Online supplementary material to Boot and Pick (2016) "Optimal forecasts from Markov switching models"

A Mathematical details

A.1 Derivations conditional on states

A.1.1 Weights for two-state Markov switching model

In order to derive weights (10)–(13), define $\lambda = \frac{\beta_2-\beta_1}{\sigma_2}$ and $q = \frac{\sigma_1}{\sigma_2}$, $\pi_1 = \frac{1}{T}\sum_{t=1}^{T} s_{1t}$, and $\pi_2 = \frac{1}{T}\sum_{t=1}^{T} s_{2t}$. Then we have

$$\begin{aligned}\mathbf{M} &= \mathbf{Q} + \tilde{\mathbf{S}}'\boldsymbol{\lambda}\boldsymbol{\lambda}'\tilde{\mathbf{S}} \\ &= q^2\mathbf{S}_1 + \mathbf{S}_2 + \lambda^2\mathbf{s}_2\mathbf{s}_2'\end{aligned}$$

where $\mathbf{S}_i$ is a $T\times T$ diagonal matrix with typical t,t-element $s_{i,t}$. The inverse of $\mathbf{M}$ is

$$\begin{aligned}\mathbf{M}^{-1} &= (q^2\mathbf{S}_1 + \mathbf{S}_2)^{-1} - \frac{\lambda^2(q^2\mathbf{S}_1+\mathbf{S}_2)^{-1}\mathbf{s}_2\mathbf{s}_2'(q^2\mathbf{S}_1+\mathbf{S}_2)^{-1}}{1+\lambda^2\mathbf{s}_2'(q^2\mathbf{S}_1+\mathbf{S}_2)^{-1}\mathbf{s}_2} \\ &= \frac{1}{q^2}\mathbf{S}_1 + \mathbf{S}_2 - \frac{\lambda^2(\frac{1}{q^2}\mathbf{S}_1+\mathbf{S}_2)\mathbf{s}_2\mathbf{s}_2'(\frac{1}{q^2}\mathbf{S}_1+\mathbf{S}_2)}{1+\lambda^2\mathbf{s}_2'(\frac{1}{q^2}\mathbf{S}_1+\mathbf{S}_2)\mathbf{s}_2} \\ &= \frac{1}{q^2}\mathbf{S}_1 + \mathbf{S}_2 - \frac{\lambda^2\mathbf{s}_2\mathbf{s}_2'}{1+\lambda^2T\pi_2}\end{aligned}$$

The weights are given by

$$\mathbf{w} = \lambda^2\mathbf{M}^{-1}\mathbf{s}_2 s_{2,T+1} + \frac{\mathbf{M}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\mathbf{M}^{-1}\boldsymbol{\iota}}\left(1-\lambda^2\boldsymbol{\iota}'\mathbf{M}^{-1}\mathbf{s}_2 s_{2,T+1}\right)$$

The various components needed to calculate the weights are given by

$$\begin{aligned}\mathbf{M}^{-1}\mathbf{s}_2 &= \mathbf{s}_2 - \frac{\lambda^2T\pi_2}{1+\lambda^2T\pi_2}\mathbf{s}_2 \\ &= \frac{1}{1+\lambda^2T\pi_2}\mathbf{s}_2 \\ \mathbf{M}^{-1}\boldsymbol{\iota} &= \frac{1}{q^2}\mathbf{s}_1 + \mathbf{s}_2 - \frac{\lambda^2T\pi_2}{1+\lambda^2T\pi_2}\mathbf{s}_2 \\ &= \frac{\mathbf{s}_1(1+\lambda^2T\pi_2)+q^2\mathbf{s}_2}{q^2(1+\lambda^2T\pi_2)}\end{aligned}$$

and

$$\boldsymbol{\iota}'\mathbf{M}^{-1}\mathbf{s}_2 = \frac{T\pi_2}{1+\lambda^2T\pi_2}, \quad \boldsymbol{\iota}'\mathbf{M}^{-1}\boldsymbol{\iota} = T\frac{\pi_1+\lambda^2T\pi_1\pi_2+q^2\pi_2}{q^2(1+\lambda^2T\pi_2)}$$

This yields the weights

$$\begin{aligned}\mathbf{w} &= \lambda^2 \frac{1}{1+\lambda^2 T\pi_2}\mathbf{s}_2 s_{2,T+1} + \frac{1}{T}\frac{\mathbf{s}_1(1+\lambda^2 T\pi_2)+q^2\mathbf{s}_2}{\pi_2 q^2 + \pi_1(1+T\pi_2\lambda^2)}\left(1-\lambda^2\frac{T\pi_2 s_{2,T+1}}{1+\lambda^2 T\pi_2}\right)\\ &= \frac{1}{1+\lambda^2 T\pi_2}\left\{\mathbf{s}_2 s_{2,T+1} + \frac{1}{T}\frac{\mathbf{s}_1(1+\lambda^2 T\pi_2)+q^2\mathbf{s}_2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\left[1+\lambda^2 T\pi_2(1-s_{2,T+1})\right]\right\}\end{aligned}$$

