TI 2026-096/III
Tinbergen Institute Discussion Paper

Economic Tail Risk Dynamics: Using Panel Data and Double-Pooling

Yicong Lin [1]
Andre Lucas [2]
Shiqi Ye [3]

1 Vrije Universiteit Amsterdam

2 Vrije Universiteit Amsterdam

3 State Key Laboratory of Mathematical Sciences, AMSS, Chinese Academy of Sciences; and AMSS Center for Forecasting Science, Chinese Academy of Sciences; and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS

Economic Tail Risk Dynamics Using Panel Data and Double-Pooling*

Yicong Lin[a] Andre Lucas[a] Shiqi Ye[b,c,d]

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*
[b] *State Key Laboratory of Mathematical Sciences, AMSS, Chinese Academy of Sciences*
[c] *AMSS Center for Forecasting Science, Chinese Academy of Sciences*
[d] *MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS*

November 20, 2026

Abstract

We propose a novel Double-Pooling (DP) framework for jointly estimating time-varying economic tail risk measures, e.g., Growth-at-Risk (GaR) and Growth Expected Shortfall (GES), for connected economic regions using a panel of low-frequency economic time series. The new method filters these risk dynamics simultaneously for different levels of extreme risk by pooling information across both time, the cross-section, and tail areas. It does so using a nonlinear three-factor model motivated by extreme value theory, with score-driven dynamics and a penalized distribution-free objective function. An application of the DP model to 14 European economies over the period 1961–2026 shows that the new approach has better out-of-sample performance than a range of alternative models and can produce more timely early-warning signals for increases in economic risk.

Keywords: dynamic tail estimation, score-driven models, Extreme Value Theory, Growth-at-Risk, quantile and expected shortfall modeling.

*The authors are grateful to participants at the Inaugural AE^2 Conference (2025, Sydney), the 19th International Joint Conference on Computational and Financial Econometrics (2025, London), the Conference on Score-Driven and Nonlinear Time Series Models (2026, Venice), the Chinese Economic Association UK/Europe Annual Conference (2026, Beijing), and the Econometric Society Australasia Meeting (2026, Adelaide) for valuable comments and suggestions. We particularly thank Yi Ding (discussant), Brendan Beare, Peter C. B. Phillips, Yongmiao Hong, Andrew Patton, Peter Hansen, Dario Palumbo, Christian M. Hafner, Jiti Gao, Anders Rahbek, and Liudas Giraitis for helpful discussions and feedback. Correspondence to Shiqi Ye (Email: shiqi.ye.c@gmail.com). Authors are listed in alphabetical order. Ye gratefully acknowledges support from the National Natural Science Foundation of China (No. 723B2020, No. 7198101), the China Postdoctoral Science Foundation (No. 2024M763465, No. 2025T181016, No. GZC20252267), and the MOE Social Sciences Innovative Group on Complex Systems Modeling in Economic Management in the Era of Digital Intelligence, University of Chinese Academy of Sciences.

1 Introduction

A range of high-impact events over the past two decades, including financial and pandemic crises, tariffs and trade wars and geopolitical tensions, have underlined the importance of accurate and timely assessments of macroeconomic tail risk, e.g., by measures such as Growth-at-Risk (GaR; Adrian et al., 2019, 2022; Marfè and Pénasse, 2024; Loria et al., 2025; Adrian et al., 2026), (Growth-and)-Inflation-at-Risk ((G)IaR; Lopez-Salido and Loria, 2024; Corradi and Llorens-Terrazas, 2026), or Housing-Price-at-Risk (HaR; Alter and Mahoney, 2021). The expectation is that the accurate measurement of such tail risk measures will contribute to avoiding the disproportionate economic costs of extreme downturns by triggering a timely mitigating policy response. The computation of economic tail risk measures like GaR and GES, however, is challenging for several reasons. Foremost, they have to be computed using macroeconomic data that are typically only available at low frequencies like monthly, quarterly, or annually. This reduces their precision compared to the financial markets' setting, where daily or intraday data are typically available (see, e.g., Liu and Tse, 2015; Patton et al., 2019). The limited data problems are further exacerbated by the fact that tail risk measures, by their nature, only consider the most extreme outcomes of the (low-frequency) economic time series, thus discarding most of the observations. Finally, macroeconomic tail risk obviously need not be constant over time, as it varies with economic and political circumstances, imposing further limits on data availability as observations may only reflect tail behavior locally in time.

One can distinguish three main approaches for quantifying economic tail risk. The first approach models both extreme and non-extreme growth outcomes in a single framework, typically using a parametric distribution class (see, e.g., Brownlees and Souza, 2021). The quantiles and tail expectations of the fitted distribution then serve as the estimated risk measures. A disadvantage of this approach is that it also uses observations from the center of the distribution, which might be less revealing about economic tail risk and can even bias tail risk estimates. In the second approach, no distributional assumptions are made and risk quantiles and tail expectations are estimated jointly using weighted-average quantile estimation (Chetverikov et al., 2025), quantile regressions (Brownlees et al., 2026), or by minimizing a so-called proper scoring rule (e.g., Fissler and Ziegel, 2016; Nolde and Ziegel, 2017; Patton et al., 2019). An advantage of this second approach is its simplicity and intuitive appeal: one directly concentrates on the objects of interest, which are the tail risk measures at some pre-specified specific risk level. The main disadvantage of the approach, however, is that risk estimates may remain unreliable if the tail area is small in relation to the available sample size, which could already be at 5% or 1% for

low-frequency macroeconomic data. For example, D'Innocenzo et al. (2026) show that far-out risk quantiles and Expected Shortfall (ES) dynamics based on proper scoring rules (as in Patton et al., 2019) may exhibit undesirable behavior with incidental large jumps and long periods of geometric decay toward an overall mean. The third approach uses Extreme Value Theory (EVT) to describe the entire tail above some threshold, as well as its dynamics (e.g., Adrian et al., 2026; D'Innocenzo et al., 2024, 2026). A drawback of this approach is that one first has to choose a (possibly time-varying) starting point for the EVT approximation to become applicable. One way of doing so is to make use of a quantile estimator based on the second approach above.

With each of the above approaches having its own drawback, this paper sets out to develop a new, complementary perspective for modeling dynamic tail risk measures for macroeconomic data. The *Double Pooling* (DP) framework proposed in this paper directly concentrates on tail risk dynamics for a specific tail area using a distribution free framework. In this way, it avoids any parametric distributional choice required for the tails. At the same time, it retains the intuitive appeal of the second approach of directly concentrating the empirical analysis on the object of interest: the GaR and GES at a specific risk level. In order to avoid the undesirable tail risk dynamics for low-frequency macro data of the second approach, however, we *regularize* the estimated tail dynamics by a penalty term that reflects the behavior of the tail over a *pool* of less extreme quantiles. The regularization term uses EVT-inspired restrictions to limit the number of free time-varying parameters, while at the same time avoiding the problem of (dynamic) quantile crossings (see the explanation in Catania and Luati, 2023). We do so without the need to specify the starting point of the tail in an ad hoc way. Rather, the estimation of the quantiles is an integral part of the whole framework. The new DP framework leads to a three-component representation for tail risk dynamics across different risk levels. This representation, in turn, yields a dynamic three-factor model for regularized tail risk dynamics. The regularization induces more modest reactions to intermediate adverse economic signals, rather than only incidental extreme reactions to outcomes far out in the tails (compare Patton et al., 2019; D'Innocenzo et al., 2026).

The new three-factor structure for tail dynamics also directly allows us to generalize the framework in a second direction: cross-sectional pooling. To see why such pooling is plausible, Figure 1 plots the GDP history of the 14 European economies also used later in this paper. The figure shows that common movements in GDP are clearly not only restricted to the center of the distribution, but also extend to the tail area, i.e., to periods of stress. Such common extremes can, for instance, be due to the existence of joint production or trade networks (see, e.g., Herskovic et al., 2020; Dew-Becker, 2023; Huo et al., 2025), including the joint exposure to external major trading partners. If such

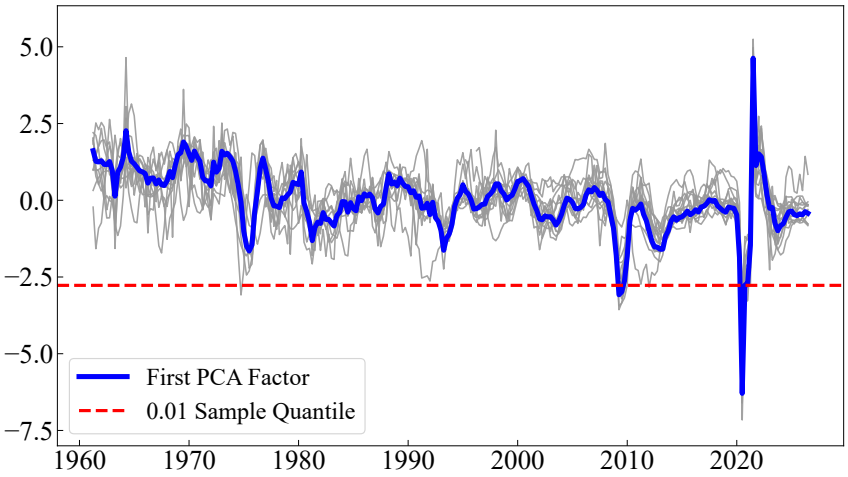


Figure 1: Real GDP growth of 14 European economies. The gray lines represent standardized GDP growth rates for individual countries. The blue solid line denotes the common factor extracted from the balanced panel without missing observations, and the red dashed line reports the pooled 1% sample quantile obtained from all country observations.

common extreme movements are prevalent, one can expect to obtain more information on extreme economic risk by pooling data cross-sectionally. In particular, an extreme outcome in one country might be predictive for increases in tail risk of another country. Pooling such information might provide more timely signals than studying macroeconomic tail risk for a single country only (e.g., Adrian et al., 2019; Delle Monache et al., 2020; Adrian et al., 2026), or by only combining such measures into joint risk regions as in Brownlees and Souza (2021); see also the arguments for panel quantile regressions in Adrian et al. (2022) and Lang et al. (2026). The nonlinear DP three-factor structure obtained in this paper facilitates such pooling and decomposes cross-sectional tail risk dynamics into a few common time-varying components with cross-sectional heterogeneity.

To estimate the model, we generalize the approach of Patton et al. (2019) from a single time series and a single risk level to a panel data setting with multiple tail areas. Our estimation principle builds directly on the proper scoring rule framework of Fissler and Ziegel (2016) without making distributional assumptions. It jointly and consistently elicits Expected Shortfall (ES) and Value-at-Risk (VaR), while regularizing the extreme tail with the implied behavior at less extreme tail areas within a single optimization framework. Familiar EVT results (de Haan and Ferreira, 2006) are used to express individual tail risk measures in terms of their common time-varying components and their cross-sectional deviations. The resulting nonlinear three-factor structure is subsequently endowed with a score-driven dynamic specification to capture the time-variation (see also Creal et al., 2013; Harvey, 2013; Patton et al., 2019). Cross-sectional heterogeneity is accommodated through country-specific parameters that shift the level of risk per country.

This yields a parsimonious, highly nonlinear, but analytically tractable updating recursion that permits direct estimation of the model's parameters. The model's structure also directly implies that the risk dynamics in one country not only react to growth realizations in the country itself around the prespecified risk level, but also to risk realizations in other countries and at other risk levels.

We apply the DP methodology to 1% Growth-at-Risk (GaR) and extreme Growth Expected Shortfall (GES) dynamics for a panel of 14 European economies, using quarterly year-on-year GDP growth from 1961Q1 to 2026Q2. The European panel exhibits several features that motivate double pooling: limited information in the extreme tails, and strong co-movement in downturns. We obtain three main findings. First, the DP model performs well against a range of competing benchmark models. Within the class of DP specifications, it appears most important to allow the tail location to vary, more than its scale or shape, particularly from an out-of-sample perspective. Second, GaR displays substantial time-variation and cross-country heterogeneity that are automatically captured by the DP's three-factor structure. Third, decomposing GES(1%) into a tail location and a tail shape component reveals that the former accounts for much of the risk increases during the global financial and the COVID-19 crises, while the tail shape component plays a relatively larger role during the euro sovereign debt crisis for some specific countries. Taken together, the results show that the DP model uncovers persistent European-wide drivers of downside economic risk, quantifies economically meaningful cross-country asymmetries in tail exposures and shock transmission, and improves the accuracy of out-of-sample tail-risk forecasting in macroeconomic panels.

We contribute to several related strands of literature. First, our approach closely relates to a growing literature on macroeconomic risk assessment, particularly the literature on GaR (Adrian et al., 2019, 2022; Marfè and Pénasse, 2024; Loria et al., 2025; Brownlees et al., 2026). Existing GaR methods typically estimate lower quantiles (e.g., the 5% or 10% quantile) of GDP growth distributions either directly via quantile regressions or indirectly by extracting quantiles from predictive parametric distributions fitted on all the (tail and non-tail) data. The DP approach, by contrast, does not impose explicit parametric assumptions on the predictive distribution, such as skewed-t or other. Instead, our approach is semiparametric in nature and focuses on the tail region only, identifying the dynamics in the more extreme tails using a regularization penalty based on a pool of less extreme GaR levels. Our approach is also related to the growing literature on dynamic tail risk modeling for ES and VaR (Patton et al., 2019) and CoVaR (Fissler and Hoga, 2024). Unlike these earlier papers, however, our method regularizes the more extreme tail by a penalty term across a range of tail regions, including both less and more extreme quantile levels. This regularization is particularly relevant for extracting

extreme risk dynamics from macroeconomic panels with a low frequency time series dimension. Other related papers estimate time-varying extremes by EVT-implied dynamic tail parameters, where extreme VaR and ES arise as by-products (D'Innocenzo et al., 2024, 2026). The DP framework differs in four key aspects: (i) it directly concentrates on the risk measures of interest, which can be an advantage from a policy perspective; (ii) the DP model directly provides a three-component tail representation that allows the extraction of threshold, scale, and tail-shape components in a one-step procedure rather than having to determine the starting point of the tail separately; (iii) due to the three-component structure, the DP model directly generalizes to a panel data setting with both time and cross-sectional effects; and (iv) by introducing the less extreme quantiles via the regularization term, the fit around the quantile of interest versus the other quantiles can be traded off explicitly and in a data-driven way, rather than the EVT tail approximation having to fit all quantiles beyond the threshold equally well. Finally, the DP framework also relates to an emerging literature that exploits cross-quantile and cross-sectional information for dynamic tail estimation. For instance, Chen et al. (2021) minimize a standard check loss function to estimate conditional quantiles across panel units. Catania and Luati (2023) develop an observation-driven approach that models inter-quantile distances, jointly estimating non-crossing dynamic quantiles via a check loss. All these methods, however, primarily focus on less extreme quantiles, whereas our method within the context of low-frequency macro data focuses on the more extreme tail quantiles, while regularizing these with the less extreme quantiles in a data-driven way. In addition, the nonlinear three-factor structure of the DP model becomes more parsimonious than its counterparts if the number of dynamic tail quantiles increases, while still avoiding quantile crossings via an EVT-based constraint rather than via an exponential link function for all quantile differences (Catania and Luati, 2023). Also Müller and Watson (2024) propose a panel approach where parameters of a Generalized Pareto Distribution evolve as random walks and estimation is conducted via Hamiltonian Monte Carlo methods. With respect to this line of research, the DP approach is embedded in the score-driven updating scheme of Creal et al. (2013) and Harvey (2013), allowing for direct estimation without any sampling techniques, while having a robust propagation mechanism for the tail risk parameters in the presence of incidentally large observations.

The remainder of this paper is structured as follows. Section 2 introduces the DP framework and discusses its key properties. Section 3 provides Monte Carlo simulations to illustrate the model's finite-sample performance. Section 4 applies the proposed framework empirically by estimating and analyzing Growth-at-Risk and Growth-Expected-Shortfall for 14 European economies. Section 5 concludes. The appendices provide further theoretical, simulation, and empirical results and details.

2 Double-pooling dynamic tail modeling

In Section 2.1, we first briefly recapitulate the proper scoring rule approach for joint VaR and ES elicitation. In Section 2.2, we then introduce the Double-Pooling (DP) approach and the three-component nonlinear factor model to jointly elicit VaR and ES across different cross-sectional units and different risk quantiles for an unbalanced panel of economic time series. Section 2.3 proceeds by formulating the corresponding filter, while parameter estimation is discussed in Section 2.4.

2.1 Joint VaR and ES elicitation

Let $Y_{i,t}$ denote our observed (macro)economic variable of interest, such as GDP growth or inflation, for unit (e.g., country) $i = 1, \dots, N$ at time $t = 1, \dots, T$. The collection $\{Y_{i,t}\}$ forms a panel. The panel may be unbalanced in practice, and we discuss how the model can deal with missing data in Section 2.3. Let $\mathcal{F}_{t-1}$ denote the information set available up to time $t-1$. This includes the past history of the panel $\{Y_{i,s} : s \leq t-1,\ i = 1, \dots, N\}$, as well as any additional predetermined covariates, if appropriate.

Our interest lies in the more extreme outcomes of $Y_{i,t}$, summarized in measures such as growth or inflation at risk (referred to as VaR from now on) and their Expected Shortfall (ES) counterparts. These measures are defined in terms of a tail probability $\alpha \in (0, 1)$. Rather than sticking to a single α, we immediately introduce the general notation for $J \geq 1$ left-tail areas in order to introduce our regularized objective function later on. Let $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_J < 0.5$ denote the J tail probabilities, with associated VaR ($v_{i,j,t}$) and ES ($e_{i,j,t}$) levels for unit i at risk level α_j and time t, both conditional on $\mathcal{F}_{t-1}$ and defined as

$$\begin{aligned} v_{i,j,t} &= \inf\left\{v \,:\, \mathbb{P}\left[Y_{i,t} \leq v \,\middle|\, \mathcal{F}_{t-1}\right] \geq \alpha_j\right\}, \\ e_{i,j,t} &= \mathbb{E}\left[Y_{i,t} \,\middle|\, Y_{i,t} \leq v_{i,j,t},\, \mathcal{F}_{t-1}\right]. \end{aligned} \tag{1}$$

Fissler and Ziegel (2016) establish that VaR and ES are jointly elicitable: the pair can be characterized as the unique minimizer of a strictly consistent scoring function within a general class. Let $\mathbb{1}\{\cdot\}$ be the indicator function. Following Patton et al. (2019), we use the scoring function

$$L_{FZ}(Y, v, e; \alpha) = -\frac{1}{\alpha e}\,\mathbb{1}\{Y \leq v\}(v - Y) + \frac{v}{e} + \log(-e) - 1, \tag{2}$$

with $e < 0$, which we refer to as the FZ loss. At the population level, jointly minimizing the expected FZ loss recovers the true VaR and ES at level α_j, conditional on $\mathcal{F}_{t-1}$, i.e.,

$$(v_{i,j,t}, e_{i,j,t}) = \underset{(v,e)}{\operatorname{argmin}}\ \mathbb{E}[L_{FZ}(Y_{i,t}, v, e; \alpha_j) \mid \mathcal{F}_{t-1}]. \tag{3}$$

The FZ elicitation procedure in (2)–(3) has several features that are useful for our purpose. First, unlike indirect approaches that recover VaR and ES as by-products of a fully specified (typically parametric) predictive distribution for $Y_{i,t}$, the FZ loss targets the tail risk measures directly without any distributional assumptions. This is important in our setting, where the objective is to model the tail region rather than the entire conditional distribution of $Y_{i,t}$. In particular, we do not make any claims about the distributional shape of $Y_{i,t}$ up to this stage. Second, strict consistency for VaR and ES is obtained under weak conditions, including continuity of the distribution of $Y_{i,t}$, existence of the first moment, uniqueness of the α_j-quantile, and $e < 0$ (see Fissler, 2017, for a comprehensive discussion of FZ loss functions). Third, for our specific choice in Eq. (3), the loss difference $L_{FZ}(Y, v^{(1)}, e^{(1)}; \alpha) - L_{FZ}(Y, v^{(2)}, e^{(2)}; \alpha)$, between two forecasts $(v^{(1)}, e^{(1)})$ and $(v^{(2)}, e^{(2)})$ is homogeneous of degree zero, implying the VaR and ES comparisons between models are invariant to how the data are scaled. This is a desirable property for forecast evaluation (see Proposition 1 in Patton et al., 2019, for the proof and further discussion).

