Tractable Bayesian Estimation of Smooth Transition Vector Autoregressive Models

-

SOME DERIVATIONS, updated file as of 2025

Martin Bruns[*] Michele Piffer[†]

December 1, 2025

[*]University of East Anglia, UK, School of Economics.
[†]King's Business School, King's College London, UK. Corresponding author e-mail: m.b.piffer@gmail.com

Contents

A Model and inference

This section derives the likelihood function and the posterior distribution of the heteroskedastic smooth transition model outlined in the paper. By construction, if one limits the model to the special case of homoskedasticity ($\psi = 0$) and calibrates rather than estimates γ and c, then the derivations of the posterior distribution are identical to the standard case of linear Bayesian VAR models (see for instance Canova, 2007 and Koop and Korobilis, 2010 for a reference to the latter). In this appendix, we build on standard derivations from linear Bayesian VAR models and extend the analysis to allow for inference on (γ, c, ψ).

A.1 Likelihood function

The model used in the paper is

$$\boldsymbol{y}_t = \quad g(z_{t-1}, \gamma, c)\big(\Pi_a \boldsymbol{x}_{t-1} + B_a \boldsymbol{m}_t + D_a \boldsymbol{q}_t\big) + \qquad \text{(A.1a)}$$

$$+ (1 - g(z_{t-1}, \gamma, c))\big(\Pi_b \boldsymbol{x}_{t-1} + B_b \boldsymbol{m}_t + D_b \boldsymbol{q}_t\big) + \boldsymbol{u}_t,$$

$$\boldsymbol{u}_t | z_{t-1} \sim \; N(\boldsymbol{0}, \Sigma_t), \qquad \text{(A.1b)}$$

$$\Sigma_t = h(z_{t-1}, \gamma, c, \psi) \cdot \Sigma, \qquad \text{(A.1c)}$$

$$h(z_{t-1}, \gamma, c, \psi) = g(z_{t-1}, \gamma, c) + e^{\psi}\big(1 - g(z_{t-1}, \gamma, c)\big), \qquad \text{(A.1d)}$$

$$g(z_{t-1}, \gamma, c) = \frac{1}{1 + e^{-\gamma(z_{t-1} - c)}}, \quad \gamma > 0, \qquad \text{(A.1e)}$$

where z_{t-1} is potentially a function of the endogenous variables of the model until time $t-1$. For $\psi = 0$ the model features homoskedasticity, with $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma)$. The model can be rewritten as

$$\boldsymbol{y}_t = \Theta \boldsymbol{w}_t + \boldsymbol{u}_t, \qquad \text{(A.2)}$$

with $\boldsymbol{u}_t$ heteroskedastic, or in the homoskedastic-equivalent form

$$\frac{\boldsymbol{y}_t}{\sqrt{h(z_{t-1},\gamma,c,\psi)}} = \Theta \frac{\boldsymbol{w}_t}{\sqrt{h(z_{t-1},\gamma,c,\psi)}} + \frac{\boldsymbol{u}_t}{\sqrt{h(z_{t-1},\gamma,c,\psi)}}, \tag{A.3}$$

with $\frac{\boldsymbol{u}_t}{\sqrt{h(z_{t-1},\gamma,c,\psi)}} \sim N(\boldsymbol{0},\Sigma)$, and

$$\Theta = [\Pi_a, B_a, D_a, \Pi_b, B_b, D_b], \tag{A.4}$$

$$\boldsymbol{x}_{t-1} = \begin{pmatrix} \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{pmatrix} \qquad \underset{half}{\boldsymbol{w}_t} = \begin{pmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{m}_t \\ \boldsymbol{q}_t \end{pmatrix}, \tag{A.5}$$

$$\boldsymbol{w}_t(\gamma,c) = \begin{pmatrix} g(z_{t-1},\gamma,c)\,\underset{half}{\boldsymbol{w}_t} \\ \big(1-g(z_{t-1},\gamma,c)\big)\,\underset{half}{\boldsymbol{w}_t} \end{pmatrix} = \begin{pmatrix} g(z_{t-1},\gamma,c) \begin{pmatrix} \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p} \\ \boldsymbol{m}_t \\ \boldsymbol{q}_t \end{pmatrix} \\ \big(1-g(z_{t-1},\gamma,c)\big) \begin{pmatrix} \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p} \\ \boldsymbol{m}_t \\ \boldsymbol{q}_t \end{pmatrix} \end{pmatrix}, \tag{A.6}$$

with $\boldsymbol{y}_t$ of dimensions $k \times 1$, $\boldsymbol{m}_t$ of dimensions $k_m \times 1$, $\boldsymbol{q}_t$ of dimensions $k_q \times 1$, $\boldsymbol{x}_t$ of dimensions $kp \times 1$, $m = kp + k_m + k_q$, $\underset{half}{\boldsymbol{w}_t}$ of dimensions $m \times 1$, $\boldsymbol{w}_t(\gamma,c)$ of dimensions $2m \times 1$, Θ of dimensions $k \times 2m$.

The likelihood function of the model can be factorized as

$$p(Y|\Theta,\Sigma,\gamma,c,\psi) = \prod_{t=1}^{T} p\big(\boldsymbol{y}_t|\boldsymbol{w}_t, z_{t-1}, \Theta, \Sigma, \gamma, c, \psi\big), \tag{A.7}$$

with

$$\begin{aligned} p\big(\boldsymbol{y}_t|\boldsymbol{w}_t, z_{t-1}, \Theta, \Sigma, \gamma, c, \psi\big) = \quad & (2\pi)^{-\frac{k}{2}}\Big|\det\Big(h(z_{t-1},\gamma,c,\psi)\cdot\Sigma\Big)\Big|^{-\frac{1}{2}}\cdot \\ & \cdot\, e^{-\frac{1}{2}(\boldsymbol{y}_t-\Theta\boldsymbol{w}_t)'\left(h(z_{t-1},\gamma,c,\psi)\cdot\Sigma\right)^{-1}(\boldsymbol{y}_t-\Theta\boldsymbol{w}_t)}. \end{aligned} \tag{A.8}$$