Suppose $s_{2,T+1} = s_{2,t} = 1$, then

$$\begin{aligned}w_{(2,2)} &= \frac{1}{1+\lambda^2 T\pi_2}\left(\lambda^2 + \frac{1}{T}\frac{q^2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\right)\\ &= \frac{1}{1+\lambda^2 T\pi_2}\frac{1}{T}\frac{1}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\left[q^2(1+\lambda^2 T\pi_2)+\lambda^2 T\pi_1(1+\lambda^2 T\pi_2)\right]\\ &= \frac{1}{T}\frac{q^2+\lambda^2 T\pi_1}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\end{aligned}$$

when $s_{2,T+1} = 1$, $s_{2,t} = 0$, then

$$\begin{aligned}w_{(2,1)} &= \frac{1}{1+\lambda^2 T\pi_2}\left(\frac{1}{T}\frac{1+\lambda^2 T\pi_2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\right)\\ &= \frac{1}{T}\frac{1}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\end{aligned}$$

when $s_{2,T+1} = 0, s_{2,t} = 1$, then

$$\begin{aligned}w_{(1,2)} &= \frac{1}{T}\frac{1}{1+\lambda^2 T\pi_2}\frac{q^2(1+\lambda^2 T\pi_2)}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\\ &= \frac{1}{T}\frac{q^2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\end{aligned}$$

finally, when $s_{2,T+1} = 0, s_{2,t} = 0$, then

$$\begin{aligned}w_{(1,1)} &= \frac{1}{T}\frac{1}{1+\lambda^2 T\pi_2}\frac{(1+\lambda^2 T\pi_2)^2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\\ &= \frac{1}{T}\frac{1+\lambda^2 T\pi_2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\end{aligned}$$

In order to show the symmetry of the weights, consider the definition of λ and q conditional on the regime $s_{i,T+1}$. If $s_{2,T+1} = 1$, define $\lambda = \frac{\beta_2-\beta_1}{\sigma_2}$ and $q = \frac{\sigma_1}{\sigma_2}$, but if $s_{1,T+1} = 1$, define $\lambda_* = \frac{\beta_1-\beta_2}{\sigma_1}$ and $q_* = \frac{\sigma_2}{\sigma_1}$. Then, $\lambda^2 = \lambda_*^2/q_*^2$ and we have for $w_{(1,2)}$ and $w_{(1,1)}$

$$\begin{aligned}w_{(1,2)} &= \frac{1}{T}\frac{q^2}{\pi_2 q^2+\pi_1(1+T\pi_2\lambda^2)}\\ &= \frac{1}{T}\frac{1/q_*^2}{\pi_2/q_*^2+\pi_1(1+1/q_*^2 T\pi_2\lambda_*^2)}\\ &= \frac{1}{T}\frac{1}{\pi_1 q_*^2+\pi_2(1+T\pi_1\lambda_*^2)}\end{aligned}$$

$$
\begin{aligned}
w_{(1,1)} &= \frac{1}{T}\frac{1+\lambda^2 T\pi_2}{\pi_2 q^2 + \pi_1(1+T\pi_2\lambda^2)} \\
&= \frac{1}{T}\frac{1+1/q_*^2\lambda_*^2 T\pi_2}{\pi_2/q_*^2 + \pi_1(1+1/q_*^2 T\pi_2\lambda_*^2)} \\
&= \frac{1}{T}\frac{q_*^2+\lambda_*^2 T\pi_2}{\pi_1 q_*^2 + \pi_2(1+T\pi_1\lambda_*^2)}
\end{aligned}
$$

The symmetry of the weights is a natural consequence of the fact that the Markov Switching model is invariant under a relabeling of the states.

A.1.2 Weights and MSFE for m-state Markov switching model

To derive weights for an m-state Markov switching model, we will concentrate on $s_{k,T+1} = 1$ as we have shown above that the weights are symmetric. In this case, define $\lambda_i = (\beta_i - \beta_k)/\sigma_k$ and $q_i = \sigma_i/\sigma_k$. The model is given by

$$
\begin{aligned}
y_t &= \sum_{i=1}^{m} \beta_i s_{it} + \sum_{i=1}^{m} \sigma_i s_{it}\varepsilon_t \\
&= \beta_k + \sum_{i=1}^{m}(\beta_i - \beta_k)s_{it} + \sum_{i=1}^{m}\sigma_i s_{it}\varepsilon_t \\
&= \sigma_k\left(\frac{\beta_k}{\sigma_k} + \sum_{i=1}^{m}\lambda_i s_{it} + \sum_{i=1}^{m} q_i s_{it}\varepsilon_t\right)
\end{aligned}
$$

For the observation at $T+1$ we have

$$
\frac{1}{\sigma_k} y_{T+1} = \frac{\beta_k}{\sigma_k} + \varepsilon_{T+1}
$$

The forecast error is

$$
\frac{1}{\sigma_k}(y_{T+1} - \mathbf{w}'\mathbf{y}) = \varepsilon_{T+1} - \sum_{i=1}^{m}\lambda_i \mathbf{w}'\mathbf{s}_i - \sum_{i=1}^{m} q_i \mathbf{w}'\mathbf{S}_i\boldsymbol{\varepsilon}
$$

Squaring and taking expectations gives

$$
\mathrm{E}\left[\sigma_k^{-2}(y_{T+1} - \mathbf{w}'\mathbf{y})^2\right] = 1 + \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j \mathbf{w}'\mathbf{s}_i\mathbf{s}_j'\mathbf{w} + \sum_{i=1}^{m} q_i^2 \mathbf{w}\mathbf{S}_i\mathbf{w}
$$