2.2 Three-component tail representation

Without further parametric assumptions on the dynamics of $(v_{i,j,t}, e_{i,j,t})$, the specification in (3) contains too many parameters to estimate. For a single time series ($N = 1$), Patton et al. (2019), therefore, introduce score-driven dynamics (see Creal et al., 2013; Harvey, 2013) for the time-variation in VaR and ES, thus making their approach semiparametric. They then use Eq. (3) to estimate the static parameters that govern the new risk measure dynamics. In a likelihood context, Beutner et al. (2026) show that such score-driven steps result in a consistent estimate of the path of the time-varying parameter, in this case the risk measures, even if the score-driven model itself is heavily misspecified. The argument is generalized to apply to proper scoring rule based score-driven dynamics in Livieri and Palmari (2026).

There are, however, two major challenges for generalizing the approach for a single time series of Patton et al. (2019) to our current panel setting. First, as mentioned in the introduction, economic tail risk measures appear to exhibit quite some commonality over time. It is as yet unclear how to include such commonality in a nonlinear factor type structure for tail risk measures without making further distributional assumptions on $Y_{i,t}$. Second, as shown in D'Innocenzo et al. (2026), using score-driven dynamics based on the FZ loss function may result in undesirable dynamics for the risk measures if the tail quantile is small relative to the number of observations. As the number of time series observations in typical macroeconomic panels is limited, this drawback is directly relevant in our current setting and some regularization seems to be called for.

In the current Double-Pooling (DP) approach, we tackle both of the above challenges at the same time. We do so using an Extreme Value Theory (EVT) based regularization of the objective function. The regularization term is constructed from a pool of different tail risk measures at risk levels α_j and combined into a single, coherent objective, while only imposing mild regularity conditions on the conditional tail shape of $Y_{i,t}$. This approach also directly provides us with a factor type structure for the tail risk dynamics that can easily be generalized in the cross-sectional direction to provide a coherent model for both different tail areas and different cross-sectional units.

Our starting point for the DP approach is the Pickands–Balkema–de Haan peaks-over-threshold (PoT) result for tails (see, e.g., de Haan and Ferreira, 2006, Theorem 1.2.5), which leads to the following approximation for the tail beyond the VaR level $u_{i,t} := v_{i,J,t}$ corresponding to α_J:

$$\mathbb{P}\left[Y_{i,t} \leq y \mid Y_{i,t} \leq u_{i,t}, \mathcal{F}_{t-1}\right] \approx \left(1 - \gamma_{i,t}\frac{y - u_{i,t}}{c_{i,t}}\right)^{-1/\gamma_{i,t}}, \tag{4}$$

for y below the threshold $u_{i,t}$ ($y < u_{i,t}$), where $\gamma_{i,t} > 0$ characterizes the tail shape, and $c_{i,t} > 0$ the tail-scale beyond the threshold $u_{i,t}$.[1] In essence, this states that the extreme left conditional tail of $Y_{i,t}$ can be approximated by a generalized Pareto distribution. Using the above approximation, we obtain the following relations; see Appendix A.1.[2]

Proposition 2.1 (Three-component mapping)**.** *If the left conditional tail of $Y_{i,t}$ beyond $u_{i,t} := v_{i,J,t}$ has an exact generalized Pareto form (4) with $0 < \gamma_{i,t} < 1$, then we obtain the following three-component mapping:*

$$u_{i,t} = v_{i,J,t} = \inf\left\{v \,:\, \mathbb{P}\big[Y_{i,t} \leq v \,\big|\, \mathcal{F}_{t-1}\big] \geq \alpha_J\right\}, \tag{5a}$$

$$v_{i,j,t} = u_{i,t} - \frac{c_{i,t}}{\gamma_{i,t}}\left(\left(\frac{\alpha_J}{\alpha_j}\right)^{\gamma_{i,t}} - 1\right), \quad \text{for } j = 1, \ldots, J-1, \tag{5b}$$

$$e_{i,1,t} = u_{i,t} - \frac{u_{i,t} - v_{i,1,t} + c_{i,t}}{1 - \gamma_{i,t}}, \tag{5c}$$

where $u_{i,t}$ is the threshold or VaR level for the largest tail area α_J, i.e., the quantile that is least far out in the tails, and $c_{i,t} > 0$ and $0 < \gamma_{i,t} < 1$ are tail-scale and tail-shape components that govern the magnitude of losses beyond the threshold $u_{i,t}$.

Proposition 2.1 states that downside risk arises via three economically distinct chan-

[1] Following de Haan and Ferreira (2006), the parameters are also related in the limit via $\lim_{u_{i,t}\to-\infty} c_{i,t}/(-u_{i,t}) = \gamma_{i,t}$. We abstract from this limiting relation to provide the model with more flexibility in the less extreme tail.

[2] It is worthwhile to note that the PoT approximation could be carried further to second or higher order (see, e.g., Hall, 1982; Geluk et al., 1997). Experiments with such further refinements did not yield any further sizeable improvements compared to the current first-order approach.

nels: (i) shifts in tail location, i.e., in the starting point of the tail ($u_{i,t}$), (ii) changes in the scale of tail losses ($c_{i,t}$), and (iii) changes in tail thickness ($\gamma_{i,t}$), where the assumption $\gamma_{i,t} < 1$ ensures the existence of the first moment and, hence, a finite ES. In particular, Proposition 2.1 provides an expression for the more extreme risk quantiles $v_{i,j,t}$ for $j = 1, \ldots, J-1$ and the most extreme ES $e_{i,1,t}$ in terms of a baseline tail risk quantile $u_{i,t} = v_{i,J,t}$. The particular form of the mapping in (5) only requires that the tail of $Y_{i,t}$ is sufficiently fat ($\gamma_{i,t} > 0$) and allows for a first moment ($\gamma_{i,t} < 1$), such that the tail expectation (ES) is well-defined. This is a mild assumption that is commonly adopted both here and elsewhere (Cai et al., 2015; Daouia et al., 2018; Hoga, 2025) and covers a large class of distributional shapes for $Y_{i,t}$.

We also note that neither $u_{i,t}$, nor $c_{i,t}$ and $\gamma_{i,t}$ depend on j. The EVT-inspired mapping in Proposition 2.1 thus pools the information across quantiles α_j for $j = 1, \ldots, J$. It does so while safeguarding the ordering of the extreme quantiles in a theory-based way via the PoT result. In other contexts, such a consistent ordering typically has to be enforced separately, whereas here it follows automatically (see, for instance, Catania and Luati, 2023, who use exponential transformations for the differences of adjacent quantiles to enforce correct ordering over the full distributional support).

The three-component structure in Proposition 2.1 also immediately provides us with the appropriate nonlinear factor structure that can be used for pooling information over the cross-section when using panel data. We specify the following decomposition into cross-sectional heterogeneity and common time effects:

$$u_{i,t} = -\exp\{f_t^u + \bar{f}_i^u\}, \qquad c_{i,t} = \exp\{f_t^c + \bar{f}_i^c\}, \qquad \gamma_{i,t} = \frac{\exp(f_t^\gamma + \bar{f}_i^\gamma)}{1 + \exp(f_t^\gamma + \bar{f}_i^\gamma)}. \tag{6}$$

The common factors f_t^u, f_t^c, and f_t^γ now capture the common time variation in the extreme tail location, scale, and shape that is shared across the cross-sectional units. The quantities $\bar{f}_i^u$, $\bar{f}_i^c$, and $\bar{f}_i^\gamma$, by contrast, capture the static cross-sectional heterogeneity in the tails. The parameterization $u_{i,t}$ ensures that $e_{i,1,t} < 0$, as required by the FZ loss, as well as that all $v_{i,j,t} < 0$, as required further below. Together, Proposition 2.1 and Eq. (6) provide a consistent, EVT-based framework where extreme economic risk dynamics can be filtered from a heterogeneous, unbalanced panel using both the time series, cross-section, and cross-quantile information in the data. The next section discusses how to set up such a filter.

2.3 Filtering and missing values

To filter the tail risk measured directly from the data without imposing further distributional assumptions on $Y_{i,t}$, we proceed as follows. Let $\boldsymbol{Z}_{t-1} \in \mathcal{F}_{t-1}$ collect all lagged

information on $Y_{i,t}$ as well as any predetermined covariates. To allow the tail risk measures to vary over time, we assume that

$$v_{i,j,t} = v_i(\boldsymbol{Z}_{t-1}; \alpha_j, \boldsymbol{\theta}), \qquad e_{i,1,t} = e_i(\boldsymbol{Z}_{t-1}; \alpha_1, \boldsymbol{\theta}),$$

for some parameter vector $\boldsymbol{\theta}$, where we impose the three-component mapping of Proposition 2.1 with the panel decomposition from Eq. (6). Given its dependence on the observable $\boldsymbol{Z}_{t-1} \in \mathcal{F}_{t-1}$, the approach is observation-driven.[3] More specifically, below we use the score-driven dynamic framework of Creal et al. (2013) and Harvey (2013); compare Patton et al. (2019) for the case $N = J = 1$.

To estimate $\boldsymbol{\theta}$, we use a (penalized) FZ loss function, pooled across cross-sectional units. Before formulating the criterion, we first introduce additional notation to account for missing values. Missing values constitute a core challenge for panels of macroeconomic variables. To account for this, let $\boldsymbol{Y}_t^\dagger \in \mathbb{R}^{N_t}$ denote the vector of *observed* cross-sectional units at time t, with corresponding index set $I_t \subseteq \{1, \dots, N\}$ for the elements of $\boldsymbol{Y}_t = (Y_{1,t}, \dots, Y_{N,t})^\top \in \mathbb{R}^N$ that are observed. We now define the DP objective function

$$\begin{aligned}
&L_{DP}(\boldsymbol{Y}_t^\dagger; \boldsymbol{V}_t, \boldsymbol{e}_{1,t}, \boldsymbol{\alpha}) \;= \sum_{i \in I_t} L_{DP,i}(Y_{i,t}; \boldsymbol{v}_{i,t}, e_{i,1,t}, \boldsymbol{\alpha}), \\
&L_{DP,i}(Y_{i,t}; \boldsymbol{v}_{i,t}, e_{i,1,t}, \boldsymbol{\alpha}) \\
&= (1-\rho)\left[-\mathbb{1}\{Y_{i,t} \le v_{i,1,t}\}\,\frac{v_{i,1,t} - Y_{i,t}}{e_{i,1,t}} + \alpha_1\left(\frac{v_{i,1,t}}{e_{i,1,t}} + \log(-e_{i,1,t}) - 1\right)\right] \\
&\quad + \rho\left[\frac{1}{J}\sum_{j=2}^{J}\left(\alpha_j \log(-v_{i,j,t}) + \mathbb{1}\{Y_{i,t} \le v_{i,j,t}\}\log\left(\frac{Y_{i,t}}{v_{i,j,t}}\right)\right)\right],
\end{aligned} \tag{7}$$

with $\boldsymbol{v}_{i,t} = (v_{i,1,t}, \dots, v_{i,J,t})^\top$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)^\top$, $\boldsymbol{V}_t = \{\boldsymbol{v}_{i,t}\}_{i \in I_t}$, and $\boldsymbol{e}_{1,t} = \{e_{i,1,t}\}_{i \in I_t}$, and where the requirements $v_{i,j-1,t} < v_{i,j,t} < 0$ and $e_{i,1,t}$ follow directly from the model's parameterization in Proposition 2.1 and Eq. (6). Under the convention that $\sum_{j=2}^{1} \cdot \equiv 0$, the DP loss in (7) directly generalizes the FZ loss in Eq. (2) by a summation across all available cross-sectional units at time t, and by adding penalty terms for the less extreme VaR levels $v_{i,j,t}$, $j = 2, \dots, J$. If $\rho = 0$, we directly obtain a cross-sectionally pooled version of the original $J = N = 1$ objective function of Patton et al. (2019), with common time components and cross-sectional heterogeneity as specified in (6). For $\rho \in (0, 1)$, the joint elicitation for the most extreme GaR and GES levels $v_{i,1,t}$ and $e_{i,1,t}$ is penalized by the (lack of) fit of the tail shape at less extreme quantiles. As the GPD approximation

[3] An alternative would be a parameter-driven (state-space) type of approach; compare Müller and Watson (2024) who, in a simpler tail context, specify the tail parameters as (unobserved) random walk processes. This typically requires Bayesian or other sampling techniques for estimation. The advantage of an observation-driven paradigm is that standard estimation techniques apply.

may be less accurate in the less extreme tails, the loss at the less extreme quantiles may be weighted differently compared to the fit at the (α_1) quantile of interest. This can be done by setting the regularization parameter ρ appropriately.[4] The penalization prevents overfitting of the extreme tail risk dynamics of the tail parameters $u_{i,t}$, $c_{i,t}$, and $\gamma_{i,t}$. We comment on how to set the penalty parameter ρ in a data-driven way in Section 4, similar to the penalty parameter in a LASSO approach. Note that the penalized objective still fits the general proper scoring rule framework of Fissler and Ziegel (2016). Moreover, its loss differences between different models remain homogeneous of degree zero, which is a desirable property for model forecast comparisons.

As mentioned earlier, we model the dynamics of f_t^u, f_t^c, and f_t^γ in the panel decomposition of VaR and ES using the score-driven dynamics of Creal et al. (2013) and Harvey (2013),

$$\boldsymbol{f}_t = \boldsymbol{B}\boldsymbol{f}_{t-1} + \boldsymbol{A}\boldsymbol{s}_{t-1}, \qquad \boldsymbol{f}_t = (f_t^u, f_t^c, f_t^\gamma)^\top, \qquad \boldsymbol{s}_t = (s_t^u, s_t^c, s_t^\gamma)^\top, \tag{8}$$

where we use the DP loss function in (7) to compute the scaled scores $\boldsymbol{s}_t$, with

$$\begin{aligned} s_t^u &= (H_t^u)^{-1/2}\, \nabla_t^u, \qquad \nabla_t^u = \sum_{i \in I_t} \nabla_{i,t}^u, \qquad H_t^u = \sum_{i \in I_t} H_{i,t}^u, \\ \nabla_{i,t}^u &= -\frac{\partial L_{DP,i}(Y_{i,t}; \boldsymbol{v}_{i,t}, e_{i,1,t}, \boldsymbol{\alpha})}{\partial f_t^u}, \qquad H_{i,t}^u = \mathbb{E}\left[\frac{\partial^2 L_{DP,i}(Y_{i,t}; \boldsymbol{v}_{i,t}, e_{i,1,t}, \boldsymbol{\alpha})}{\partial (f_t^u)^2} \,\middle|\, \mathcal{F}_{t-1} \right], \end{aligned} \tag{9}$$

and similar definitions for the other elements of $\boldsymbol{s}_t$. No intercepts are included in the recursion in (8), given these are already accounted for by the cross-sectionally heterogeneous terms $\bar{f}_i^u$, $\bar{f}_i^c$, and $\bar{f}_i^\gamma$ in (6). The use of the score improves the three quantities f_t^u, f_t^c, and f_t^γ at each moment in a steepest descent direction given the DP loss function (compare Patton et al., 2019; Catania and Luati, 2023). For likelihood and GMM objective functions, such score steps have already been shown to reduce their respective criterion functions in expectation upon each step (Blasques et al., 2015; Creal et al., 2024; De Punder et al., 2024). For proper scoring rules, see Livieri and Palmari (2026). The scaling by H_t^u further accounts for the curvature of the score steps in terms of the time-varying parameters and further stabilizes the model (see Creal et al., 2013; Beutner et al., 2026). Using these definitions, we obtain the following result.