To simplify the derivations, rewrite model (A.2) in compact form as

$$Y = \Theta W(\gamma,c) + U, \tag{A.9}$$

with $Y = \big[\boldsymbol{y}_1, ..., \boldsymbol{y}_t, ..., \boldsymbol{y}_T\big]$ of dimensions $k \times T$,

$$W(\gamma,c) = \big[\boldsymbol{w}_1(\gamma,c), ..., \boldsymbol{w}_t(\gamma,c), ..., \boldsymbol{w}_T(\gamma,c)\big] \tag{A.10}$$

$$= \left[\begin{pmatrix} g(z_0,\gamma,c)\,\underset{half}{\boldsymbol{w}_1} \\ \\ (1-g(z_0,\gamma,c))\,\underset{half}{\boldsymbol{w}_1} \end{pmatrix}, \cdots, \begin{pmatrix} g(z_{t-1},\gamma,c)\,\underset{half}{\boldsymbol{w}_t} \\ \\ (1-g(z_{t-1},\gamma,c))\,\underset{half}{\boldsymbol{w}_t} \end{pmatrix}, \cdots, \begin{pmatrix} g(z_{T-1},\gamma,c)\underset{half}{\boldsymbol{w}_T} \\ \\ (1-g(z_{T-1},\gamma,c))\underset{half}{\boldsymbol{w}_T} \end{pmatrix}\right], \tag{A.11}$$

of dimensions $2m \times T$ and $U = \big[\boldsymbol{u}_1, ..., \boldsymbol{u}_t, ..., \boldsymbol{u}_T\big]$ of dimensions $k \times T$. Then make use of the formula $\text{vec}(ABC) = (C' \otimes A)\cdot\text{vec}(B)$ (see **?**, mathematical appendix) with vec(.) the operator that stacks the columns of a matrix vertically and obtain

$$\tilde{\boldsymbol{y}} = Z(\gamma,c)\boldsymbol{\theta} + \tilde{\boldsymbol{u}}, \tag{A.12}$$

with $\tilde{\boldsymbol{y}} = \mathrm{vec}(Y)$ of dimensions $kT \times 1$,

$$Z(\gamma, c) = \big(W(\gamma, c)' \ \otimes \ I_k\big), \tag{A.13}$$

of dimensions $kT \times 2mk$, $\tilde{\boldsymbol{u}} = \mathrm{vec}(U)$ of dimensions $kT \times 1$ and $\boldsymbol{\theta} = \mathrm{vec}(\Theta)$ of dimensions $2mk \times 1$. Since $\tilde{\boldsymbol{u}} \sim N\big(\mathbf{0}, H(\gamma, c, \psi) \otimes \Sigma\big)$ with

$$H(\gamma, c, \psi) = \mathrm{diag}\big(h(z_0, \gamma, c, \psi),\ ..\ , h(z_{t-1}, \gamma, c, \psi),\ ..\ , h(z_{T-1}, \gamma, c, \psi)\big), \tag{A.14}$$

of dimensions $T \times T$, the likelihood function of the model is

$$p(Y|\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \propto |\det(H(\gamma, c, \psi) \otimes \Sigma)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)'\left(H(\gamma,c,\psi)\otimes\Sigma\right)^{-1}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)}, \tag{A.15}$$

$$\propto |\det(\Sigma)|^{-\frac{T}{2}} \cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}} \cdot e^{-\frac{1}{2}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)'\left(H(\gamma,c,\psi)\otimes\Sigma\right)^{-1}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)}. \tag{A.16}$$

Postmultiplying model (A.9) by $H(\gamma, c, \psi)^{-\frac{1}{2}}$ gives the equivalent homoskedastic model

$$YH(\gamma, c, \psi)^{-\frac{1}{2}} = \Theta W(\gamma, c) H(\gamma, c, \psi)^{-\frac{1}{2}} + UH(\gamma, c, \psi)^{-\frac{1}{2}}, \tag{A.17}$$

from which the Maximum Likelihood estimators of (Θ, Σ) conditioning on (γ, c, ψ) can be derived as

$$\hat{\Theta}(\gamma, c, \psi) = YH(\gamma, c, \psi)^{-1} W(\gamma, c)' \big(W(\gamma, c) H(\gamma, c, \psi)^{-1} W(\gamma, c)'\big)^{-1}, \tag{A.18}$$

$$\hat{\Sigma}(\gamma, c, \psi) = \frac{\big(Y - \hat{\Theta}(\gamma, c, \psi) W(\gamma, c)\big) H(\gamma, c, \psi)^{-1} \big(Y - \hat{\Theta}(\gamma, c, \psi) W(\gamma, c)\big)'}{T - 2m}. \tag{A.19}$$

Under the restriction $\psi = 0$ (homoskedasticity) the likelihood function simplifies to

$$p(Y|\boldsymbol{\theta}, \Sigma, \gamma, c) \propto |\det(\Sigma)|^{-\frac{T}{2}} \cdot e^{-\frac{1}{2}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)'\left(I_T \otimes \Sigma\right)^{-1}\left(\tilde{\boldsymbol{y}} - Z(\gamma,c)\boldsymbol{\theta}\right)}, \tag{A.20}$$

and the Maximum Likelihood estimators of (Θ, Σ) conditioning on (γ, c) simplify to

$$\hat{\Theta}(\gamma, c) = YW(\gamma, c)'\left(W(\gamma, c)W(\gamma, c)'\right)^{-1}, \tag{A.21}$$

$$\hat{\Sigma}(\gamma, c) = \frac{\left(Y - \hat{\Theta}(\gamma, c)W(\gamma, c)\right)\left(Y - \hat{\Theta}(\gamma, c)W(\gamma, c)\right)'}{T - 2m}. \tag{A.22}$$

A.2 Posterior distribution

Our analysis uses the joint prior distribution

$$p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi), \tag{A.23}$$

$$\propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma) \cdot \tilde{p}(\Sigma) \cdot \tilde{p}(\gamma, c, \psi), \tag{A.24}$$

$$\tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma) = N(\boldsymbol{\mu}, \bar{V}(\gamma) \otimes \Sigma), \tag{A.25}$$

$$\tilde{p}(\Sigma) = iW(S, d), \tag{A.26}$$

$$\tilde{p}(\gamma, c, \psi) = \text{free}. \tag{A.27}$$

The prior $\tilde{p}(\gamma, c, \psi)$ is freely set by the researcher, $\tilde{p}(\Sigma)$ is restricted to the inverse Wishart distribution, and $\tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma)$ is restricted to the Normal distribution under the standard conjugate Kronecker structure on the covariance matrix. The hyperparameters $(\boldsymbol{\mu}, \bar{V}, S, d)$ can potentially all depend on (γ, c, ψ). To simplify the notation, we limit the analysis to the case discussed in the paper, which features prior independence of Σ from (γ, c, ψ), and prior dependence of $\boldsymbol{\theta}$ only on γ, via $\bar{V}(\gamma)$. The indicator function $\mathrm{I}\{\boldsymbol{\theta}\}$ takes value one if additional restrictions introduced by the researcher on $\boldsymbol{\theta}$ are satisfied, for instance restrictions on the stationarity of the model (in the form of restrictions on the eigenvalues of Π_a, Π_b), or sign restrictions on the impact effects of the shocks of interest (in the form of restrictions on B_a, B_b).

