# Missing Data Density Imputation using Time-Varying High-Dimensional Hierarchical Density Models with Applications to Credit Curves

Yonas Khanna[1,2], Andre Lucas[2,3]

[1]Integrated Risk Model Development, ING Bank, The Netherlands

[2]School of Business and Economics, Vrije Universiteit Amsterdam, The Netherlands

[3]Tinbergen Institute, The Netherlands

# Missing Data Density Imputation using Time-Varying High-Dimensional Hierarchical Density Models with Applications to Credit Curves*

Yonas Khanna[1,2], Andre Lucas[2,3]

[1] *ING Bank, The Netherlands*
[2] *Vrije Universiteit Amsterdam, The Netherlands*
[3] *Tinbergen Institute, The Netherlands*

This version: March 19, 2025

**Abstract**

We propose an observation-driven dynamic hierarchical factor model for missing value imputation in high-dimensional panel data. Both the mean and variance of the cross-sectional units are modeled with multiple common and idiosyncratic factors. The model exploits both serial and cross-sectional information in the data and can easily cope with time-variation in conditional means and variances, as well as with either isolated or long patches of missing values. The novel approach not only provides point forecasts, but also density forecasts for the missing data, thus allowing the researcher to quantify imputation uncertainty. The model's static parameters can be estimated by standard maximum likelihood methods due to the model's score-driven structure. We apply the model on high-dimensional panels of daily global credit default swap (CDS) curves between 2011-2020, which contain large gaps of missing values. Our empirical findings demonstrate the significant importance of firm-specific location-scale effects in imputing densities and risk quantiles of CDS term-structures.

**Keywords:** missing value imputation, dynamic hierarchical factor models, forecasting, imputation uncertainty, high-dimensional panel data.
**JEL classifications: C32, C33, C58, G32, G17.**

*Khanna is grateful for the financial support provided by ING Bank. The views expressed in this paper are those of the authors and do not necessarily reflect the views ING Bank.

# 1 Introduction

The problem of missing data is a widespread challenge in many areas of empirical research. In social sciences, large longitudinal surveys and biological cohort studies often have missing values due to participant attrition, whereas economic panel data can be unbalanced when variables are measured at different frequencies or are not fully available for all spatial groups (see, e.g., Foroni and Marcellino 2014; Aruoba et al. 2009; Schumacher and Breitung 2008). Also in finance, typical data like asset returns, bond yields, and credit curves can be incomplete. This can be due to forms not yet or no longer being listed, bankruptcies, illiquidity and stale prices, or asynchronous data releases (see e.g., Gregory 2020; van der Merwe et al. 2018). As a result, such datasets often contain a significant number of missing values.

Instead of discarding cross-sectional units, which would lead to information loss, missing values are often imputed. A notable imputation example in the financial industry is found for credit default swap (CDS) term-structures, which are crucial for risk management and regulatory purposes. CDS curves serve as a key input for calculating Credit Value Adjustments (see, e.g., Gregory 2020; Pallavicini et al. 2011; Green 2015; Morini and Prampolini 2011). The are also needed to quantify Basel III capital requirements for OTC transaction profits and losses (see BCBS, 2010). However, since CDS curves are often highly illiquid or completely unavailable for many firms, regulators require financial institutions to construct synthetic curves (proxies) based on spatial factors (e.g., region, sector and rating) extracted from traded CDS instruments (EBA, 2015). In practice, building synthetic curves boils down to a large-scale time-series imputation exercise of significant industrial importance. Since imputation involves the prediction of unknown values without any means of direct verification (as the true data value is not observed), selecting an appropriate imputation method is crucial for ensuring the accuracy and reliability of the imputed data.

