Online Appendix: A tale of two tails: 130 years of growth-at-risk

Martin Gächter* Elias Hasler† Florian Huber‡

August 21, 2024

Appendix A: Technical appendix

We use a Bayesian approach to estimation and inference. This requires choosing suitable shrinkage priors on the coefficients and latent states of the model. To do so, we exploit the non-centered parameterization of the state space model (Frühwirth-Schnatter and Wagner (2010)) in Eq. (1):

$$y_{t+h} = \boldsymbol{\beta}_0' \boldsymbol{x}_t + \tilde{\boldsymbol{\beta}}_t \sqrt{\boldsymbol{V}}_\beta \boldsymbol{x}_t' + \varepsilon_{t+h},$$

with $\boldsymbol{\beta}_0$ denoting a set of K time-invariant coefficeints, $\tilde{\boldsymbol{\beta}}_t$ is a vector of normalized states with j^{th} element $\tilde{\beta_{jt}} = (\beta_{jt} - \beta_{j0})/\pm\sqrt{v_j^2}$ and $\sqrt{\boldsymbol{V}}_\beta$ is a diagonal matrix with $\pm\sqrt{v_j^2}$ in the $(j,j)^{th}$ position.

The state equation of this reparameterized model is given by:

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

The non-centered parameterization can be used to elicit shrinkage priors that allow us to answer the question whether certain elements in $\boldsymbol{\beta}_t$ should be constant or time-varying. This is achieved by using Gaussian priors on β_{j0} and $\pm\sqrt{v_j^2}$:

$$\pm\sqrt{v_j^2}|\tau_{j,v}^2 \sim \mathcal{N}(0, \tau_{j,v}^2), \quad \tau_{j,v}^2|a_v, \lambda_{j,v} \sim \mathcal{G}\left(a_v, \frac{a_v \lambda_{j,v}}{2}\right), \quad \lambda_{j,v}|c_v, \kappa_v \sim \mathcal{G}\left(c_v, \frac{c_v}{\kappa_v}\right),$$
$$\beta_{j0}|\tau_{j,\beta}^2 \sim \mathcal{N}(0, \tau_{j,\beta}^2), \quad \tau_{j,\beta}^2|a_\beta, \lambda_{\beta j} \sim \mathcal{G}\left(a_\beta, \frac{a_\beta \lambda_{\beta,j}}{2}\right), \quad \lambda_{\beta,j}|c_\beta, \kappa_\beta \sim \mathcal{G}\left(c_\beta, \frac{c_\beta}{\kappa_\beta}\right).$$

*Liechtenstein Financial Market Authority, Liechtenstein; University of Innsbruck, Austria; e-mail: martin.gaechter@fma-li.li

†Corresponding author: Liechtenstein Financial Market Authority, Liechtenstein; University of Innsbruck, Austria; e-mail: Elias.Hasler@fma-li.li.

‡University of Salzburg, Austria; e-mail: florian.huber@plus.ac.at.

Here, the parameters $a_v, a_\beta, c_v, c_\beta$ and κ_v, κ_β can be either set by the researcher or additional hyperparameters can be used to infer them from the data. We follow the second approach and specify yet another set of hyperpriors on these parameters. On the rescaled λ_v and λ_β we use F distributed priors and on the remaining hyperparameters we use Beta distributions. The precise prior setting then follows the standard setup discussed in Knaus et al. (2021) and exercised in the shrinkTVP package.

This prior can be used to discriminate between the following cases. If a given regressor is insignificant for all t, the prior shrinks β_{j0} and v_j towards zero. If a given regressor is initially unimportant but grows in importance over time, the prior forces $\beta_{j0} \approx 0$ and $v_j > 0$. If a regressor is important and has a time-invariant effect on y_t, the prior would imply that $\beta_{j0} \neq 0$ and $v_j \approx 0$. Finally, if a regressor is time-varying and important, the prior would allow for non-zero of β_{j0} and $v_j > 0$. All this is achieved automatically through the different scaling parameters.

We simulate from the posterior distribution of the latent states and coefficients using an Markov chain Monte Carlo (MCMC) algorithm. Since all steps are standard (and implemented in the `R` package `shrinkTVP`), we only summarize the main steps involved and refer to Knaus et al. (2021) for more details.