The joint posterior distribution

$$p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y) \propto p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \cdot p(Y|\boldsymbol{\theta}, \Sigma, \gamma, c, \psi), \tag{A.28}$$

$$\propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \cdot p(Y|\boldsymbol{\theta}, \Sigma, \gamma, c, \psi), \tag{A.29}$$

$$\propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y), \tag{A.30}$$

is more conveniently explored by analyzing $\tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y)$ and then constructing the sampler to account for the additional restrictions coming from $\mathrm{I}\{\boldsymbol{\theta}\}$. It holds that

$$\tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y) \propto p(Y|\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \cdot \tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma) \cdot \tilde{p}(\Sigma) \cdot \tilde{p}(\gamma, c, \psi),$$

$$\propto |\det(\Sigma)|^{-\frac{T}{2}} \cdot \left(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\right)^{-\frac{k}{2}} \cdot e^{-\frac{1}{2}\left(\tilde{\boldsymbol{y}}-Z(\gamma,c)\cdot\boldsymbol{\theta}\right)'\left(H(\gamma,c,\psi)\otimes\Sigma\right)^{-1}\left(\tilde{\boldsymbol{y}}-Z(\gamma,c)\cdot\boldsymbol{\theta}\right)}. \tag{A.31}$$

$$\cdot |\det(\bar{V}(\gamma) \otimes \Sigma)|^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'(\bar{V}(\gamma)^{-1}\otimes\Sigma^{-1})(\boldsymbol{\theta}-\boldsymbol{\mu})} \cdot |\det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}tr\left[\Sigma^{-1}S\right]} \cdot \tilde{p}(\gamma, c, \psi),$$

$$\propto e^{-\frac{1}{2}\left\{\boldsymbol{\theta}'\left(W(\gamma,c)H(\gamma,c,\psi)^{-1}W(\gamma,c)'\otimes\Sigma^{-1}\right)\boldsymbol{\theta} \;-\; 2\boldsymbol{\theta}'\left(W(\gamma,c)H(\gamma,c,\psi)^{-1}\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}} \;+\; \tilde{\boldsymbol{y}}'\left(H(\gamma,c,\psi)^{-1}\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}\right\}}.$$

$$\cdot\, e^{-\frac{1}{2}\left\{\boldsymbol{\theta}'(\bar{V}(\gamma)^{-1}\otimes\Sigma^{-1})\boldsymbol{\theta} \;-\; 2\boldsymbol{\theta}'(\bar{V}(\gamma)^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} \;+\; \boldsymbol{\mu}'(\bar{V}(\gamma)^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}\right\}}.$$

$$\cdot |\det(\bar{V}(\gamma))|^{-\frac{k}{2}} \cdot |\det(\Sigma)|^{-m} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot \left(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\right)^{-\frac{k}{2}}. \tag{A.32}$$

$$\cdot |\det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}tr\left[\Sigma^{-1}S\right]} \cdot \tilde{p}(\gamma, c, \psi).$$

In what follows, we will often simplify notation and replace $\bar{V}(\gamma)$ with $\bar{V}$ unless when we must stress its dependence on γ. We will follow the same practice for $h(z_{t-1}, \gamma, c, \psi) = h_{t-1}$, $H(\gamma, c, \psi) = H$ and $W(\gamma, c) = W$.

Using formulas $\boldsymbol{a}'B\boldsymbol{a} + \boldsymbol{a}'C\boldsymbol{a} = \boldsymbol{a}'(B + C)\boldsymbol{a}$ and $\boldsymbol{a}'\boldsymbol{b} + \boldsymbol{a}'\boldsymbol{c} = \boldsymbol{a}'(\boldsymbol{b} + \boldsymbol{c})$, we can rewrite the term in the exponent as

$$\begin{aligned}
&\boldsymbol{\theta}'\big((\bar{V}^{-1}\otimes\Sigma^{-1})+(WH^{-1}W'\otimes\Sigma^{-1})\big)\boldsymbol{\theta} \;-\; 2\boldsymbol{\theta}'((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} \;+\\
&\quad+(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}})+\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\big[\Sigma^{-1}S\big],\\
&=\boldsymbol{\theta}'\big([\bar{V}^{-1}+WH^{-1}W']\otimes\Sigma^{-1}\big)\boldsymbol{\theta} \;-\; 2\boldsymbol{\theta}'((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} \;+\\
&\quad+(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}})+\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\big[\Sigma^{-1}S\big],\\
&=\boldsymbol{\theta}'\,\underbrace{\big([\bar{V}^{-1}+WH^{-1}W']\otimes\Sigma^{-1}\big)}_{V^{*-1}}\,\boldsymbol{\theta}\;+\\
&\;-2\boldsymbol{\theta}'\,\underbrace{([\bar{V}^{-1}+WH^{-1}W']\otimes\Sigma^{-1})}_{V^{*-1}}\cdot\\
&\;\cdot\underbrace{([\bar{V}^{-1}+WH^{-1}W']\otimes\Sigma^{-1})^{-1}((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}})}_{\boldsymbol{\mu}^*}+\\
&\;+\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\big[\Sigma^{-1}S\big]+\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*-\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*,\\
&=\boldsymbol{\theta}'\,\underbrace{\big(\bar{V}^{*-1}\otimes\Sigma^{-1}\big)}_{V^{*-1}}\,\boldsymbol{\theta}\;-2\boldsymbol{\theta}'\,\underbrace{(\bar{V}^{*-1}\otimes\Sigma^{-1})}_{V^{*-1}}\cdot\\
&\;\cdot\underbrace{(\bar{V}^{*-1}\otimes\Sigma^{-1})^{-1}((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}})}_{\boldsymbol{\mu}^*}+\\
&\;+\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\big[\Sigma^{-1}S\big]+\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*-\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*,\\
&=\boldsymbol{\theta}'\,\underbrace{\big(\bar{V}^{*-1}\otimes\Sigma^{-1}\big)}_{V^{*-1}}\,\boldsymbol{\theta}\;-2\boldsymbol{\theta}'\,\underbrace{(\bar{V}^{*-1}\otimes\Sigma^{-1})}_{V^{*-1}}\cdot\\
&\;\cdot\underbrace{(\bar{V}^{*}\bar{V}^{-1}\otimes I_k)^{-1}\boldsymbol{\mu}+(\bar{V}^{*}W\otimes I_k)\tilde{\boldsymbol{y}})}_{\boldsymbol{\mu}^*}+\\
&\;+\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\big[\Sigma^{-1}S\big]+\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*-\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*,
\end{aligned}$$

Hence

$$\tilde{p}(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi | Y) \propto \underbrace{|\det(\bar{V}^*)|^{-\frac{k}{2}} \cdot |\det(\Sigma)|^{-m} \cdot e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}^*)'(\bar{V}^{*-1}\otimes\Sigma^{-1})(\boldsymbol{\theta}-\boldsymbol{\mu}^*)}}_{\text{kernel of } \ \tilde{p}(\boldsymbol{\theta}|Y,\Sigma,\gamma,c,\psi)} \cdot$$

$$\cdot |\det(\Sigma)|^{-\frac{d+T+k+1}{2}} \cdot |\det(\bar{V}^*)|^{\frac{k}{2}} \cdot |\det(\bar{V}(\gamma))|^{-\frac{k}{2}}$$

$$\cdot\, e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}-\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^*+\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\left[\Sigma^{-1}S\right]\right\}}\cdot$$

$$\cdot\, \tilde{p}(\gamma, c, \psi) \cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}},$$

with

$$V^* = V^*(\Sigma, \gamma, c, \psi) = \bar{V}^*(\gamma, c, \psi) \otimes \Sigma, \tag{A.33}$$

$$\bar{V}^*(\gamma, c, \psi) = \Big[\bar{V}(\gamma)^{-1} + W(\gamma, c)H(\gamma, c, \psi)^{-1}W(\gamma, c)'\Big]^{-1}, \tag{A.34}$$

