Shrinkage of time-varying effects in panel data models

Helga Wagner[1], Roman Pfeiler[1]

[1] Department of Applied Statistics, Johannes Kepler University Linz, Austria, (e-mail: helga.wagner@jku.at, Roman.Pfeiler@jku.at)

Abstract: We consider regression models for panel data with time-varying effects in a Bayesian framework. We implement shrinkage of regression effects and the process variances of the effects to distinguish between effects that are practically zero, constant or time-varying via shrinkage priors. Longitudinal dependence is taken into account by including a subject specific random factor with weights that may also vary over time. The model is applied to analyse panel data on annual incomes of mothers returning to the job market after maternity leave.

Keywords: dynamic effects; factor model; shrinkage prior

1 Introduction

Panel data where subjects are observed at several time points provide richer information than cross sectional data but pose additional challenges as correlation of observations within subjects has to be taken into account. The multiple measurements per subject allow to model their heterogeneity and the longitudinal structure provides information on development over time. A standard way to take into account heterogeneity in panel data regression analysis is by including subject specific random effects in the linear predictor and development over time can be modelled by allowing for time-varying regression effects. However, modelling all regression effects as time-varying will result in an overspecified model if actually one or more effects are time-constant or even 0. In a Bayesian approach, based on an adaquate model formulation, appropriate prior distributions allow to identify constant or zero effects in time series regression models (Frühwirth-Schnatter & Wagner, 2010). In this paper we will use the shrinkage priors recently proposed in Bitto & Frühwirth-Schnatter, 2019 for time series and investigate their performance for panel data where the number of subjects is larger than 1 but time series are short, e.g. in our application we have individual time series of length 8.

2 Model specification and inference

2.1 Regression model with time-varying effects

To keep notation simple, we assume balanced panel data where $i = 1,...,n$ subjects are observed at time points $t = 1,...,T$. Let y_{it} denote the response of subject i at time t and $\mathbf{x}_{it}$ is the $p \times 1$ vector of covariates. We consider the following regression specification

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta}_t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \tag{1}$$

where $\boldsymbol{\beta}_t$ is the $p \times 1$ vector of regression effects at time t and $\boldsymbol{\Omega}$ is a $T \times T$ covariance matrix.

To model time-varying parameters we assume that the regression effects follow a random walk

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

with independent increments, $\mathbf{Q} = \text{diag}(\theta_1^2, \ldots, \theta_p^2)$, and starting values

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_0).$$

The process variance θ_j^2, $j = 1, \ldots, p$ carries information on the evolvement of the regression effect β_{jt} over time.

To allow shrinkage to time-constant or zero effects we use shrinkage priors on the effects and process standard deviations in the non-centered parameterization (Frühwirth-Schnatter & Wagner, 2010), which is given as

$$\boldsymbol{\beta}_t = \boldsymbol{\beta} + \boldsymbol{\theta} \tilde{\boldsymbol{\beta}}_t.$$

Here $\boldsymbol{\theta} = \text{diag}(\theta_1, \ldots, \theta_p)$ is the vector of process standard deviations and $\tilde{\boldsymbol{\beta}}_t$ is defined as

$$\tilde{\boldsymbol{\beta}}_t = \tilde{\boldsymbol{\beta}}_{t-1} + \tilde{\boldsymbol{\omega}}_t, \quad \tilde{\boldsymbol{\omega}}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Hence, the regression model (1) in its non-centered parameterization is given as

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{x}_{it}^T \boldsymbol{\theta} \tilde{\boldsymbol{\beta}}_t + \epsilon_{it}.$$

Shrinkage of elements of $\boldsymbol{\beta}$ as well as $\boldsymbol{\theta}$ is induced by appropriate prior distributions.

2.2 Modelling longitudinal association

To allow for longitudinal association within subjects we specify the error term ϵ_{it} in terms of a subject specific latent factor f_i and the idiosyncratic error ε_{it} as

$$\epsilon_{it} = \lambda_t f_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_t^2)$$

and hence

$$\boldsymbol{\Omega} = \boldsymbol{\lambda}\boldsymbol{\lambda}^T + \boldsymbol{\Sigma}$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_T^2)$. This model encompasses as special case compound symmetry structure of $\boldsymbol{\Omega}$ when $\lambda_t = \lambda$. To model time-varying factor loadings we again model the evolvement of the factor loadings by a random walk

$$\lambda_t = \lambda_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, \psi^2).$$

parameterization.

2.3 Prior Distributions

To encourage shrinkage of constant effects β_j and their process variances θ_j^2, $j = 1, \ldots, p$, following Bitto & Frühwirth-Schnatter, 2019 we specify the priors on β_j as independent Normal-Gamma and on the process variances θ_j as independent double Gamma-priors. The same specification is used for the priors on the factor loading parameters in the noncentered parameterisation.

For the error variances σ_t^2 of the idiosyncratic errors we use independent uninformative Inverse Gamma priors.

2.4 Inference

Inference is performed by MCMC methods extending the Gibbs sampling proposed in Bitto & Frühwirth-Schnatter, 2019 by the additional steps to sample the subject specific factors and the factor loadings in the non-centered parameterization.

3 Application

We apply the developed methods to analyse earnings of mothers in Austria after their return to the labor market after their last maternity leave. The data set comprises earnings for $n = 8877$ mothers after return to labour market observed for $T = 8$ panel periods.

Covariates in the regression model are categorical predictors of the number of children (baseline: 1 child, dummy for 2 children, dummy for 3 or more children), binary variables for type of contract (baseline: white collar), leave duration and working experience (for both the baseline is *below the median*) as well as the log-earnings before the maternity leave. All regression parameters, except the effect of 3 or more children and also the factor loadings vary over time. Figure 1 compares the estimated time-varying intercept and the effects of 3 or more children under the shrinkage priors to the estimated effects in a random intercept model with unstructured time-varying effects. The shrinkage prior results in smoother effects which can also be effectively reduced to zero, see the lower panel of Figure 1.

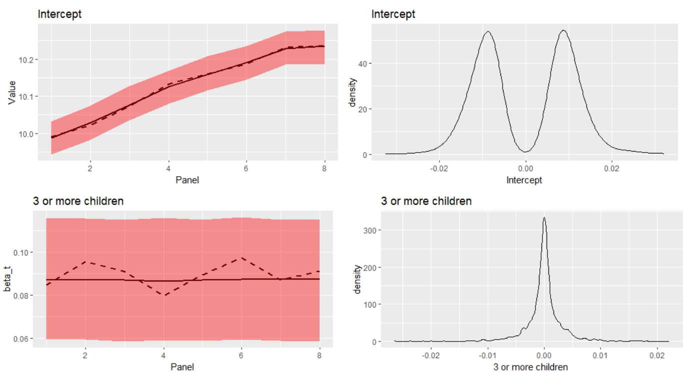

Figure 1. *Results for intercept and effect of 3 or more children. Left: Posterior mean estimates and 95%-HPD intervals of the regression effects. Dotted lines are the estimated time-varying regression effects from a random intercept model without smoothing. Right: Posterior of the process standard deviations.*

References

BITTO, A., & FRÜHWIRTH-SCHNATTER, S. 2019. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics.*, **210**, 75–97.

FRÜHWIRTH-SCHNATTER, S., & WAGNER, H. 2010. Stochastic model specification search for Gaussian and partially Non-Gaussian state space models. *Journal of Econometrics.*, **154**, 85–100.