One of the common ways used in practice to address missing values in high-dimensional time-series panels is that of factor-based imputation methods. These methods exploit the co-movement of observed cross-sectional units to extract common components (factors), which are then used to impute the missing data. Factors are typically estimated using principal components (PCs) and recent studies, including those by Cahan et al. (2023), Xiong and Pelger (2023), Bai and Ng (2021), and Jin et al. (2021), have introduced various PC-based approaches to consistently estimate common components when the factor

structure is strong in unbalanced panel data. Although these factor-based methods are computationally efficient and scalable for large data panels, they do not exploit the serial correlation embedded in sparse panel data, which is particularly relevant for time-series modeling. This is where dynamic time-series models can prove useful, as they exploit both serial and cross-sectional information to impute (or forecast) the missing values. For example, studies by Jin et al. (2021), Stock and Watson (2016), and Jungbacker et al. (2011) explore dynamic factor models to handle missing values, while Chan et al. (2023) propose conditionally Gaussian state-space models for the same purpose. However, these existing methods focus solely on predicting the mean of missing observations, while not fully dealing with their time-varying variances.[1] It is well known that economic and financial variables often exhibit time-varying volatility, and hence, assuming constant variances is inadequate for risk management. An inadequate treatment of volatilities may even lead to biased estimates of risk measures such as value-at-risk and expected shortfall, for which quantification of the time-varying volatility of missing data is crucial. In addition, most research focuses on imputing missing values using common components only, overlooking the potential significance of idiosyncratic mean and variance components in certain applications. Consequently, high-dimensional dynamic mean-variance models that incorporate both common and idiosyncratic components and that can handle complex missing data patterns remain underdeveloped.

This paper overcomes the above shortcomings and provides three key improvements to handle missing values in high-dimensional unbalanced panels of financial data. First, we relax the constant variance matrix assumption for panel data by developing a hierarchical multivariate Gaussian model with time-varying means and variances based on the generalized autoregressive score framework of Creal et al. (2013) and Harvey (2013). In this model, both moments follow a hierarchical factor structure with a predefined exogenous design matrix that links score-driven factors to each of the time series' first two moments. Specifically, the design matrix consists of effect-coded dummy variables and acts as a factor-loading matrix that decomposes cross-sections of means and variances into interpretable location and scale factors. It accommodates multiple common

[1]The so-called factor-based 'residual overlay' algorithm of Cahan et al. (2023) can be used to consistently estimate the covariance matrix of panels with missing units, but only if the respective matrix is assumed to be constant. The multivariate state-space models of Chan et al. (2023) on the other hand do include a single common stochastic volatility factor, but do not cover model specifications with time-varying idiosyncratic volatilities.

and idiosyncratic dummies as well as complex hierarchical spatial structures (e.g., region, sector, or rating dummies and their interactions), and it may vary over time to capture structural shifts (e.g., changes in firm ratings). The model's static parameters can be estimated by standard maximum likelihood methods due to the model's observation-driven structure. Therefore, this dynamic hierarchical model offers substantial flexibility for extracting moments from sparse multivariate time-series and to impute the full density of missing values.

Second, our modeling approach leverages on the hierarchical structure of the data to improve imputation in sparse multivariate time-series. In score-driven models, the time-varying parameters are updated using the score of the predictive density of the observed data only, see Blasques et al. (2021) and Lucas et al. (2016). However, when some observations are missing, this score update typically still induces a change in all parameters, even those linked exclusively to the missing units. This can to the approach not fully exploiting the information in the available data to capture for instance the common components, because some information 'leaks' or is transferred to the identification of idiosyncratic factors for which no data is available at that moment. In situations where the common factor is a dominant driver of the time-series dynamics, this is a sub-optimal outcome. If the common factor is dominant, it is crucial to estimate it as accurately as possible for a good imputation result, even if this comes at the expense of a less good imputation result for the idiosyncratic components that are solely linked to the units that are not observed at that time.

To address this issue, we exploit the hierarchical structure in the dataset and introduce a new filtering mechanism that allows the user to enforce mean-reversion on (e.g., idiosyncratic) factors that are deemed less relevant for the overall imputation outcome and that are affected by the missing values. Through an extensive Monte Carlo study using the model as the true data generating process (DGP), we demonstrate that this new filter accurately captures common location and scale signals as long as at least one observation is present in the panel's cross-section. This even holds when data are historically block or randomly missing 75% of the time, thus enabling accurate prediction of missing value densities even with limited data. The standard score-driven filtering approach, by contrast, does not fare well in this context. This paper therefore substantially extends the univariate score-driven filter for missing values of Blasques et al. (2021) to large-scale multifactor time-varying score-driven location-scale models, including those with non-continuous and non-Gaussian distributions.