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}^*(\gamma, c, \psi) = (\bar{V}^* \otimes \Sigma)((\bar{V}^{-1} \otimes \Sigma^{-1})\boldsymbol{\mu} + (W(\gamma, c)H(\gamma, c, \psi)^{-1} \otimes \Sigma^{-1})\tilde{\boldsymbol{y}}), \tag{A.35}$$

$$= \big(\bar{V}^*(\gamma, c)\bar{V}(\gamma)^{-1} \otimes I_k\big)\boldsymbol{\mu} + \Big(\bar{V}^*(\gamma, c)W(\gamma, c)H(\gamma, c, \psi)^{-1} \otimes I_k\Big)\tilde{\boldsymbol{y}}, \tag{A.36}$$

$$= \big(\bar{V}^*(\gamma, c)\bar{V}(\gamma)^{-1} \otimes I_k\big)\boldsymbol{\mu} + \tag{A.37}$$

$$+ \Big(\bar{V}^*(\gamma, c)W(\gamma, c)H(\gamma, c, \psi)^{-1}W(\gamma, c)' \otimes I_k\Big)\mathrm{vec}(\hat{\Theta}(\gamma, c, \psi)).$$

It follows that

$$\tilde{p}(\Sigma, \gamma, c, \psi | Y) \propto |\det(\Sigma)|^{-\frac{d+T+k+1}{2}} \cdot |\det(\bar{V}^*(\gamma, c, \psi))|^{\frac{k}{2}}\cdot$$

$$\cdot\, e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'(H(\gamma,c,\psi)^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}-\boldsymbol{\mu}^{*\prime}V^*(\Sigma,\gamma,c,\psi)^{-1}\boldsymbol{\mu}^*+\boldsymbol{\mu}'(\bar{V}^{-1}(\gamma)\otimes\Sigma^{-1})\boldsymbol{\mu}+tr\left[\Sigma^{-1}S\right]\right\}}\cdot$$

$$\cdot\, \tilde{p}(\gamma, c, \psi) \cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}} \cdot |\det(\bar{V}(\gamma))|^{-\frac{k}{2}}. \tag{A.38}$$

One can now simplify $\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^{*}$. Note that

$$\begin{aligned}
\boldsymbol{\mu}^{*\prime}V^{*-1}\boldsymbol{\mu}^{*} &= \big((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} + (WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}\big)'(\bar{V}^{*}\otimes\Sigma)\big((\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu}+ \\
&\quad + (WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}\big), \\
&= \boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})(\bar{V}^{*}\otimes\Sigma)(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} + \\
&\quad + \tilde{\boldsymbol{y}}'(WH^{-1}\otimes\Sigma^{-1})'(\bar{V}^{*}\otimes\Sigma)(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} + \\
&\quad + 2\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})(\bar{V}^{*}\otimes\Sigma)(WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}, \\
&= \boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} \;+\; \tilde{\boldsymbol{y}}'(H^{-1}W'\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} + \\
&\quad + \;\; 2\boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}}.
\end{aligned}$$

We can now rewrite the terms in the exponent as

$$\begin{aligned}
&\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} \;+\; \boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} + \\
&\quad -\; \boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} \;-\; \tilde{\boldsymbol{y}}'(H^{-1}W'\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} \;-\; 2\boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} \\
&\quad + tr\big[\Sigma^{-1}S\big].
\end{aligned}$$

Now, using formulas

$$\begin{aligned}
\mathrm{vec}(A)'(D\otimes B)\mathrm{vec}(C) &= tr(A'BCD'), \\
tr(AB) &= tr(BA),
\end{aligned}$$

after defining $\underset{k\times 2m}{N}$ such that $\boldsymbol{\mu} = \mathrm{vec}(N)$, we get

$$\begin{aligned}
\tilde{\boldsymbol{y}}'(H^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} &= tr\big\{Y'\Sigma^{-1}YH^{-1}\big\} = tr\big\{\Sigma^{-1}YH^{-1}Y'\big\},\\
\boldsymbol{\mu}'(\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} &= tr\big\{N'\Sigma^{-1}N\bar{V}^{-1}\big\} = tr\big\{\Sigma^{-1}N\bar{V}^{-1}N'\big\},\\
\boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}\otimes\Sigma^{-1})\boldsymbol{\mu} &= tr\big\{N'\Sigma^{-1}N\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}\big\} = tr\big\{\Sigma^{-1}N\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}N'\big\},\\
\tilde{\boldsymbol{y}}'(H^{-1}W'\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} &= tr\big\{Y'\Sigma^{-1}YH^{-1}W'\bar{V}^{*}WH^{-1}\big\} = tr\big\{\Sigma^{-1}YH^{-1}W'\bar{V}^{*}WH^{-1}Y'\big\},\\
2\boldsymbol{\mu}'(\bar{V}^{-1}\bar{V}^{*}WH^{-1}\otimes\Sigma^{-1})\tilde{\boldsymbol{y}} &= 2tr\big\{N'\Sigma^{-1}YH^{-1}W'\bar{V}^{*}\bar{V}^{-1}\big\} = 2tr\big\{\Sigma^{-1}YH^{-1}W'\bar{V}^{*}\bar{V}^{-1}N'\big\}.
\end{aligned}$$

The term in the exponent hence rewrites as

$$tr\big\{\Sigma^{-1}\cdot[S+YH^{-1}Y'+N\bar{V}^{-1}N'-N\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}N'-YH^{-1}W'\bar{V}^{*}WH^{-1}Y'-2YH^{-1}W'\bar{V}^{*}\bar{V}^{-1}N']\big\}. \tag{A.39}$$

