes \Gamma_j\big)F_i(k,\psi_j,2)\big(I_n \otimes \Gamma_j'\big)$	5
11	$\sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\Big(\psi_j' \otimes \varphi\big(Q_i(t)\big) \otimes \psi_j'\Big)$	6
13	$\sum_i p_{ij}\big(\Gamma_j \otimes I_n\big)\big(\psi_j' \otimes Q_i(t) \otimes \psi_j\big)\big(I_n \otimes \Gamma_j'\big)$	7
14	$\sum_i p_{ij}\big(\Gamma_j q_i(t)\psi_j'\big) \otimes \big(\psi_j\psi_j'\big)$	8
18	$\sum_i p_{ij}\big(\Gamma_j q_i(t)\psi_j'\big) \otimes \big(G_j G_j'\big)$	9
21	$\sum_i p_{ij}\big(\Gamma_j \otimes \Gamma_j\big)\sum_{kl}(E_{kl}' \otimes I_n)(I_n \otimes Q_i(t))(I_n \otimes E_{kl}')\big(G_j \otimes G_j\big)'$	10
24	$\sum_i p_{ij}\big(\Gamma_j \otimes I_n\big)\Big(q_i(t) \otimes \varphi\big(I_n\big)' \otimes \psi_j\Big)\big(G_j \otimes G_j\big)'$	11
25	$\sum_i p_{ij}\big(\Gamma_j \otimes G_j\big)\sum_{kl}(E_{kl}' \otimes I_n)(Q_i(t) \otimes I_n)(E_{kl}' \otimes I_n)K_{nn}\big(G_j \otimes \Gamma_j\big)'$	12
29	$\sum_i p_{ij}\big(\Gamma_j \otimes G_j\big)\Big(q_i(t) \otimes I_n \otimes \psi_j'\Big)\big(G_j' \otimes I_n\big)$	13
28	$\sum_i p_{ij}\big(I_n \otimes \Gamma_j\big)F_i(k,\psi_j,1)\big(\Gamma_j \otimes \Gamma_j\big)'$	14
29	$\sum_i p_{ij}\big(I_n \otimes \Gamma_j\big)\Big(\psi_j \otimes Q_i(t) \otimes \psi_j'\Big)\big(\Gamma_j' \otimes I_n\big)$	15
31	$\sum_i p_{ij}\Big(\big(\psi_j \otimes \psi_j\big) \otimes \varphi\big(Q_i(t)\big)'\Big)\big(\Gamma_j \otimes \Gamma_j\big)'$	16
32	$\sum_i p_{ij}\big(\psi_j q_i(t)'\Gamma_j'\big) \otimes \big(\psi_j\psi_j'\big)$	17
36	$\sum_i p_{ij}\big(I_n \otimes G_j\big)\Big(\big(\psi_j q_i(t)'\big) \otimes I_n\Big)\big(\Gamma_j \otimes G_j\big)'$	18
37	$\sum_i p_{ij}\big(\psi_j\psi_j'\big) \otimes \big(\Gamma_j Q_i(t)\Gamma_j'\big)$	19
38	$\sum_i p_{ij}\big(\psi_j\psi_j'\big) \otimes \big(\Gamma_j q_i(t)\psi_j'\big)$	20

Table 2: Second twenty terms

Term	Value	Number
40	$\sum_i p_{ij}\big(\psi_j\psi_j'\big)\otimes\big(\psi_j q_i(t)'\Gamma_j'\big)$	21
41	$\sum_i p_{ij}\pi_i(t)\big(\psi_j\psi_j'\big)\otimes\big(\psi_j\psi_j'\big)$	22
45	$\sum_i p_{ij}\pi_i(t)\big(\psi_j\psi_j'\big)\otimes\big(G_jG_j'\big)$	23
48	$\sum_i p_{ij}\big(I_n\otimes\Gamma_j\big)\Big(\psi_j\otimes\varphi\big(I_n\big)'\otimes q_i(t)\Big)\big(G_j\otimes G_j\big)'$	24
51	$\sum_i p_{ij}\pi_i(t)\Big(\psi_j\otimes\varphi\big(I_n\big)'\otimes\psi_j\Big)\big(G_j\otimes G_j\big)'$	25
52	$\sum_i p_{ij}\big(I_n\otimes G_j\big)\Big(\psi_j\otimes I_n\otimes q_i(t)'\Big)\big(G_j\otimes\Gamma_j\big)'$	26
53	$\sum_i p_{ij}\pi_i(t)\big(I_n\otimes G_j\big)\Big(\psi_j\otimes I_n\otimes\psi_j'\Big)\big(G_j'\otimes I_n\big)$	27
57	$\sum_i p_{ij}\big(G_j\otimes\Gamma_j\big)\sum_{kl}(E_{kl}'\otimes I_n)(I_n\otimes Q_i(t))(E_{kl}'\otimes I_n)K_{nn}\big(\Gamma_j\otimes G_j\big)'$	28
60	$\sum_i p_{ij}\big(G_j\otimes I_n\big)\Big(q_i(t)'\otimes I_n\otimes\psi_j\Big)\big(\Gamma_j\otimes G_j\big)'$	29
61	$\sum_i p_{ij}\big(G_j\otimes G_j\big)\sum_{kl}(E_{kl}'\otimes I_n)(Q_i(t)\otimes I_n)(I_n\otimes E_{kl}')\big(\Gamma_j\otimes\Gamma_j\big)'$	30
62	$\sum_i p_{ij}\big(G_j\otimes G_j\big)\Big(q_i(t)'\otimes\varphi\big(I_n\big)\otimes\psi_j'\Big)\big(\Gamma_j'\otimes I_n\big)$	31
66	$\sum_i p_{ij}\big(G_j\otimes\Gamma_j\big)\Big(\psi_j'\otimes I_n\otimes q_i(t)\Big)\big(I_n\otimes G_j'\big)$	32
69	$\sum_i p_{ij}\pi_i(t)\big(G_j\otimes I_n\big)\Big(\psi_j'\otimes I_n\otimes\psi_j\Big)\big(I_n\otimes G_j'\big)$	33
70	$\sum_i p_{ij}\big(G_j\otimes G_j\big)\Big(\psi_j'\otimes\varphi\big(I_n\big)\otimes q_i(t)'\Big)\big(I_n\otimes\Gamma_j'\big)$	34
71	$\sum_i p_{ij}\pi_i(t)\big(G_j\otimes G_j\big)\big(\psi_j'\otimes\varphi\big(I_n\big)\otimes\psi_j'\big)$	35
73	$\sum_i p_{ij}\big(G_jG_j'\big)\otimes\big(\Gamma_jQ_i(t)\Gamma_j'\big)$	36
74	$\sum_i p_{ij}\big(G_j\otimes\Gamma_j\big)\big(I_n\otimes q_i(t)\psi_j'\big)\big(G_j'\otimes I_n\big)$	37
76	$\sum_i p_{ij}\big(G_j\otimes I_n\big)\big(I_n\otimes\psi_j q_i(t)'\big)\big(G_j\otimes\Gamma_j\big)'$	38
77	$\sum_i p_{ij}\pi_i(t)\big(G_j\otimes I_n\big)\big(I_n\otimes\psi_j\psi_j'\big)\big(G_j'\otimes I_n\big)$	39
81*	$\sum_i p_{ij}\pi_i(t)\big(G_j\otimes G_j\big)\Omega_{WW}\big(G_j\otimes G_j\big)'$	40

* Ω_{WW} described in the text

As with the lower moments, given the terms above, wich define $S(t)$, it is possible to solve

$$\begin{aligned}H(t+1) &= \mathcal{L}\big(H(t)\big)+S(t)\\ \Rightarrow H &= \varphi^{-1}\Big(\big(I_{Nn^4}-\mathcal{M}_4\big)^{-1}\varphi(S)\Big)\\ \mathbb{H} &= \sum_i H_i\end{aligned}$$

where S replaces all terms $q(t)$, $Q(t)$ and $F(t)$ in $S(t)$ with their long run equivalents.