Implementing the constraint $\sum_{t=1}^{T} w_t = 1$ by a Lagrange multiplier and taking the derivative gives

$$
\begin{aligned}
\mathbf{w} &= \left(\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j \mathbf{w}'\mathbf{s}_i\mathbf{s}_j' + \sum_{i=1}^{m} q_i^2 \mathbf{w}\mathbf{S}_i\right)^{-1}(-\theta\boldsymbol{\iota}) \\
&= -\theta\mathbf{M}^{-1}\boldsymbol{\iota}
\end{aligned} \tag{36}
$$

The inverse can be expressed analytically through the Sherman Morrison formula as

$$\begin{aligned}
\mathbf{M}^{-1} &= \sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{S}_i - \frac{\left(\sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{s}_i\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i\lambda_j \mathbf{s}_i\mathbf{s}_j'\right)\left(\sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{s}_i\right)}{1+\left(\sum_{j=1}^{m} \lambda_j \mathbf{s}_j'\right)\left(\sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{s}_i\right)\left(\sum_{i=1}^{m} \lambda_j \mathbf{s}_i\right)} \\
&= \sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{S}_i - \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\lambda_i}{q_i^2}\frac{\lambda_j}{q_j^2}\mathbf{s}_i\mathbf{s}_j'}{1+T\sum_{i=1}^{m}\frac{\lambda_i^2}{q_i^2}\pi_i}
\end{aligned}$$

Multiplying with $\boldsymbol{\iota}$ as in equation (36) gives

$$\mathbf{M}^{-1}\boldsymbol{\iota} = \sum_{i=1}^{m} \frac{1}{q_i^2}\mathbf{s}_i - \frac{T\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\lambda_i}{q_i^2}\frac{\lambda_j}{q_j^2}\pi_j\mathbf{s}_i}{1+T\sum_{i=1}^{m}\frac{\lambda_i^2}{q_i^2}\pi_i}$$

Since the weights should sum up to one, we have

$$\begin{aligned}
\boldsymbol{\iota}'\mathbf{w} &= \left(T\sum_{i=1}^{m}\frac{1}{q_i^2}\pi_i - \frac{T^2\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\lambda_i}{q_i^2}\frac{\lambda_j}{q_j^2}\pi_j\pi_i}{1+T\sum_{i=1}^{m}\frac{\lambda_i^2}{q_i^2}\pi_i}\right)(-\theta) \\
&= 1
\end{aligned}$$

which gives

$$\begin{aligned}
\theta &= \frac{1+T\sum_{j=1}^{m}\frac{\lambda_i^2}{q_i^2}\pi_i}{T}\left[\sum_{i=1}^{m}\frac{1}{q_i^2}\pi_i + T\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{1}{q_i^2}\frac{\lambda_j}{q_j^2}\pi_i\pi_j - \frac{\lambda_i}{q_i^2}\frac{\lambda_j}{q_j^2}\pi_j\pi_i\right)\right]^{-1} \\
&= \frac{1+T\sum_{j=1}^{m}\frac{\lambda_i^2}{q_i^2}\pi_i}{T}\left[\sum_{i=1}^{m}\frac{1}{q_i^2}\pi_i + T\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{q_i^2}\frac{1}{q_j^2}\pi_i\pi_j\lambda_j(\lambda_j-\lambda_i)\right]^{-1}
\end{aligned}$$

The weights are then given by

$$\mathbf{w} = \frac{1}{T}\frac{\sum_{i=1}^{m}\frac{1}{q_i^2}\mathbf{s}_i + T\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{q_i^2}\frac{1}{q_j^2}\pi_j\lambda_j(\lambda_j-\lambda_i)\mathbf{s}_i}{\sum_{i=1}^{m}\frac{1}{q_i^2}\pi_i + T\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{q_i^2}\frac{1}{q_j^2}\pi_i\pi_j\lambda_j(\lambda_j-\lambda_i)}$$

So that if $s_{lt} = 1$ the weight at time t is

$$w_t = \frac{1}{T}\frac{\frac{1}{q_l^2} + T\sum_{j=1}^{m}\frac{1}{q_l^2}\frac{1}{q_j^2}\pi_j\lambda_j(\lambda_j-\lambda_l)}{\sum_{i=1}^{m}\frac{1}{q_i^2}\pi_i + T\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{q_i^2}\frac{1}{q_j^2}\pi_i\pi_j\lambda_j(\lambda_j-\lambda_i)}$$

The MSFE is easy to derive by noting that we can substitute the first order condition for the weights

$$\begin{aligned}
\mathrm{E}\left[\sigma_k^{-2}(y_{T+1}-\mathbf{w}'\mathbf{y})^2\right] &= 1+\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j\mathbf{w}'\mathbf{s}_i\mathbf{s}_j'\mathbf{w} + \sum_{i=1}^{m} q_i^2\mathbf{w}\mathbf{S}_i\mathbf{w} \\
&= 1-\theta \\
&= 1+w_{(k,k)}
\end{aligned}$$

where $w_{(k,k)}$ is the weight when $s_{k,T+1} = s_{kt} = 1$.