Finally, rewrite

$$\begin{aligned}
|\det(\bar{V}^{*})|^{\frac{k}{2}} &= |\det\big((\bar{V}^{-1}+WH^{-1}W')^{-1}\big)|^{\frac{k}{2}},\\
&= |\det(\bar{V}^{-1}+WH^{-1}W')|^{-\frac{k}{2}}.
\end{aligned}$$

$$\begin{aligned}
S(\gamma,c,\psi)^{*} = {} & S+YH^{-1}Y'+N\bar{V}^{-1}N'-N\bar{V}^{-1}\bar{V}^{*}\bar{V}^{-1}N'+\\
& -YH^{-1}W'\bar{V}^{*}WH^{-1}Y'-2YH^{-1}W'\bar{V}^{*}\bar{V}^{-1}N', \qquad \text{(A.40)}\\
= {} & S+YH(\gamma,c,\psi)^{-1}Y'+N\bar{V}^{-1}(\gamma)N'+\\
& -\big(\bar{V}(\gamma)^{-1}N'+W(\gamma,c)H(\gamma,c,\psi)^{-1}Y'\big)'\bar{V}(\gamma,c,\psi)^{*}\big(\bar{V}(\gamma)^{-1}N'+W(\gamma,c)H(\gamma,c,\psi)^{-1}Y'\big).
\end{aligned} \tag{A.41}$$

$$\tilde{p}(\Sigma,\gamma,c,\psi|Y) \propto |\det(\Sigma)|^{-\frac{(d+T)+k+1}{2}} \cdot e^{-\frac{1}{2}\left\{tr\left[\Sigma^{-1}S(\gamma,c,\psi)^*\right]\right\}}\cdot$$

$$\cdot\ \tilde{p}(\gamma,c,\psi)\cdot|\det(\bar{V}^*(\gamma,c,\psi))|^{\frac{k}{2}}\Big(\prod_{t=1}^{T}h(z_{t-1},\gamma,c,\psi)\Big)^{-\frac{k}{2}}\cdot|\det(\bar{V}(\gamma))|^{-\frac{k}{2}}, \tag{A.42}$$

$$\propto \underbrace{|\det(S(\gamma,c,\psi)^*)|^{\frac{d+T}{2}}\cdot|\det(\Sigma)|^{-\frac{(d+T)+k+1}{2}}\cdot e^{-\frac{1}{2}\left\{tr\left[\Sigma^{-1}S(\gamma,c,\psi)^*\right]\right\}}}_{\text{kernel of } \tilde{p}(\Sigma|Y,\gamma,c,\psi)}. \tag{A.43}$$

$$\cdot\underbrace{\tilde{p}(\gamma,c,\psi)\cdot|\det(S(\gamma,c,\psi)^*)|^{-\frac{d+T}{2}}\cdot\left|\det(\bar{V}^*(\gamma,c,\psi))\right|^{\frac{k}{2}}\Big(\prod_{t=1}^{T}h(z_{t-1},\gamma,c,\psi)\Big)^{-\frac{k}{2}}\cdot|\det(\bar{V}(\gamma))|^{-\frac{k}{2}}}_{\text{kernel of } \tilde{p}(\gamma,c,\psi|Y)}.$$

All in all, this confirms that the joint posterior distribution (A.28) satisfies

$$\tilde{p}(\boldsymbol{\theta}|Y,\Sigma,\gamma,c,\psi) = N\Big(\boldsymbol{\mu}^*(\gamma,c,\psi),\bar{V}^*(\gamma,c,\psi)\otimes\Sigma\Big), \tag{A.44}$$

$$\tilde{p}(\Sigma|Y,\gamma,c,\psi) = iW\big(S^*(\gamma,c,\psi),d^*\big), \tag{A.45}$$

$$\tilde{p}(\gamma,c,\psi|Y)\propto\tilde{p}(\gamma,c,\psi)\cdot|\det(S^*(\gamma,c,\psi))|^{-\frac{d+T}{2}}\cdot$$

$$\cdot\left|\det(\bar{V}^*(\gamma,c,\psi))\right|^{\frac{k}{2}}\cdot\Big(\prod_{t=1}^{T}h(z_{t-1},\gamma,c,\psi)\Big)^{-\frac{k}{2}}\cdot|\det(\bar{V}(\gamma))|^{-\frac{k}{2}}, \tag{A.46}$$

$$\text{with}\quad \boldsymbol{\mu}^*(\gamma,c,\psi)=\big(\bar{V}^*(\gamma,c,\psi)\bar{V}(\gamma)^{-1}\otimes I_k\big)\boldsymbol{\mu}+\Big(\bar{V}^*(\gamma,c,\psi)W(\gamma,c)H(\gamma,c,\psi)^{-1}\otimes I_k\Big)\tilde{\boldsymbol{y}}, \tag{A.47}$$

$$\bar{V}^*(\gamma,c,\psi)=\Big[\bar{V}(\gamma)^{-1}+W(\gamma,c)H(\gamma,c,\psi)^{-1}W(\gamma,c)'\Big]^{-1}, \tag{A.48}$$

$$S(\gamma,c,\psi)^*=S+YH(\gamma,c,\psi)^{-1}Y'+N\bar{V}(\gamma)^{-1}N'+ \tag{A.49}$$

$$-\big(\bar{V}(\gamma)^{-1}N'\ +W(\gamma,c)H(\gamma,c,\psi)^{-1}Y'\big)'\bar{V}^*(\gamma,c,\psi)\big(\bar{V}(\gamma)^{-1}N'+W(\gamma,c)H(\gamma,c,\psi)^{-1}Y'\big),$$

$$d^*=d+T. \tag{A.50}$$

In the special case of homoskedasticity ($\psi = 0$) the posterior distribution simplifies to