The MCMC algorithm cycles between the following steps:

- The latent states are simulated all without a loop from a multivariate Gaussian posterior distribution. This can be achieved efficiently by exploiting sparse algorithms.
- The time invariant parameters $\boldsymbol{\beta}_0$ and the diagonal elements of $\sqrt{V}_\beta$ are simulated within a single step from multivariate Gaussian posteriors.
- To improve sampling efficiency, an ancillarity-sufficiency interweaving step is introduced that redraws $\boldsymbol{\beta}_0$ from a sequence of Gaussian distributions and the diagonal elements of $\boldsymbol{V}_\beta$ from a generalized inverse Gaussia distribution.
- The prior variances and the hyperparameters are simulated either from well known full conditionals detailed in Cadonna et al. (2020) or by entertaining a Metropolis Hastings update.
- The error variances are obtained using the algorithm outlined in Kastner and Frühwirth-Schnatter (2014).

Appendix B: Further Results

One quarter ahead

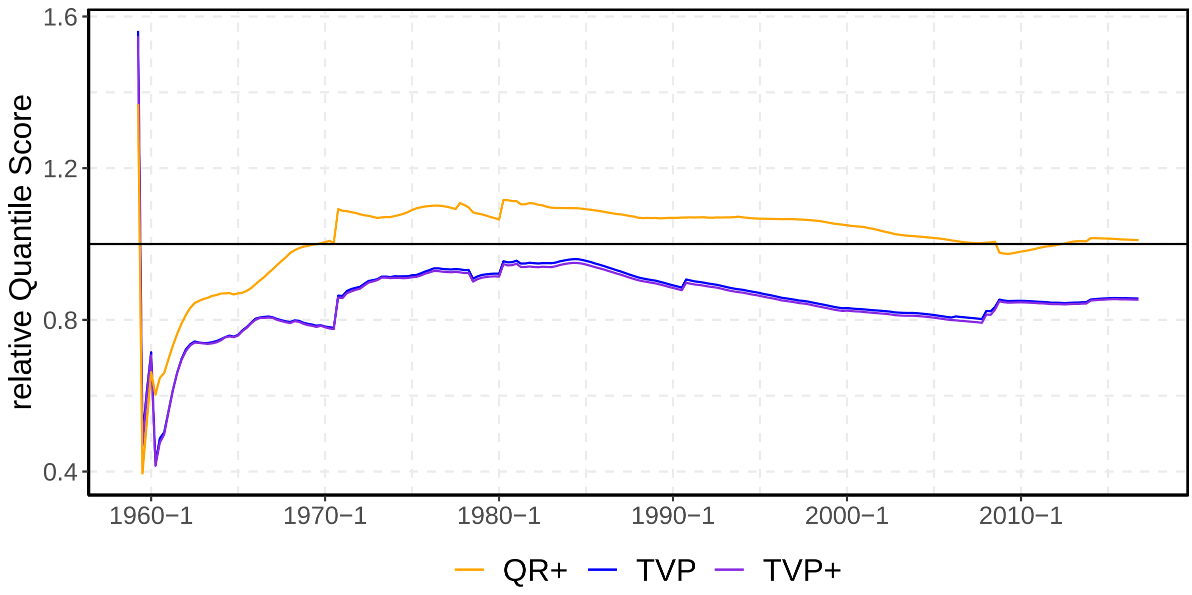

Figure A.1: Recursive mean quantile scores (relative to the QR model) one quarter ahead post WWII

Note: This figure shows the cumulative out-of-sample quantile score against the QR model post WWII.