A.2 Derivations conditional on state probabilities

A.2.1 Large T approximation for optimal weights

Rewrite (21) as

$$w_t = \frac{1}{T}\frac{d_t[\frac{1}{T} + \lambda^2 \frac{1}{T}\sum_{t'=1}^T d_{t'}(\xi_{2,T+1} - \xi_{2t'})(\xi_{2t} - \xi_{2t'})]}{\frac{1}{T}\left(\frac{1}{T}\sum_{t'=1}^T d_{t'}\right) + \lambda^2\left[\frac{1}{T}\sum_{t'=1}^T d_{t'}\xi_{2t'}^2 \frac{1}{T}\sum_{t'=1}^T d_{t'} - \left(\frac{1}{T}\sum_{t'=1}^T d_{t'}\xi_{2t'}\right)^2\right]} \tag{37}$$

where

$$d_t = \left[\lambda^2 \xi_{2t}(1-\xi_{2t}) + q^2 + (1-q^2)\xi_{2t}\right]^{-1}$$

To perform the large sample approximation we need to establish that $\frac{1}{T}\sum_{t=1}^T d_t < \infty$, $\frac{1}{T}\sum_{t=1}^T \xi_{2t} d_t < \infty$ and $\frac{1}{T}\sum_{t=1}^T \xi_{2t}^2 d_t < \infty$. Proving the first of these relations implies the other two, since $0 \leqslant \xi_{2t} \leqslant 1$. Define $a_t = \frac{1}{d_t}$. We then need to prove that $a_t > 0$. The only scenario where $a_t = 0$ is when $\xi_{2t} = 0$ and $q^2 = 0$, so the only restriction that we must impose to obtain $a_t > 0$ is that $q^2 > 0$. Then

$$\frac{1}{T}\sum_{t=1}^T d_t = \frac{1}{T}\sum_{t=1}^T \frac{1}{a_t} \leqslant \frac{1}{T} T \frac{1}{a_{\min}} = \frac{1}{a_{\min}} < \infty$$

where $a_{\min}$ is the minimum value of a_t over $t = 1, 2, \ldots, T$.

Denote $\bar{d} = \frac{1}{T}\sum_{t=1}^T d_t$, $\overline{d\xi} = \frac{1}{T}\sum_{t=1}^T d_t \xi_{2t}$, and $\overline{d\xi^2} = \frac{1}{T}\sum_{t=1}^T d_t \xi_{2t}^2$, then (37) can be written as

$$\begin{aligned} w_t &= \frac{1}{T} d_t \left[\frac{\frac{1}{T}}{\frac{1}{T}\bar{d} + \lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)} + \frac{\lambda^2(\xi_{2t}\xi_{2,T+1}\bar{d} - \xi_{2t}\overline{d\xi} - \xi_{2,T+1}\overline{d\xi} + \overline{d\xi^2})}{\frac{1}{T}\bar{d} + \lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)}\right] \\ &= \frac{1}{T} d_t \left[\frac{1}{T}\frac{1}{\lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)}\frac{1}{1+\frac{\theta}{T}} + \frac{\lambda^2(\xi_{2t}\xi_{2,T+1}\bar{d} - \xi_{2t}\overline{d\xi} - \xi_{2,T+1}\overline{d\xi} + \overline{d\xi^2})}{\lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)}\frac{1}{1+\frac{\theta}{T}}\right] \\ &= \frac{1}{T} d_t \frac{\lambda^2(\xi_{2t}\xi_{2,T+1}\bar{d} - \xi_{2t}\overline{d\xi} - \xi_{2,T+1}\overline{d\xi} + \overline{d\xi^2})}{\lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)} + \mathcal{O}\left(T^{-2}\right) \end{aligned}$$

where $\theta = \frac{\bar{d}}{\lambda^2\left(\overline{d\xi^2}\bar{d} - \overline{d\xi}^2\right)} = \frac{1}{\lambda^2\sum_{t=1}^T \tilde{d}_t(\xi_{2t} - \frac{1}{T}\sum_{t'=1}^T \tilde{d}_{t'}\xi_{2t'})^2}$ where $\tilde{d}_t = d_t/\sum_{t'} d_{t'}$. The numerator is nonzero unless for the trivial case when ξ_{2t} is constant for all t. Using this and the result that $\bar{d}$, $\overline{d\xi}$ and $\overline{d\xi^2}$ are finite for any T proves that we can apply the expansion in terms of θ/T. Dividing w_t by $\sum_{t=1}^T d_t$ yields (23).

A.2.2 Weights and MSFE for standard Markov switching model

The Markov switching weights can be written as

$$
\begin{aligned}
\mathbf{w}_{\mathrm{MS}} &= \frac{\xi_{1,T+1}\boldsymbol{\xi}_1}{\sum_{t=1}^{T}\xi_{1t}} + \frac{\xi_{2,T+1}\boldsymbol{\xi}_2}{\sum_{t=1}^{T}\xi_{2t}} \\
&= \frac{1}{T}\frac{\xi_{2,T+1}\boldsymbol{\xi}_2}{\bar{\xi}_2} + \frac{1}{T}\frac{(1-\xi_{2,T+1})(\boldsymbol{\iota}-\boldsymbol{\xi}_2)}{(1-\bar{\xi}_2)} \\
&= \frac{1}{T}\frac{1}{\bar{\xi}_2(1-\bar{\xi}_2)}(\xi_{2,T+1}\boldsymbol{\xi}_2(1-\bar{\xi}_2+\bar{\xi}_2)+\bar{\xi}_2\boldsymbol{\iota}-\bar{\xi}_2\xi_{2,T+1}\boldsymbol{\iota}-\bar{\xi}_2\boldsymbol{\xi}_2) \\
&= \frac{1}{T}\frac{1}{\bar{\xi}_2(1-\bar{\xi}_2)}(\xi_{2,T+1}-\bar{\xi}_2)(\boldsymbol{\xi}_2-\bar{\xi}_2\boldsymbol{\iota})+\bar{\xi}_2(1-\bar{\xi}_2) \\
&= \frac{1}{T}+\frac{1}{T}\frac{(\xi_{2,T+1}-\bar{\xi}_2)(\boldsymbol{\xi}_2-\bar{\xi}_2\boldsymbol{\iota})}{\bar{\xi}_2(1-\bar{\xi}_2)} \qquad (38)
\end{aligned}
$$