Proposition 2.2 (Scaled score-driven steps)**.** *Given the three-component mapping from Proposition 2.1 and the two-way factor decomposition in* (6)*, we obtain the score expres-*

[4]The current approach opts for a linear weighting of the two components in (7). Further refinements are also possible, e.g., an exponential penalization scheme that replaces $1-\rho$ and ρ by $\rho(\alpha;\alpha_1) = (\alpha_1/\alpha)^k$ for $\alpha \geq \alpha_1$ and some $k > 0$.

sions:

$$\nabla^u_{i,t} = (1-\rho)u_{i,t}\left[\frac{(v_{i,1,t}-e_{i,1,t})\lambda^v_{i,1,t}}{e^2_{i,1,t}} + \frac{\alpha_1\lambda^e_{i,t}}{e^2_{i,1,t}}\right] - \frac{\rho}{J}\sum_{j=2}^{J}\frac{u_{i,t}\lambda^v_{i,j,t}}{v_{i,j,t}},$$

$$\nabla^c_{i,t} = (1-\rho)\left[\frac{u_{i,t}(e_{i,1,t}-v_{i,1,t})\lambda^v_{i,1,t}}{e^2_{i,1,t}} - \frac{\alpha_1(u_{i,t}-e_{i,1,t})\lambda^e_{i,t}}{e^2_{i,1,t}}\right] + \frac{\rho}{J}\sum_{j=2}^{J}\frac{(u_{i,t}-v_{i,j,t})\lambda^v_{i,j,t}}{v_{i,j,t}},$$

$$\nabla^\gamma_{i,t} = (1-\rho)\left[\frac{\left[(\gamma_{i,t}u_{i,t}+c_{i,t})\lambda^\gamma_{i,1,t} - v_{i,1,t}\gamma_{i,t}(u_{i,t}-e_{i,1,t})\right]\lambda^v_{i,1,t}}{e^2_{i,1,t}} + \frac{\alpha_1\left[\lambda^\gamma_{i,1,t}-\gamma_{i,t}(u_{i,t}-e_{i,1,t})\right]\lambda^e_{i,t}}{e^2_{i,1,t}}\right] - \frac{\rho}{J}\sum_{j=2}^{J}\frac{(1-\gamma_{i,t})\lambda^\gamma_{i,j,t}\lambda^v_{i,j,t}}{v_{i,j,t}},$$

where $\lambda^v_{i,j,t} = \alpha_j - \mathbb{1}\{Y_{i,t} \le v_{i,j,t}\}$, $\lambda^e_{i,t} = \alpha_1^{-1}\mathbb{1}\{Y_{i,t} \le v_{i,1,t}\}Y_{i,t} - e_{i,1,t}$, *and* $\lambda^\gamma_{i,j,t} = \big(1+\gamma_{i,t}\log(\alpha_j/\alpha_J)\big)(u_{i,t}-v_{i,j,t}) + c_{i,t}\log(\alpha_j/\alpha_J)$. *For the scaling factors of the score,*

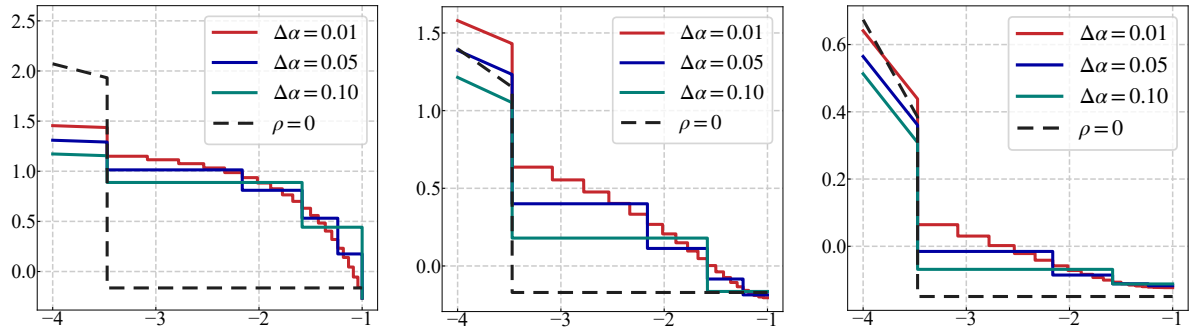
$$H^u_t = \sum_{i\in I_t} u^2_{i,t}\left[(1-\rho)\left(\frac{\alpha_1}{e^2_{i,1,t}} - \frac{p_{i,1,t}}{e_{i,1,t}}\right) + \frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}}{(-v_{i,j,t})}\right],$$

$$H^c_t = \sum_{i\in I_t}\left[(1-\rho)\left(\frac{\alpha_1(u_{i,t}-e_{i,1,t})^2}{e^2_{i,1,t}} - \frac{p_{i,1,t}(u_{i,t}-v_{i,1,t})^2}{e_{i,1,t}}\right) + \frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}(u_{i,t}-v_{i,j,t})^2}{(-v_{i,j,t})}\right],$$

$$H^\gamma_t = \sum_{i\in I_t}\left[(1-\rho)\left(\frac{\alpha_1\left[\lambda^\gamma_{i,1,t}-\gamma_{i,t}(u_{i,t}-e_{i,1,t})\right]^2}{e^2_{i,1,t}} - \frac{p_{i,1,t}(1-\gamma_{i,t})^2(\lambda^\gamma_{i,1,t})^2}{e_{i,1,t}}\right) + \frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}(1-\gamma_{i,t})^2(\lambda^\gamma_{i,j,t})^2}{(-v_{i,j,t})}\right],$$

where $p_{i,j,t} = \frac{\alpha_j}{c_{i,t}}\left(\frac{\alpha_j}{\alpha_J}\right)^{\gamma_{i,t}}$ *is the generalized Pareto distribution (GPD) implied conditional density evaluated at* $v_{i,j,t}$.

The scores in (9) automatically account for missing observations by only summing over the N_t observations $i \in I_t$ for which $Y_{i,t}$ is available. The expressions in Proposition 2.2 also pool across quantiles by summing the scores over $j = 1, \ldots, J$. Only if the extreme tail is not regularized, i.e., if $\rho = 0$, these terms drop from the expressions. In this way, the score-driven steps exploit all available information in the panel on the tail dynamics via the EVT-inspired nonlinear three-factor structure of the tail, regularizing the extreme tail dynamics via its implications for the less extreme tail. Note that the DP approach can easily allow for more dynamics in (8), including more lags of $\boldsymbol{f}_t$ or $\boldsymbol{s}_t$ or even fractionally



(a) s_t^u (b) s_t^c (c) s_t^γ

Figure 2: Score functions determining the News Impact Curve for f_t^u, f_t^c, and f_t^γ in the two-way panel decomposition in (6) using Proposition 2.2 for three different numbers $J = 3, 5, 21$ of pooled tail area probabilities and $\alpha_1 = 0.05$ and $\alpha_J = 0.25$ with $\rho = 0.5$.

integrated dynamics for $\boldsymbol{f}_t$ (for the latter, see, e.g., Opschoor and Lucas, 2019). Similarly, we can incorporate additional covariates in the tail dynamics by specifying

$$\boldsymbol{f}_t = \boldsymbol{B}\boldsymbol{f}_{t-1} + \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{C}\boldsymbol{x}_t,$$

for a vector of covariates $\boldsymbol{x}_t$ and a matrix of coefficients $\boldsymbol{C}$. All of this can be done without materially changing any of the expressions in Proposition 2.2.

To obtain some intuition for how the score-driven dynamics change the time-varying parameters when new data arrives, we plot the score expressions as a function of $Y_{i,t}$, a plot also known as the News Impact Curve (NIC). We concentrate on the NICs for one cross-sectional unit i, realizing that the full NIC is just an average of similar NICs over the cross-section. The results are presented in Figure 2.

The patterns in Figure 2 directly follow the result in Proposition 2.2. The proposition states that the NICs contain two types of reactions to incoming data, namely a step-wise (or check-loss) movement via the quantities $\lambda^v_{i,j,t} = \alpha_j - \mathbb{1}\{Y_{i,t} \leq v_{i,j,t}\}$, and a piecewise linear (or ReLU) reaction via $\lambda^e_{i,t} = \alpha_1^{-1}\,\mathbb{1}\{Y_{i,t} \leq v_{i,1,t}\}\,Y_{i,t} - e_{i,1,t}$. All the other quantities do not depend on the current data $Y_{i,t}$. The check-loss reactions are located at each of the VaR levels $v_{i,j,t}$ for $j = 1, \ldots, J$, whereas the linear reaction only sets in beyond the most extreme conditional quantile $v_{i,1,t}$. The effect is clearly visible in Figure 2. We fix $\rho = 0.5$, $u_{i,t} = -1$, $c_{i,t} = 1$, and $\gamma_{i,t} = 0.5$, and plot the effect of a hypothetical realization of $Y_{i,t}$ on s_t^u, s_t^c, and s_t^γ, which is the core part of the NIC. We use three different equally spaced grids for the α_js, ranging from $\alpha_1 = 0.05$ to $\alpha_J = 0.25$, with $J = 3, 5, 21$. At each of the J different VaR levels, the scores make a discrete jump, while beyond the deepest VaR level, the NIC continues linearly. As the spacing between neighboring quantile levels decreases, the NIC of the regularized DP filter becomes smoother. Without such regularization, the NIC in Figure 2 for $\rho = 0$ shows that the model only reacts to extreme,

highly incidental growth realizations, and that the corresponding reaction is large. This is in line with the undesirable dynamics also highlighted in D'Innocenzo et al. (2026). The robustness of the DP approach is further extended by pooling in the cross-sectional direction. Cross-sectional pooling causes extremes in a single cross-sectional unit to have less of an effect on the time-varying parameter, whereas joint exceedances result in a stronger reaction. Jointly, the two ways of pooling in the DP model provide a much more stable and economically interpretable pattern for economic tail risk dynamics.

2.4 Parameter estimation

Let $\boldsymbol{\theta} = \left(\bar{\boldsymbol{f}}^{u\top}, \bar{\boldsymbol{f}}^{c\top}, \bar{\boldsymbol{f}}^{\gamma\top}, \mathrm{vec}(\boldsymbol{B})^\top, \mathrm{vec}(\boldsymbol{A})^\top\right)^\top \in \boldsymbol{\Theta}$ collect all unknown parameters, where $\bar{\boldsymbol{f}}^u$, $\bar{\boldsymbol{f}}^c$, and $\bar{\boldsymbol{f}}^\gamma$ collect the unit-specific long-run components in the two-way decomposition in (6), and $\boldsymbol{A}$ and $\boldsymbol{B}$ are the 3×3 coefficient matrices in the score-driven filtering equation (8). We initialize the filter recursion at a fixed non-random value $\widehat{\boldsymbol{f}}_0$, and write the resulting initialized filter as $\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) = \widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}, \widehat{\boldsymbol{f}}_0)$, for $t = 1, \ldots, T$. The static parameters are then estimated as

$$\widehat{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T(\widehat{\boldsymbol{f}}_0) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} L_{DP}\Big(\tilde{\boldsymbol{Y}}_t; \boldsymbol{V}_t\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big), \boldsymbol{e}_{1,t}\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big), \boldsymbol{\alpha}\Big), \tag{10}$$

where $\tilde{\boldsymbol{Y}}_t = \boldsymbol{Y}_t - \boldsymbol{C}$, with $\boldsymbol{C}$ denoting a constant shift vector chosen to ensure that $u_{i,t} < 0$.[5] Moreover, $\boldsymbol{V}_t(\cdot)$ and $\boldsymbol{e}_{1,t}(\cdot)$ are obtained by combining (i) the three-component tail representation in Proposition 2.1, (ii) the two-way effects decomposition for the tail parameters in Eq. (6), and (iii) the score-driven recursions from Eq. (8) and Proposition 2.2. Note that this requires no specific assumptions on the conditional distribution of $Y_{i,t}$. Combined with the parametric part for describing the dynamics of the tail-risk measures $v_{i,j,t}$ and $e_{i,1,t}$, our approach thus becomes semiparametric.

For a given value of $\boldsymbol{\theta}$, the criterion in (10) is straightforward to evaluate, because the filter is generated recursively. Compared with much of the existing literature on time-varying EVT-based tail estimation, the DP estimator is obtained in one step; see also Patton et al. (2019). Alternative approaches typically proceed in two steps by first estimating a threshold quantile and then modeling the time-varying excess distribution above this threshold (D'Innocenzo et al., 2024; Müller and Watson, 2024; Fan et al., 2025). The current unified way of estimating the parameters has two advantages. First, it mitigates the potential propagation of estimation errors from the thresholds to the tail parameters that is typical in a two-stage approach. Second, by using one unified,

[5] Empirically, we set $\boldsymbol{C}_i = \hat{\mathcal{Q}}_{0.99}(\boldsymbol{Y}_i)$, where $\hat{\mathcal{Q}}_{0.99}(\boldsymbol{Y}_i)$ denotes the empirical 0.99 quantile for individual i in the training sample for $i = 1, \ldots, N$.

single-step estimation procedure, we can exploit information from the entire tail region (multiple quantiles) and the cross-sectional dimension. These features are particularly valuable in macroeconomic applications, where the time dimension is often short and extreme-tail observations are scarce.

3 Simulations

We first investigate the performance of the new method in a simulation setting. In particular, we investigate the contributions of cross-sectional and cross-quantile (regularization) pooling in cases where the model is incorrectly specified. This seems empirically the most relevant setting. In Appendix B, we also consider a setting where the DP model is correctly specified and show that RMSEs of static parameter estimates decrease as T grows from 50 to 200, suggesting the consistency of the estimation procedure.

For the misspecified setting, we consider the following data generating process for a population of $N = 14$ countries:

$$\begin{aligned} Y_{it} &= \sigma_{it}\varepsilon_{it}, \qquad & \varepsilon_{it} &= \sqrt{\frac{\nu_i - 2}{\nu_i}}\, Z_{it}, \qquad & Z_{it} &\sim t_{\nu_i}, \\ \sigma_{it}^2 &= \exp(a_i + f_t), \quad & f_t &= 0.8 f_{t-1} + 0.6\eta_t, \quad & \eta_t &\overset{\text{iid}}{\sim} \mathcal{N}(0,1), \end{aligned}$$

where t_ν denotes the Student-t distribution with ν degrees of freedom. The variables Z_{it} are independent across both i and t, and are independent of the volatility innovations η_t. Note that $\mathbb{V}\text{ar}(\varepsilon_{it}) = 1$. We set $a_i = 0.7 + 0.6(i-1)/13$ and $\nu_i = 3 + 8(i-1)/13$ for $i = 1, \ldots, 14$, such that the data do *not* follow a simple multivariate Student's t specification. The common state f_t is initialized from a standard $\mathcal{N}(0,1)$, with a burn-in period of 1,000 observations before starting the training and testing samples. This design generates common movements in volatility and tail risk while allowing individual differences in scale and tail thickness.

We generate 300 independent samples, each containing $T + 100$ observations for training and testing, with $T \in \{50, 100, 200\}$. For each sample, we estimate the model with $N \in \{1, 7, 14\}$. For $N = 1$, we estimate the model for each of the 14 simulated series separately. For $N = 7$, we estimate the model for series 1–7 and 8–14, respectively, while for $N = 14$ we estimate the model for all data simultaneously. Every simulation thus uses the same full set of data. We estimate the model on observations $1, \ldots, T$, and evaluate the one-step-ahead predictions for observations $T+1, \ldots, T+100$. The quantile regularization grid is $\{\alpha_1, \alpha_2, \ldots, \alpha_J\} = \{0.01, 0.02, \ldots, 0.20\}$. For the simulations we focus on the DP model only. In our empirical study in Section 4, by contrast, we include a range of competing benchmark models.

Table 1: Tail risk under model misspecification

		GaR (1%)		GaR (5%)		GaR (10%)		GaR (20%)		GES (1%)	
T	N	RMSE	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE	Corr.
50	1	3.204	0.259	1.696	0.262	1.246	0.263	0.874	0.254	6.148	0.257
50	7	2.928	0.496	1.549	0.495	1.125	0.492	0.776	0.479	5.372	0.495
50	14	**2.761**	**0.570**	**1.468**	**0.569**	**1.066**	**0.566**	**0.738**	**0.555**	**5.038**	**0.569**
100	1	2.892	0.293	1.588	0.296	1.172	0.296	0.799	0.286	5.050	0.290
100	7	2.546	0.532	1.394	0.535	1.025	0.533	0.692	0.521	4.323	0.529
100	14	**2.418**	**0.600**	**1.320**	**0.603**	**0.968**	**0.601**	**0.658**	**0.591**	**4.098**	**0.596**
200	1	2.820	0.308	1.549	0.312	1.140	0.313	0.760	0.302	4.634	0.304
200	7	2.406	0.548	1.339	0.555	0.987	0.555	0.655	0.541	3.795	0.543
200	14	**2.274**	**0.614**	**1.265**	**0.621**	**0.932**	**0.621**	**0.620**	**0.609**	**3.616**	**0.609**

Notes. $N = 1$ estimates the model for each individual separately; $N = 7$ uses two fixed groups (1–7 and 8–14), and $N = 14$ estimates the entire panel jointly. RMSE is the square root of the mean squared path error, averaged over the T training observations, individuals, and replications. Corr. is the within-individual time-series correlation between the fitted and oracle paths, averaged across individuals and replications. The oracle conditions on the true latent SV state. For each T and risk measure, boldface identifies the lowest RMSE and the highest correlation across N.

Table 1 reports the in-sample RMSEs and time-series correlations for four GaR levels and the extreme GES using a regularization parameter $\rho = 0.5$. Increasing N from 1 to 7 and then to 14 lowers RMSE and raises correlation for every risk measure and time span. For each N, longer estimation samples also reduce RMSEs and increase correlations across all five measures. The gains in tracking tail-risk dynamics are already visible at $T = 50$, where the GaR and GES correlations increase from 0.259 and 0.257 for $N = 1$ to 0.570 and 0.569 for $N = 14$. These results indicate that cross-sectional information may be helpful for recovering both the levels and the time variation of tail risk, even when the conditional distribution and volatility dynamics are misspecified.

We next evaluate the one-step-ahead forecasting performance of the model over the 100 observations out-of-sample period. Parameters are not re-estimated during this period, but the filtered tail risk measures are updated as new observations become available. During estimation, we consider different values for the regularization parameter ρ in the grid $\{0.10, 0.25, 0.50, 0.75, 0.90\}$ to study the effect of this parameter on the out-of-sample results.

Figure 3 reports the forecast RMSEs of 1% GaR and GES and the mean unregularized out-of-sample FZ loss. The gains from pooling across individuals clearly extend to out-of-sample prediction. For every combination of T and ρ, both forecast RMSEs and the mean FZ loss decline as N increases from 1 to 7 and then to 14. Mean FZ loss also decrease with T for every group size and pooling weight. The additional cross-sectional

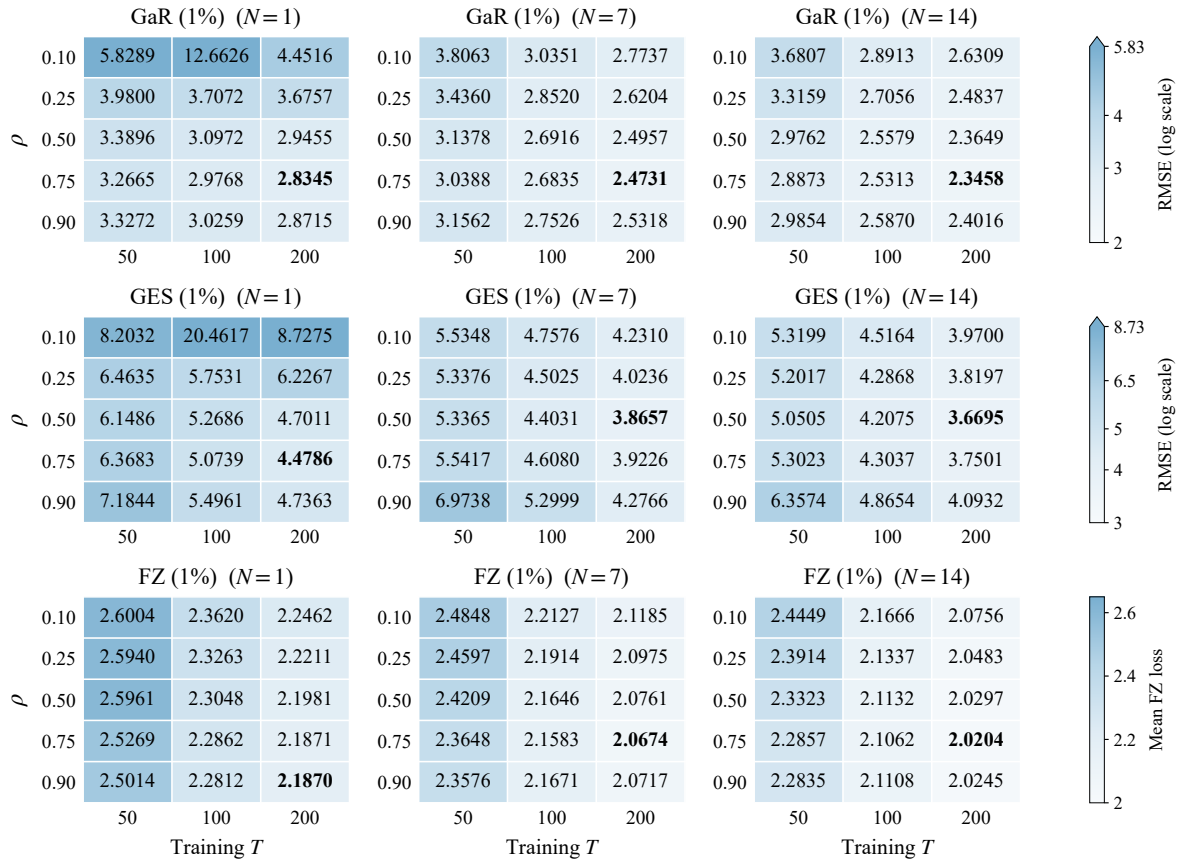


Figure 3: Out-of-sample tail-risk accuracy across regularization weights. Rows report the RMSEs of 1% GaR and GES and the mean FZ loss at $\alpha = 0.01$, respectively. Columns correspond to the estimation-group sizes $N = 1, 7, 14$. Entries summarize 300 replications evaluated over the subsequent 100 observations. Boldface identifies the lowest value in each panel. Lower values indicate better performance.

information therefore improves predictive performance under the current setting with misspecified tail dynamics and a common volatility component.