As a third contribution, this paper not only focuses on the imputation predition itself, but also on the inference or uncertainty about the missing values in score-driven models. Although it is rather straightforward to generate in-sample point-forecasts with the score-driven models, generating reliable interval forecasts requires simulation-based techniques. We extend the out-of-sample simulation algorithm of Blasques et al. (2016) to the in-sample data imputation problem. Whenever missing data is encountered during filtering, the algorithm imputes these values by simulation based on the model's distributional assumptions and combined with observed data to form a complete cross-section, which updates the time-varying parameters. This process is repeated multiple times throughout the sample to produce simulation bands for both the missing values and the underlying time-varying parameters and in this way captures the imputation uncertainty. Additionally, following Blasques et al. (2016), the algorithm is further extended to also incorporate parameter uncertainty using the asymptotic covariance of the parameters. Monte Carlo experiments reveal that the simulation bands perform well and capture the imputation uncertainty at nominal levels for both the factors and various risk quantiles of the true DGP. Therefore, the outputs of our imputation framework are likely to be valuable not only to academics, but also to practitioners.

In our empirical analysis, we demonstrate the strong performance of our dynamic hierarchical density model for imputing missing values in time-series of CDS term-structures through two applications: single-name and multi-name curve imputation. For both applications we use daily CDS data of global financial institutions, spanning the period from Jan 2011 to Dec 2022. In the first application, we assess various score-driven model specifications using CDS time-series of JPMorgan Chase & Co. (JPM). We simulate artificial missing data (either in blocks or randomly missing) into JPM's credit curves and then compare the score-driven model's density forecasts with those from a state-space model using the Kalman filter (Durbin and Koopman, 2012). Our results reveal similar in-sample forecasting performance between the two models, which complements the out-of-sample findings of Koopman et al. (2016). In the multi-name application, we build a hierarchical CDS curve model to simultaneously impute six tenors for 24 global financial institutions (a total of 144 CDS time series over ten years) using a design matrix that incorporates nearly 100 common and idiosyncratic location and scale factors. We find that the GAS model not only scales well to high dimensions, but that hierarchical models with firm-specific factors also produce more accurate risk quantiles of the panel compared to models without such idiosyncratic components.

The remainder of the paper is organized as follows. Section 2 introduces the modeling framework for imputing densities using hierarchical score-driven location-scale models and also describes the simulation algorithm to construct in-sample forecast bands. In Section 3, an extensive Monte Carlo experiment is performed to study the performance of our model across a range of missing data mechanisms. Section 4 presents the empirical applications. Finally, Section 5 concludes.

# 2 Modeling Framework

## 2.1 Hierarchical dynamic location-scale factor models

Let $y_t = (y_{1,t}, \ldots, y_{p,t})' \in \mathbb{R}^p$ be a $p$-dimensional time-series process observed for $t = 1, \ldots, n$. Not all elements of $y_t$ are observed at all times. Let $y_t^\dagger = S_t y_t \in \mathbb{R}^{p_t}$ denote the the observed part of $y_t$ , where $S_t$ is a $p_t \times p$ selection matrix of ones and zeros holding the appropriate rows from the unit matrix $\mathrm{I}_p$. The dimension $p_t$ of $y_t^\dagger$ can vary between $p_t = p$ for a fully observed $y_t$ with $S_t = \mathrm{I}_p$, and a completely missing $y_t$ with $p_t = 0$. Components of $y_t$ may be missing at random (MAR), for instance for an i.i.d. selection process $S_t$ that is also independent of $\{y_t\}_{t=1}^n$. In many settings, however, data may be missing in a more structured way. For example, time-series may include gaps at the start, because one of the components was not recorded or did not exist at the time. Gaps may also occur at the end of the sample because of attrition effects, or in the middle for components that were temporarily not observed. In our empirical application in Section 4 we indeed have longer stretch of (partially) missing values either at the start, middle, or end of the sample where specific series are unavailable due to a lack of trading.