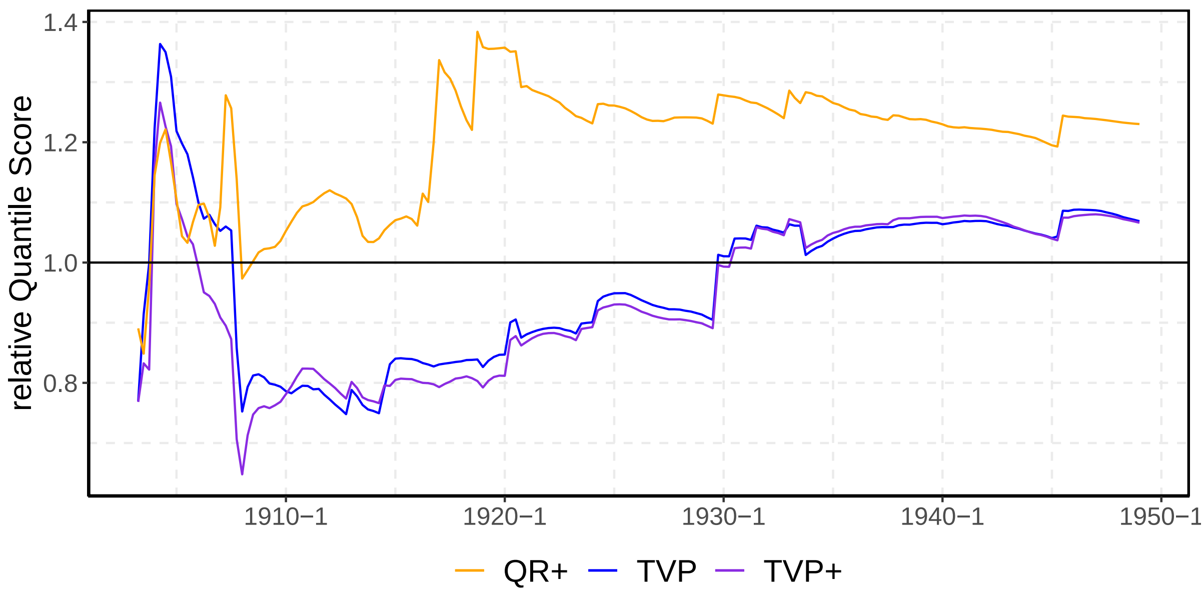

Figure A.2: Recursive mean quantile scores (relative to the QR model) one quarter ahead pre WWII

Note: This figure shows the cumulative out-of-sample quantile score against the QR model pre WWII.