3 Unconditional moments of U.S. GDP growth

This section examines the behaviour of U.S. GDP growth. The results in Section 2 are applied to linear and Markov-switching autoregressions. Considering the higher moments suggests that, in the light of the financial crisis, the Great Moderation can be better understood as an exchange of variance for kurtosis. The results below are reported in terms of the first four standardised moments, which can be constructed from the unconditional moments according to Table 3, which also gives the sample approximations employed.

Table 3: Raw, standardised and sample momets

Std. moment	Definition	Sample	Construction from raw
1	$\mu = \mathbb{E}y_t$	$\bar{y} = \frac{1}{T}\sum_t y_t$	$\mathbb{E}y_t$
2	$\sigma^2 = \mathbb{E}(y_t - \mu)^2$	$\hat{\sigma}^2 = \frac{1}{T}\sum_t (y_t - \bar{y})^2$	$\mathbb{E}y_t^2 - (\mathbb{E}y_t)^2$
3	$\frac{\mathbb{E}(y_t-\mu)^3}{\sigma^3}$	$\frac{\frac{1}{T}\sum_t (y_t-\bar{y})^3}{(\hat{\sigma}^2)^{3/2}}$	$(\mathbb{E}y_t^3 - 3\mu\sigma^2 - \mu^3)/\sigma^3$
4	$\frac{\mathbb{E}(y_t-\mu)^4}{\sigma^4}$	$\frac{\frac{1}{T}\sum_t (y_t-\bar{y})^4}{(\hat{\sigma}^2)^2}$	$(\mathbb{E}y_t^4 - 4\mathbb{E}y_t^3\mu + 6\mathbb{E}y_t^2\mu^2 - 3\mu^4)/\sigma^4$

3.1 The whole sample 1954-2011

US GDP growth since 1954 is plotted in Figure 1; the second column Table 4 gives the first four standardised sample moments. There is some evidence of negative skew and excess kurtosis.

Table 4: Standardised Moments of US GDP growth 1954-2011

Moment	Sample moment	Linear AR(1)	MSI(3)-AR(1)
mean	0.7591	0.7561	0.7462
variance	0.8547	0.8613	0.8016
skew	-0.4314	*eps**	-0.4956
kurtosis	4.5735	3.000	4.6463

*Machine zero

Although very simple, the AR(1) remains a popular model for GDP growth. Inspection of sample ACF and PACF often suggest an AR(1) structure, and it performs well in forecasting relative to more complex ARMA models. The standardized moments implied by the linear AR(1) are given in the third column of Table 4. While the linear AR(1) is able to match the first two moments well, it is by construction unable to match the higher moments. The AR(1) is a Method of Moments estimator, using only the information contained in the first two moments.

The final column of Table 4 shows the standardised moments implied by the MSI(3)-AR(1) model. In this estimation the intercept and variance parameter are regime dependent, while the autoregressive parameter is constant across regimes. Some accuracy may be lost in fitting the first two moments, but the Markov-switching model is clearly able to replicate key features of the data which the linear model is not. The MSI-AR(1) is not - to my knowledge - a method of moments estimator, rather the flexible Markov-mixture of normals assumed for the errors allows the model to capture the skewness and kurtosis of the sample much better than the linear model. In itself this could be a useful result for long-horizon planning, suggesting there will be more *bad* shocks coming from the business cycle than the linear model would indicate. However, a review of the smoothed probabilities in Figure 2 suggests this is not the only story.

Fitting all the cycles in the period 1954-2011 with a Markov-Switching regression is not as easy as it once seemed. The natural two-state approach, pioneered by Hamilton [1989], does not fare well in in the longer sample, especially if switching variance is allowed, Clements and Krolzig [1998]. A brief glance at Figure 1 suggests switching variance is at least as important as switching in the mean or intercept of the model, and many authors now insist on this. Sims and Zha [2006] discuss the importance of switching variance further. Figure 2 demonstrates that a three-regime model with switching intercept and variance fits the NBER pattern of recessions quite well. However, the plot also suggests a one off change in the variance of the high-growth regime around 1984.

What is the picture of the business cycle which emerges from this estimation? Turning to the estimation results in Table 7 we see the intercepts are all positive. However in regime two, which is the regime most likely to occur during NBER recessions, the intercept is statistically zero, and the variance is high. In particular the variance of this low (statistically zero) growth regime is 2.8 times that of the regime most likely to hold in the expansionary phases of the business cycle 1954-84 and 8 times that of the high growth regime likely to hold 1984-2011. While regime 2 is not a negative growth regime, we can interpret it as a 'high risk' regime, with low average growth and large shocks, negative outcomes are likely in this state. Returning to the expansionary regimes, the intercept of the 54-84 regime is nearly twice that of the 84-11 regime. However, 54-84 regime is associated with a 14.5 percent chance of moving into the low (zero) growth regime in one period, while for the 84-11 regime this one-step transition probability is only 4.5 percent.

3.2 A change in the variance and the higher moments

Figure 2 suggests a change in the variance of the high-growth regime around 1984. To make the picture clearer, Figure 3 displays the most likely regime to have held on each date in the sample according to the MSI(3)-AR(1) model. The one-time switch between regimes 1 and 3 in 1984.3 is clear. Both regimes are expansion regimes, but the variance of the second expansion regime (regime 3) is much lower. This decline in variance is well known and the split in the smoothed regime probabilities between the 1-2 and 3-2 transitions invites us to look again at the sample moments in the two subperiods identified by the Markov switching model. These results are displayed in Table 5.

Table 5: Sample Moments of US GDP growth in subperiods

Moment	1954-1984.3	1984.4-2011
mean	0.8682	0.6522
variance	1.2731	0.3989
skew	-0.4145	-1.5601
kurtosis	3.3432	8.4844

The great exchange of variance for kurtosis and the resulting decline in the accuracy of the lognormal approximation for GDP growth is very clear in Table 5. The Markov- switching VAR identifies these well known subperiods quite clearly, but does not manage to match this change in the moment structure. Although there are no transitions between regimes 1 and 3, the three-regime structure ensures that all regimes have positive weight in the long-run probabilities: the state variable can move in principle from regime 3 to 2 to 1. There is no irreversible change in regimes and the resulting moments, like their whole-sample empirical equivalents, average those for the two sub-periods.