We assume that the time-series behavior of $y_t^\dagger$ can be described by a multivariate location-scale model with a hierarchical structure for both the locations and scales:

$$
\begin{aligned}
y_t &= \mu_t + \Sigma_t^{1/2}\varepsilon_t, & \varepsilon_t &\overset{\text{iid}}{\sim} \mathrm{N}(0, \mathrm{I}_p) & t &= 1, \ldots, n, \\
\mu_t &= F_t f_t^\mu, & \sigma_t &= \exp\left(\tfrac{1}{2} F_t f_t^{\log\sigma^2}\right), & \Sigma_t &= \mathrm{diag}(\sigma_t)^2, \\
y_t^\dagger &= S_t y_t, & \varepsilon_t^\dagger &= S_t \varepsilon_t, & & \\
\mu_t^\dagger &= S_t \mu_t, & \sigma_t^\dagger &= \mathrm{diag}(S_t \sigma_t), & \Sigma_t^\dagger &= S_t \Sigma_t S_t'.
\end{aligned}
\tag{1}
$$

where the exponential for $\sigma_t$ works element-wise. The matrix $F_t \in \mathbb{R}^{p \times k}$ is a known design matrix that imposes a hierarchical structure on the means and (log) variances as driven by the time varying parameters $f_t^\mu \in \mathbb{R}^k$ and $f_t^{\log\sigma^2} \in \mathbb{R}^k$, respectively. For

instance, in Section 4 we observe time-series for a number of financial instruments (level 4), each instrument being characterized by the firm it relates to (level 3), and each firm characterized by its credit quality (rating), its industry, and its region (level 2), which in turn hinge on the global common developments that affect all firms and instruments (level 1). Note that the model can easily be extended to allow for different hierarchical structures $F_t^\mu$ and $F_t^\sigma$ in the means and volatilities, respectively, as well as for different dimensions of $f_t^\mu$ and $f_t^{\log \sigma^2}$.

To fix the idea, consider for example an $F_t$ such that

$$\underbrace{\begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \vdots \\ \mu_{p,t} \end{bmatrix}}_{\mu_t \ (p\times 1)} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & \ddots & 0 \\ \vdots & 0 & 0 & 1 \\ 1 & -1 & \dots & -1 \end{bmatrix}}_{F_t = F \ (p\times p)} \underbrace{\begin{bmatrix} f_{c,t}^\mu \\ f_{1,t}^\mu \\ \vdots \\ f_{p-1,t}^\mu \end{bmatrix}}_{f_t^\mu \ (p\times 1)}, \tag{2}$$

or

$$\begin{aligned} &\mu_{i,t} = f_{c,t}^\mu + f_{i,t}^\mu, \qquad \text{for } i = 1, \dots, p, \\ &f_{1,t}^\mu + \dots + f_{p,t}^\mu = 0, \end{aligned}$$

which decomposes the means $\mu_{i,t}$ into a common factor $f_{c,t}^\mu$ and an idiosyncratic component $f_{i,t}^\mu$ for $i = 1, \dots, p$. The second equation is needed for identification and enforces the idiosyncratic components to always sum to zero to avoid multicollinearity. From this, we easily derive

$$\widetilde{f}_t^\mu = \underbrace{\begin{bmatrix} f_{c,t}^\mu \\ f_{1,t}^\mu \\ \vdots \\ f_{p-1,t}^\mu \\ f_{p,t}^\mu \end{bmatrix}}_{\widetilde{f}_t^\mu \ ((p+1)\times 1)} = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ 0 & -1 & \dots & -1 \end{bmatrix}}_{\widetilde{F}_t \ ((p+1)\times p)} \underbrace{\begin{bmatrix} f_{c,t}^\mu \\ f_{1,t}^\mu \\ \vdots \\ f_{p-1,t}^\mu \end{bmatrix}}_{f_t^\mu \ (p\times 1)} = \widetilde{F}_t \, f_t^\mu, \tag{3}$$