Regarding the choice of ρ, it is clear from Figure 3 that regularization improves the estimation quality of tail risk dynamics. All cases favor a value of $\rho \geq 0.25$. The precise effect of cross-quantile regularization nevertheless differs across evaluation criteria. The GaR forecast RMSE is lowest at $\rho = 0.75$ in all nine (N, T) designs. GES forecast errors generally favor more moderate pooling, with $\rho = 0.50$ delivering the lowest RMSE in seven of the nine designs, including all specifications with $N = 7$ and $N = 14$. The joint FZ loss also favors $\rho = 0.75$ or even $\rho = 0.90$ across the nine designs. All three criteria therefore favor substantial cross-quantile pooling rather than weak regularization ($\rho = 0.10$), highlighting the value of pooling across quantile levels. The differences between different performance measures also suggest that ρ needs to be set in a data-driven way. We turn to this in the next section.

4 Downside growth risk in Europe

4.1 Data and preliminary evidence

In our empirical analysis, we consider 14 European economies: Austria (AUT), Belgium (BEL), Czech Republic (CZE), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Hungary (HUN), Italy (ITA), Netherlands (NLD), Poland (POL), Portugal (PRT), and Spain (ESP); see Appendix C for more details. We focus on year-on-year real GDP growth using seasonally adjusted real GDP levels $L_{i,t}$. The sample covers the period from March 31, 1961 to June 30, 2026, although the effective starting dates differ across countries.

Figure 4 displays a heatmap of quarterly GDP growth rates. The figure highlights several features that are directly relevant for dynamic tail estimation. First, the panel exhibits a ragged-edge structure, reflecting differences in data availability, particularly for Eastern European economies due to the Cold War era. Second, growth dynamics display a clear transition from rapid expansions in the 1960s and early 1970s to more moderate growth in later decades. This broad pattern coexists with periods of rapid growth in CZE, HUN, and POL in the late 1990s and early 2000s, followed by more moderate growth in the 2010s, consistent with the broader European slowdown. Third, the COVID-19 shock in 2020 appears as a sharp and nearly synchronous contraction across countries, followed by a heterogeneous rebound. Finally, the panel exhibits persistent cross-sectional heterogeneity: economies such as DEU, FRA, NLD, AUT, BEL, and FIN tend to display steadier growth, whereas economies such as GRC, PRT, ESP, and to some extent ITA, experience larger volatility and deeper downturns, particularly during the euro area sovereign debt crisis. These features motivate a framework that exploits both time series common effects and cross-sectional heterogeneity when measuring macroeconomic tail risk.

To further summarize the information contained in the panel, we return to the GDP growth series shown in Figure 1. The gray lines represent standardized country-level GDP growth rates, while the blue solid line denotes the common factor extracted from the balanced subsample with no missing observations, and the red dashed line reports the pooled 1% sample quantile obtained by pooling all country observations across all time points. Two findings are worth emphasizing. First, GDP growth rates exhibit strong cross-country comovement, especially during major downturns, indicating that pooling across individuals can help identify common dynamics in macroeconomic tail risk. Second, even after pooling observations across all countries, the absolute number of observations in the 1% tail remains small. This illustrates the information scarcity problem in the far tail and underscores the value of borrowing information from neighboring

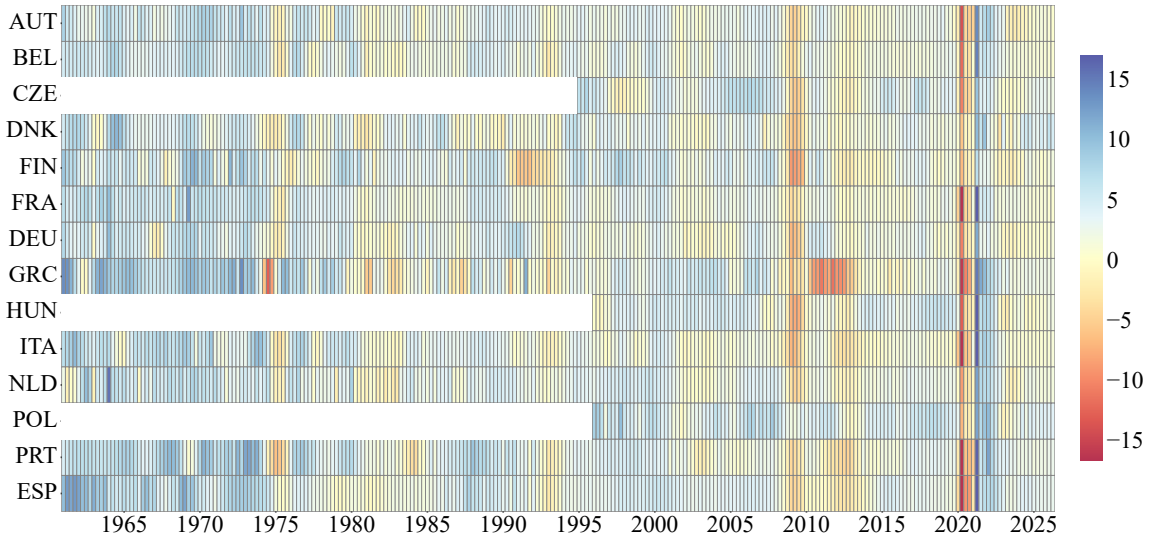


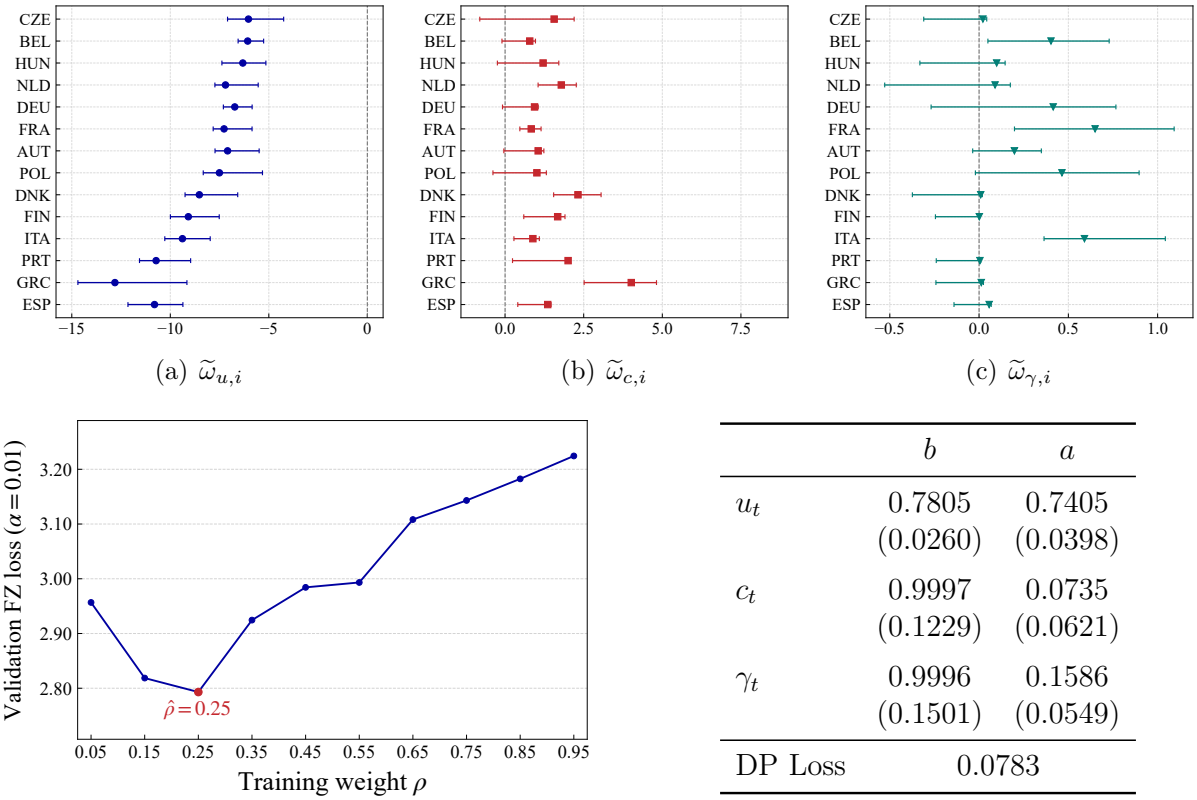
Figure 4: Heatmap of quarterly GDP growth across 14 European economies.

quantile levels. Together, these patterns provide direct empirical motivation for applying the DP framework to country-level Growth-at-Risk and extreme expected shortfall.

4.2 In-sample estimation results

To determine the country-level Growth-at-Risk (GaR) and Growth Expected Shortfall (GES) dynamics, we estimate the DP specification. We focus on the 1% GaR and GES tail risk measures, using the GaR at $\alpha_j = j/100$ for $j = 2, \ldots, 20$ for regularization, similar to Section 3. The tuning parameter $\hat{\rho} = 0.25$ is selected by splitting the samples into a 70% training set and a 30% validation set, then choosing ρ that minimizes the out-of-sample FZ($\rho = 0$) loss at the $\alpha_1 = 0.01$ quantile level in the validation set, see the lower-left panel in Figure 5.

Figure 5 reports the full-sample estimates. The common European tail components exhibit different dynamics. The threshold $u_t = \exp(f_t^u)$ is persistent ($b_u \approx 0.78$) with sizable time-variation ($a_u \approx 0.74$). The standard error of a_u is small compared to the size of the coefficient, supporting the dynamics for macroeconomic tail risk. The persistence parameters for $c_t = \exp(f_t^c)$ and $\gamma_t = \exp(f_t^\gamma)$ are close to one ($b_c \approx 1.00$ and $b_\gamma \approx 1.00$) and difficult to estimate reliably. The corresponding score loading parameters a_c and a_γ, also have large standard errors compared to their coefficient sizes compared to the threshold case (a_u). The standard error for a_c is particularly large, raising the question of whether the tail scale genuinely varies over time or not. Given the nonidentification problem under $H_0: \ a_c = a_\gamma = 0$ for the persistence parameters b_c and b_γ (see Lin and Lucas, 2027) and the non-Lipschitz score-driven updates as shown in Proposition 2.2, a formal testing procedure is currently unavailable. We therefore investigate the time



(a) $\widetilde{\omega}_{u,i}$ (b) $\widetilde{\omega}_{c,i}$ (c) $\widetilde{\omega}_{\gamma,i}$

	b	a
u_t	0.7805	0.7405
	(0.0260)	(0.0398)
c_t	0.9997	0.0735
	(0.1229)	(0.0621)
γ_t	0.9996	0.1586
	(0.1501)	(0.0549)
DP Loss	0.0783	

Figure 5: Full-sample estimation results

Note: Top panels present the intercept estimates and corresponding bootstrapped intervals; see Appendix C.1. The FZ loss in the lower-left panel is computed for different values of the regularization parameter ρ over the test sample for a 70/30 train/test sample split.

variation of all tail parameters in our out-of-sample forecasting analysis in Section 4.3 by also including versions of the DP model with constant c_t, γ_t, and u_t.

A second take-away from Figure 5 is that the country-specific components reveal substantial heterogeneity in long-run tail risk. The threshold component $\tilde{\omega}_{u,i}$ ranges from comparatively mild values for CZE −6.03 and BEL −6.07, to substantially lower tail locations for GRC −12.81, ESP −10.80, and PRT −10.72. More negative threshold components indicate that the left-tail region is located further below the country-specific 99th percentile of real GDP growth. For instance, countries hit hard during the European sovereign debt crisis (GRC, ESP, PRT, ITA) have the most negative values for $\tilde{\omega}_{u,i}$. These differences remain even after accounting for common time-variation of the tail.

Third, long-run tail scale estimates $\tilde{\omega}_{c,i}$ vary markedly across countries. Greece has the highest value of $\tilde{\omega}_{c,i} = \exp\{\bar{f}_i^c\}$, with GRC 4.01, indicating larger tail loss magnitudes. By contrast, the Netherlands, the Czech Republic, and Hungary exhibit much smaller tail scales, with NLD 1.79, CZE 1.56, and HUN 1.21. Estimation uncertainty remains substantial for some countries, particularly CZE. The long-run tail-shape parameters $\tilde{\omega}_{\gamma,i}$ are most challenging to estimate and generally exhibit greater uncertainty relative to their magnitudes, particularly where the estimates are close to zero.

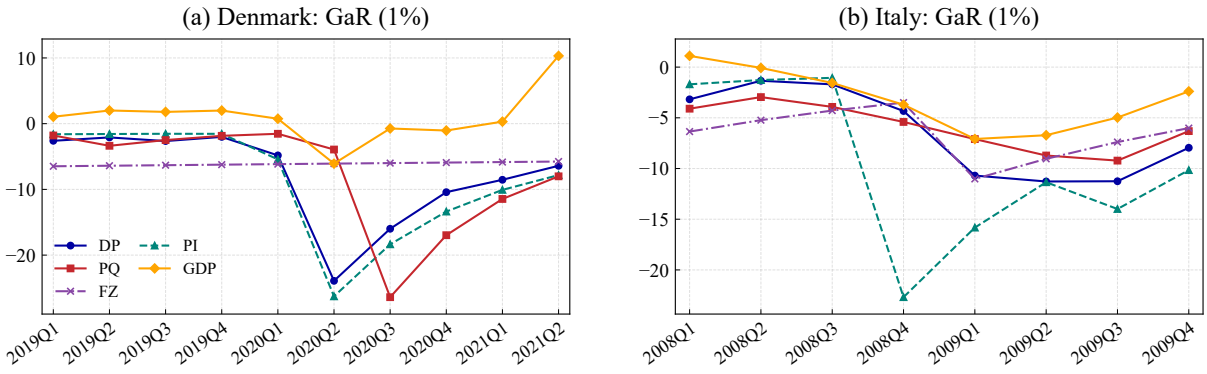


Figure 6: Cross-country and cross-quantile pooling in growth-at-risk estimation. Panel (a) presents Denmark during 2019Q1–2021Q2, and panel (b) presents Italy during 2008Q1–2009Q4. The figure compares full-sample fitted GaR(1%) paths from DP, PI, and PQ with realized year-on-year GDP growth.

More interesting than the parameter estimates themselves, however, are the fitted risk measure dynamics. As an example, Figure 6 compares the full-sample fitted GaR(1%) paths with GDP growth in Denmark during the COVID-19 episode and Italy during the global financial crisis (full results for all countries are available in Appendix C.2). Next to the DP model that jointly pools information across countries and quantile levels, we also provide the GaR(1%) for a PQ model that only pools across quantiles for a single country, for a PI model that pools across countries (i.e., across the i dimension), but reduces the cross-quantile pooling to $\rho = 0.01$ and $J = 2$, and finally for the FZ model of Patton et al. (2019), which does not use any pooling.

Panel (a) in Figure 6 highlights the contribution of cross-country information to the timing of tail-risk adjustments. Both DP and PI indicate a pronounced increase in risk in 2020Q2, which coincides with the drop in real GDP growth that is actually realized in that quarter. Pooling across quantiles only (PQ) is not sufficient here, and only indicates an increase in downside risk a quarter later in 2020Q3, *after* the contraction of Denmark's own economy. This illustrates the value of pooling information across countries to obtain more timely warning signals.

Panel (b), by comparison, shows how pooling across quantiles (or regularization) helps to stabilize extreme tail risk estimates during severe downturns in some countries. In this case, only using cross-sectional pooling (PI) results in a major shift in GaR(1%) from close to zero in 2008Q3 to −22.69% in 2008Q4. The DP and PQ approaches only show modest declines to −4.33% and −5.41%, respectively, by regularizing the tail risk dynamics via the pool of less extreme tail outcomes. Moreover, the PQ and particularly the DP GaR paths more closely follow the actual real GDP growth realizations for ITA during all of 2009, including the gradual improvements in growth. We also note that no pooling (FZ) appears problematic in both settings: for Denmark, there are no dynamics for extreme

risk at all, not even during the COVID-19 crisis. For Italy, we see that the FZ yields a dynamic GaR, but the GaR still goes up in 2008Q3, where all pooled counterparts (DP, PQ, PI) already signal an increase in risk. Taken together, these examples support the potential usefulness of the double pooling approach for capturing tail risk dynamics. Also the full results in Appendix C.2 show similar patterns.

Figure 7 compares the in-sample 1% GaR paths across the 14 countries. As a point of reference, we also plot the 1% GaR paths using the univariate approach of Patton et al. (2019). For most countries, the DP estimates remain broadly aligned with the empirical 1% quantiles, while capturing pronounced increases in downside risk during major downturns. The FZ paths are estimated without any pooling. As a result, they face the challenge of very few (1%) observations to estimate the dynamics of the extreme tail risk measures. The FZ paths therefore exhibit sharper jumps and prolonged deviations, particularly for Hungary and Poland, making the evolution of tail risk harder to interpret. At the 1% level, sparse tail exceedances clearly provide limited information for estimating country-specific tail risk dynamics, especially in shorter samples. As a result, the DP GaR paths can also be more responsive, such as for Denmark, Italy and Portugal during the early 1990s, without resulting in the large spikes of the unpooled FZ results. The pooling of information across individuals and quantile levels thus makes the GaR paths more stable and economically interpretable.

Finally, Figure 8 decomposes extreme growth expected shortfall (GES(1%)) in a tail threshold ($u_{i,t}$) component, and a remainder $e_{i,1,t} - u_{i,t}$, where the latter mainly captures the movements in tail shape. We do so using the expressions from Proposition 2.1. The black solid line in Figure 8 shows the GES(1%), i.e., $e_{i,1,t}$, the blue area represents the contribution of the threshold component $u_{i,t}$, and the yellow area captures the contribution of the tail shape component $e_{i,1,t} - u_{i,t}$. There are clear differences in the contributions of the two components across countries and across time. Episodes where the threshold component dominates for all countries include the global financial crisis and the 2020 COVID pandemic. This suggests that these major shocks primarily induce a leftward shift of the left tail of the overall conditional distribution.