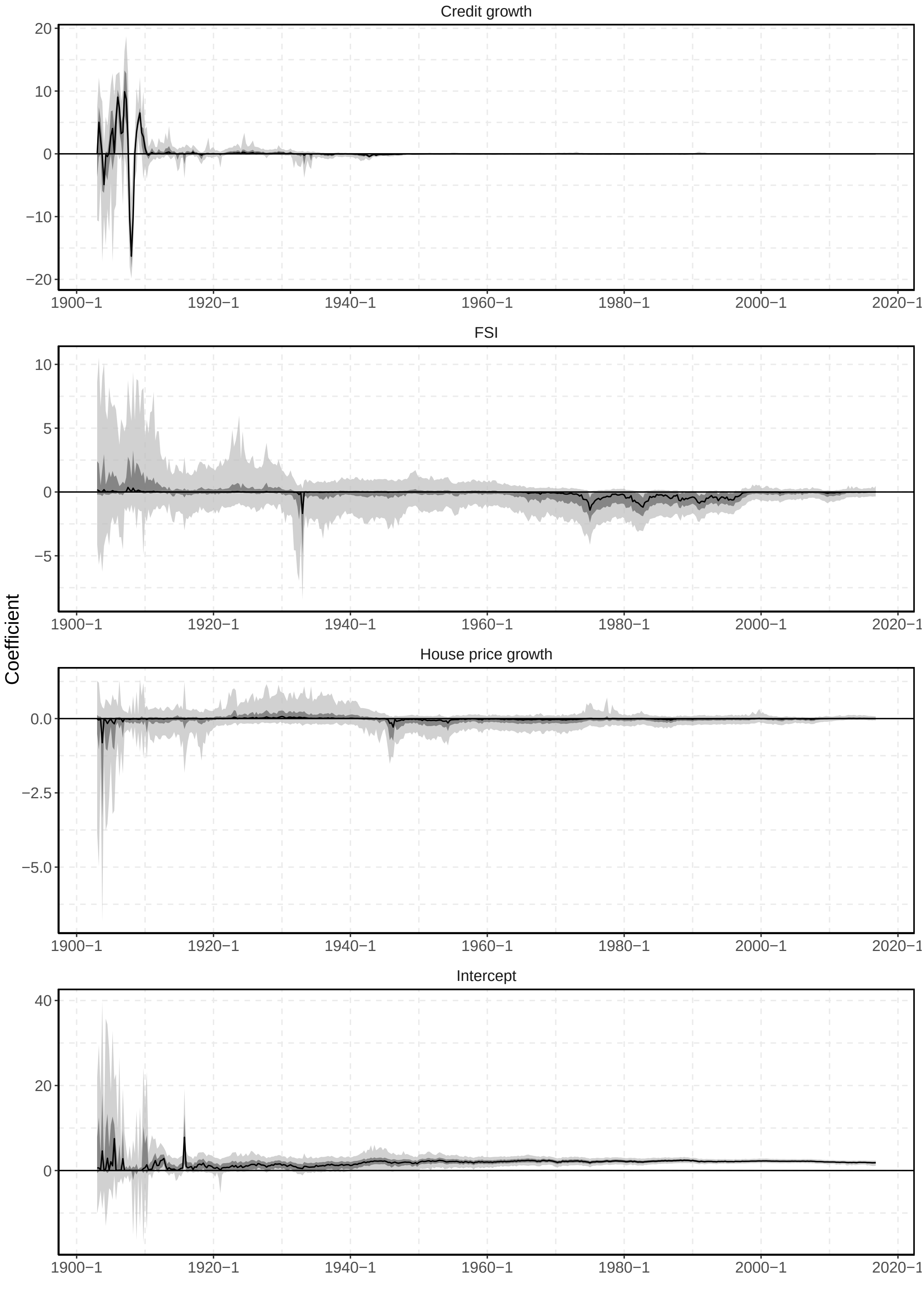

Figure A.3: Time series evolution of the coefficients. Recursive estimation one quarters ahead

Note: This figure shows the time-varying parameters one quarter ahead. The median is displayed as a black line, and the shaded areas indicate the pointwise 90% and 50% posterior credible intervals.

Four quarters ahead

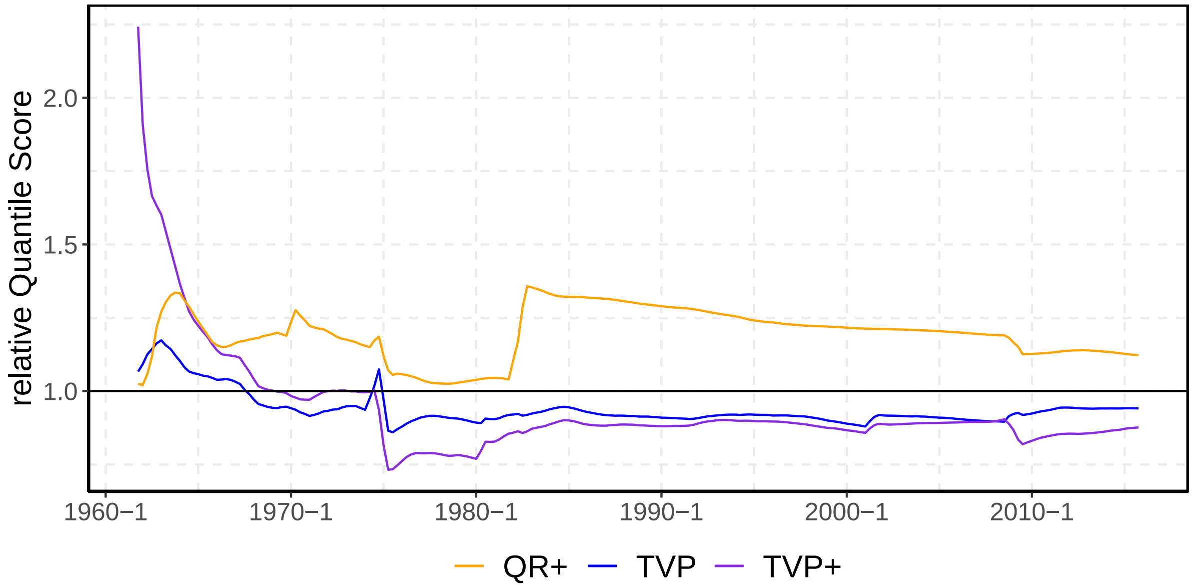

Figure A.4: Recursive mean (relative to QR FSI) four quarters ahead post WWII

Note: This figure shows the cumulative out-of-sample Quantile Score against the QR model post WWII.