For a general vector of weights $\mathbf{w}$, subject to $\sum_{t=1}^{T} w_t = 1$, and assuming a constant error variance, we have the following MSFE

$$
\begin{aligned}
\mathrm{E}[\sigma^{-2}e_{T+1}^2] &= 1+\lambda^2\xi_{2,T+1}+\mathbf{w}'\mathbf{M}\mathbf{w}-2\lambda^2\mathbf{w}'\boldsymbol{\xi}\xi_{2,T+1} \\
&= 1+\lambda^2\xi_{2,T+1}+\lambda^2(\mathbf{w}'\boldsymbol{\xi})^2+\mathbf{w}'\mathbf{D}\mathbf{w}-2\lambda^2\mathbf{w}'\boldsymbol{\xi}\xi_{2,T+1}
\end{aligned} \qquad (39)
$$

where $\mathbf{D} = (1+\lambda^2\sigma_\xi^2)\mathbf{I}$.

Using (38) we have that

$$
\begin{aligned}
\mathbf{w}_{\mathrm{MS}}'\boldsymbol{\xi} &= \bar{\xi}_2 + \frac{\xi_{2,T+1}-\bar{\xi}_2}{(1-\bar{\xi}_2)\bar{\xi}_2}\left(\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 - T\bar{\xi}_2^2\right) \\
&= \bar{\xi}_2 + \frac{\xi_{2,T+1}-\bar{\xi}_2}{(1-\bar{\xi}_2)\bar{\xi}_2}\left[\bar{\xi}_2(1-\bar{\xi}_2)-\sigma_\xi^2\right] \\
&= \xi_{2,T+1} - \frac{\xi_{2,T+1}-\bar{\xi}_2}{\bar{\xi}_2(1-\bar{\xi}_2)}\sigma_\xi^2
\end{aligned}
$$

where we have used (24), and

$$
\mathbf{w}_{\mathrm{MS}}'\mathbf{D}\mathbf{w}_{\mathrm{MS}} = (1+\lambda^2\sigma_\xi^2)\left\{\frac{1}{T}+\frac{(\xi_{2,T+1}-\bar{\xi}_2)^2}{T\bar{\xi}_2^2(1-\bar{\xi}_2)^2}\left[\bar{\xi}_2(1-\bar{\xi}_2)-\sigma_\xi^2\right]\right\}
$$

So that the MSFE is

$$
\begin{aligned}
\mathrm{E}[\sigma^{-2}e^2_{T+1}]_{\mathrm{MS}} &= 1 + \lambda^2\xi_{2,T+1} + \lambda^2\left[\xi^2_{2,T+1} - 2\frac{\xi_{2,T+1}(\xi_{2,T+1}-\bar{\xi}_2)\sigma^2_\xi}{\bar{\xi}_2(1-\bar{\xi}_2)} + \frac{(\xi_{2,T+1}-\bar{\xi}_2)^2\sigma^4_\xi}{\bar{\xi}^2_2(1-\bar{\xi}_2)^2}\right] \\
&\quad - \lambda^2\left[2\xi^2_{2,T+1} - 2\frac{\xi_{2,T+1}(\xi_{2,T+1}-\bar{\xi}_2)\sigma^2_\xi}{\bar{\xi}_2(1-\bar{\xi}_2)}\right] \\
&\quad + (1+\lambda^2\sigma^2_\xi)\frac{1}{T}\left\{1 + \frac{(\xi_{2,T+1}-\bar{\xi}_2)^2}{\bar{\xi}^2_2(1-\bar{\xi}_2)^2}\left[\bar{\xi}_2(1-\bar{\xi}_2) - \sigma^2_\xi\right]\right\} \\
&= 1 + \lambda^2\xi_{2,T+1}(1-\xi_{2,T+1}) + \lambda^2\frac{(\xi_{2,T+1}-\bar{\xi}_2)^2\sigma^4_\xi}{\bar{\xi}^2_2(1-\bar{\xi}_2)^2} \\
&\quad + (1+\lambda^2\sigma^2_\xi)\frac{1}{T}\left\{1 + \frac{(\xi_{2,T+1}-\bar{\xi}_2)^2}{\bar{\xi}^2_2(1-\bar{\xi}_2)^2}\left[\bar{\xi}_2(1-\bar{\xi}_2) - \sigma^2_\xi\right]\right\} \\
&= 1 + \lambda^2\xi_{2,T+1}(1-\xi_{2,T+1}) + (1+\lambda^2\sigma^2_\xi)\frac{1}{T} \\
&\quad + \frac{(\xi_{2,T+1}-\bar{\xi}_2)^2}{\bar{\xi}^2_2(1-\bar{\xi}_2)^2}\left\{\lambda^2\sigma^4_\xi + (1+\lambda^2\sigma^2_\xi)\frac{1}{T}\left[\bar{\xi}_2(1-\bar{\xi}_2) - \sigma^2_\xi\right]\right\}
\end{aligned}
$$