The decomposition also highlights heterogeneity in the mechanisms generating extreme downside risk. For example, during the 2010–2014 euro area sovereign debt crisis, the contribution of the (yellow) tail shape component increases for countries like Denmark, Greece, the Netherlands, and the Czech Republic. This suggests that their expected losses are amplified by tail rescaling and tail thickening, and not only by shifts in the tail location $u_{i,t}$. By contrast, economies like Spain and Italy only experience a small tail shape component, indicating that most of their GES risks are already captured by the GaR and threshold $u_{i,t}$ downward shift. Such differences indicate that European down-

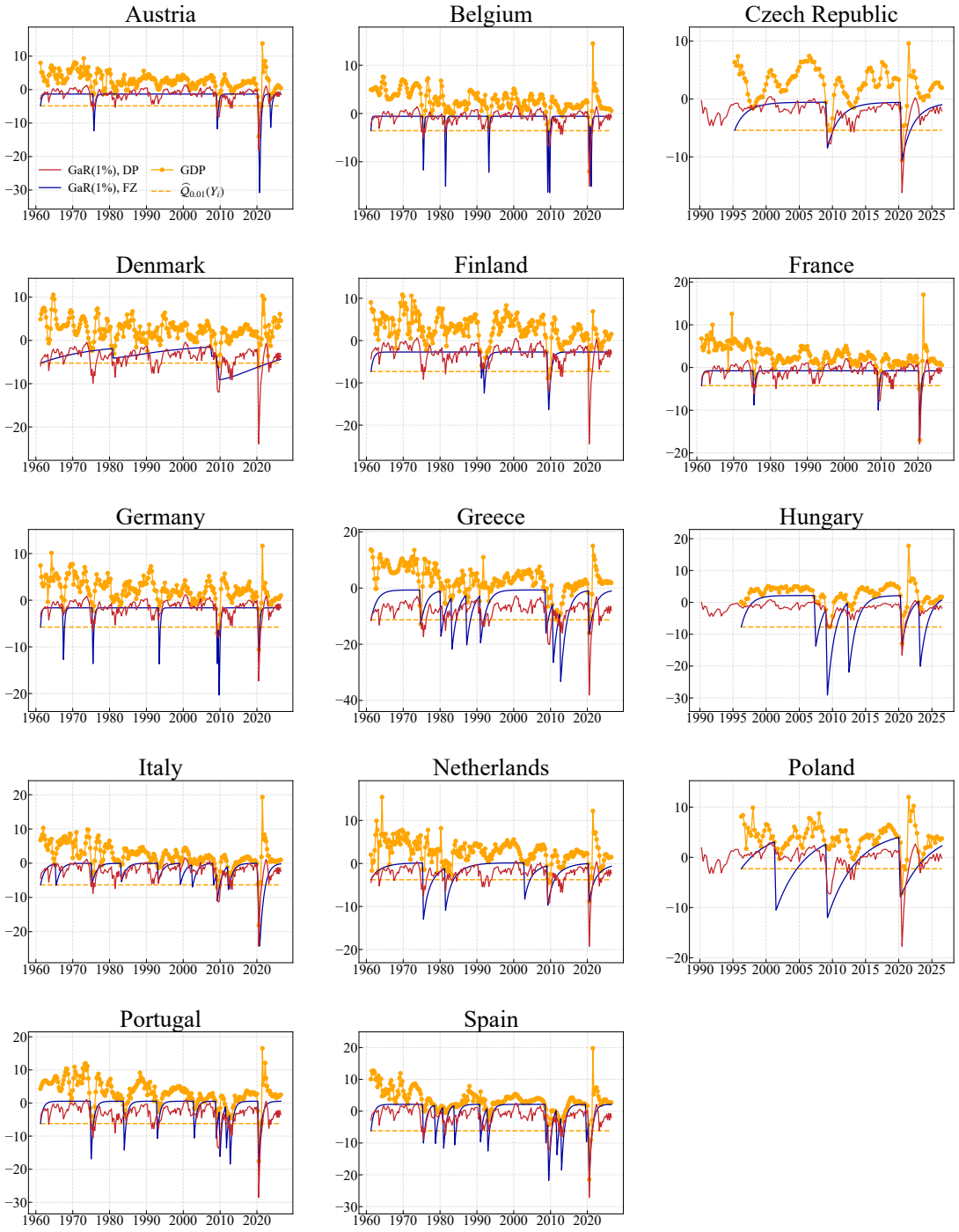


Figure 7: Estimated 1% Growth-at-Risk. The red and blue solid lines show the full-sample fitted GaR paths from DP and the FZ model (Patton et al., 2019), respectively. The orange solid line depicts realized GDP growth, while the orange dashed line denotes the empirical 1% quantile, $\widehat{\mathcal{Q}}_{0.01}(Y_i)$.

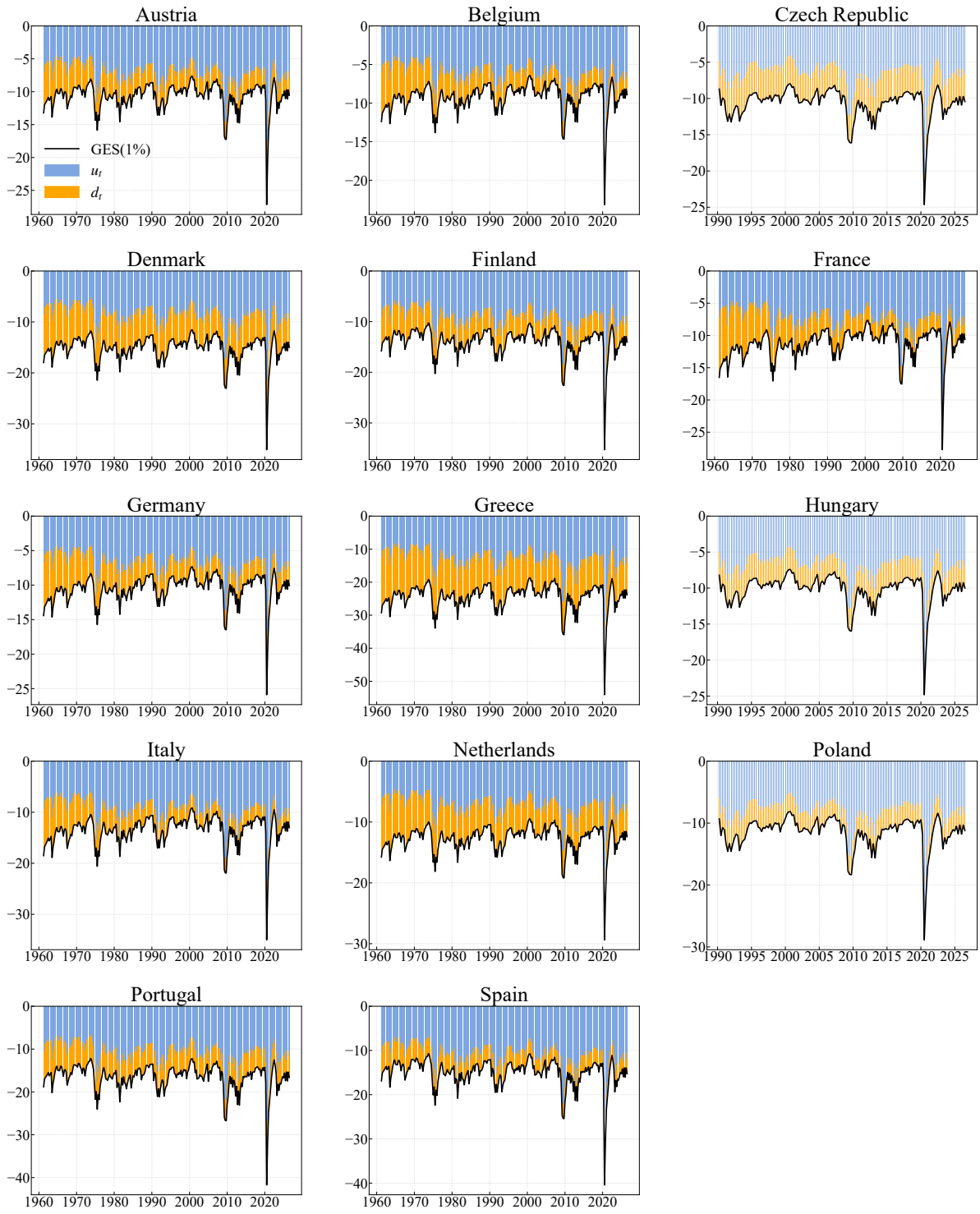


Figure 8: Estimated Growth Expected Shortfall. The black solid line is the GES(1%) or $e_{i,1,t}$, the blue area represents the portion attributable to movements in $u_{i,t}$, and the yellow area captures the remainder contribution, i.e., $e_{i,1,t} - u_{i,t}$.

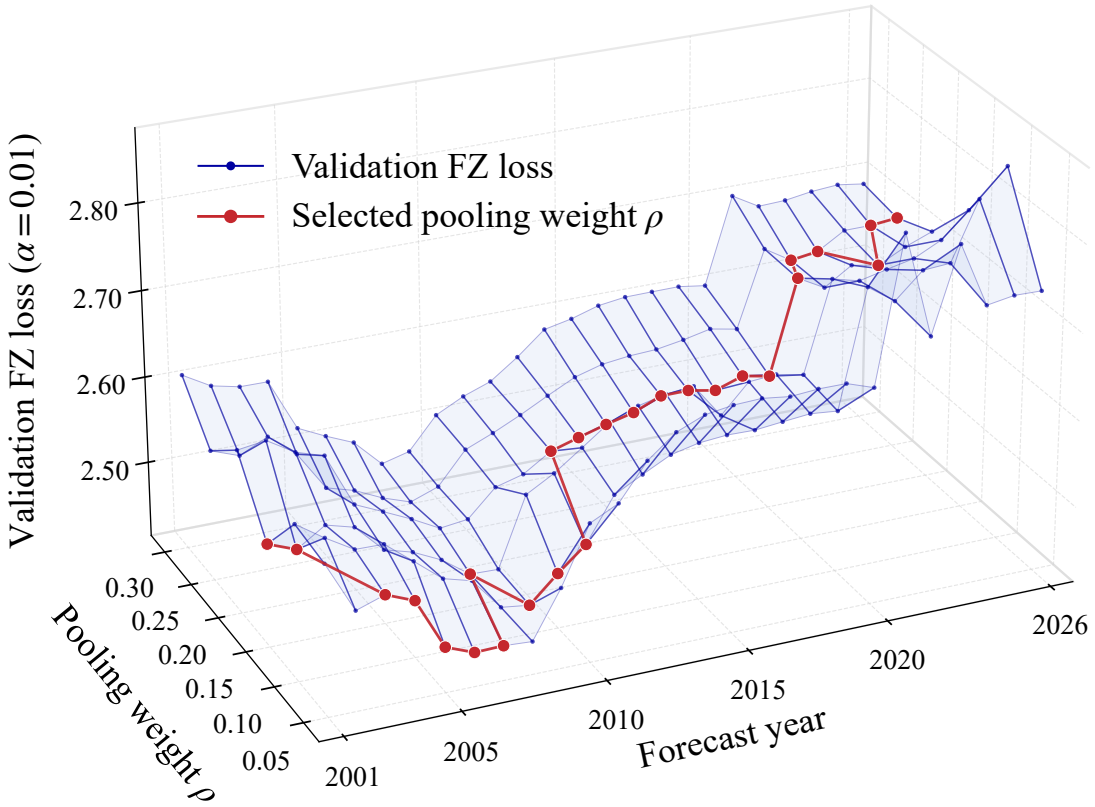


Figure 9: Validation FZ loss (1%) as a function of the training weight ρ

turns are transmitted through different combinations of location shifts and tail-shape adjustments across countries.

4.3 Out-of-sample analysis

We conclude our empirical analysis of the DP model with an out-of-sample forecasting analysis to assess the predictive performance of the DP method. We use 1961Q1 to 2000Q4 as our initial estimation sample and then re-estimate the model at the beginning of each year. For each estimated model and estimation window, we obtain one-step-ahead forecasts of $(\boldsymbol{v}_t, \boldsymbol{e}_{1,t})$ for all 14 countries over the subsequent year. For each estimation window, we also set the regularization parameter ρ in a data-driven way by splitting the estimation window into a 70/30 training–validation sample, following the same procedure as in the full-sample analysis. Forecast accuracy is evaluated using the out-of-sample unregularized FZ loss for the 1% tail area.[6] We report results both at the country level and averaged across countries.

Figure 9 plots the DP model's mean out-of-sample FZ loss at $\alpha = 0.01$ over time

[6]As a robustness check, Table C.1 in Appendix C.4 reports an evaluation comparison in terms of the check loss only rather than the full FZ loss, where check loss is defined as $L_{QR}(Y_t, v_t; \alpha) = (\alpha - \mathbb{1}\{Y_t \leq v_t\})(Y_t - v_t)$ for the 1% tail probability. This concentrates on the GaR only, rather than on (GaR, GES) jointly.

and across different values of ρ, with red markers indicating the selected ρ for each forecast year. There is a clear shift towards stronger cross-quantile pooling in the post-financial-crisis period. Specifically, ρ rises from 0.05 in 2003–2007 to 0.20 in 2012–2021, reaching 0.25 in some of the more recent years. This pattern is consistent with cross-quantile information becoming more valuable when financial stress and macroeconomic uncertainty make downside growth risk harder to assess, highlighting that information from less extreme quantiles can help to stabilize tail risk estimation in sparsely observed (1%) tail areas.

To benchmark the DP approach, we consider a range of seventeen competing models in our out-of-sample analysis. Full descriptions are provided in Appendix C.3. First, we include (i) our current (DP) double-pooling approach, along with four sub-variants that use (ii) a constant tail shape ($\text{DP}_{\gamma 0}$), (iii) a constant tail scale ($\text{DP}_{\text{c}0}$), (iv) a constant tail scale and shape ($\text{DP}_{\text{c}\gamma 0}$), and (v) a constant tail scale, tail shape, and threshold ($\text{DP}_{\text{uc}\gamma 0}$). The latter are used to investigate whether all components in the nonlinear three-factor set-up are genuinely time-varying. Second, we use three benchmark models that use less pooling, including (vi) a no-pooling benchmark (FZ) that fits the 1% GaR and GES separately for each country as in Patton et al. (2019), (vii) a benchmark that only pools across the cross-sectional dimension i (PI), and (viii) a benchmark (PQ) with full regularization (quantile pooling) but no cross-sectional pooling, fitted to each country separately. We also include three regression benchmarks, namely (ix) the quantile-regression approach (GaR) of Adrian et al. (2019) using the one quarter lagged GDP growth and Country-Level Index of Financial Stress (CLIFS) from the ECB data portal as predictors and a skew-t distribution fitted to the predicted quantiles, and (x) a panel extension (PGaR) of the previous, using a panel quantile regression with country fixed effects in the first stage, and (xi) a country-specific conditional Pareto-tail model (EGaR) adapted from Adrian et al. (2026). Third, we include three parametric and EVT-based approaches, including (xii) a Gaussian (GARCH) and (xiii) Student-t (GARCH-t) GARCH(1,1) model, fitted separately for each country, (xiv) a generalized Pareto distribution (MLE) fitted separately for each country with the threshold fixed at the recursively estimated 20th percentile. Finally, we also include three non-parametric benchmarks, namely (xv) a simple equally weighted (EW) recursive estimate of GaR and GES; (xvi) a 40 quarter or 10 year (RW-40) and (xvii) a 60 quarter or 15 year (RW-60) rolling-window analogue;

Table 2 reports the main results for GaR and GES quality in terms of 1% FZ loss, while Table C.1 in the appendix provides a robustness check for GaR only using a check-loss function. Across both criteria, the proposed DP method and its sub-variants deliver the most stable overall performance. The DP models achieve the lowest cross-country average losses, and also are the most dominant model for most of the individual countries. The

PQ and PI benchmarks that only pool across quantiles and countries, respectively, also perform well on average, more so than the FZ approach, which does not use any pooling. This again confirms the value-added of pooling, where the DP approach exploits the best of both worlds by pooling in both directions simultaneously. This is particularly clear for countries such as AUT, HUN, and POL, where DP outperforms both PQ and the GARCH-type benchmarks under both loss criteria.

Comparisons within the DP family highlight that the dynamics of the extreme tail scale $c_{i,t}$ and tail shape $\gamma_{i,t}$ may not be easily identifiable. In particular, putting these components to constant values c_i and γ_i as in the $\text{DP}_{\text{c}\gamma 0}$ model often performs similarly or better out-of-sample than a fully dynamic DP specification. It does seem important, by contrast, to allow the tail location component $u_{i,t}$ to be time-varying; typically the $\text{DP}_{\text{c}\gamma\text{u}0}$ specification performs substantially worse than both the DP and $\text{DP}_{\text{c}\gamma 0}$ specifications.

One might mistake the importance of the time-varying threshold $u_{i,t}$ for a standard time-varying volatility effect. However, though standard GARCH-type benchmarks occasionally perform well for specific countries, these gains are not systematic across countries or loss criteria, illustrating that the downside tail dynamic threshold $u_{i,t}$ picks up more than a common volatility effect. We also note that the no-pooling FZ benchmark sometimes performs poorly: although it performs competitively in a few cases such as France (FRA), it generates relatively big losses in countries such as Greece (GRC) and Poland (POL). The evidence is consistent with the benefits from combining cross-quantile and cross-country information when forecasting dynamic macroeconomic tail risk.

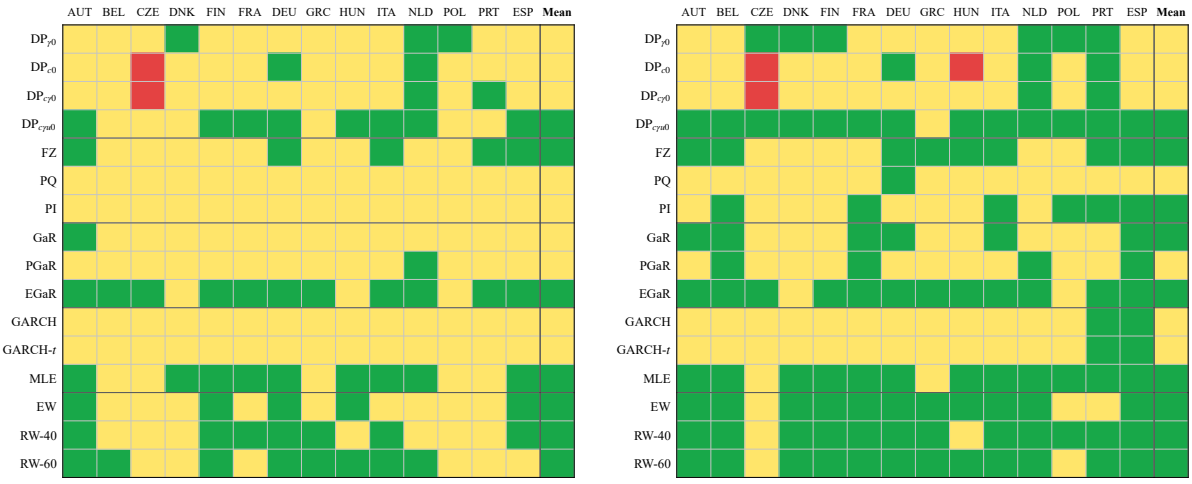
The results for the quantile regression benchmarks are also informative. PGaR improves upon GaR in terms of FZ loss, again suggesting that pooling regression coefficients across countries can improve forecast accuracy, even without introducing additional information, such as the financial-side information considered in the GaR framework of Adrian et al. (2019) and related studies. Also EGaR achieves a lower average FZ loss than GaR, but its average check loss exceeds those of both GaR and PGaR. DP records lower average losses than all three regression-based benchmarks under both criteria.