The AR(1)-MSI(3) structure may be overly restrictive. The AIC suggests the linear model prefers two lags to one. In an estimation with two lags an extra regime becomes a plausible feature of the data generating process. The AR(2)-MSI(4) model splits the sample in 1984, with the ergodic probability confined to regimes 1 and 4, the blue and cyan regimes in Figure 4. 1984q3 is taken as the date of the irreversible change in the the MSI(4)-AR(2) estimation. This is slightly later than 1984q1 identified by McConnell and Perez-Quiros [2000] as the break date for U.S. GDP. This discrepancy arises because I date the break at the point where the new regime, regime 1 in the estimation table, becomes dominant, rather than at the point where it begins to receive positive probability, which is closer to the McConnel and Perez-Quiros date.

The estimated moments from this model are reported in Table 6, with the estimation results in Table

Table 6: Estimated Moments of US GDP growth: MSI(4)-AR(2)

Moment	Estimate
mean	0.6480
variance	0.4058
skew	-1.5387
kurtosis	6.9132

8. The estimates for all four standardised moments are much closer to the empirical moments in the the post 1984 period than for the linear model or the MSI(3)-AR(1). The unconditional moments required to calculate the standardised moments were found by applying the expressions in Tables 1 and 2 to the companion form of the MSI(4)-AR(2).

While the model is effective in terms of unconditional moments in the second half of the sample, the separation between regimes in the first half is quite poor. In the second half of the sample there is a sharp division between a low variance expansionary regime, and a contractionary regime estimated to have held during the 1990 and 2009 recessions. In the second half of the sample therefore, the traditional contraction/expansion interpretation of the hidden states applies. More interesting in the current context may be the interaction between the estimated variances and the unconditional moments reported in Tabel 6. In the second half of the sample the variance of the good and of bad regime *both* decline. However, the kurtosis in the second half of the sample is very high. The complex interaction between the model parameters described in Section 2 allows the Markov-switching VAR to accommodate both declining variance and an increased probability of tail events.

4 Conclusions and directions

The unconditional moments for the MSI(M)-VAR(p) model have been derived under a stability condition that allows for the possibility of one-off changes in the parameter structure. These results are likely to be useful to researchers concerned with non-linearity in modern business cycles and can be applied to a wide range of VAR models. As a special case, the unconditional moments of the AR(1) version of the model were also derived under the stronger stationarity condition more common in the econometrics literature.

The AR(1) model is a reasonable model of GDP growth and its unconditional moments are relatively easy to derive even when all parameters are switching. Table 4 shows that the MSI(3)-AR(1) is able to capture the first four moments of GDP growth more accurately than the linear model. However, the smoothed probabilities from this estimation suggest a one-off change in the model parameters which the ergodic probabilities of the three-regime model cannot account for. The sample moments change dramatically if they are calculated on the two sub periods suggested by the smoothed probabilities of the Markov-switching estimations. The near lognormal result for the 1954-1984 period is in sharp contrast to the low-variance, high-kurtosis character of the 1984-2011 period.

The Markov-switching model with two lags and four regimes identifies a permanent change in the model parameters in 1984q3 and replicates the low-variance, high-kurtosis behaviour of recent data far more accurately than the simpler models. The view of the business cycle that emerges is one in which large shocks are important, and far more likely to be negative than positive. Rather than signal the return of the traditional high-variance near Gaussian business cycle, the financial crisis reveals new information about the change in the moment structure of U.S. GDP growth that occurred in the mid 1980s.

Considering the higher moments of GDP growth closes the gap between the Great Moderation and Financial Crisis literatures, but the estimations presented only scratch the surface. A priority for future work is to find a specification which can make plausible conditional forecasts over the medium term, while also matching th higher moments well. The results for GDP growth suggest that the non-Gaussian behaviour of the modern business cycle is important, and that Markov-switching models provide a simple way of modelling this. The derivation of the higher moments could be applied to the generation of Markov-switching DSGE models now being developed, giving these researchers the ability to exploit non-Gaussian features of the data in designing and choosing between models.

5 Appendix A: First and Second Moment

This section applies the recursive method of Costa et al. [2005] to the first and second moments of the MSI(M)-VAR(p). The timing convention in econometrics changes the first and second moment operators slightly, relative to their benchmark model. In particular, Costa et. al.'s matrix $\mathcal{B}$ for the first moment is replaced below by the matrix $\mathcal{M}_1$, while the second moment operator is thier $\mathcal{V}$ rather than their $\mathcal{T}$.

The stability concept employed is Mean Square Stability (MSS). If system (1) is MSS then, for $\mu(t) :=$

$\mathbb{E}[y(t)]$ and $\mathbb{Q}(t) := \mathbb{E}[y(t)y(t)']$, as $t \to \infty$

$$\|\mu(t) - \mu\| \to 0$$
$$\|\mathbb{Q}(t) - \mathbb{Q}\| \to 0$$

and μ and $\mathbb{Q}$ are taken as the unconditional first and second moments of the system. To find the limiting values μ and $\mathbb{Q}$, we first define recursions for the joint moments $\mathbb{E}[y(t), \theta(t) = i]$ and $\mathbb{E}[y(t)y(t)', \theta(t) = i]$; if such recursions converge then $\mu(t)$ and $\mathbb{Q}(t)$ will converge to the unconditional moments.

The (hidden) state variable $\theta(t)$ follows an N-state Markov chain with transition matrix P, where elemets $p_{ij} = \Pr(\theta(t+1) = j|\theta(t) = i)$ and $\sum_{i=1}^{N} p_{ij} = 1$. Further, let $\pi_i(t) := \Pr(\theta(t) = i)$. As the approach of Costa et al. [2005] does not require the stationarity of $\theta(t)$, these unconditional probabilities can vary with time. However, we do require the ergodicity of the discrete Markov chain, that is, $\lim_{t\to\infty} \pi_i(t) = \pi_i$, where the ergodic probabilities π_i solve the equation $\pi = P'\pi$.

Introduce the following definitions related to the first moment. To aid comparison with results in Costa et. al., the joint expectation of the pair $(y(t), \theta(t))$ is written in terms of the indicator function.

$$\begin{aligned}
q_i(t) :&= \mathbb{E}[y(t), \theta(t) = i] = \mathbb{E}[y(t)\, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
&= \mathbb{E}[y(t)|\theta(t) = i]\, \Pr(\theta(t) = i) \\
q(t) :&= \begin{bmatrix} q_1(t) \\ \vdots \\ q_N(t) \end{bmatrix} \\
\mu(t) :&= \mathbb{E}[y(t)] = \sum_i q_i(t)
\end{aligned}$$

To apply the approach in Costa et. al. plug the system equation (1) into the definition of the first moment of the joint process $(y(t), \theta(t))$

$$
\begin{aligned}
q_j(t+1) &= \mathbb{E}[y(t+1)\,1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\
&= \sum_{i=1}^{N} \mathbb{E}[y(t+1)\,1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}\,1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
&= \sum_{i=1}^{N} \mathbb{E}[\big(\psi_{\theta(t+1)} + \Gamma_{\theta(t+1)} y(t) + G_{\theta(t+1)} w(t+1)\big)\,1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}\,1_{\{\theta(\mathrm{t})=\mathrm{i}\}}] \\
&= \sum_{i=1}^{N} \psi_j p_{ij}\pi_i(t) + \sum_{i=1}^{N} \Gamma_j \mathbb{E}[y(t)\,1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]p_{ij} + \sum_{i=1}^{N} p_{ij}\pi_i(t) G_j \mathbb{E}[w(t+1)] \\
&= \psi_j \sum_{i=1}^{N} p_{ij}\pi_i(t) + \Gamma_j \sum_{i=1}^{N} p_{ij} q_i(t) \qquad (10)
\end{aligned}
$$