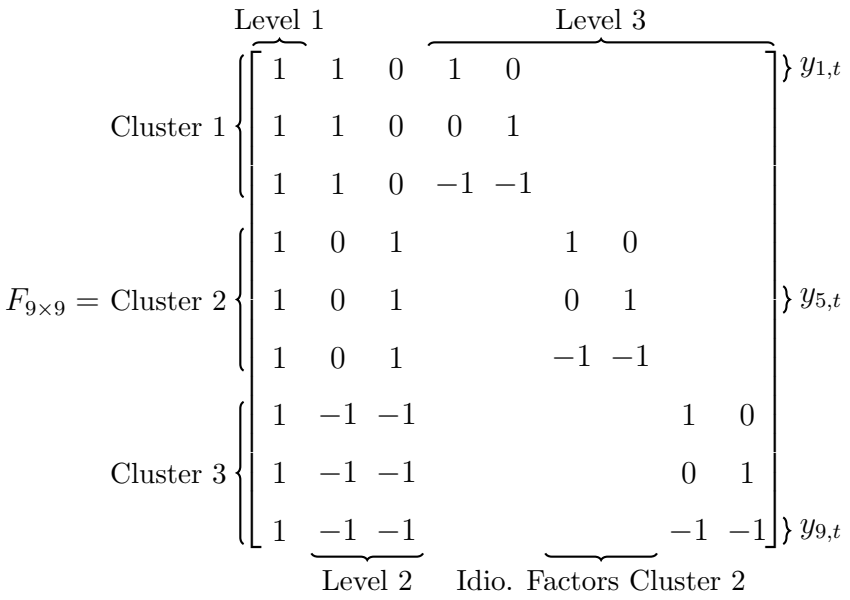
with $\widetilde{f}_t^\mu \in \mathbb{R}^{k_c}$ denoting the vector containing all constrained and unconstrained factors, and $\widetilde{F}_t \in \mathbb{R}^{k^c \times k}$ the corresponding known design matrix. Unlike plain vanilla dummy encoding where the common mean component $f_{c,t}^\mu$ reflects the mean of the *omitted* category to avoid multicollinearity, the effect dummy coding in (2) specifies $f_{c,t}^\mu$ as the overall grand mean of the components in $y_t$. This overall grand mean component can still be estimated as long as there is at least one observation left in the cross-section, and can

subsequently be used to anchor the missing components in $y_t$. This eases the interpretation in our hierarchical setting compared to plain vanilla dummy encoding. Another advantage of effect coded matrix $F_t$ is that we can now estimate idiosyncratic factors $f^{\mu}_{i,t}$ over and above the grand mean component $f^{\mu}_{c,t}$ for *all* series in the system. This is a useful feature when interpreting the output of the model. A similar use of effect coded dummies can be found in the structural time-series literature, where it is typically used to model time-varying seasonal components; see for instance Harvey (1990). It is precisely this hierarchical structure that we exploit in our approach when coping with missing data: the structure allows the user to select which component to prioritize when inferring the values of $F_t$ in the face of missing observations.

The hierarchical decomposition of means and log variances can of course be taken a step further. Figure 1 provides an example for 9 financial instruments with a common (Level 1) effect. The 9 instruments can be grouped in 3 clusters, each cluster corresponding to a different firm (Level 2). Finally, for each firm, each instrument has an idiosyncratic mean and log variance component (Level 3). Moreover, in the examples thus far, the design matrices were all time-invariant. However, situations where $F_t$ is time-varying can easily be thought of as well. For instance, in our empirical applications the time-series depend on the credit quality of the underlying firms. As the credit quality changes over time, so may the design matrix $F_t$.

