REGRESSION MODELS WITH AR(1) ERROR STRUCTURE

AAEC 6564
INSTRUCTOR: KLAUS MOELTNER

Introduction

Adding an AR(1) error structure to the standard regression model is generally straightforward, though draws of the correlation parameter ρ require a Metropolis-Hasting (MH) step. Also, we need to pay attention to the fact that ρ is restricted to the $[-1, 1]$ interval. This will be reflected in the choice of prior for this parameter. In theory, this restriction could also be imposed within the MH step when drawing ρ, but in many cases this will not be necessary if the candidate generating density (CGD) is carefully chosen.

The regression model

The resulting regression model for a sample of T observations can be written as:

$$\begin{aligned}
&\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \text{where} \\
&E(\boldsymbol{\epsilon}) = \mathbf{0} \\
&E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 * \boldsymbol{\Omega} = \\
&\sigma^2 * \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}
\end{aligned} \tag{1}$$

Thus, our model parameters about which we wish to learn are $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}' & \sigma^2 & \rho \end{bmatrix}'$.

The likelihood function

Noting that $|\sigma^2 * \boldsymbol{\Omega}|^{-1/2} = \sigma^{2(-T/2)} |\boldsymbol{\Omega}|^{-1/2}$, we can write:

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}) = (2\pi)^{-T/2} \sigma^{2(-T/2)} * |\boldsymbol{\Omega}|^{-1/2} exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \tag{2}$$

The priors

Letting the dimension of $\boldsymbol{\beta}$ be k, we have:

$$
\begin{aligned}
p(\boldsymbol{\beta}) &= (2\pi)^{-k/2} |\mathbf{V}_0|^{-1/2} exp\left(-\tfrac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'(\mathbf{V}_0)^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right) \\
p(\sigma^2) &= \frac{\tau_0^{\nu_0}}{\Gamma(\nu_0)}(\sigma^2)^{-(\nu_0+1)} exp\left(-\frac{\tau_0}{\sigma^2}\right) \\
p(\rho) &= \frac{\phi\left(\frac{\rho-\mu_\rho}{\sqrt{V_\rho}}\right)}{\sqrt{V_\rho}\left(\Phi\left(\frac{1-\mu_\rho}{\sqrt{V_\rho}}\right)-\Phi\left(\frac{-1-\mu_\rho}{\sqrt{V_\rho}}\right)\right)},
\end{aligned} \tag{3}
$$

where $\phi(.)$ and $\Phi(.)$ denote, respectively, the *pdf* and *cdf* of the standard normal density. That is, the prior for ρ is a normal density with mean μ_ρ (usually set to zero), and variance V_ρ, truncated to the [-1,1] interval. The priors for the remaining parameters are the same as for the standard regression model, i.e. normal for $\boldsymbol{\beta}$ and inverse-gamma for σ^2.

The posterior

Combining priors and likelihood function, we obtain the following joint posterior kernel:

$$
\begin{aligned}
& p(\boldsymbol{\beta}, \sigma^2, \rho | \mathbf{y}, \mathbf{X}) \propto \\
& exp\left(-\tfrac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'(\mathbf{V}_0)^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right) * \\
& (\sigma^2)^{-(\nu_0+1)} exp\left(-\frac{\tau_0}{\sigma^2}\right) * \\
& \frac{\phi\left(\frac{\rho-\mu_\rho}{\sqrt{V_\rho}}\right)}{\sqrt{V_\rho}\left(\Phi\left(\frac{1-\mu_\rho}{\sqrt{V_\rho}}\right)-\Phi\left(\frac{-1-\mu_\rho}{\sqrt{V_\rho}}\right)\right)} * \\
& \sigma^{2(-T/2)} * |\boldsymbol{\Omega}|^{-1/2} exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)
\end{aligned} \tag{4}
$$

The Gibbs Sampler

For draws of $\boldsymbol{\beta}$, conditional on the remaining parameters, the relevant portion of the joint posterior kernel is given as:

$$
\begin{aligned}
& p(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}, \sigma^2, \rho) \propto \\
& exp\left(-\tfrac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'(\mathbf{V}_0)^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right) * \\
& exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)
\end{aligned} \tag{5}
$$