Notice that the second summation can be written in vector form,

$$
\Gamma_j \sum_{i=1}^{N} p_{ij} q_i(t) = \begin{bmatrix} p_{1j}\Gamma_j & \ldots & p_{Nj}\Gamma_j \end{bmatrix} \begin{bmatrix} q_1(t) \\ \vdots \\ q_N(t) \end{bmatrix}
$$

Introduce the following definitions related to the non-homogenous term and the matrix summation

$$
\begin{aligned}
\tilde{\psi}_j(t) &= \sum_{i=1}^{N} p_{ij}\psi_j\pi_i(t) \\
\tilde{\psi}(t) &= \begin{bmatrix} \tilde{\psi}_1(t) \\ \vdots \\ \tilde{\psi}_N(t) \end{bmatrix} \\
\mathcal{M}_1 &= \mathrm{diag}[\Gamma_i][P' \otimes I_n]
\end{aligned}
$$

Again $\mathrm{diag}[\Gamma_j]$ refers to the direct sum of the Γ_j matrices, $j = 1 \ldots N$. Using these definitions, (10) can be re-written giving a recursion for the stack of joint expectations, $q(t)$

$$
\begin{aligned}
q_j(t+1) &= \Gamma_j \sum_{i=1}^{N} p_{ij} q_i(t) + \tilde{\psi}_j(t) \\
\Rightarrow q(t+1) &= \mathcal{M}_1 q(t) + \tilde{\psi}(t) \qquad (11)
\end{aligned}
$$

Next, define limiting values for the non-homogeneous term $\tilde{\psi}_i(t)$

$$\tilde{\psi}_j = \sum_{i=1}^{N} p_{ij}\psi_j\pi_i$$
$$\tilde{\psi} = \begin{bmatrix} \tilde{\psi}_1 \\ \vdots \\ \tilde{\psi}_N \end{bmatrix}$$

By the ergodicity of the Markov chain $\theta(t)$, as $t \to \infty$, $\tilde{\psi}_i(t) \to \tilde{\psi}_i$, so that we can solve for $q := \lim_{t\to\infty} q(t)$ by employing Proposition 2.9 of Costa et al. [2005]. The unconditional expectation is then striaghtforward. As $t \to \infty$, (11) tends to

$$q(t+1) = \mathcal{M}_1 q(t) + \tilde{\psi}$$
$$\Rightarrow q = (I_{Nn} - \mathcal{M}_1)^{-1}\tilde{\psi}$$
$$\Rightarrow \mu = \sum_{i=1}^{N} q_i$$

For the second moment we study

$$Q_i(t) := \mathbb{E}[y(t)y(t)' 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]$$
$$Q(t) := \left[Q_1(t), \ldots, Q_N(t)\right]$$
$$\mathbb{Q}(t) := \mathbb{E}[y(t)y(t)'] = \sum_i Q_i(t)$$

Again, begin by plugging the system equation (1) into the definition of the joint second moment $Q_j(t+1)$

$$
\begin{aligned}
Q_j(t+1) =& \mathbb{E}[y(t+1)y(t+1)' \, 1_{\{\theta(\mathrm{t+1})=\mathrm{j}\}}] \\
=& \sum_{i=1}^{N} \Gamma_j \mathbb{E}[y(t)y(t)' \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]\Gamma_j' p_{ij} + \sum_{i=1}^{N} \Big(\Gamma_j \mathbb{E}[y(t) \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]\psi_j' p_{ij} \\
& + \psi_i \mathbb{E}[y(t) \, 1_{\{\theta(\mathrm{t})=\mathrm{i}\}}]'\Gamma_j' + (\psi_j\psi_j' + G_jG_j')p_{ij}\pi_i(t) \Big) \\
=& \sum_{i=1}^{N} \Gamma_j Q_i(t)\Gamma_j' p_{ij} + \sum_{i=1}^{N} p_{ij} \Big(\Gamma_j q_i(t)\psi_j' + \psi_j q_i(t)'\Gamma_j' + (\psi_j\psi_j' + G_jG_j')\pi_i(t) \Big)
\end{aligned} \tag{12}
$$

Again, define a sequence for the non-homogeneous term

$$
R_j(t,q) = \sum_{i=1}^{N} p_{ij} \Big(\Gamma_j q_i(t)\psi_j' + \psi_j q_i(t)'\Gamma_j' + (\psi_j\psi_j' + G_jG_j')\pi_i(t) \Big)
$$

which is very similar to (3.33) in Costa et al. [2005] except for the slight differences in the way the intercept behaves, and the restriction that the covariance of the $w(t)$ innovations is identity. We let $R(t,q) = [R_1(t,q), \ldots, R_N(t,q)]$ stack these terms in an $n.Nn$ matrix. Further, notice that the first matrix summation in (12) corresponds to the operator $\mathcal{V}_j$ as defined in (3.9) of Costa et al. [2005].[2] We can then rewrite (12), and so derive a recusion for the block-row of second moments, $Q \in \mathbb{R}^{n,Nn}$.

$$
\begin{aligned}
Q_j(t+1) &= \sum_{i=1}^{N} \Gamma_j Q_i(t)\Gamma_j' p_{ij} + R_j(t,q) \\
&= \mathcal{V}_j(Q(t)) + R_j(t,q) \\
\Rightarrow Q(t+1) &= \mathcal{V}(Q(t)) + R(t,q)
\end{aligned} \tag{13}
$$

In order to solve (13) for the unconditional second moment $\mathbb{Q}$, the sequence $\{R(t,q)\}$ must converge. Given the ergodicity of the Markov chain, set

$$
\begin{aligned}
R_j(q) &= \sum_{i=1}^{N} p_{ij} \Big(\Gamma_j q_i \psi_j' + \psi_j q_i' \Gamma_j' + (\psi_j\psi_j' + G_jG_j')\pi_i \Big) \\
R(q) &= \Big[R_1(q) \quad \ldots \quad R_N(q) \Big]
\end{aligned}
$$

[2] For reference if $X = [X_1, \ldots, X_N]$ where X_i is an $n \times n$, p.d.s. matrix, then $\mathcal{V}_j(X) = \sum_i p_{ij}\Gamma_j X_i \Gamma_j'$ and $\mathcal{V}(X) = [\mathcal{V}_1(X), \ldots, \mathcal{V}_N(X)]$.

then we have

$$R(t,q) \to R(q)$$

From equation (13) and Proposition 2.9 of Costa et al. [2005] we can find the limiting value of the $\{Q(t)\}$ sequence.

$$Q = (\mathcal{I} - \mathcal{V})^{-1} R(q)$$

The limiting moments q and Q can only exist if the spectral radii of the matrix $\mathcal{M}_1$ and of the operator $\mathcal{V}$ are less than 1. This is striaghtforward to calculate numerically for $\mathcal{M}_1 = \text{diag}[\Gamma_i][P' \otimes I_n]$. To find the radius of $\mathcal{V}$ first let $\varphi(.)$ denote the standard vec operator, and $\hat{\varphi}(.)$ its operation on the block-row matrices Q, R etc. If X_i is any $n.n$ matrix with $X = [X_1, \ldots, X_N]$, and individual element x^i_{jk} on the j^{th} row and k^{th} column of the i^{th} matrix, then

$$\begin{aligned}
\varphi(X_i) &:= \begin{bmatrix} x^i_{11} & x^i_{2,1} & \ldots & x^i_{n,1} & x^i_{1,2} & x^i_{2,2} & \ldots & x^i_{n,2} & \ldots & x^i_{n,n} \end{bmatrix}' \\
\hat{\varphi}(X) &:= \begin{bmatrix} \varphi(X_1) \\ \vdots \\ \varphi(X_N) \end{bmatrix} = \varphi(X)
\end{aligned}$$