Figure 10 complements the results from Table 2 with country-specific Diebold–Mariano (DM) tests of the DP model against its competitors using the traffic light approach of Nolde and Ziegel (2017): green indicates the DM exceeds its critical value in favor of the DP model, red in favor of the competing model, and yellow if the result is insignificant. Significant differences predominantly favor the DPapproach. The only two significant losses are found for CZE against DP subversions, like with constant tail scale (DP_{c0}) or constant tail scale and shape ($DP_{c\gamma 0}$). At the 10% level, DP achieves lower aggregate FZ losses than seven competing specifications. At the 20% level, this number rises to nine out of sixteen, while significant gains also become more widespread across countries,

Table 2: Out-of-Sample Extreme FZ Loss ($\alpha = 0.01$)

Models	AUT	BEL	CZE	DNK	FIN	FRA	DEU	GRC	HUN	ITA	NLD	POL	PRT	ESP	Mean
DP	3.019	***2.864***	2.952	3.091	3.061	3.137	***2.972***	3.372	3.120	**3.103**	**2.937**	***2.821***	**2.964**	***3.015***	***3.031***
$DP_{\gamma 0}$	3.063	2.932	2.958	3.148	3.190	***3.108***	3.067	3.306	2.985	3.150	2.959	2.850	***2.980***	**3.001**	3.050
DP_{c0}	***3.007***	2.892	**2.912**	3.108	3.113	3.213	3.015	3.491	**2.907**	3.210	2.974	**2.809**	2.984	3.172	3.058
$DP_{c\gamma 0}$	**2.918**	**2.850**	***2.931***	3.092	**3.005**	**3.067**	**2.900**	3.440	***2.965***	***3.123***	2.956	2.835	2.993	3.061	**3.010**
$DP_{c\gamma u0}$	4.168	3.521	3.429	3.504	3.505	4.193	3.963	3.422	3.527	4.210	3.438	3.792	3.515	4.694	3.777
FZ	3.923	3.604	3.363	3.401	3.241	3.204	3.530	4.186	3.485	3.706	3.371	3.904	3.479	3.741	3.581
PQ	3.227	3.261	3.488	3.183	3.254	3.464	3.347	***3.302***	3.775	3.521	3.126	3.387	3.231	3.865	3.388
PI	3.453	3.231	3.321	3.131	***3.043***	3.961	3.184	3.347	3.427	3.937	3.044	3.016	3.309	3.811	3.372
GaR	3.827	3.631	3.338	***3.064***	3.125	3.673	4.018	3.417	6.644	4.286	***2.947***	3.602	6.702	5.518	4.128
PGaR	3.288	3.866	3.311	**3.021**	3.147	4.236	3.431	**3.266**	3.818	3.666	2.995	3.426	3.689	3.913	3.505
EGaR	4.441	3.633	3.501	3.150	3.469	3.557	3.199	4.547	4.203	3.523	3.042	3.897	4.070	3.565	3.700
GARCH	3.267	3.239	3.326	3.284	3.079	3.260	3.034	3.505	4.359	3.220	3.530	3.606	3.622	3.424	3.411
GARCH-t	3.244	3.261	3.327	3.280	3.077	3.259	3.033	3.508	4.336	3.216	3.283	3.573	3.621	3.420	3.388
MLE	4.213	3.595	3.442	3.566	3.564	4.282	4.034	3.629	3.576	4.302	3.517	3.987	3.732	4.951	3.885
EW	3.961	3.499	3.413	3.456	3.426	4.009	3.851	3.674	3.506	4.088	3.440	3.684	3.692	4.580	3.734
RW-40	4.122	3.939	3.893	3.677	4.003	4.270	4.063	4.094	4.183	4.277	3.754	3.863	3.860	4.552	4.039
RW-60	3.848	3.711	3.440	3.412	3.930	3.957	3.672	4.085	3.528	4.070	3.568	3.604	3.869	4.357	3.789

Notes: This table reports average out-of-sample FZ loss at $\alpha = 0.01$. The evaluation sample begins in 2001Q1 for eleven countries, 2010Q1 for CZE, and 2011Q1 for HUN and POL, and ends in 2026Q2. For each column, the smallest loss is in **bold**, and the second smallest is in ***bold italic***.



(a) 10% significance level

(b) 20% significance level

Figure 10: Diebold–Mariano comparisons of out-of-sample FZ losses at $\alpha = 0.01$. Each cell compares DP with the model in the corresponding row. Green indicates significantly lower loss for DP, red significantly higher loss, and yellow no significant difference. The last column in each panel uses the mean FZ loss across countries.

covering a broad range of alternative forecasting models. Also note that power of the tests may be limited given the small tail area (1%) on which our GaR and GES estimates concentrate.

Taken together, the results show that the forecasting gains of DP are not confined to a single evaluation metric or a single country. Overall, the DP performs favorably by combining cross-country with cross-quantile information for tail regularization, thus yielding better forecasts of dynamic macroeconomic tail risk when extreme observations are scarce and individual time series information is limited.

5 Conclusion

This paper introduced the Double Pooling (DP) approach for jointly estimating time-varying tail risk dynamics in typical macroeconomic panel data. The method combines two complementary sources of information by (i) pooling across cross-sectional units and (ii) regularizing extreme tail dynamics by pooling information from less extreme quantile levels. A key ingredient of the approach is a semiparametric three-component tail representation, building on familiar results from extreme value theory, that expresses VaR and the associated extreme ES through a threshold component, a tail scale component, and a tail-shape component. This representation provides an interpretable mapping of tail risk measures, preserves the natural ordering between Growth-at-Risk (GaR) and Growth Expected Shortfall (GES), and easily lends itself to a dynamic factor set-up to study the joint dynamics of economic tail risk across different economies. By combining this tail

representation with a penalized, yet still proper loss function using score-driven dynamics, the framework directly estimates multiple GaRs and the associated extreme ES in one step, while accommodating sparse tail observations, tail dynamics, and missing data. The less extreme GaR levels are used in this approach to regularize the most extreme one, where the amount of regularization can be set in a data-driven way. Simulations highlight the importance of such regularization.

In an empirical application to 14 European economies, we illustrated the relevance of the double-pooling mechanism for relatively short samples of quarterly observations. The estimates reveal persistent common European tail dynamics. The decomposition of extreme GES shows that major crisis episodes, including the global financial crisis, the euro area sovereign debt crisis, and the COVID-19 crisis, differ not only in severity but also in mechanism. Whereas the financial and COVID-19 crises are mainly location-driven through shifts in the threshold component for all countries simultaneously, the sovereign debt crisis appears mainly tail-shape-driven, with increasing tail scale and thicker tails for some countries, but not for others. Overall, by combining the two sources of pooling, the DP model can produce tail risk dynamics that are more timely and generally more stable than those obtained using one type of pooling only. The out-of-sample analysis further supported the value of the DP approach: its forecasts give the lowest average (proper scoring rule) losses across countries compared to a range of competing methods, as well as stable, solid performance across all countries in the empirical study.

Future work could extend the framework to richer forms of cross-sectional dependence, incorporating forward-looking predictors into the score-driven recursion, or applying the method to other macro-financial variables, such as inflation, asset returns, or financial conditions. Taken together, all results point to the double pooling approach offering a tractable and robust way to estimate dynamic extreme tail risks when direct tail information is scarce.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–1289. 2, 4, 5, 27, 28, 45

Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics 14*(3), 283–323. 2, 4, 5, 45

Adrian, T., Y. Sasaki, and Y. Wang (2026). Robust econometrics for growth-at-risk. *Journal of Econometrics 255*, 106235. 2, 3, 4, 27, 45

Alter, A. and E. M. Mahoney (2021). Local house-price vulnerability: Evidence from the US and Canada. *Journal of Housing Economics 54*, 101791. 2

Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *Annals of Probability*, 792–804. 35

Beutner, E. A., Y. Lin, and A. Lucas (2026). Consistency, distributional convergence, and optimality of time-varying parameters in score-driven models. *Journal of Econometrics 255*, 106218. 8, 12

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika 102*(2), 325–343. 12

Brownlees, C. and A. B. M. Souza (2021). Backtesting global growth-at-risk. *Journal of Monetary Economics 118*, 312–330. 2, 4

Brownlees, C. T., C. Pavanello, and A. B. M. Souza (2026). Which factors drive downside risk in the U.S. economy? *SSRN Paper 6548438*. 2, 5

Cai, J.-J., J. H. Einmahl, L. Haan, and C. Zhou (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society Series B: Statistical Methodology 77*(2), 417–442. 10

Catania, L. and A. Luati (2023). Semiparametric modeling of multiple quantiles. *Journal of Econometrics 237*(2), 105365. 3, 6, 10, 12

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910. 6

Chetverikov, D., Y. Liu, and A. Tsyvinski (2025). Weighted-average quantile regression. *Journal of Econometrics 252*, 106115. 2

Corradi, V. and J. Llorens-Terrazas (2026). Monitoring joint tail risks: An application to growth and inflation. *Journal of Econometrics 255*, 106237. 2

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795. 4, 6, 8, 11, 12

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2024). Observation-driven filtering of time-varying parameters using moment conditions. *Journal of Econometrics 238*(2), 105635. 12

Daouia, A., S. Girard, and G. Stupfler (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(2), 263–292. 10

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York: Springer. 4, 9

De Punder, R., T. Dimitriadis, and R.-J. Lange (2024). Kullback-Leibler-based characterizations of score-driven updates. *arXiv preprint arXiv:2408.02391*. 12

Delle Monache, D., A. De Polis, and I. Petrella (2020). Modeling and forecasting macroeconomic downside risk. *Journal of Business and Economic Statistics 40*(1), 542–554. 4

Dew-Becker, I. (2023). Tail risk in production networks. *Econometrica 91*(6), 2089–2123. 3

D'Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang (2024). Modeling extreme events: Time-varying extreme tail shape. *Journal of Business & Economic Statistics 42*(3), 903–917. 3, 6, 15

D'Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang (2026). Joint extreme value-at-risk and expected shortfall dynamics with a single integrated tail shape parameter. *Journal of Business & Economic Statistics*, to appear. 3, 6, 8, 15

Fan, N., C. Zhang, and Z. Zhang (2025). Dynamic modeling via autoregressive conditional GB2 for cross-sectional maxima of financial time series data. *Journal of Business & Economic Statistics 43*(4), 968–978. 15

Fissler, T. (2017). *On higher order elicitability and some limit theorems on the Poisson and Wiener space*. Ph. D. thesis, Philosophisch-naturwissenschaftliche Fakultät der Universität Bern. 8

Fissler, T. and Y. Hoga (2024). Backtesting systemic risk forecasts using multi-objective elicitability. *Journal of Business & Economic Statistics 42*(2), 485–498. 5

Fissler, T. and J. F. Ziegel (2016). Higher order elicitability and Osband's principle. *Annals of statistics 44*(4), 1680–1707. 2, 4, 7, 12

Geluk, J., L. de Haan, S. Resnick, and C. Stărică (1997). Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications 69*(2), 139–159. 9

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 44*(1), 37–42. 9

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Volume 52. Cambridge University Press. 4, 6, 8, 11, 12

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020). Firm volatility in granular networks. *Journal of Political Economy 128*(11), 4097–4162. 3

Hoga, Y. (2025). The estimation risk in extreme systemic risk forecasts. *Econometric Theory 41*(2), 341–390. 10

Huo, Z., A. A. Levchenko, and N. Pandalai-Nayar (2025). International comovement in the global production network. *Review of Economic Studies 92*(1), 365–403. 3

Lang, J. H., M. Rusnák, and M. Greiwe (2026). Medium-term growth-at-risk in the Euro area. *IMF Economic Review 74*(2), 382–424. 4

Lin, Y. and A. Lucas (2027). Testing for univariate score-driven parameter dynamics. *Econometric Theory*. 20

Liu, S. and Y.-K. Tse (2015). Intraday value-at-risk: An asymmetric autoregressive conditional duration approach. *Journal of Econometrics 189*(2), 437–454. 2

Livieri, G. and G. Palmari (2026). Proper-score observation-driven filters: Local geometry, estimation, and continuous-time limits. *arXiv preprint arXiv:2608.02828*. 8, 12

Lopez-Salido, D. and F. Loria (2024). Inflation at risk. *Journal of Monetary Economics 145*, 103570. 2

Loria, F., C. Matthes, and D. Zhang (2025). Assessing macroeconomic tail risk. *Economic Journal 135*(665), 264–284. 2, 5

Marfè, R. and J. Pénasse (2024). Measuring macroeconomic tail risk. *Journal of Financial Economics 156*, 103838. 2, 5

Müller, U. K. and M. W. Watson (2024). Time varying extremes. *Review of Economics and Statistics*, 1–37. 6, 11, 15

Nolde, N. and J. F. Ziegel (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics 11*(4), 1833–1874. 2, 28

Opschoor, A. and A. Lucas (2019). Fractional integration and fat tails for realized covariance kernels. *Journal of Financial Econometrics 17*(1), 66–90. 14

Patton, A. J., J. F. Ziegel, and R. Chen (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics 211*(2), 388–413. 2, 3, 4, 5, 7, 8, 11, 12, 15, 22, 23, 24, 27, 46

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 119–131. 35

A Some Derivations

A.1 Derivation of the three-component tail map

This appendix provides the motivation for the three-component tail map used in the main text. Let $F_t(y) = \mathbb{P}(Y_t \le y \mid \mathcal{F}_{t-1})$ denote the conditional distribution of Y_t. For a left-tail threshold u with $F_t(u) > 0$, define the conditional lower-tail excess distribution

$$G_{t,u}(x) \equiv \mathbb{P}(u - Y_t \le x \mid Y_t \le u, \mathcal{F}_{t-1}) = 1 - \frac{F_t(u-x)}{F_t(u)}, \qquad x \ge 0.$$

The Pickands–Balkema–de Haan peaks-over-threshold result (Balkema and de Haan, 1974; Pickands III, 1975) implies that, under standard max-domain-of-attraction conditions applied to the lower tail, $G_{t,u}$ converges to a generalized Pareto form as u moves sufficiently far into the left tail. In the paper, we use the corresponding limit form as a semiparametric representation of lower-tail excesses.

Let u_t satisfy $\bar{\alpha} = \mathbb{P}(Y_t \le u_t \mid \mathcal{F}_{t-1})$ and define $X_t = u_t - Y_t$ conditional on $Y_t \le u_t$. The generalized Pareto representation is

$$G_t(x) \equiv \mathbb{P}(X_t \le x \mid Y_t \le u_t, \mathcal{F}_{t-1}) = 1 - \left(1 + \gamma_t \frac{x}{c_t}\right)^{-1/\gamma_t}, \qquad x \ge 0, \tag{A.1}$$

where $c_t > 0$ and $0 < \gamma_t < 1$. For any $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, let $v_t(\alpha)$ be the conditional VaR and define $q_t(\alpha) = u_t - v_t(\alpha)$. Then

$$\alpha = \mathbb{P}(Y_t \le v_t(\alpha) \mid \mathcal{F}_{t-1}) = \bar{\alpha}\, \mathbb{P}(X_t \ge q_t(\alpha) \mid Y_t \le u_t, \mathcal{F}_{t-1}).$$

Using (A.1), this becomes

$$\alpha = \bar{\alpha} \left(1 + \gamma_t \frac{q_t(\alpha)}{c_t}\right)^{-1/\gamma_t}.$$

Solving for $q_t(\alpha)$ gives

$$q_t(\alpha) = \frac{c_t}{\gamma_t} \left[\left(\frac{\alpha}{\bar{\alpha}}\right)^{-\gamma_t} - 1 \right].$$

This is the first part of the tail map in (5).

The ES representation follows from the mean excess formula of the generalized Pareto distribution. For $q_t = q_t(\alpha)$,

$$\mathbb{E}[X_t \mid X_t \ge q_t, Y_t \le u_t, \mathcal{F}_{t-1}] = \frac{q_t + c_t}{1 - \gamma_t}, \qquad 0 < \gamma_t < 1.$$

Since

$$e_t(\alpha) = \mathbb{E}[Y_t \mid Y_t \le v_t(\alpha), \mathcal{F}_{t-1}] = u_t - \mathbb{E}[X_t \mid X_t \ge q_t, Y_t \le u_t, \mathcal{F}_{t-1}],$$

we obtain

$$d_t(\alpha) \equiv u_t - e_t(\alpha) = \frac{q_t(\alpha) + c_t}{1 - \gamma_t}.$$

Combining $v_t(\alpha) = u_t - q_t(\alpha)$ and $e_t(\alpha) = u_t - d_t(\alpha)$ yields the closed-form VaR and ES expressions.

A.2 Derivation of $\boldsymbol{H}_t$

We define the diagonal curvature terms by differentiating the conditional expected loss, evaluated at the conditional VaRs and ES:

$$H_{i,t}^k := \frac{\partial^2}{\partial (f_t^k)^2} \mathbb{E}[L_{DP,i}(Y_{i,t}; \boldsymbol{v}_{i,t}, e_{i,1,t}, \boldsymbol{\alpha}) \,|\, \mathcal{F}_{t-1}], \qquad k \in \{u, c, \gamma\}.$$

Fix i and t and suppress these subscripts for notation simplicity. Write $F(y) = \mathbb{P}(Y \leq y \mid \mathcal{F}_{t-1})$, $p(y) = F'(y)$, and $\mu(v) = \mathbb{E}[Y \mathbb{1}\{Y \leq v\} \mid \mathcal{F}_{t-1}]$. The individual DP loss is

$$\begin{aligned} L_{DP}(Y; \boldsymbol{v}, e_1, \boldsymbol{\alpha}) &= (1-\rho)\alpha_1 L_{FZ}(Y, v_1, e_1; \alpha_1) + \frac{\rho}{J} \sum_{j=2}^{J} L_{Q,j}(Y, v_j) \\ &= (1-\rho) \left[-\mathbb{1}\{Y \leq v_1\} \frac{v_1 - Y}{e_1} + \alpha_1 \left(\frac{v_1}{e_1} + \log(-e_1) - 1 \right) \right] \\ &\quad + \frac{\rho}{J} \sum_{j=2}^{J} \left[\alpha_j \log(-v_j) + \mathbb{1}\{Y \leq v_j\} \log\left(\frac{Y}{v_j} \right) \right]. \end{aligned}$$

For the FZ component, direct differentiation gives

$$\begin{aligned} \frac{\partial}{\partial v} \mathbb{E}[L_{FZ}(Y, v, e; \alpha_1) \,|\, \mathcal{F}_{t-1}] &= \frac{\alpha_1 - F(v)}{\alpha_1 e}, \\ \frac{\partial}{\partial e} \mathbb{E}[L_{FZ}(Y, v, e; \alpha_1) \,|\, \mathcal{F}_{t-1}] &= -\frac{v\{\alpha_1 - F(v)\} + \mu(v) - \alpha_1 e}{\alpha_1 e^2}. \end{aligned}$$

At the conditional VaR and ES, $F(v_1) = \alpha_1$ and $\mu(v_1) = \alpha_1 e_1$, so both derivatives vanish. Since $\mu'(v) = vp(v)$, differentiating once more yields

$$\frac{\partial^2}{\partial (v,e) \partial (v,e)'} \mathbb{E}[L_{FZ}(Y, v, e; \alpha_1) \,|\, \mathcal{F}_{t-1}] \Big|_{(v,e)=(v_1,e_1)} = \begin{pmatrix} -\dfrac{p(v_1)}{\alpha_1 e_1} & 0 \\ 0 & \dfrac{1}{e_1^2} \end{pmatrix}.$$

The off-diagonal element vanishes because it is proportional to $\alpha_1 - F(v_1)$. Moreover, $-p(v_1)/(\alpha_1 e_1) > 0$ since $p(v_1) > 0$ and $e_1 < 0$.