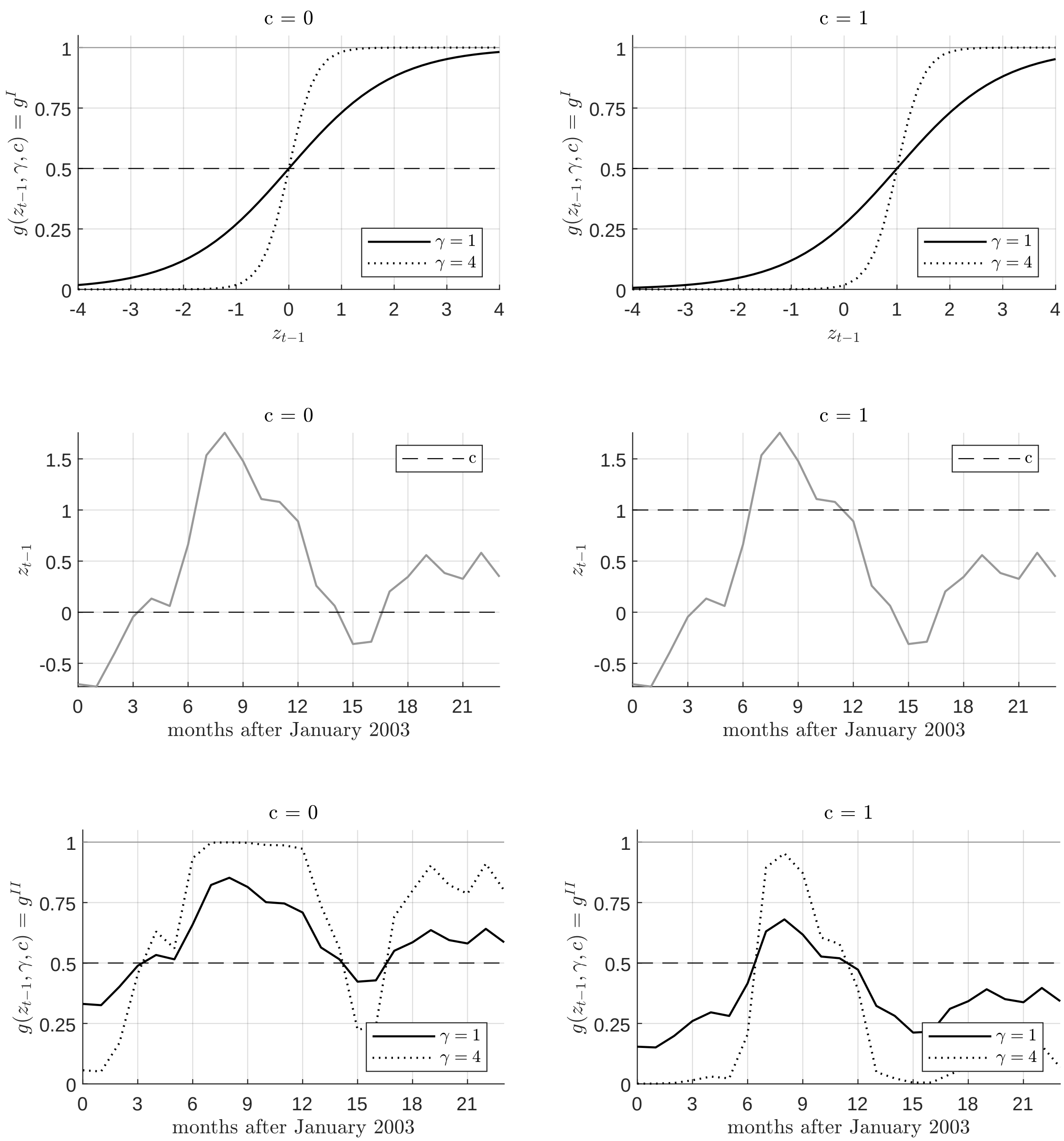
$$
\begin{aligned}
\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c) &= N\Big(\boldsymbol{\mu}^*(\gamma, c), \bar{V}^*(\gamma, c) \otimes \Sigma\Big), && \text{(A.51)} \\
\tilde{p}(\Sigma|Y, \gamma, c) &= iW\big(S^*(\gamma, c), d^*\big), && \text{(A.52)} \\
\tilde{p}(\gamma, c|Y) &\propto \tilde{p}(\gamma, c) \cdot |\det(S^*(\gamma, c))|^{-\frac{d+T}{2}} \cdot \left|\det\big(\bar{V}(\gamma)^{-1} + W(\gamma, c)W(\gamma, c)'\big)\right|^{-\frac{k}{2}} \cdot |\det(\bar{V}(\gamma))|^{-\frac{k}{2}}, && \text{(A.53)}
\end{aligned}
$$

$$
\begin{aligned}
\text{with} \quad \boldsymbol{\mu}^*(\gamma, c) &= \big(\bar{V}^*(\gamma, c)\bar{V}(\gamma)^{-1} \otimes I_k\big)\boldsymbol{\mu} + \Big(\bar{V}^*(\gamma, c)W(\gamma, c) \otimes I_k\Big)\tilde{\boldsymbol{y}}, && \text{(A.54)} \\
\bar{V}^*(\gamma, c) &= \Big[\bar{V}(\gamma)^{-1} + W(\gamma, c)W(\gamma, c)'\Big]^{-1}, && \text{(A.55)} \\
S(\gamma, c)^* &= S + YY' + N\bar{V}(\gamma)^{-1}N' + && \text{(A.56)} \\
&\quad - \big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)Y'\big)'\bar{V}^*(\gamma, c)\big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)Y'\big), \\
d^* &= d + T. && \text{(A.57)}
\end{aligned}
$$

B Some key properties of the transition function

In this section, we first illustrate some key properties of the transition function and then discuss their implications for inference in our model. The properties of the logistic transition functions in smooth transition models have been well documented in the literature, see, for instance, **?** and **?**. We summarize them here to guide the discussion on the implications when calibrating rather than estimating (γ, c), and the discussion of the posterior sampler. As for inference problems associated with the smooth transition model, see also **?**, who discusses identification issues in exponential (rather than logistic) smooth transition models.

Figure B.1: Properties of the logistic transition function



The (logistic) transition function used in the paper,

$$g(z_{t-1}, \gamma, c) \ = \frac{1}{1 + e^{-\gamma(z_{t-1} - c)}}, \quad \gamma > 0, \tag{B.58}$$

can be inspected in two ways: as a function of non-decreasing values of z_{t-1} (we will refer to it as g^I) or as a function of *data* on z_{t-1} (we will refer to it as g^{II}). Figure B.1 shows g^I in the top row, shows data on z_{t-1} from the application of the paper in the subperiod 2003M1-2004M12 in the middle row, and shows g^{II} in the bottom row evaluated for z_{t-1} from the middle row. The illustrative parameter values used are $\gamma = (1, 4)$ and $c = (0, 1)$. The figure helps appreciate that:

1) $g(z_{t-1} = c, \gamma, c) = 0.5, \ \forall \ \gamma$: g^I equals its halfpoint 0.5 whenever the transition function equals c. In the illustration from Figure B.1, this occurs whenever $z_{t-1} = 0$ (left column) or $z_{t-1} = 1$ (right column);

2) by construction, $g(z_{t-1}, \gamma, c) > 0.5$ (and is hence closer to the regime associated with $g(z_{t-1}, \gamma, c) = 1$) in periods in which $z_{t-1} > c$, and vice versa in periods in which $z_{t-1} < c$. In the left column of Figure B.1, z_{t-1} lies above $c = 0$ for 18 months out of 24, and below $c = 0$ for the remaining 6 months. By construction, the associated transition function g^{II} lies above and below 0.5 for the same number of months (18 and 6, respectively) irrespectively of the value of γ. By contrast, in the illustration from the right column, z_{t-1} lies above $c = 1$ for 5 months and below for the remaining 19 months, and the same holds for g^{II} relative to 0.5;

3) increasing values of c shift g^I rightwards, and do not shift g^{II}, but flatten it towards 0. Hence, for the same data on z_{t-1}, increasing values of c make the model interpret the data more through the extreme model associated with $g(z_{t-1}, \gamma, c) = 0$ than with $g(z_{t-1}, \gamma, c) = 1$. The opposite holds for decreasing values of c;

4) for any value of c, increasing values of γ make both g^I and g^{II} steeper. The number of periods for which g^{II} lies above or below 0.5 remains the same irrespectively of γ: 18 versus 6 months for $c = 0$, 5 versus 19 months for $c = 1$. However, the number of periods in which g^{II} lies above an arbitrary value $g^H > 0.5$ and below an arbitrary value $g^L < 0.5$ *both* increase. For instance, when $c = 0$, the number of periods for

which $g^{II} > 0.6$ increases from 9 to 16 when γ increases from 1 to 4. However, the same increase in γ also increases the number of periods for which $g^{II} < 0.4$, which moves from 2 to 5. Hence, holding c constant, movements in γ will affect the number of periods close to $g(z_{t-1}, \gamma, c) = 0$ and the number of periods close to $g(z_{t-1}, \gamma, c) = 1$ symmetrically.