With these definitions we can see

$$\begin{aligned}
\hat{\varphi}(Q) &= \left(I_{Nn^2} - \text{diag}[\Gamma_i \otimes \Gamma_i][P' \otimes I_{n^2}]\right)^{-1} \hat{\varphi}(R(q)) \\
Q &= \hat{\varphi}^{-1}\Big(\left(I_{Nn^2} - \mathcal{M}_2\right)^{-1} \hat{\varphi}(R(q))\Big)
\end{aligned} \tag{14}$$

The second line of (14) simply defines the matrix $\mathcal{M}_2$, which is identical to the matrix $\mathcal{A}_3$ in the discussion of the operator $\mathcal{V}(.)$ in Costa et al. [2005], Chapter 3. We have that the spectral radius of $\mathcal{V}$ is therefore equal to that of $\mathcal{M}_2$, which can again be calculated, for a given set of parameters in (1).

To summarize the results on the first two moments,

$$
\begin{aligned}
\mu(t) &= \sum_{i=1}^{N} q_i(t) \\
\mu &= \sum_{i=1}^{N} q_i \qquad (15) \\
\mathbb{Q}(t) &= \sum_{i=1}^{N} Q_i(t) \\
\mathbb{Q} &= \sum_{i=1}^{N} Q_i \qquad (16) \\
\text{var}(y) &= \mathbb{Q} - \mu\mu'
\end{aligned}
$$

The second moment matrix, $\mathcal{M}_2 = \text{diag}[\Gamma_i \otimes \Gamma_i][P' \otimes I_{n^2}]$, determines the stability of the system. If the largest eigenvalue of $\mathcal{M}_2 < 1$ then both first and second moment recursions converge and the unconditional mean and variance of the MSVAR are well defined. See Costa et. al., Chapter 3 for more on this stability concept. This completes the discussion of the first two moments.

6 Appendix B: The first-order autoregressive model

Consider the Markov-switching autoregression with N regimes and one lag. In deriving the moments all parameters are assumed to change. The model is given by

$$y_t = \psi_{\theta_t} + \Gamma_{\theta_t} y_{t-1} + G_{\theta_t}\epsilon_t \qquad (17)$$

where the intercept ψ_t, autoregressive parameter Γ_t and regime-conditional variance, G_t^2 are functions of the hidden state varible θ_t. θ_t follows a first-order, discrete, Markov chain with N states, where $p_{ij} = \Pr(\theta_t = j|\theta_{t-1} = i)$, are the transition probabilities, collected in P, the rows of which sum to one. Finally, it is assumed that $\epsilon_t \sim N(0,1)$.

There are three approaches to the unconditional moments in the literature of closely related models. First, Timmermann [2000] studies the Markov-switching mean (MSM-AR(p)) model and derives its first four unconditional, centred moments. Francq and Zakoïan [2001] on the other hand find expressions for the first two moments of the intercept-switching form of the MS-VARMA(p,q) process, of which the model (17) is a particular case. Finally Costa et al. [2005] study the stability properties and first two moments of a general Markov-Jump Linear System similar to the MSI(M)-VAR(p), though with different timing conventions between the continuous and discrete components. This section follows Timmermann [2000] and Francq and Zakoïan [2001] in assuming that $\mathbb{E}[y_t^r] = \mathbb{E}[y_{t-1}^r]$, provided these moments exist. The notation, however, is chosen to be most similar to Costa et. al. Their approach is followed closely in Section 2, for the general MSI(M)-VAR(p) specification.

Timmerman and Francq and Zakoïan differ slightly in their approach to the uconditional moments. The relation between their methods is examined in the derivation of the unconditional mean. The higher moments are then derived using the Fancq and Zakoïan method. In the approach of Timmerman, begin from the conditional expectations $\mathbb{E}[y_t|\theta_t = j]$ and use the law of iterated expectations $\mathbb{E}[y_t] = \mathbb{E}[\mathbb{E}[y_t|\theta_t]] = \pi'\mathbb{E}[y_t|\theta_t]$ to find the unconditional first moment. Here π are the steady-state probabilites of the transition matrix, P, and $\mathbb{E}[y_t|\theta_t]$ stacks the N conditional probabilities, $\mathbb{E}[y_t|\theta_t = j]$, in a column vector. Timmerman employs the 'backward transition matrix' B which naturally arises in this approach. The backward transition probabilities are given by

$$\begin{aligned} b_{ji} = \Pr(\theta_{t-1} = i|\theta_t = j) &= \frac{\Pr(\theta_t = j, \theta_{t-1} = i)}{\Pr(\theta_t = j)} \\ &= \frac{\Pr(\theta_t = j|\theta_{t-1} = i)\Pr(\theta_{t-1} = i)}{\Pr(\theta_t = j)} \\ &= \frac{p_{ij}\pi_i}{\pi_j} \end{aligned} \tag{18}$$

Using (18) the regime-conditional first moments follow

$$\begin{aligned} \mathbb{E}[y_t|\theta_t = j] &= \mathbb{E}[\psi_{\theta_t} + \Gamma_{\theta_t} y_{t-1} + G_{\theta_t}\epsilon_t|\theta_t = j] \\ &= \psi_j + \Gamma_j \mathbb{E}[y_{t-1}|\theta_t = j] \\ &= \psi_j + \Gamma_j \sum_i \mathbb{E}[y_{t-1}, \theta_{t-1} = i|\theta_t = j] \\ &= \psi_j + \Gamma_j \sum_i \mathbb{E}[y_{t-1}|\theta_{t-1} = i]\Pr(\theta_{t-1} = i|\theta_t = j) \\ &= \psi_j + \Gamma_j \sum_i \mathbb{E}[y_{t-1}|\theta_{t-1} = i] b_{ji} \end{aligned} \tag{19}$$

The N conditional first moments can be stacked up as follows

$$\begin{bmatrix} \mathbb{E}[y_t|\theta_t=1] \\ \vdots \\ \mathbb{E}[y_t|\theta_t=N] \end{bmatrix} = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_N \end{bmatrix} + \begin{bmatrix} \Gamma_1 & 0 & \dots & 0 \\ 0 & \Gamma_2 & & 0 \\ \vdots & & \ddots & \\ 0 & & \dots & \Gamma_N \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1N} \\ b_{21} & b_{22} & \dots & b_{2N} \\ \vdots & & \ddots & \\ b_{N1} & b_{N2} & \dots & b_{NN} \end{bmatrix} \begin{bmatrix} \mathbb{E}[y_{t-1}|\theta_{t-1}=1] \\ \vdots \\ \mathbb{E}[y_{t-1}|\theta_{t-1}=N] \end{bmatrix}$$