This leads again to a well-understood normal conditional posterior, form which we can draw $\boldsymbol{\beta}$:

$$\begin{aligned}
&\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\sigma^2,\rho \sim n\left(\boldsymbol{\mu}_1,\mathbf{V}_1\right), \quad \text{with} \\
&\mathbf{V}_1 = \left(\mathbf{V}_0{}^{-1} + \mathbf{X}'\left(\sigma^2 * \boldsymbol{\Omega}\right)^{-1}\mathbf{X}\right)^{-1}, \\
&\boldsymbol{\mu}_1 = \mathbf{V}_1\left(\mathbf{V}_0{}^{-1}\boldsymbol{\mu}_0 + \mathbf{X}'\left(\sigma^2 * \boldsymbol{\Omega}\right)^{-1}\mathbf{y}\right)
\end{aligned} \tag{6}$$

The conditional posterior kernel for σ^2 is given as:

$$\begin{aligned}
&p\left(\sigma^2|\mathbf{y},\mathbf{X},\boldsymbol{\beta},\rho\right) \propto \\
&\left(\sigma^2\right)^{-(\nu_0+1)} exp\left(-\frac{\tau_0}{\sigma^2}\right) * \\
&\sigma^{2(-T/2)} * exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right) = \\
&\sigma^{2-\left(\frac{T+2\nu_0}{2}+1\right)} * exp\left(-\frac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\right)
\end{aligned} \tag{7}$$

Thus,

$$\begin{aligned}
&\sigma^2|\mathbf{y},\mathbf{X},\boldsymbol{\beta},\rho \sim ig\left(\nu_1,\tau_1\right), \quad \text{where} \\
&\nu_1 = \frac{T+2\nu_0}{2}, \\
&\tau_1 = \tau_0 + \tfrac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)
\end{aligned} \tag{8}$$

For draws of ρ we need to design a Metropolis-Hastings step. The conditional posterior kernel is given as:

$$\begin{aligned}
&p\left(\rho|\mathbf{y},\mathbf{X},\boldsymbol{\beta},\sigma^2\right) \propto \\
&p\left(\rho\right) * |\boldsymbol{\Omega}|^{-T/2} * exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right),
\end{aligned} \tag{9}$$

where $p\left(\rho\right)$ is given in (3).

We will use a tailored independence-chain MH approach where we first find the mode of the relevant posterior kernel given in (9) using MLE. Importantly in this case, we need to restrict ρ to the [-1,1] interval within the MLE procedure. This is easily accomplished by sending into MLE as starting value the *inverse hyperbolic tangent (atanh)* of the current draw of *rho* in the GS. Inside the MLE function, this procedure is immediately reversed by taking the hyperbolic tangent of the starting draw. When the output is sent back to the main GS, we need to perform this reversion again. See script `AR1_ic` and MLE function `rho_mode`.

Note that, strictly speaking, the Hessian coming out of the MLE routine is for $arctanh\,(\rho)$, not for (ρ) itself, but rather than bothering with the Delta method to obtain the actual variance of ρ we can simply adjust the MH tuners to achieve desirable acceptance rates in this simple scalar case.

Denoting the MLE mode as ρ^*, we then take a candidate draw ρ^c from a t-distribution with mean ρ^*, standard deviation $tc * \sqrt{H^{-1}}$, and degrees of freedom tv, where tc and tv are tuning parameters, and H is the Hessian matrix (here just a scalar) from the MLE routine evaluated at the mode ρ^*.

Thus:
$$q\left(\rho^c|\rho^0, \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2\right) = q\left(\rho^c|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2\right) = t\left(\rho^*, tc * \sqrt{H^{-1}}, tv\right) \tag{10}$$

The new draw ρ^c is then accepted with probability

$$\alpha\left(\rho^0, \rho^c\right) = min(1, \frac{p\left(\rho^c|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2\right) * t\left(\rho^0|\rho^*, tc * \sqrt{H^{-1}}, tv\right)}{p\left(\rho^0|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2\right) * t\left(\rho^c|\rho^*, tc * \sqrt{H^{-1}}, tv\right)}) \tag{11}$$

Note that the CGD is no longer symmetric, so all terms in (11) need to be included.

Matlab implementation with simulated data: `AR1_data` (data generation), `AR1_ic` (main script for the GS), `gs_AR1_ic` (GS), `rho_mode` (function to be optimized in the MLE routine);