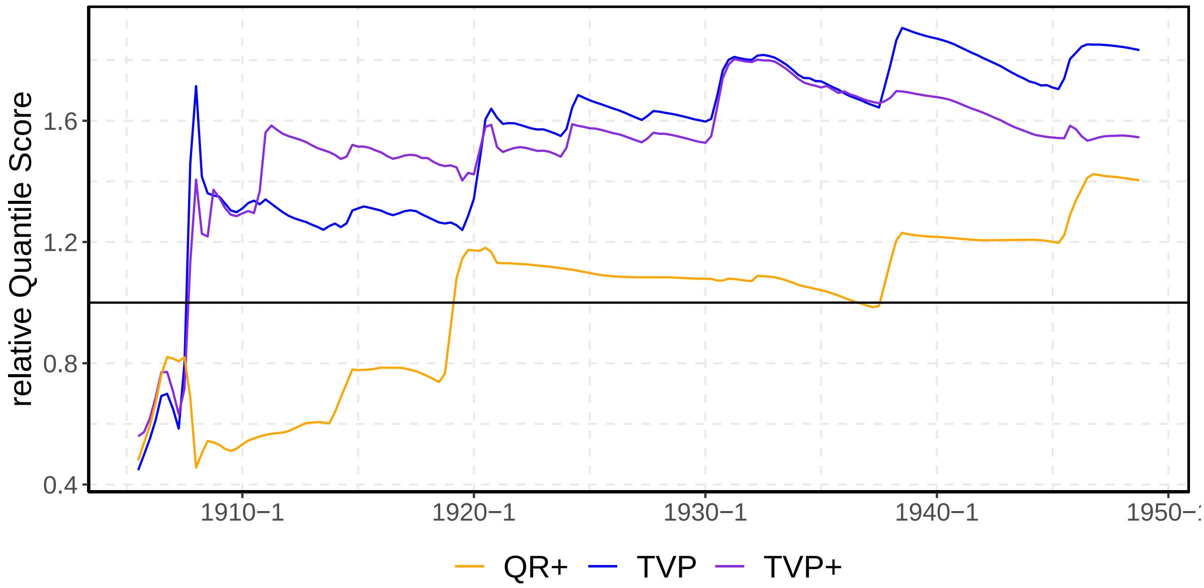

Figure A.5: Recursive mean (relative to QR FSI) four quarters ahead pre WWII

Note: This figure shows the cumulative out-of-sample Quantile Score against the QR model pre WWII.