5) very high values of γ imply that g^I approaches the indicator function $\mathrm{I}\{z_{t-1} \geq c\}$, leading g^{II} to take values only approximately close to 0 or 1. Hence, further increases in γ imply variations in g^{II} that are hard to detect.

The above features of the logistic transition function have the following implications for inference:

high values of γ are hard to identify, because $W(\gamma, c)$ becomes insensitive to further increases in γ, and so is $\tilde{p}(\boldsymbol{\theta}, \Sigma | Y, \gamma, c, \psi)$; because $\gamma \approx 0$ leads to a relatively flat transition function at 0.5 that is insensitive to variations in z_{t-1}, then

2) 2a) (Π_a, B_a, D_a) and (Π_b, B_b, D_b) are not separately identified, although $\underset{half}{\Pi}, \underset{half}{B}, \underset{half}{D}$ defined as

$$\underset{half}{\Pi} = \frac{\Pi_a + \Pi_b}{2} \qquad \underset{half}{B} = \frac{B_a + B_b}{2} \qquad \underset{half}{D} = \frac{D_a + D_b}{2}, \tag{B.59}$$

are identified;

2b) the expected value $\boldsymbol{\mu}^*$ of $\tilde{p}(\boldsymbol{\theta} | Y, \Sigma, \gamma, c, \psi)$ is still well-defined, but it must be computed from equation (A.36) rather than (A.37) because $\hat{\Theta}(\gamma, c, \psi)$ becomes degenerate;

2c) $\tilde{p}(\gamma, c, \psi|Y, \boldsymbol{\theta}, \Sigma)$ is still well-defined, as long as $\boldsymbol{\theta}$ is not degenerate, for instance because drawn using (A.37) rather than (A.36);

2d) $\tilde{p}(\gamma, c, \psi|Y)$ is still well-defined thanks to the prior dependence introduced via $\bar{V}(\gamma)$.

To see points 2a) and 2b), note that for $\gamma \approx 0$ the model approaches multicollinearity, because $g_t = \frac{1}{1+e^{-\gamma \cdot (z_{t-1}-c)}} \approx 0.5$, $\forall\ z_{t-1}$. This implies

$$\begin{aligned}\boldsymbol{y}_t = \quad & g(z_{t-1}, \gamma, c)\big(\Pi_a \boldsymbol{x}_{t-1} + B_a \boldsymbol{m}_t + D_a \boldsymbol{q}_t\big) + \\ & + (1 - g(z_{t-1}, \gamma, c))\big(\Pi_b \boldsymbol{x}_{t-1} + B_b \boldsymbol{m}_t + D_b \boldsymbol{q}_t\big) + \boldsymbol{u}_t, \end{aligned} \tag{B.60}$$

$$\approx 0.5\big(\Pi_a \boldsymbol{x}_{t-1} + B_a \boldsymbol{m}_t + D_a \boldsymbol{q}_t\big) + 0.5\big(\Pi_b \boldsymbol{x}_{t-1} + B_b \boldsymbol{m}_t + D_a \boldsymbol{q}_t\big) + \boldsymbol{u}_t, \tag{B.61}$$

$$\approx \frac{\Pi_a + \Pi_b}{2}\boldsymbol{x}_{t-1} + \frac{B_a + B_b}{2}\boldsymbol{m}_t + \frac{D_a + D_b}{2}\boldsymbol{q}_t + \boldsymbol{u}_t, \tag{B.62}$$

and the regressors in the model approach perfect multicollinearity, because the $T \times 2m$

matrix $W(\gamma, c)'$ of the regressors in model (A.9) simplifies to

$$W' = \begin{bmatrix} g_0 \underset{half}{\boldsymbol{w}_0}', & (1-g_0)\underset{half}{\boldsymbol{w}_0}', \\ \vdots & \vdots \\ g_t \underset{half}{\boldsymbol{w}_t}', & (1-g_t)\underset{half}{\boldsymbol{w}_t}', \\ \vdots & \vdots \\ g_T \underset{half}{\boldsymbol{w}_T}', & (1-g_T)\underset{half}{\boldsymbol{w}_T}', \end{bmatrix} = \begin{bmatrix} 0.5 \underset{half}{\boldsymbol{w}_0}', & 0.5 \underset{half}{\boldsymbol{w}_0}', \\ \vdots & \vdots \\ 0.5 \underset{half}{\boldsymbol{w}_t}', & 0.5 \underset{half}{\boldsymbol{w}_t}', \\ \vdots & \vdots \\ 0.5 \underset{half}{\boldsymbol{w}_T}', & 0.5 \underset{half}{\boldsymbol{w}_T}', \end{bmatrix}, \tag{B.63}$$

$$= 0.5 \begin{bmatrix} \underset{half}{\boldsymbol{w}_0}', & \underset{half}{\boldsymbol{w}_0}', \\ \vdots & \vdots \\ \underset{half}{\boldsymbol{w}_t}', & \underset{half}{\boldsymbol{w}_t}', \\ \vdots & \vdots \\ \underset{half}{\boldsymbol{w}_T}', & \underset{half}{\boldsymbol{w}_T}', \end{bmatrix} = 0.5 \cdot \left(\boldsymbol{\iota}_2' \otimes \underset{half}{W}' \right), \tag{B.64}$$

with $\boldsymbol{\iota}_2$ the 2×1 unit vector, $\underset{half}{\boldsymbol{w}_t}$ defined in equation (A.5) and

$$\underset{half}{W} = \left[\underset{half}{\boldsymbol{w}_0}, .., \underset{half}{\boldsymbol{w}_t}, .., \underset{half}{\boldsymbol{w}_T} \right], \tag{B.65}$$

of dimensions $m \times T$. The same can be seen using the notation from the homoskedastic-equivalent model (A.3), i.e.

$$\frac{\boldsymbol{y}_t}{\sqrt{h_{t-1}}} \approx \frac{\Pi_a + \Pi_b}{2} \frac{\boldsymbol{x}_{t-1}}{\sqrt{h_{t-1}}} + \frac{B_a + B_b}{2} \frac{\boldsymbol{m}_t}{\sqrt{h_{t-1}}} + \frac{D_a + D_b}{2} \frac{\boldsymbol{q}_t}{\sqrt{h_{t-1}}} + \frac{\boldsymbol{u}_t}{\sqrt{h_{t-1}}}, \tag{B.66}$$

with the $T \times 2m$ matrix of multicollinear regressors

$$H(\gamma, c, \psi)^{-\frac{1}{2}} W(\gamma, c)' = 0.5 \cdot \left(\boldsymbol{\iota}_2' \otimes H(\gamma, c, \psi)^{-\frac{1}{2}} \underset{half}{W}' \right). \tag{B.67}$$