A.2.3 MSFE for Markov switching model using optimal weights

Equation (22) for an arbitrary number of states is derived as follows

$$
\begin{aligned}
\mathrm{E}[\sigma^{-2}e^2_{T+1}] &= (\boldsymbol{\iota}'\mathbf{M}^{-1}\boldsymbol{\iota})^{-1}(1-\boldsymbol{\iota}'\mathbf{M}^{-1}\tilde{\boldsymbol{\xi}}\tilde{\xi}_{T+1})^2 + \\
&\quad + \sum_{j=2}^{m}\lambda^2_j\xi_{j,T+1} - \tilde{\xi}^2_{T+1}\tilde{\boldsymbol{\xi}}'\mathbf{M}^{-1}\tilde{\boldsymbol{\xi}} + \sum_{j=1}^{m}q^2_j\xi_{j,T+1} \\
&= \frac{1+\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}}{\boldsymbol{\iota}'\mathbf{D}^{-1}\boldsymbol{\iota}(1+\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}) - (\boldsymbol{\iota}\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}})^2}\left[1 + \frac{\tilde{\xi}^2_{T+1}(\boldsymbol{\iota}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}})^2}{(1+\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}})^2} + \right. \\
&\quad \left. -2\frac{\tilde{\xi}_{T+1}\boldsymbol{\iota}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}}{1+\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}}\right] + \tilde{\xi}^2_{T+1} - \frac{\tilde{\xi}^2_{T+1}\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}}{1+\tilde{\boldsymbol{\xi}}'\mathbf{D}^{-1}\tilde{\boldsymbol{\xi}}} + \frac{1}{d_{T+1}} \\
&= \frac{1+\sum_{t=1}^T\tilde{\xi}^2_t - 2\tilde{\xi}_{T+1}\sum_{t=1}^T d_t\tilde{\xi}_t + \tilde{\xi}^2_{T+1}\sum_{t=1}^T d_t}{\sum_{t=1}^T d_t(1+\sum_{t'=1}^T d_{t'}\tilde{\xi}^2_{t'}) - (\sum_{t=1}^T d_t\tilde{\xi}_t)^2} + \frac{1}{d_{T+1}} \\
&= \frac{w_{T+1}}{d_{T+1}} + \frac{1}{d_{T+1}} \\
&= \frac{1}{d_{T+1}}(1+w_{T+1})
\end{aligned}
$$

A.2.4 Derivation of (32)

To save on notation, in the following we use $p(s_{jt}|s_{i,t+m},\Omega_T)$ to write $p(s_{jt} = 1|s_{i,t+m} = 1,\Omega_T)$. To derive (32), take for example a three state model and calculate

$$\begin{aligned}
p(s_{jt}|s_{i,t+3},\Omega_T) &= \sum_{k=0}^{2} p(s_{jt}|s_{k,t+1},s_{i,t+3},\Omega_T)p(s_{k,t+1}|s_{i,t+3},\Omega_T) \\
&= \sum_{k=0}^{2} p(s_{jt}|s_{k,t+1},\Omega_t)\sum_{l=0}^{2} p(s_{k,t+1}|s_{l,t+2},\Omega_{t+1})p(s_{l,t+2}|s_{i,t+3},\Omega_{t+2}) \\
&= \sum_{k=0}^{2} \frac{p_{jk}p(s_{jt}|\Omega_t)}{p(s_{k,t+1}|\Omega_t)}\sum_{l=0}^{2} \frac{p_{kl}p(s_{k,t+1}|\Omega_{t+1})}{p(s_{l,t+2}|\Omega_{t+1})}\frac{p_{li}p(s_{l,t+2}|\Omega_{t+2})}{p(s_{i,t+3}|\Omega_{t+2})} \\
&= \frac{p(s_{jt}|\Omega_t)}{p(s_{i,t+3}|\Omega_{t+2})}\sum_{k=0}^{2}\sum_{l=0}^{2} p_{jk}a_{t+1}^{k}p_{kl}a_{t+2}^{l}p_{li} \\
&= \frac{p(s_{jt}|\Omega_t)}{p(s_{i,t+3}|\Omega_{t+2})}\left(\mathbf{P}'\mathbf{A}_{t+1}\mathbf{P}'\mathbf{A}_{t+2}\mathbf{P}'\right)_{j,i}
\end{aligned}$$

where $a_{t+1}^{k} = \frac{p(s_{k,t+1}=1|\Omega_{t+1})}{p(s_{k,t+1}=1|\Omega_t)}$. On the second line we use that the regime s_t depends on future observations only through s_{t+1}.

A.3 The MSFE with exogenous regressors

The expected MSFE is given by

$$\begin{aligned}
\mathrm{E}(\sigma_m^{-2}e_{T+1}^2) &= \sum_{i=1}^{m}\mathrm{E}(s_{i,T+1})\mathbf{x}_{T+1}'\boldsymbol{\Lambda}_{ij}\mathbf{x}_{T+1} + \sum_{i=1}^{m}\mathrm{E}(s_{i,T+1})q_i^2\varepsilon_{T+1}^2 \\
&+\mathbf{x}_{T+1}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{E}[(\mathbf{X}'\mathbf{W}\mathbf{S}_i\mathbf{X})\boldsymbol{\Lambda}_{ij}(\mathbf{X}'\mathbf{S}_j\mathbf{W}\mathbf{X})](\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{x}_{T+1} \\
&+\mathbf{x}_{T+1}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\sum_{i=1}^{m}q_i^2\mathbf{X}'\mathbf{W}\mathrm{E}(\mathbf{S}_i)\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{x}_{T+1} \\
&-2\mathbf{x}_{T+1}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{E}\left(\mathbf{X}'\mathbf{W}\mathbf{S}_i\mathbf{X}\boldsymbol{\Lambda}_{ij}s_{j,T+1}\right)\mathbf{x}_{T+1}
\end{aligned}$$

B Additional Monte Carlo results

B.1 Monte Carlo results for $T = 50$ and 100

Tables 8 and 9 report the results for the mean only model for $T = 50$ and 100 and complement the results in Tables 3 and 4 in the paper.