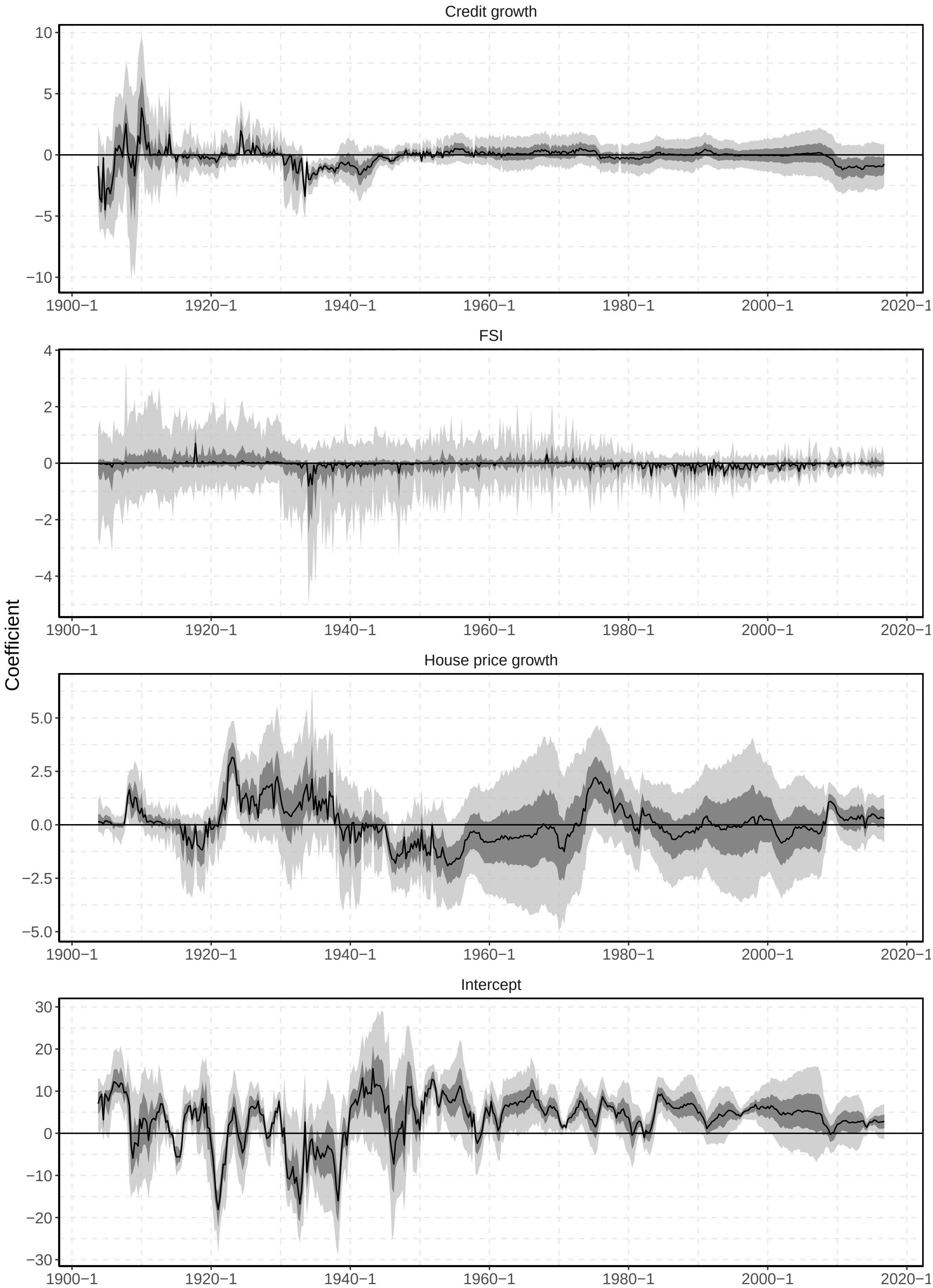

Figure A.6: Time series evolution of the coefficients. Recursive estimation four quarters ahead

Note: This figure shows the time-varying parameters four quarters ahead. The median is displayed as a black line, and the shaded areas indicate the pointwise 90% and 50% posterior credible intervals.

Further model evaluations

Figure A.7 displays the predicted 5^{th} percentile for all models. This figure is helpful for assessing whether the predicted tail risks of the models are well-calibrated. To complement the visual inspection, we also analyze the actual over expected (AE) ratio, which measures how often the growth rate falls below the predicted 5^{th} percentile.

For this analysis, we exclude both world wars and the surrounding years, as these periods are not the primary focus and are typically excluded in long historical data studies (e.g. Jordà et al., 2015).

For one-quarter-ahead predictions, the AE ratios are as follows: QR, 0.88; QR+, 0.95; TVP, 1.21; TVP+, 1.21. For four-quarter-ahead predictions, the AE ratios are: QR, 1.01; QR+ models: 1.42; TVP: 2.06; TVP+, 1.56.

The visual inspection of Figure A.7 indicates that the TVP+ model is less well-calibrated before the First World War, which inflates the AE ratio. When excluding this initial part of the sample, the AE ratio of the TVP+ model improves to 1.25 for the four-quarter-ahead prediction and 1.12 for the one-quarter-ahead prediction.