For $j = 2, \ldots, J$, let $L_{Q,j}(Y, v_j) = \alpha_j \log(-v_j) + \mathbb{1}\{Y \leq v_j\} \log(Y/v_j)$, where the indicator-weighted logarithm is evaluated only when $Y \leq v_j$. Direct differentiation of its

conditional expectation gives

$$\frac{\partial}{\partial v_j}\mathbb{E}[L_{Q,j}(Y,v_j)\,|\,\mathcal{F}_{t-1}] = \frac{\alpha_j - F(v_j)}{v_j}, \qquad \frac{\partial^2}{\partial v_j^2}\mathbb{E}[L_{Q,j}(Y,v_j)\,|\,\mathcal{F}_{t-1}]\bigg|_{F(v_j)=\alpha_j} = -\frac{p(v_j)}{v_j}.$$

Write $p_j = p(v_j)$ and $\kappa_1 = -p_1/(\alpha_1 e_1)$. Define $r_j = \alpha_j/\alpha_J$, $q_j = u - v_j$, $m_j = q_j + (\gamma q_j + c)\log r_j$, and $d = u - e_1$. The transformations in (6) imply $\partial u/\partial f^u = u$, $\partial c/\partial f^c = c$, and $\partial\gamma/\partial f^\gamma = \gamma(1-\gamma)$. Combining these derivatives with the three-component tail mapping gives

$$\boldsymbol{a}_j \coloneqq \frac{\partial v_j}{\partial \boldsymbol{f}} = \begin{pmatrix} u \\ -q_j \\ (1-\gamma)m_j \end{pmatrix}, \qquad \boldsymbol{b} \coloneqq \frac{\partial e_1}{\partial \boldsymbol{f}} = \begin{pmatrix} u \\ -d \\ m_1 - \gamma d \end{pmatrix}.$$

In particular, $\partial e_1/\partial f^c = e_1 - u = -d$ and $\partial e_1/\partial f^\gamma = m_1 - \gamma d$. Let a_j^k and b^k denote the corresponding components of $\boldsymbol{a}_j$ and $\boldsymbol{b}$, for $k \in \{u, c, \gamma\}$.

At the conditional VaRs and ES, the first derivatives and mixed second derivatives of the conditional expected loss with respect to the forecasts vanish. The terms involving second derivatives of the tail mapping therefore drop out of the coordinatewise chain rule, yielding

$$H_{i,t}^k = (1-\rho)\alpha_1\left[\kappa_{i,1,t}(a_{i,1,t}^k)^2 + \frac{(b_{i,t}^k)^2}{e_{i,1,t}^2}\right] + \frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}}{-v_{i,j,t}}(a_{i,j,t}^k)^2. \tag{A.2}$$

Under the GPD assumption, the conditional cdf and density below $u_{i,t} = v_{i,J,t}$ are

$$\begin{aligned} F_{i,t}(y) &= \alpha_J\left(1 + \gamma_{i,t}\frac{u_{i,t}-y}{c_{i,t}}\right)^{-1/\gamma_{i,t}}, \\ p_{i,t}(y) &= \frac{\alpha_J}{c_{i,t}}\left(1 + \gamma_{i,t}\frac{u_{i,t}-y}{c_{i,t}}\right)^{-1/\gamma_{i,t}-1}, \qquad y \le u_{i,t}. \end{aligned}$$

Since $1 + \gamma_{i,t}q_{i,j,t}/c_{i,t} = (\alpha_j/\alpha_J)^{-\gamma_{i,t}}$, evaluation at $y = v_{i,j,t}$ gives

$$p_{i,j,t} = \frac{\alpha_j}{c_{i,t}}\left(\frac{\alpha_j}{\alpha_J}\right)^{\gamma_{i,t}}, \qquad \kappa_{i,1,t} = -\frac{1}{c_{i,t}e_{i,1,t}}\left(\frac{\alpha_1}{\alpha_J}\right)^{\gamma_{i,t}}.$$

Both quantities are strictly positive under $c_{i,t} > 0$ and $e_{i,1,t} < 0$. For $j = J$, the density expression uses differentiability of the conditional cdf at the threshold; absent smooth density matching, the GPD expression gives only the left-hand density at $u_{i,t}$.

Define $z_{i,j,t} = (1-\gamma_{i,t})m_{i,j,t}$ and $h_{i,t} = m_{i,1,t} - \gamma_{i,t}d_{i,t}$. The diagonal elements used for

score scaling are

$$
\begin{aligned}
H^{u}_{i,t} &= u^{2}_{i,t}\left[(1-\rho)\alpha_1\left(\kappa_{i,1,t}+\frac{1}{e^{2}_{i,1,t}}\right)+\frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}}{-v_{i,j,t}}\right],\\
H^{c}_{i,t} &= (1-\rho)\alpha_1\left(\kappa_{i,1,t}q^{2}_{i,1,t}+\frac{d^{2}_{i,t}}{e^{2}_{i,1,t}}\right)+\frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}q^{2}_{i,j,t}}{-v_{i,j,t}},\\
H^{\gamma}_{i,t} &= (1-\rho)\alpha_1\left(\kappa_{i,1,t}z^{2}_{i,1,t}+\frac{h^{2}_{i,t}}{e^{2}_{i,1,t}}\right)+\frac{\rho}{J}\sum_{j=2}^{J}\frac{p_{i,j,t}z^{2}_{i,j,t}}{-v_{i,j,t}}.
\end{aligned}
$$

All three diagonal elements are strictly positive under $J \geq 2$, $0 < \alpha_1 < \alpha_J$, $0 < \rho < 1$, $u_{i,t} < 0$, $c_{i,t} > 0$, $e_{i,1,t} < 0$, and $0 < \gamma_{i,t} < 1$.

Finally, additivity of the conditional expected DP loss gives

$$
H^{k}_{t} = \sum_{i \in I_t} H^{k}_{i,t}, \qquad s^{k}_{t} = (H^{k}_{t})^{-1/2}\nabla^{k}_{t}, \qquad k \in \{u, c, \gamma\}.
$$

B Additional simulation results under correct model specification

In this experiment, we assume the DP model is correctly specified and evaluate its in-sample parameter and tail-risk recovery. We consider $N = 14$ and $T \in \{50, 100, 200\}$, focusing on short time series of a length relevant to the empirical application. We are interested in the 1% GaR and GES and use the tail region $\boldsymbol{\alpha} = \{0.01, 0.02, \ldots, 0.20\}$ for regularization with a pooling weight ρ fixed at $\rho = 0.5$ in both the data generating process and in model estimation. With the components ordered as (u, c, γ), the dynamic coefficient matrices are $\boldsymbol{A} = \mathrm{diag}(0.20, 0.16, 0.07)$ and $\boldsymbol{B} = \mathrm{diag}(0.95, 0.90, 0.90)$. The common states follow score-driven recursion in (8)–(9) and are initialized at zero. For each individual, we set $\bar{f}_i^u = 0$, $\bar{f}_i^c = \log(0.5)$, and $\bar{f}_i^\gamma = \log(0.1/0.9)$, so that the initial threshold, scale, and shape are -1, 0.5, and 0.1, respectively. In particular, $u_{it} = -\exp(\bar{f}_i^u + f_t^u)$, $c_{it} = \exp(\bar{f}_i^c + f_t^c)$, and $\gamma_{it} = \min\{\Lambda(\bar{f}_i^\gamma + f_t^\gamma), 0.99\}$, where $\Lambda(x) = (1 + \exp(-x))^{-1}$.

At each date, observations are generated from a spliced conditional distribution: with probability $p = 0.20$, the observation Y_{it} falls in the left tail and $u_{it} - Y_{it}$ follows a generalized Pareto distribution $\mathrm{GPD}(c_{it}, \gamma_{it})$ with zero location. With probability $1 - p$, it lies above the threshold and follows a scaled half Student-t distribution with degrees of freedom $\nu = 8$. Specifically,

$$Y_{it} = \begin{cases} u_{it} - Q_{\mathrm{GPD}}(U_{it}; c_{it}, \gamma_{it}), & \text{with probability } p, \\ u_{it} + s_{it} Q_{t_8}((U_{it} + 1)/2), & \text{with probability } 1 - p, \end{cases} \qquad s_{it} = \frac{2(1-p) f_{t_8}(0) c_{it}}{p},$$

where $U_{it} \sim \mathrm{Unif}(0, 1)$ and the draws are independent across individuals conditional on the common time-varying parameters. Moreover, $Q_{\mathrm{GPD}}(\,\cdot\,; c_{it}, \gamma_{it})$ and $Q_{t_8}(\cdot)$ denote the quantile functions of $\mathrm{GPD}(c_{it}, \gamma_{it})$ and t_8 distributions, respectively, while $f_{t_8}(\cdot)$ denotes the density of the t_8 distribution. The scale s_{it} is chosen such that the left and right-hand conditional densities match at the threshold. This construction preserves the model's three-component left-tail representation while allowing a different distribution for observations above the threshold.

Table B.1 reports the results across 500 Monte Carlo replications with $N = 14$. Panel A shows more accurate recovery of the common dynamics as the time dimension increases. Increasing T from 50 to 200 lowers the bias and RMSE of all six common dynamic coefficients. The gains are particularly pronounced for the threshold dynamics, where the RMSE of a_u falls from 0.110 to 0.042, and that of b_u from 0.187 to 0.064, reductions of approximately 62% and 66%, respectively. Panel B also shows declining RMSEs for all three individual-level parameters. For comparison, the RMSE of the individual threshold

Table B.1: Parameter and tail-risk recovery under correct specification

		$T = 50$			$T = 100$			$T = 200$		
Panel A. Common dynamic parameters										
Par	True	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
a_u	0.200	0.153	-0.047	0.110	0.196	-0.004	0.070	0.208	0.008	**0.042**
a_c	0.160	0.068	-0.092	0.106	0.097	-0.063	0.085	0.123	-0.037	**0.070**
a_γ	0.070	0.029	-0.041	0.046	0.037	-0.033	0.037	0.047	-0.023	**0.030**
b_u	0.950	0.808	-0.142	0.187	0.860	-0.090	0.127	0.904	-0.046	**0.064**
b_c	0.900	0.787	-0.113	0.171	0.809	-0.091	0.149	0.833	-0.067	**0.140**
b_γ	0.900	0.795	-0.105	0.167	0.788	-0.112	0.169	0.803	-0.097	**0.162**
Panel B. Individual-level parameters										
$u_{i,0}$	-1.000	-1.019	-0.019	0.230	-1.010	-0.010	0.178	-1.001	-0.001	**0.130**
$c_{i,0}$	0.500	0.464	-0.036	0.207	0.490	-0.010	0.147	0.500	0.000	**0.111**
$\gamma_{i,0}$	0.100	0.088	-0.012	0.156	0.086	-0.014	0.124	0.084	-0.016	**0.098**
Panel C. In-sample tail-risk paths										
Risk measure		RMSE		Corr.	RMSE		Corr.	RMSE		Corr.
GaR (1%)		0.781		0.876	0.545		0.936	**0.411**		**0.962**
GaR (2%)		0.556		0.885	0.389		0.944	**0.293**		**0.969**
GaR (3%)		0.451		0.891	0.317		0.949	**0.238**		**0.973**
GaR (5%)		0.339		0.897	0.241		0.955	**0.178**		**0.978**
GaR (10%)		0.226		0.904	0.162		0.962	**0.115**		**0.984**
GaR (15%)		0.191		0.907	0.135		0.963	**0.094**		**0.985**
GaR (20%)		0.186		0.917	0.130		0.963	**0.090**		**0.984**
GES (1%)		1.595		0.862	1.035		0.925	**0.743**		**0.953**

Notes. Results summarize 500 Monte Carlo replications with $N = 14$ and $\rho = 0.5$. Panels A and B report the mean estimate, bias, and RMSE for the common dynamic parameters and individual-level parameters, respectively. Panel C reports the RMSE and correlation for the tail-risk paths. RMSE is the square root of the mean squared error across replications and, where applicable, individuals and dates. Corr. is the within-individual time-series correlation between fitted and true paths, averaged across individuals and replications. For each parameter or risk measure, boldface identifies the lowest RMSE and, where applicable, the highest correlation across T.

level falls from 0.230 to 0.130, a reduction of approximately 43%. The larger proportional gains for the common threshold coefficients are consistent with the potential benefits of pooling information across individuals to estimate common dynamics. Finally, Panel C shows that the improvements extend across all reported quantile levels, from 1% to 20%, as well as the extreme GES. RMSEs decline and temporal correlations increase with T for every reported risk measure, indicating more accurate recovery of both the levels and the time variation of tail risk as T increases.

C Additional empirical details

The data in the application are primarily obtained from the OECD Quarterly National Accounts, except for France in 1961Q1–1980Q4, Denmark in 1992Q1–1995Q4, and Austria and the Czech Republic in 1996. For these periods, we use the corresponding historical series from INSEE, Statistics Denmark, and Eurostat, respectively.

The remainder of this appendix describes the bootstrap procedure in Appendix C.1, and provides additional figures and tables for tail risk dynamics during the global financial and the COVID-19 crises (Appendix C.2) and for the out-of-sample analysis (Appendix C.4). A list of the models used for the out-of-sample analysis is provided in Appendix C.3.

C.1 A circular bootstrap procedure

We assess estimation uncertainty using the following (recursive-design) circular block bootstrap procedure. We hold the regularization parameter ρ and the country-specific shifts fixed (as defined below Eq. (10)). Let $T_{\text{sub}} = 122$ denote the number of consecutive quarters for which observations are available for all countries in our dataset, and let $\mathbb{T}_{\text{sub}}$ denote the resulting set of time indices.

Step 1. Obtain the estimate $\widehat{\boldsymbol{\theta}}_T = (\hat{\theta}_{T,j})$ of the static parameter as in (10) and, for $i = 1, \dots, N$ and $t = 1, \dots, T$, recursively construct $(\widehat{c}_{i,t}, \widehat{\gamma}_{i,t}, \widehat{u}_{i,t})$ following (6), (8), and the score-driven constructions in Proposition 2.2. Construct the standardized residuals $\widehat{Z}_{i,t} = (Y_{i,t} - \widehat{u}_{i,t})/\widehat{c}_{i,t}$ and let $\widehat{H}_i$ denote the empirical CDF of the positive residuals $\widehat{Z}_{i,t} > 0$ for each $i = 1, \dots, N$.

Step 2. Compute the probability integral transform values as

$$\widehat{U}_{i,t} = \begin{cases} p\big(1 - \widehat{\gamma}_{i,t}\widehat{Z}_{i,t}\big)^{-1/\widehat{\gamma}_{i,t}}, & \widehat{Z}_{i,t} \le 0, \\ p + (1-p)\widehat{H}_i\big(\widehat{Z}_{i,t}\big), & \widehat{Z}_{i,t} > 0, \end{cases}$$

for $i = 1, \dots, N$ and $t \in \mathbb{T}_{\text{sub}}$. Then, for each country i and $t \in \mathbb{T}_{\text{sub}}$, associate $\widehat{U}_{i,t}$ with the rank interval

$$\widehat{R}_{i,t} = \left[\frac{1}{T_{\text{sub}}} \sum_{s \in \mathbb{T}_{\text{sub}}} \mathbb{1}\Big\{\widehat{U}_{i,s} < \widehat{U}_{i,t}\Big\}, \quad \frac{1}{T_{\text{sub}}} \sum_{s \in \mathbb{T}_{\text{sub}}} \mathbb{1}\Big\{\widehat{U}_{i,s} \le \widehat{U}_{i,t}\Big\} \right].$$

Step 3. Let $\ell \in \mathbb{Z}^+$, with $\ell < T_{\text{sub}}$, denote the block length. Define $\widehat{\boldsymbol{R}}_t := \big(\widehat{R}_{1,t}, \dots, \widehat{R}_{N,t}\big)$

and set $\widehat{\boldsymbol{R}}_{t+T_{\text{sub}}} = \widehat{\boldsymbol{R}}_t$ for $t \in \mathbb{T}_{\text{sub}}$ to construct T_{sub} circular overlapping blocks

$$\widehat{\boldsymbol{B}}_s = (\widehat{\boldsymbol{R}}_s, \ldots, \widehat{\boldsymbol{R}}_{s+\ell-1}), \qquad s \in \mathbb{T}_{\text{sub}}.$$

Independently draw $\pi(1), \ldots, \pi(K)$, with $K = \lceil T/\ell \rceil$, from the discrete uniform distribution on $\mathbb{T}_{\text{sub}}$. Concatenate the blocks $\widehat{\boldsymbol{B}}_{\pi(1)}, \ldots, \widehat{\boldsymbol{B}}_{\pi(K)}$ to obtain $(\boldsymbol{R}^*_1, \ldots, \boldsymbol{R}^*_{K\ell})$, where $\boldsymbol{R}^*_t := (R^*_{1,t}, \ldots, R^*_{N,t})$, and retain the first T vectors. Conditional on $\boldsymbol{R}^*_1, \ldots, \boldsymbol{R}^*_T$, independently draw $U^*_{i,t} \sim \mathrm{U}(R^*_{i,t})$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

Step 4. Keeping $\widehat{\boldsymbol{\theta}}_T$ fixed, generate the bootstrap observations sequentially over $t = 1, \ldots, T$. At each time t, obtain $(c^*_{i,t}, \gamma^*_{i,t}, u^*_{i,t})$ from the model recursions evaluated at $\widehat{\boldsymbol{\theta}}_T$ using the previously generated bootstrap observations, and generate

$$Y^*_{i,t} = \begin{cases} u^*_{i,t} - \dfrac{c^*_{i,t}}{\gamma^*_{i,t}} \left[\left(\dfrac{p}{U^*_{i,t}} \right)^{\gamma^*_{i,t}} - 1 \right], & U^*_{i,t} \le p, \\ u^*_{i,t} + c^*_{i,t} \widehat{H}_i^{-1} \left(\dfrac{U^*_{i,t} - p}{1 - p} \right), & U^*_{i,t} > p. \end{cases}$$

Step 5. Repeat the bootstrap procedure B times to obtain the bootstrap estimates $\widehat{\boldsymbol{\theta}}^*_{T,b}$, $b = 1, \ldots, B$. Let $q^*_j(\xi)$ denote the empirical ξ-quantile of the jth component of $\sqrt{T}(\widehat{\boldsymbol{\theta}}^*_{T,b} - \widehat{\boldsymbol{\theta}}_T)$, $b = 1, \ldots, B$. Construct componentwise $(1 - \eta)$ bootstrap confidence intervals as

$$\left[\hat{\theta}_{T,j} - \frac{q^*_j(1 - \xi/2)}{\sqrt{T}}, \quad \hat{\theta}_{T,j} - \frac{q^*_j(\xi/2)}{\sqrt{T}} \right].$$

In practice, we set $p = 0.2$ in Step 2, $\ell = 8$ for the block length (two years) in Step 3, and $B = 500$ in Step 5.