B.2 Exogenous regressors

In this set of experiments, we use the set up of the experiments of the mean only, two state model and add an exogenous regressor to the model, such

Table 8: Monte Carlo results: two states, mean only models

		$T = 50$			$T = 100$		
λ	$\tilde{\sigma}^2_{\hat{\xi}\|T}$	$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$	$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$
		Switches in mean					
1	0.0-0.1	0.982	1.008	1.008	0.993	1.005	1.005
	0.1-0.2	0.991	1.026	1.030	0.997	1.013	1.022
	0.2-0.3	0.996	1.034	1.039	0.999	1.019	1.032
	0.3-0.4	0.999	1.036	1.042	1.000	1.024	1.037
2	0.0-0.1	0.996	1.009	1.017	0.999	1.005	1.023
	0.1-0.2	1.001	1.009	1.025	1.002	0.994	1.034
	0.2-0.3	1.004	0.983	1.002	1.003	0.977	1.004
	0.3-0.4	1.005	0.961	0.977	1.004	0.960	0.973
3	0.0-0.1	1.000	0.997	1.009	1.000	0.997	1.022
	0.1-0.2	1.004	0.969	0.999	1.005	0.961	0.993
	0.2-0.3	1.007	0.926	0.950	1.007	0.920	0.944
	0.3-0.4	1.009	0.890	0.907	1.007	0.892	0.912
		Switches in mean and variance ($q^2 = 2$)					
1	0.0-0.1	0.984	1.001	1.002	0.992	1.001	1.001
	0.1-0.2	0.990	1.016	1.018	0.996	1.009	1.013
	0.2-0.3	0.996	1.029	1.032	0.999	1.014	1.021
	0.3-0.4	1.000	1.028	1.034	1.001	1.018	1.026
2	0.0-0.1	0.993	1.009	1.011	0.998	1.005	1.019
	0.1-0.2	0.999	1.015	1.028	1.002	0.999	1.030
	0.2-0.3	1.002	1.003	1.018	1.003	0.992	1.021
	0.3-0.4	1.006	0.983	0.998	1.003	0.987	1.003
3	0.0-0.1	0.998	1.003	1.016	1.000	0.999	1.027
	0.1-0.2	1.004	0.985	1.011	1.003	0.980	1.025
	0.2-0.3	1.007	0.953	0.971	1.007	0.946	0.962
	0.3-0.4	1.009	0.929	0.942	1.007	0.920	0.939

Note: The table reports the ratio of the MSFE of the optimal weights to that of the Markov switching weights. $y_t = \beta_1 s_{1t} + \beta_2 s_{2t} + (\sigma_1 s_{1t} + \sigma_2 s_{2t})\varepsilon_t$ where $\varepsilon_t \sim \mathrm{N}(0,1)$, $\sigma_2^2 = 0.25$, $q^2 = \sigma_1^2/\sigma_2^2$. Column labels: $\lambda = (\beta_2 - \beta_1)/\sigma_2$, $\tilde{\sigma}^2_{\hat{\xi}|T}$ is the normalized variance in of the smoothed probability vector (35). $w_{\hat{s}}$: forecasts from weights based on estimated parameters and state probabilities. $w_{\hat{\xi}}$: forecasts from weights conditional on state probabilities. $w_{\hat{\mathbf{M}}}$ are the weights based on numerically inverting $\hat{\mathbf{M}}$.

Table 9: Monte Carlo results: three states, intercept only models

		$T = 50$			$T = 100$		
$\{\lambda_{31}, \lambda_{21}\}$	$\tilde{\sigma}^2_{\hat{\xi}_{\lvert T}}$	$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$	$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$
{2,1}	0.0-0.1	0.996	1.035	1.033	0.998	1.025	1.027
	0.1-0.2	0.998	1.033	1.037	0.999	1.027	1.046
	0.2-0.3	0.999	1.027	1.032	1.000	1.012	1.027
	0.3-0.4	1.001	1.017	1.025	1.001	1.007	1.018
{3,1}	0.0-0.1	0.998	1.020	1.016	0.999	1.011	1.026
	0.1-0.2	1.000	1.011	1.013	1.001	0.998	1.013
	0.2-0.3	1.002	0.991	0.993	1.002	0.971	0.986
	0.3-0.4	1.004	0.962	0.967	1.003	0.939	0.953
{3.5,2}	0.0-0.1	0.999	1.014	1.013	1.000	1.009	1.013
	0.1-0.2	1.000	1.004	1.003	1.001	0.994	1.008
	0.2-0.3	1.002	0.983	0.988	1.002	0.964	0.979
	0.3-0.4	1.004	0.946	0.947	1.003	0.933	0.946

Note: The table reports the ratio of the MSFE of the optimal weights to that of the Markov switching weights. For details see Table 3.

that $\mathbf{x}_t = [1, z_t]'$, $z_t \sim \mathrm{N}(0, \sigma_z^2)$ and $\sigma_z = 1/2$ is chosen such that the centered R^2 is of a similar magnitude to the model with a constant only. The latter requirement is due to the fact that an important determinant of the quality of the forecasts is how well identified the states are and increasing the R^2 would improve the identification.