Writing in a more condensed notation gives

$$\mathbb{E}[y_{t|\theta_t}] = \psi + \text{diag}[\Gamma_j]B\mathbb{E}[y_{t-1|\theta_{t-1}}] \tag{20}$$

where $\text{diag}[X_j]$ is the direct sum of the X_j matrices. The unconditional first moment follows by solving the recursion (20) under the stationarity assumption and applying the law of iterated expectations

$$\begin{aligned} \mathbb{E}[y_t|\theta_t] &= (I_N - \text{diag}[\Gamma_j]B)^{-1}\psi \\ \mathbb{E}[y_t] &= \pi'(I_N - \text{diag}[\Gamma_j]B)^{-1}\psi \end{aligned}$$

Alternatively, Francq and Zakoïan use the joint expectations $\mathbb{E}[y_t, \theta_t = j]$, which removes the backward probabilities altogether. This is closer to the approach in Costa et. al. and is used in the remainder of this section. The joint expections can be derived from the above by multiplying both sides by the ergodic regime probability $\Pr(\theta_t = j) := \pi_j$ and using $\mathbb{E}[y_t, \theta_t = j] = \mathbb{E}[y_t|\theta_t = j]\Pr(\theta_t = j)$.

$$\begin{aligned} \pi_j\mathbb{E}[y_t|\theta_t = j] &= \pi_j\psi_j + \Gamma_j \sum_i \mathbb{E}[y_{t-1}|\theta_{t-1} = i]p_{ij}\pi_i \\ \Rightarrow \pi \odot \mathbb{E}[y_t|\theta_t] &= \pi \odot \psi + \text{diag}[\Gamma_j]P'(\pi \odot \mathbb{E}[y_{t-1}|\theta_{t-1}]) \end{aligned} \tag{21}$$

where $\odot$ stands for element-wise multiplication. $\pi \odot \mathbb{E}[y_t|\theta_t]$ is a stack of joint expectations over the Markov process (y_t, θ_t). Solving (21) the unconditional expectation is obtained

$$\begin{aligned} \pi \odot \mathbb{E}[y_t|\theta_t] &= (I_N - \text{diag}[\Gamma_j]P')^{-1}(\pi \odot \psi) \\ \mathbb{E}[y_t] &= \mathbf{1}'_N(I_N - \text{diag}[\Gamma_j]P')^{-1}\tilde{\psi} \end{aligned} \tag{22}$$

where $\tilde{\psi} = (\pi \odot \psi)$ and $\mathbf{1}_N$ is an $N \times 1$ column of ones. The second line of (22) adds the N joint probabilities to obtain the unconditional mean of the y_t process. For future reference, introduce the notation $q_i = \mathbb{E}[y_t, \theta_t = i]$ for the joint expectation, and let $q = [q_1, \ldots, q_N]'$ be the stack of these expectations.

To derive the unconditional variance notice that

$$
\begin{aligned}
\mathbb{E}[y_t^2|\theta_t = j] &= \mathbb{E}[(\psi_{\theta_t} + \Gamma_{\theta_t} y_{t-1} + G_{\theta_t}\epsilon_t)^2|\theta_t = j] \\
&= \psi_j^2 + 2\psi_j\Gamma_j\mathbb{E}[y_{t-1}|\theta_t = j] + \Gamma_j^2\mathbb{E}[y_{t-1}^2|\theta_t = j] + G_j^2 \\
&= \psi_j^2 + 2\psi_j\Gamma_j\sum_i \mathbb{E}[y_{t-1}|\theta_{t-1} = i]\frac{p_{ij}\pi_i}{\pi_j} + \Gamma_j^2\sum_i \mathbb{E}[y_{t-1}^2|\theta_{t-1} = i]\frac{p_{ij}\pi_i}{\pi_j} + G_j^2
\end{aligned}
$$

then stack the joint expections $\mathbb{E}[y_t^2, \theta_t = j]$ and employ the stationarity condition to solve for the unconditional second moment of the y_t process. To save some notation, let $x^{\cdot 2}$ represent the vector $x \odot x$.

$$
\begin{aligned}
\pi \odot \mathbb{E}[y_t^2|\theta_t] =& \pi \odot \psi^{\cdot 2} + \mathrm{diag}[2\psi_j\Gamma_j]P'(\pi \odot \mathbb{E}[y_{t-1}|\theta_{t-1}]) \\
&+ \mathrm{diag}[\Gamma_j^2]P'(\pi \odot \mathbb{E}[y_{t-1}^2|\theta_{t-1}]) + \pi \odot G^{\cdot 2} \\
=& \big(I_N - \mathrm{diag}[\Gamma_j^2]P'\big)^{-1}\big(\pi \odot (\psi^{\cdot 2} + G^{\cdot 2}) + \mathrm{diag}[2\psi_j\Gamma_j]P'(\pi \odot \mathbb{E}[y_{t-1}|\theta_{t-1}])\big) \\
=& \big(I_N - \mathrm{diag}[\Gamma_j^2]P'\big)^{-1}\big(\pi \odot (\psi^{\cdot 2} + G^{\cdot 2}) + \mathrm{diag}[2\psi_j\Gamma_j]P'q\big)
\end{aligned} \tag{23}
$$

The final line of (23) employs the earlier definition, q, of the joint expectations. In turn define

$$
\begin{aligned}
Q_i &= \mathbb{E}[y_t^2|\theta_t = i] \cdot \pi_i \\
&= \mathbb{E}[y_t^2, \theta_t = i]
\end{aligned}
$$

Again let $Q = [Q_1, \ldots, Q_N]'$. The marginal expectation of the square then follows directly

$$
\mathbb{E}[y_t^2] = \mathbf{1}_N' Q
$$

To derive the higher moments, proceed in the same way. In the scalar case it is strightforawrd to see

$$
\begin{aligned}
\mathbb{E}[y_t^3|\theta_t=j] =&\mathbb{E}[(\psi_{\theta_t}+\Gamma_{\theta_t}y_{t-1}+G_{\theta_t}\epsilon_t)^3|\theta_t=j]\\
=&\psi_j^3+3\psi_jG_j^2+3(\psi_j^2\Gamma_j+G_j^2\Gamma_j)\mathbb{E}[y_{t-1}|\theta_t=j]\\
&+3\psi_j\Gamma_j^2\mathbb{E}[y_{t-1}^2|\theta_t=j]+\Gamma_j^3\mathbb{E}[y_{t-1}|\theta_t=j]\\
=&\psi_j^3+3\psi_jG_j^2+3(\psi_j^2\Gamma_j+G_j^2\Gamma_j)\sum_i q_ip_{ij}\\
&+3\psi_j\Gamma_j^2\sum_i Q_ip_{ij}+\Gamma_j^3\sum_i\mathbb{E}[y_{t-1}^3|\theta_{t-1}=i]p_{ij}\pi_i
\end{aligned}
$$

Stacking the joint expectations, $\mathbb{E}[y_t^3,\theta_t=i]:=F_i$, in the vector $F=[F_1,\dots,F_N]'$ and again solving under stationarity gives

$$
\begin{aligned}
F =&\Big(I_N-\operatorname{diag}[\Gamma_j^3]P'\Big)^{-1}\Big(\pi\odot(\psi^{\cdot 3}+3\psi\odot G^{\cdot 2})+\big(3\operatorname{diag}\left[\psi_j^2\Gamma_j+G_j^2\Gamma_j\right]P'\big)q\\
&+\big(3\operatorname{diag}\left[\psi_j\Gamma_j^2\right]P'\big)Q\Big)
\end{aligned}
\tag{24}
$$