$\hat{\Theta}(\gamma, c, \psi)$ from equation (A.18) is hence not well-defined, but equation (A.37) still is. Under a flat prior on Θ, for γ away from 0, $\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi)$ reaches its maximum at $\boldsymbol{\theta} = \text{vec}\big(\hat{\Theta}(\gamma, c, \psi)\big)$. For γ approaching 0, the mode of $\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi)$ approaches a ridge along

$$\Pi_a + \Pi_b = 2 \cdot \underset{half}{\hat{\Pi}}, \qquad B_a + B_b = 2 \cdot \underset{half}{\hat{B}}, \qquad D_a + D_b = 2 \cdot \underset{half}{\hat{D}}, \tag{B.68}$$

with $(\underset{half}{\Pi}, \underset{half}{B}, \underset{half}{D})$ implicitly defined in

$$\underset{half}{\hat{\Theta}}(\gamma, c, \psi) = [\underset{half}{\hat{\Pi}}, \underset{half}{\hat{B}}, \underset{half}{\hat{D}}] = \tag{B.69}$$

$$= Y H(\gamma, c, \psi)^{-1} \underset{half}{W}' \big(\underset{half}{W} H(\gamma, c, \psi)^{-1} \underset{half}{W}'\big)^{-1}. \tag{B.70}$$

Accordingly, (Π_a, Π_b) are not separately identified, nor are (B_a, B_b) and (D_a, D_b)

As for point 2c), the conditional posterior can be rewritten as

$$\begin{aligned}\tilde{p}(\gamma, c, \psi|Y, \boldsymbol{\theta}, \Sigma) \;&\propto \tilde{p}(\gamma, c, \psi) \cdot e^{-\frac{1}{2}\left(\tilde{\boldsymbol{y}} - (W(\gamma,c)' \otimes I_k)\boldsymbol{\theta}\right)' \left(H(\gamma,c,\psi) \otimes \Sigma\right)^{-1} \left(\tilde{\boldsymbol{y}} - (W(\gamma,c)' \otimes I_k)\boldsymbol{\theta}\right)} \cdot \\ &\cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}}, \qquad\qquad (B.71)\\ &\propto e^{-\frac{1}{2}\left[\tilde{\boldsymbol{y}}'(H(\gamma,c,\psi)^{-1} \otimes \Sigma^{-1})\tilde{\boldsymbol{y}} \;+\; \boldsymbol{\theta}'\left(W(\gamma,c)H(\gamma,c,\psi)^{-1}W(\gamma,c)' \otimes \Sigma^{-1}\right)\boldsymbol{\theta} \;-\; 2\boldsymbol{\theta}'\left(W(\gamma,c)H(\gamma,c,\psi)^{-1} \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}}\right]} \cdot \\ &\cdot \tilde{p}(\gamma, c, \psi) \cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}}, \qquad\qquad (B.72)\end{aligned}$$

and for $\gamma = 0$ the term

$$W \cdot \; H^{-\frac{1}{2}} H^{-\frac{1}{2}} \; W' = 0.5^2 \Big(\boldsymbol{\iota}_2 \otimes \underset{half}{W} H^{-\frac{1}{2}}\Big)\Big(\boldsymbol{\iota}_2' \otimes H^{-\frac{1}{2}} \underset{half}{W}'\Big), \tag{B.73}$$

$$= 0.5^2 \Big(\boldsymbol{\iota}_2 \boldsymbol{\iota}_2' \otimes \underset{half}{W} H^{-1} \underset{half}{W}'\Big), \tag{B.74}$$

does not diverge.

For point 2d), note that when γ approaches zero, $W(\gamma,c)H(\gamma,c,\psi)^{-1}W(\gamma,c)'$ becomes singular, so absent $\bar{V}(\gamma)$, the last term in $\tilde{p}(\gamma,c,\psi|Y)$ is not well-defined. As in **?**, the presence of $\bar{V}(\gamma)$ avoids the resulting non-invertibility issues by introducing prior information, especially as we make the prior on $\boldsymbol{\theta}$ progressively more informative as γ approaches 0.

3) very low values of c imply that the transition variable is always very close to 1, preventing any full transition of the model from one extreme to the other (see also **?**). This implies that $\boldsymbol{w}_t(\gamma,c) = \begin{pmatrix} g(z_{t-1},\gamma,c)\,\underset{half}{\boldsymbol{w}_t} \\ \big(1-g(z_{t-1},\gamma,c)\big)\,\underset{half}{\boldsymbol{w}_t} \end{pmatrix} \approx \begin{pmatrix} \underset{half}{\boldsymbol{w}_t} \\ \boldsymbol{0} \end{pmatrix}$, making (Π_b, B_b, D_b) not identified, although $(\Pi_a, B_a, D_a, \Sigma)$ remain identified. The opposite holds for very high (rather than low) values of c.

To partly address the challenges above, we follow common practice and standardize the transition variable, which makes (γ,c) scale-free (see, for instance, **?** and **?**). Call y_{it} the variable used to construct the transition variable, which is the log of real GDP in the application of the paper. Set $\Delta = 3$ as the difference in the lags used to construct the transition variable. Given the $1 \times l$ vector

$$\tilde{\boldsymbol{z}}'_{t-1,ns} = \big(y_{it-1} - y_{it-1-\Delta},\ y_{it-2} - y_{it-2-\Delta},\ \dots,\ y_{it-l} - y_{it-l-\Delta}\big), \tag{B.75}$$

we define the non-standardized transition variable as

$$z_{t-1,ns} = \frac{\boldsymbol{\iota}_l'\tilde{\boldsymbol{z}}_{t-1,ns}}{l}, \tag{B.76}$$

with $\boldsymbol{\iota}_l$ the $l \times 1$ unit vector. The scaling of $z_{t-1,ns}$ depends on the scaling of the underlying variable y_{it}, once differenced. We then define the (standardized) transition variable as

$$z_{t-1} = \frac{z_{t-1,ns} - \mu_z}{\sigma_z}, \tag{B.77}$$

with μ_z and σ_z the sample mean and standard deviation of $z_{t-1,ns}$ computed on the estimation sample. The transition function is then defined on the standardized transition variable,

$$g(z_{t-1}, \gamma, c) \;= \frac{1}{1 + e^{-\gamma(z_{t-1}-c)}}, \quad \gamma > 0. \tag{B.78}$$

Since z_{t-1} is standardized, its distribution can help define a suitable prior for c. We use a Normal prior $N(\mu_c, \sigma_c)$, truncated outside the set (τ_l, τ_h) to avoid extreme values of c, and calibrate σ_c such that 95% of the truncated prior mass is in the set (p_l, p_h). We find that the more skewed is the distribution of z_{t-1}, the better it is to set μ_c equal to the median (rather than the mean) of z_{t-1}, otherwise an excessively narrow set (p_l, p_h) fails to include the mean of the untruncated Normal distribution, making it hard to numerically calibrate σ_c. See the paper for how we set $(\tau_l, \tau_h, p_l, p_h, \mu_c)$ in the application of the paper.

References

Canova, F. (2007). *Methods for applied macroeconomic research*, Volume 13. Princeton University Press.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.