Table 10 displays the results for models that include an exogenous regressor. The optimal forecast are obtained by using an asymptotic approximation to the covariance matrix in (34). As the ratio of parameters to estimate versus the number of observations increases, the performance of the optimal weights $w_{\hat{s}}$ is less pronounced but the differences are generally small and the conclusions from experiments with mean only models carry over to the case of exogenous regressors.

B.3 Monte Carlo results for MSI and MSM models

Table 11 presents Monte Carlo results for the models that are frequently used in empirical applications. These models are the m-state Markov switching in intercept (MSI) and Markov switching in mean (MSM) models which include p lags of the dependent variable. We analyze the performance of the optimal weights for an MSI(2)-AR(2) and MSM(2)-AR(2) model. For both models, Table 11 shows that the improvements by using optimal weights are

Table 10: Monte Carlo results: two states, models with exogenous regressors

λ	$\tilde{\sigma}^2_{\hat{\xi}_{\mid T}}$	$T = 50$ $w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$	$T = 100$ $w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$
1	0.0-0.1	0.962	0.988	0.986	0.986	1.002	1.002
	0.1-0.2	0.973	1.021	1.001	0.993	1.014	1.018
	0.2-0.3	0.991	1.025	1.021	0.999	1.023	1.028
	0.3-0.4	0.995	1.030	1.028	1.000	1.026	1.032
2	0.0-0.1	0.990	1.000	1.002	0.999	1.003	1.013
	0.1-0.2	1.004	1.008	1.016	1.006	0.997	1.031
	0.2-0.3	1.011	0.999	1.013	1.011	0.978	1.009
	0.3-0.4	1.012	0.986	0.999	1.019	0.956	0.991
3	0.0-0.1	1.005	1.004	1.013	1.005	1.001	1.027
	0.1-0.2	1.018	0.998	1.026	1.020	0.979	1.033
	0.2-0.3	1.031	0.983	1.010	1.043	0.935	1.008
	0.3-0.4	1.020	0.969	0.991	1.051	0.919	0.958

Note: The table reports the ratio of the MSFE of the optimal asymptotic weights to that of the Markov switching weights. DGP: $y_t = x_t'\beta_1 + \sigma\,(x_t'\lambda s_{2t} + \varepsilon_t)$ where $\varepsilon_t \sim \text{NID}(0,1)$. Also $\sigma^2 = 0.25$, $\beta_1 = 1$ and $x_t = [1, z_t]$ where $z_t \sim \text{N}(0, 0.25)$. For the column labels see the footnote of Table 3.

consistent with the results for the Markov switching model with no lagged dependent variables. However, the additional parameter estimates imply noise that leads to slightly less pronounced differences in MSFE compared to the intercept only model.

Table 11: Monte Carlo results: MSI and MSM models

λ	$\tilde{\sigma}^2_{\hat{\xi}_{\mid T}}$	$T = 50$			$T = 100$		
		$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$	$w_{\hat{s}}$	$w_{\hat{\xi}}$	$w_{\hat{\mathbf{M}}}$
MSI							
1	0.0-0.1	0.988	1.008	1.002	0.994	1.006	1.006
	0.1-0.2	0.994	1.019	1.016	0.997	1.016	1.020
	0.2-0.3	0.997	1.018	1.018	0.999	1.017	1.026
2	0.0-0.1	0.997	1.005	1.006	0.999	1.003	1.020
	0.1-0.2	1.000	1.005	1.017	1.002	0.994	1.030
	0.2-0.3	1.003	0.993	1.007	1.003	0.985	1.018
3	0.0-0.1	1.000	0.999	1.004	1.000	0.999	1.012
	0.1-0.2	1.004	0.983	1.026	1.004	0.972	1.020
	0.2-0.3	1.005	0.970	0.986	1.005	0.944	0.981
MSM							
1	0.0-0.1	0.991	1.010	1.008	0.994	1.019	1.020
	0.1-0.2	0.994	1.023	1.017	0.996	1.033	1.042
	0.2-0.3	0.995	1.029	1.037	0.998	1.033	1.043
2	0.0-0.1	0.996	1.011	1.009	0.999	1.012	1.028
	0.1-0.2	0.998	1.015	1.019	1.000	1.010	1.034
	0.2-0.3	0.999	1.015	1.022	1.001	1.007	1.024
3	0.0-0.1	0.999	1.004	1.004	1.000	1.002	1.015
	0.1-0.2	1.000	1.002	1.013	1.002	0.991	1.012
	0.2-0.3	1.000	1.006	1.007	1.003	0.974	0.983

Note: The table reports the ratio of the MSFE of the optimal weights to that of the Markov switching weights. DGP MSI: $y_t = \beta_1 s_{1t} + \beta_2 s_{2t} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \sigma\varepsilon_t$ where $\varepsilon_t \sim \mathrm{N}(0,1)$. DGP MSM: $y_t = \beta_1 s_{1t} + \beta_2 s_{2t} + \phi_1(y_{t-1} - \beta_{s_{t-1}}) + \phi_2(y_{t-2} - \beta_{s_{t-2}}) + \sigma\varepsilon_t$, $\sigma^2 = 0.25$, $\phi_1 = 0.4$, $\phi_2 = -0.3$. Column labels as in Table 3.