Finally, for the conditional fourth moment of y_t we have

$$
\begin{aligned}
\mathbb{E}[y_t^4|\theta_t=j] =&\mathbb{E}[(\psi_{\theta_t}+\Gamma_{\theta_t}y_{t-1}+G_{\theta_t}\epsilon_t)^4|\theta_t=j]\\
=&\psi_j^4+3G_j^4+6G_j^2\psi_j^2+(4\psi_j^3\Gamma_j+12\psi_j\Gamma_jG_j^2))\mathbb{E}[y_{t-1}|\theta_t=j]+\\
&+6(\psi_j^2\Gamma_j^2+G_j^2\Gamma_j^2)\mathbb{E}[y_{t-1}^2|\theta_t=j]+4\psi_j\Gamma_j^3\mathbb{E}[y_{t-1}^3|\theta_t=j]\\
&+\Gamma_j^4\mathbb{E}[y_{t-1}^4|\theta_t=j]\\
=&\nu_j+A_{1j}\mathbb{E}[y_{t-1}|\theta_t=j]+A_{2j}\mathbb{E}[y_{t-1}^2|\theta_t=j]\\
&+A_{3j}\mathbb{E}[y_{t-1}^3|\theta_t=j]+\Gamma_j^4\mathbb{E}[y_{t-1}^4|\theta_t=j]
\end{aligned}
$$

where the final line simply defines the scalars ν_j and $A_{k,j}$, $k=1,2,3$. Again using that

$$
\mathbb{E}[y_{t-1}^r|\theta_t=j]=\sum_i\mathbb{E}[y_{t-1}^r|\theta_{t-1}=i]p_{ij}\frac{\pi_i}{\pi_j}
$$

it is clear that

$$\pi_j \mathbb{E}[y_t^4|\theta_t = j] = \pi_j \nu_j + A_{ij} \sum_i q_i p_{ij} + A_{2j} \sum_i Q_i p_{ij} \\ + A_{3j} \sum_i F_i p_{ij} + \Gamma_j^4 \sum_i \mathbb{E}[y_{t-1}^4|\theta_{t-1} = i] p_{ij} \pi_i$$

Let $H_i = \mathbb{E}[y_t^4, \theta_t = i]$ and stack these expectations in the vector $H = [H_1, \dots, H_N]'$. Then once again solve the equation for H assuming stationarity.

$$H = \Big(I_N - \mathrm{diag}[\Gamma_j^4] P'\Big)^{-1} \Big(\pi \odot \nu + \mathrm{diag}[A_{1j}] P' q + \\ + \mathrm{diag}[A_{2j}] P' Q + \mathrm{diag}[A_{3j}] P' F\Big) \tag{25}$$

From (25) we have $\mathbb{E}[y_t^4] = \mathbf{1}_N' H$.

References

F. Bianchi. Methods for measuring expectations and uncertainty in Markov-switching models. *Duke University, Working Paper*, 2013.

T. Broer and A. Kero. Great moderation or great mistake: Can rising confidence in low macro-risk explain the boom in asset prices? *Working Paper*, 2011.

M.K. Brunnermeier and Y. Sannikov. A macroeconomic model with a financial sector. *Working Paper, Department of Economics, Princeton University*, 2011.

M. P. Clements and H. M. Krolzig. A comparison of the forecast performance of Markov-switching and threshold autoregressive models of us gnp. *Econometrics Journal*, 1(1), 1998.

O.L.V. Costa, M.D. Fragoso, and R.P. Marques. *Discrete-Time Markov Jump Linear Systems.* Springer-Verlag, London, 2005.

R.E.A. Farmer, D. F. Waggoner, and T. Zha. Minimal state variable solutions to Markov-switching rational expectations models. *Journal of Economic Dynamics and Control*, 35(12):2150–2166, 2011.

C. Francq and J.-M. Zakoïan. Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, 102:339–364, 2001.

G. A. Ghazal and H. Neudecker. On second-order and fourth-order moments of jointly distributes random matrices: a survey. *Linear Algebra and its Applications*, 321:61–93, 2000.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

K. Hubrich and R. Tetlow. Financial stress and economic dynamcis: the transmission of crises. *FEDS*, 82, 2012.

C. J. Kim and C. R. Nelson. Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *The Review of Economics and Statistics*, 81(4): 608–616, 1999.

J. Magnus and H. Neudecker. The commutation matrix: Some properties and applications. *Annals of Statistics*, 7(2):381–394, 1979.

J. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, Chichester, 1988.

M. McConnell and G. Perez-Quiros. Output fluctuations in the United States: What has changed since the early 1980s? *American Economic Review*, 90(5):1464–1476, 2000.

J.R. Schott. *Matrix Analysis for Statistics*. John Wiley and Sons, New Jersey, 2005.

C.A. Sims and T. Zha. Were there regime switches in us monetary policy? *The American Economic Review*, 96(1), 2006.

J. H. Stock and M. W. Watson. Has the business cycle changed and why? In *NBER Macroeconomics Annual*. MIT Press, 2003.

L. Svensson and N. Williams. Monetary policy with model uncertainty: Distribution forecast targeting. *Princeton University Working Paper*, 2007.