This exercise shows that out TVP model of choice is well calibrated, with regards to predicted left tail risks, especially after the initial period. However, AE ratios suggest that the QR models are slightly better calibrated. Looking at Figure A.7 it is evident why. The estimated QR models are much more conservative compared to the TVP models, which could favor them regarding the AE ratio since, this measure only counts the number of violation over the expected violations, which are instances where the growth rate is below the predicted 5^{th} percentile, but says nothing about the severity of the violation. Furthermore, when measuring the predictive performance by using quantile scores, the conservative QR models are favored. Although TVP outperform the QR models when it comes to predictive performance, and are more or less equally well calibrated, even tough we believe TVP models have somewhat of a disadvantage with the measures we use. Nevertheless, since the AE ratio is slightly worse for the TVP models, we check the severity of the violations, with absolute mean and maximum deviations of violations as an additional check

This exercise shows that our TVP model of choice is well calibrated with regards to predicted left tail risks, especially after the initial part of the 20^{th} century and when compared to the literature (see e.g. Brownlees and Souza, 2021; Gächter et al., 2023). However, AE ratios suggest that the QR models are slightly better calibrated. Looking at Figure A.7, it is evident why. The estimated QR models are much more conservative compared to the TVP models, which could favor them regarding the AE ratio since this measure only counts the number of violations, which are instances where the growth rate is below the predicted 5^{th} percentile, over the expected violations, but says nothing about the severity of the violations.

Furthermore, when measuring predictive performance using quantile scores, the conservative QR models are favored. Although TVP models outperform the QR models when it comes to predictive performance and are more or less equally well calibrated, we believe the TVP models have somewhat of a disadvantage with the measures we use. Nevertheless, since the AE ratio is slightly worse for the TVP models, we check the severity of the violations using absolute mean and maximum deviations of violations as an additional check and an alternative for quantile scores.

Table A.1 shows that the average and maximum deviations of violations are smaller for the TVP+ model compared to the QR+ model. While the TVP+ model is not always strictly better than the QR model, it is either close or clearly better. Hence, taking all the evidence into account, the TVP+ model seems to be the best choice for the task at hand.

Table A.1: Additional out-of-sample model evaluation

	Pre WWII				Post WWII			
Horizon	QR FSI	QR small	TVP FSI	TVP small	QR FSI	QR small	TVP FSI	TVP small
				$h = 1$				
AD mean	4.56	8.90	4.96	5.78	1.65	1.90	1.77	1.62
AD max	15.40	22.59	16.55	15.25	3.63	5.75	4.81	5.01
				$h = 4$				
AD mean	5.01	7.31	7.21	6.08	1.72	1.66	1.26	1.03
AD max	15.78	22.11	21.85	15.68	4.04	2.91	2.64	2.44

Note: This table reports additional out-of-sample model evaluation for the forecast horizon of 1 and 4 quarters. It shows the absolute mean and maximum deviations of violations, which are instances where the growth rate is below the predicted 5^{th} percentile.

Lastly, we formally describe the quantile score and describe why it generally favors more conservative models. The quantile score is defined as follows:

$$\text{Quantile Score} = \sum_{i=1}^{n} \rho_\tau(y_i - \hat{y}_i)$$

where $u = y_i - \hat{y}_i$ and $\rho_\tau(u)$ is the quantile loss function given by:

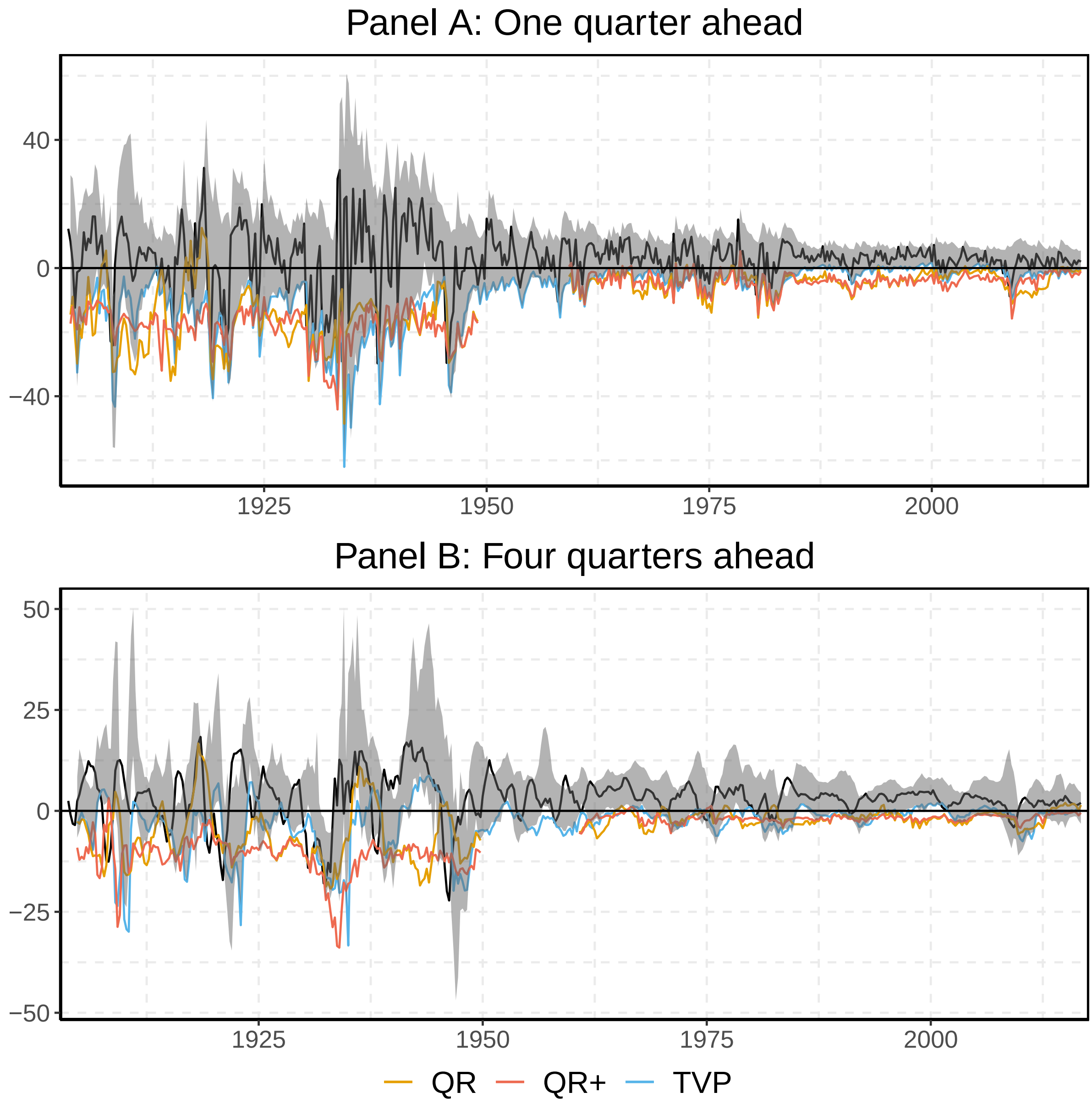

Figure A.7: Time series evolution of the predicted tail risks

Note: This figure shows the out-of-sample one quarter ahead (Panel A) and four quarters ahead (Panel B) forecast of the 5^{th} and 95^{th} percentile (gray shaded area) together with the realised growth rate (black line). The colored lines show the predicted 5^{th} percentile of the other models.

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases}$$

In this formula: - y_i is the actual value. - $\hat{y}_i$ is the predicted value for the τth quantile. - τ is the quantile being predicted (e.g., $\tau = 0.05$ for the 5th percentile).

The quantile loss function penalizes underestimates and overestimates asymmetrically. If the actual value y_i is less than the predicted quantile $\hat{y}_i$, the penalty is $(\tau - 1)(y_i - \hat{y}_i)$. If the actual value is greater than the predicted quantile, the penalty is $\tau(y_i - \hat{y}_i)$. This asymmetry leads to the characteristic that quantile scores penalize under-predictions more heavily than over-predictions when $\tau < 0.5$, resulting in more conservative or pessimistic predictions.

References

Brownlees, C. and Souza, A. B. (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics*, 118:312–330.

Cadonna, A., Frühwirth-Schnatter, S., and Knaus, P. (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics*, 8(2):20.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Gächter, M., Geiger, M., and Hasler, E. (2023). On the structural determinants of growth-at-risk. *International Journal of Central Banking*, 19(2):251–293.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2015). Leveraged bubbles. *Journal of Monetary Economics*, 76:S1–S20.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.

Knaus, P., Bitto-Nemling, A., Cadonna, A., and Frühwirth-Schnatter, S. (2021). Shrinkage in the time-varying parameter model framework using the r package shrinktvp. *Journal of Statistical Software*, 100:1–32.