C.2 Tail-Risk Dynamics during Major Crises

Figures C.1 and C.2 show the evolution of the 1% GaR for all countries in the sample during the COVID-19 and the global financial crisis, respectively. It underlines the pattern from the main text where the DP model typically performs at par or better in terms of closer alignment with GDP growth dynamics or in terms of early warning signals than versions of the model that only pool over one dimension or do not pool at all.

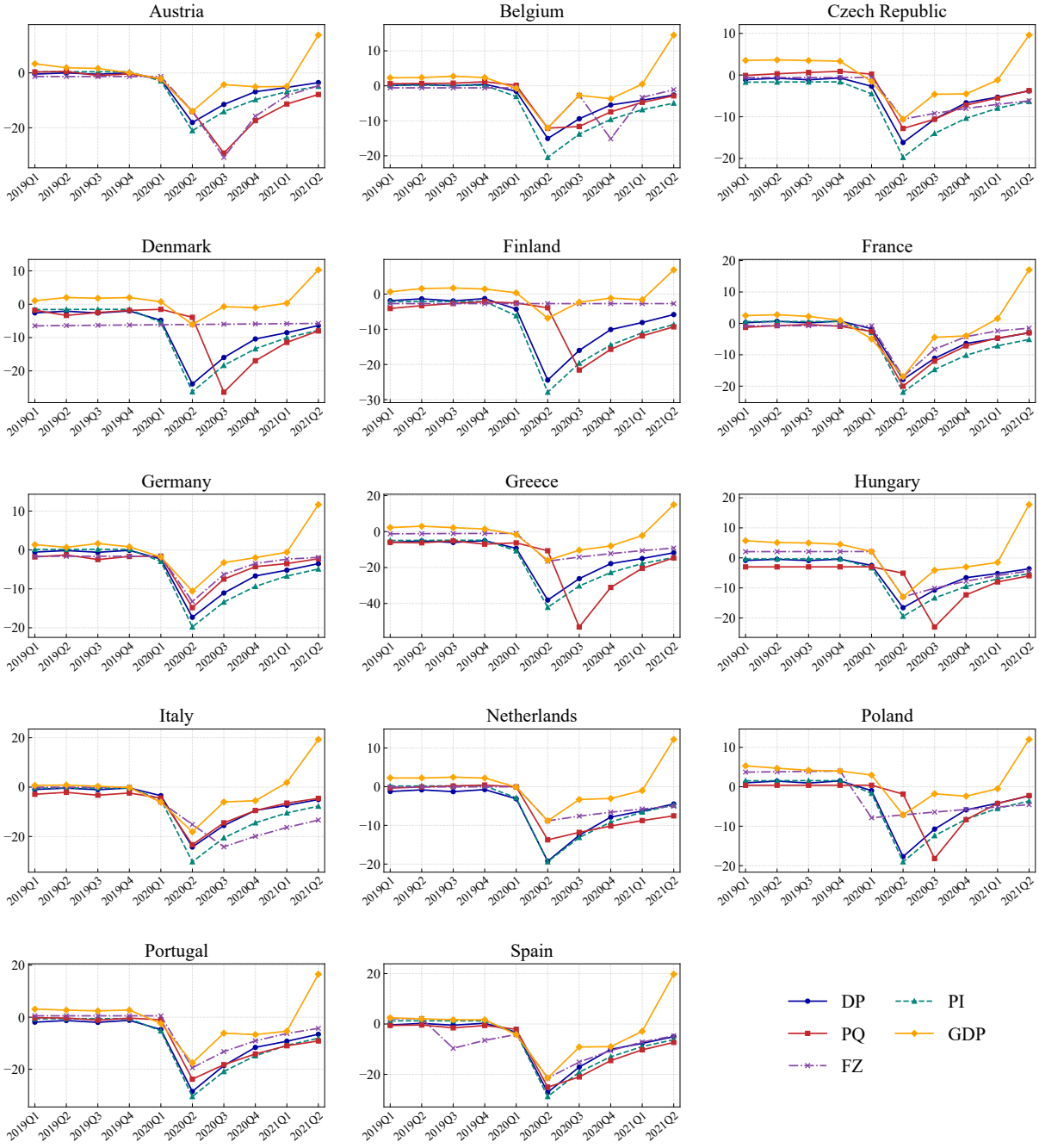


Figure C.1: Full-sample fitted GaR (1%) across 14 countries during 2019Q1–2021Q2. The panels compare DP, PI, PQ, and FZ with realized year-on-year real GDP growth.

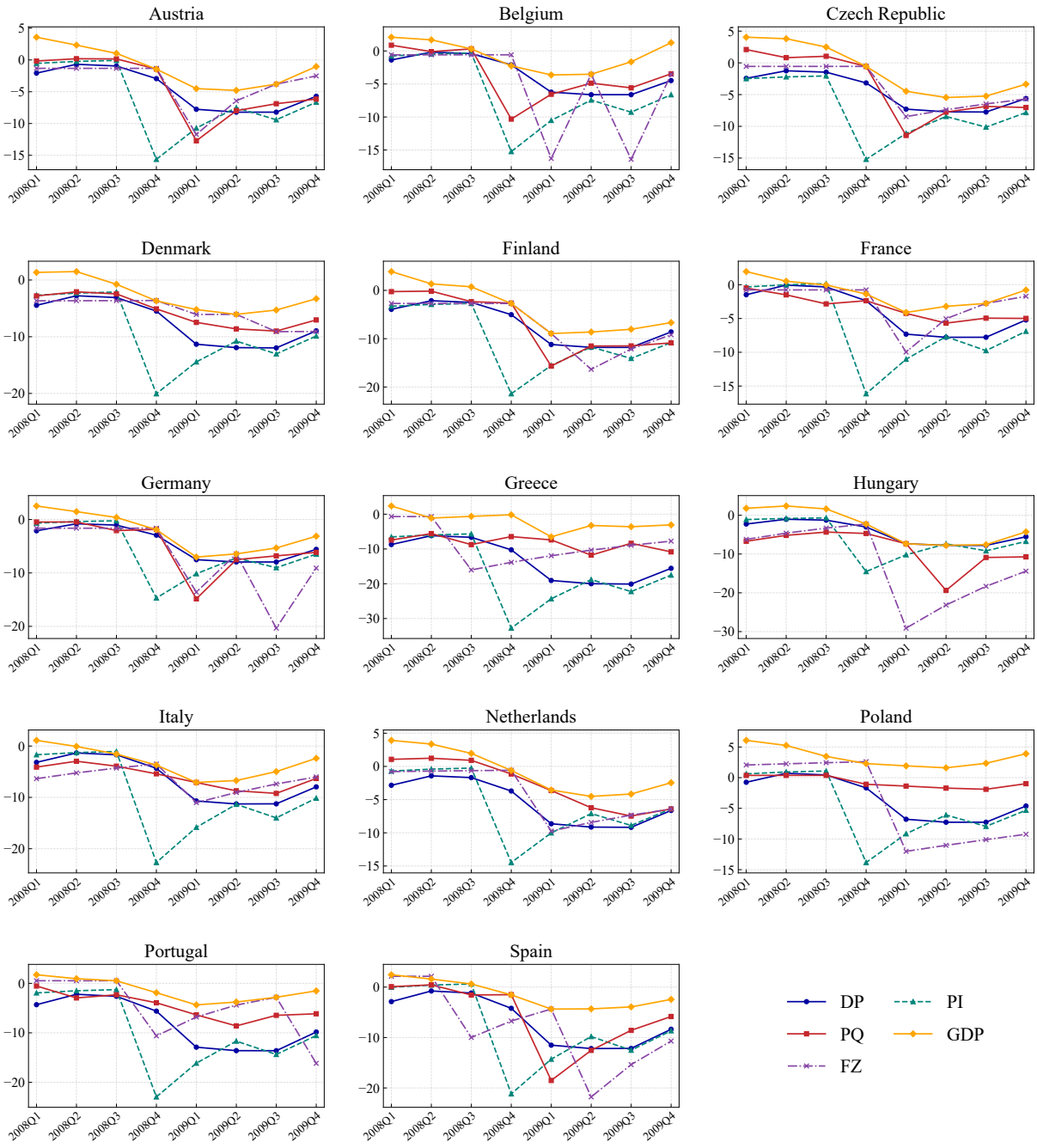


Figure C.2: Full-sample fitted GaR (1%) across 14 countries during 2008Q1–2009Q4. The panels compare DP, PI, PQ, and FZ with realized year-on-year real GDP growth.

C.3 Adopted models in the out-of-sample analysis

EW: empirical GaR and GES are re-estimated annually using an expanding sample and held fixed within the forecast year.

RW-40: a rolling-window variant that uses only the most recent 40 observations, corresponding to 10 years of quarterly data.

RW-60: a rolling-window variant that uses only the most recent 60 observations, corresponding to 15 years of quarterly data.

GARCH: a Gaussian GARCH(1,1) model with constant mean is fitted separately for each country annually at the beginning of each year, while the conditional variance is updated quarterly. GaR and GES are obtained from the one-step-ahead conditional predictive distribution.

GARCH-t: a Student-t GARCH(1,1) model with constant mean is fitted separately for each country annually at the beginning of each year, while the conditional variance is updated quarterly. GaR and GES are obtained from the one-step-ahead conditional predictive distribution.

MLE: a generalized Pareto distribution is fitted separately for each country, with the threshold fixed at $\hat{u}_t = \hat{u}$, where $\hat{u}$ is set to the empirical GaR at $\tau = 0.200$. The threshold and distributional parameters are re-estimated annually using an expanding sample and held fixed within the forecast year.

GaR: following the quantile-regression approach of Adrian et al. (2019), country-specific conditional quantiles are estimated annually using an expanding sample at levels 0.05, 0.25, 0.50, 0.75, and 0.95, using GDP growth and Country-Level Index of Financial Stress (CLIFS), obtained from ECB data portal, both lagged by one quarter.[7] For each quarter, a skew-t distribution is fitted to the predicted quantiles, and the conditional 1% GaR and GES are obtained from the fitted distribution.

PGaR: a panel extension of **GaR**, motivated by Adrian et al. (2022). The only difference is that, in the first-stage quantile regression, we use a panel quantile regression with country fixed effects.

EGaR: a country-specific conditional Pareto-tail model adapted from Adrian et al. (2026), using lagged GDP growth and, when available, lagged CLIFS. The threshold is fixed at the empirical 20% quantile of the annual estimation sample[8]. A covariate-dependent tail exponent and a kernel estimate of the conditional threshold probability

[7] The starting dates of the CLIFS series differ across countries and are all later than 1960. We therefore switch to the specification including CLIFS only when at least 40 quarters of effective CLIFS observations are available. We also considered alternative specifications using either one or four lags of GDP growth, including the PGAR and EGAR benchmarks. The key findings remain robust.

[8] We also considered a 10% threshold specification, but its out-of-sample performance was worse.

are used to extrapolate the conditional 1% GaR and GES. Model coefficients are re-estimated annually, while predictors are updated quarterly.

FZ: a no-pooling benchmark that fits each country and each quantile separately within each expanding window, with the threshold fixed at $\hat{u}_t = \hat{u}$, the empirical GaR at $\tau = 0.200$. This benchmark can be viewed as a flexible variant of the two-factor GAS specification in Patton et al. (2019).

PI: a pooling-across-country benchmark that assigns a small weight ($\rho = 0.01$) to the additional quantile terms with $J = 2$.

PQ: a pooling-across-quantiles benchmark estimated separately for each country within each expanding window, producing GaR and extreme GES forecasts.

We further consider a range of variants of the proposed DP model.

DP: the full DP model that pools information across countries and the 20 quantile levels $\{0.01, 0.02, \ldots, 0.20\}$. Parameters are re-estimated annually using an expanding sample, with ρ selected within each estimation window.

$\mathbf{DP_{\gamma 0}}$: **DP** with time-invariant tail shape parameters ($f_t^{\gamma} \equiv 0$).

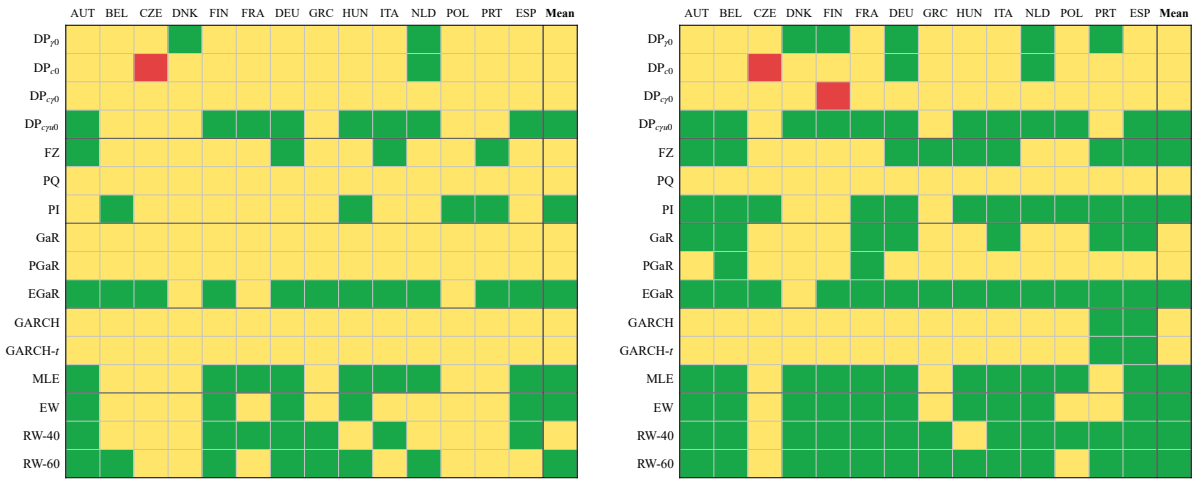
$\mathbf{DP_{c0}}$: **DP** with time-invariant scale parameters ($f_t^{c} \equiv 0$).

$\mathbf{DP_{c\gamma 0}}$: **DP** with time-invariant tail shape and scale parameters ($f_t^{\gamma} = f_t^{c} \equiv 0$).

$\mathbf{DP_{c\gamma u0}}$: **DP** with time-invariant tail shape, scale, and threshold parameters ($f_t^{\gamma} = f_t^{c} = f_t^{u} \equiv 0$).

C.4 Additional out-of-sample results

This appendix contains additional results tables for the out-of-sample analysis.



(a) 10% significance level

(b) 20% significance level

Figure C.3: Diebold–Mariano comparisons of out-of-sample quantile check losses at $\alpha = 0.01$. Each cell compares DP with the model in the corresponding row. Green indicates significantly lower loss for DP, red significantly higher loss, and yellow no significant difference.

Table C.1: Out-of-Sample Extreme Quantile Check Loss ($\alpha = 0.01$)

Models	AUT	BEL	CZE	DNK	FIN	FRA	DEU	GRC	HUN	ITA	NLD	POL	PRT	ESP	Mean
DP	0.078	***0.049***	0.064	0.088	0.084	0.106	***0.063***	0.161	0.112	***0.084***	***0.059***	***0.058***	**0.063**	0.093	***0.083***
$\text{DP}_{\gamma 0}$	0.085	0.060	0.064	0.103	0.116	***0.101***	0.083	0.154	0.073	0.099	0.068	0.060	0.072	***0.091***	0.088
$\text{DP}_{\text{c}0}$	***0.070***	0.052	**0.055**	0.094	0.097	0.104	0.072	0.192	**0.057**	0.106	0.067	**0.053**	0.068	0.106	0.085
$\text{DP}_{\text{c}\gamma 0}$	**0.050**	**0.041**	***0.059***	0.085	**0.067**	**0.076**	**0.046**	0.171	***0.069***	**0.082**	0.060	0.059	***0.064***	**0.079**	**0.072**
$\text{DP}_{\text{c}\gamma\text{u}0}$	0.278	0.166	0.164	0.163	0.176	0.291	0.244	0.166	0.203	0.310	0.150	0.213	0.184	0.408	0.223
FZ	0.236	0.176	0.144	0.136	0.107	0.104	0.165	0.321	0.182	0.181	0.130	0.221	0.163	0.237	0.179
PQ	0.120	0.122	0.167	0.101	0.117	0.170	0.134	***0.140***	0.227	0.176	0.095	0.147	0.129	0.247	0.149
PI	0.193	0.172	0.198	0.132	0.150	0.269	0.147	0.316	0.283	0.286	0.128	0.183	0.223	0.296	0.213
GaR	0.211	0.162	0.150	***0.079***	0.092	0.192	0.173	0.148	0.327	0.242	**0.054**	0.164	0.251	0.325	0.184
PGaR	0.131	0.176	0.144	**0.072**	0.097	0.231	0.127	**0.132**	0.221	0.193	0.062	0.158	0.196	0.254	0.157
EGaR	0.304	0.138	0.201	0.116	0.213	0.186	0.127	0.503	0.403	0.176	0.084	0.255	0.373	0.231	0.236
GARCH	0.129	0.120	0.138	0.115	0.082	0.125	0.078	0.183	0.295	0.117	0.152	0.175	0.193	0.184	0.149
GARCH-t	0.126	0.124	0.138	0.115	***0.081***	0.125	0.078	0.184	0.296	0.116	0.121	0.176	0.193	0.184	0.147
MLE	0.281	0.176	0.164	0.171	0.185	0.302	0.249	0.216	0.206	0.325	0.159	0.233	0.213	0.432	0.236
EW	0.244	0.157	0.154	0.154	0.156	0.252	0.229	0.227	0.187	0.280	0.147	0.193	0.206	0.391	0.213
RW-40	0.256	0.215	0.220	0.180	0.239	0.293	0.243	0.286	0.285	0.310	0.187	0.211	0.230	0.350	0.250
RW-60	0.221	0.191	0.160	0.141	0.231	0.244	0.191	0.292	0.192	0.273	0.165	0.174	0.230	0.321	0.216

Notes: This table reports average out-of-sample check loss at $\alpha = 0.01$, computed as $(\alpha - \mathbf{1}\{Y_{i,t} \leq v_{i,t}\})(Y_{i,t} - v_{i,t})$ in original GDP growth. The evaluation sample begins in 2001Q1 for eleven countries, 2010Q1 for CZE, and 2011Q1 for HUN and POL, and ends in 2026Q2. For each column, the smallest loss is in **bold**, and the second smallest is in ***bold italic***.