A. Timmermann. Moments of Markov switching models. *Journal of Econometrics*, 96:75–111, 2000.

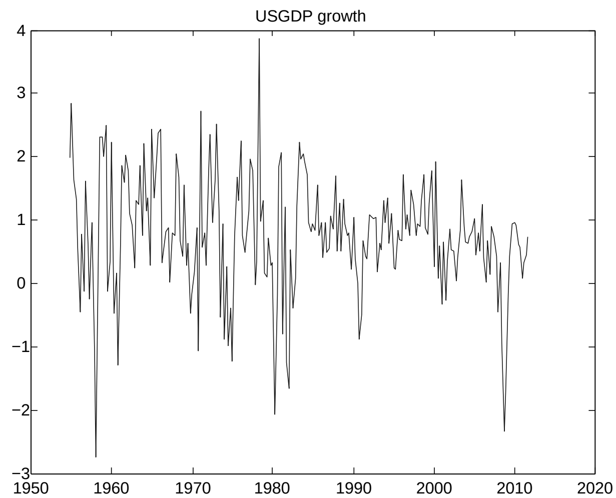

Figure 1: GDP growth 1954-2011

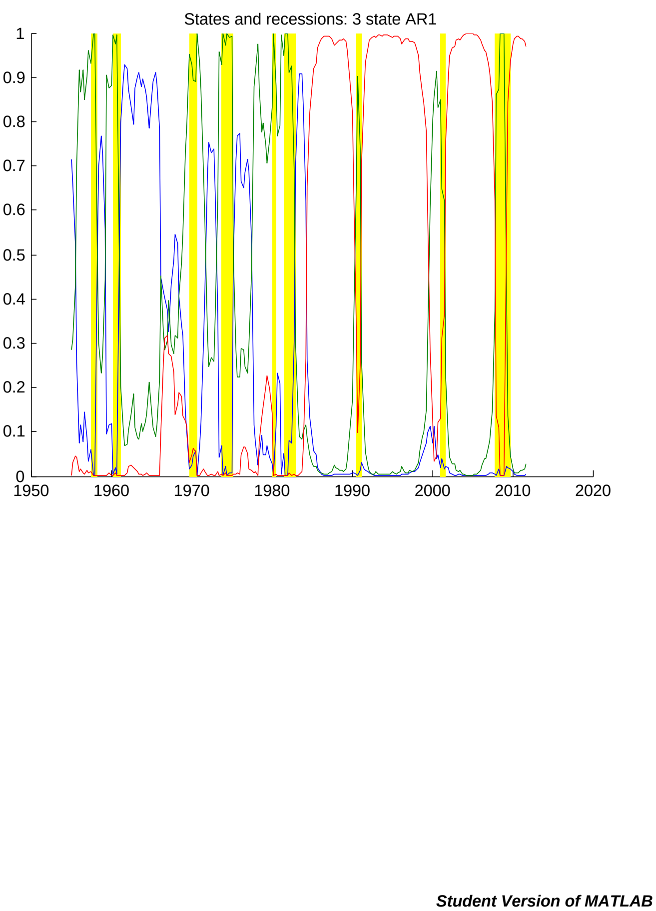

Figure 2: US recessions and smoothed state probabilites MS(3)-AR(1). State 2, green, is clearly a recession state

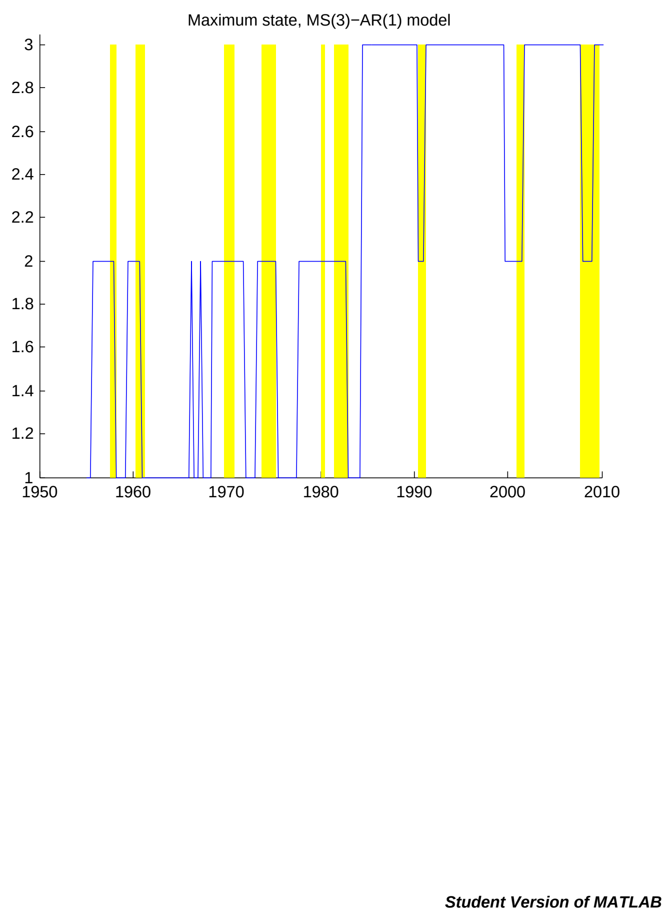

Figure 3: Most likely state from MS(3)-AR(1) model. There is a switch in expansion states in 1984.3

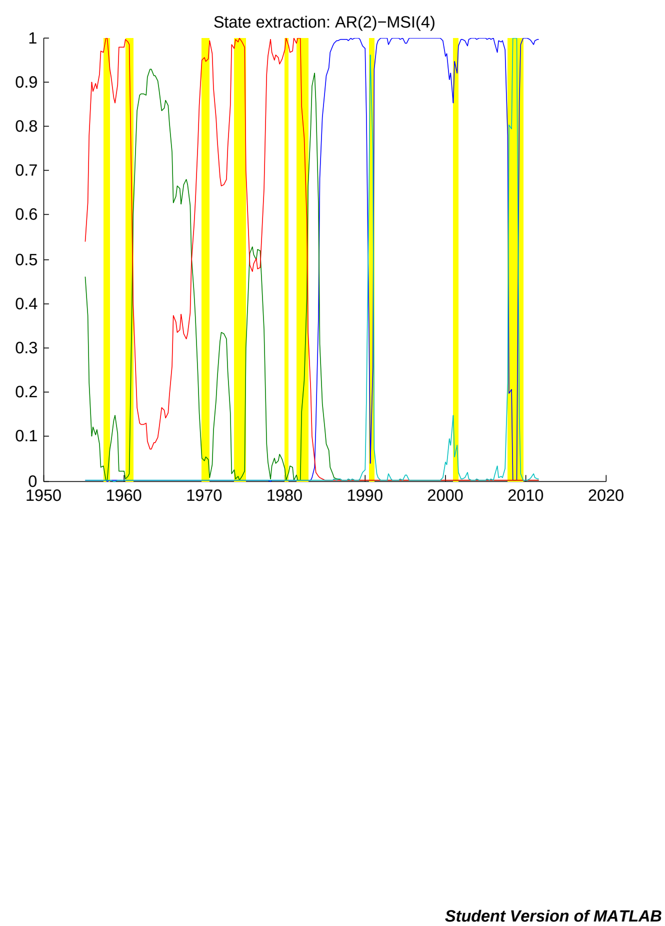

Figure 4: The MSI(4)-AR(2) model assigns a distinct pair of regimes to pre and post 1984 data.

Table 7: MSI(3)-AR(1)

Regime j	ψ_j	γ_j	σ_j	p_{j1}	p_{j2}	p_{j3}
1	1.1363	0.2406	0.4635	0.8302	0.1449	0.0250
	(0.2730)	(0.0747)	(0.0842)	(0.1193)	(0.1136)	(0.0392)
2	0.2191	0.2406	1.308	0.0935	0.8581	0.0484
	(0.1807)	(0.0747)	(0.1294)	(0.0628)	(0.0747)	(0.0334)
3	0.5913	0.2406	0.1616	0	0.045	0.9550
	(0.0772)	(0.0747)	(0.0186)	(0)	(0.0279)	(0.0279)

Table 8: MSI(4)-AR(2)

Regime j	ψ_j	γ_1^j	γ_2^j	σ_j	p_{j1}	p_{j2}	p_{j3}	p_{j4}
1	0.5401	0.1652	0.1456	0.1784	0.9747	0	0	0.0253
	(0.0731)	(0.0699)	(0.0619)	(0.0289)	(0.0200)	(0)	(0)	(0.0200)
2	0.9450	0.1652	0.1456	0.4685	0.0242	0.8787	0.0971	0
	(0.2285)	(0.0699)	(0.0619)	(0.1387)	(0.0258)	(0.0995)	(0.0961)	(0)
3	0.3725	0.1652	0.1456	1.3853	0	0.0599	0.9401	0
	(0.1750)	(0.0699)	(0.0619)	(0.2748)	(0)	(0.0597)	(0.0597)	(0)
4	-0.6415	0.1652	0.1456	0.6369	0.2944	0	0	0.7056
	(0.4700)	(0.0699)	(0.0619)	(0.3780)	(0.1849)	(0)	(0)	(0.1849)