We describe the time-varying parameters $f^{\mu}_t$ and $f^{\log\sigma^2}_t$ using the score-driven dynamics of (Creal et al., 2011, 2013) and Harvey (2013). The score-driven dynamics for $f^{\mu}_t$ and $f^{\log\sigma^2}_t$ adjust the time-varying parameters in a steepest ascent direction of the time $t$ predictive density. This has a distinct advantage over a state space approach in our application. Score-driven dynamics allow more easily for generalizations of the model, such as the time-varying volatility structure in (1), without increasing the computational burden of the approach. Such additional features are important for the type of data in Section 4. Particularly in high-dimensional settings such as ours, accommodating such features without undue computational costs is a relevant consideration. Also further generalizations such as fat-tailed densities in (1) are easy to accommodate, as are structural time-series components as the ones introduced by Harvey (2013). Score-driven dynamics for $f^{\mu}_t$ and $f^{\log\sigma^2}_t$ also improve the expected Kullback-Leibler divergence between the unknown data generating process and the statistical model as time progresses and provide consistent estimates of the time-varying parameter paths even if the model is mis-specified (Blasques et al., 2015; Beutner et al., 2023). Finally, (Koopman et al., 2016) show that

**Figure 1: An hierarchical design for factor loading matrix $F$**




*Notes*: An illustration of an effect coded multi-level loading matrix for 3 clusters among 9 time-series in the cross-section. The level 1 dummy indicates the common level among all series, level 2 dummies account for dynamic firm effects across the clusters, and level 3 dummies model idiosyncratic (firm, instrument) dynamics within each level 2 cluster.

(univariate) score-driven models exhibit a similar state prediction accuracy as state space models even if the latter are the data generating process.

The score-driven dynamics for $f_t^\mu$ and $f_t^{\log\sigma^2}$ are given by the following expressions:

$$\begin{aligned} f_{t+1}^\mu &= \kappa^\mu + B^\mu f_t^\mu + A^\mu s_t^\mu, \\ f_{t+1}^{\log\sigma^2} &= \kappa^{\log\sigma^2} + B^{\log\sigma^2} f_t^{\log\sigma^2} + A^{\log\sigma^2} s_t^{\log\sigma^2}, \end{aligned} \tag{4}$$

where $\kappa^\mu, \kappa^{\log\sigma^2} \in \mathbb{R}^k$ are vector-valued, and $A^\mu, A^{\log\sigma^2}, B^\mu, B^{\log\sigma^2} \in \mathbb{R}^{k\times k}$ are matrix-valued parameters. We collect all unknown parameters into the vector $\psi$, which we estimate by maximum likelihood later on. We also collect $f_t^\mu$ and $f_t^{\log\sigma^2}$ into the vector $f_t$, and $s_t^\mu$ and $s_t^{\log\sigma^2}$ into the vector $s_t$, respectively. The innovations $s_t$ in the transition equation (4) are given by the score of the log-predictive density $p(y_t^\dagger \mid f_t, \mathcal{F}_{t-1}; \psi)$,

$$\begin{aligned} s_t^\mu &= (\mathcal{I}_t^\mu)^+ \nabla_t^\mu, \qquad \nabla_t^\mu = \frac{\partial \log p\left(y_t^\dagger \mid f_t, \mathcal{F}_{t-1}; \psi\right)}{\partial f_t^\mu}, \\ \mathcal{I}_t^\mu &= -\mathbb{E}_{t-1}\left[\frac{\partial^2 \log p\left(y_t^\dagger \mid f_t, \mathcal{F}_{t-1}; \psi\right)}{\partial f_t^\mu \partial f_t^{\mu\prime}}\right] = \mathbb{E}_{t-1}\left[\nabla_t^\mu \nabla_t^{\mu\prime}\right], \end{aligned} \tag{5}$$

see Creal et al. (2013, 2014), where $(A)^+$ denotes the pseudo-inverse of the matrix $A$ and $\mathcal{F}_{t-1}$ is the information set of all past observations. A similar set of equations holds for $f_t^{\log\sigma^2}$. Note that we take the scores of the predictive densities of the observed data

$y_t^\dagger$ only, which makes sense, as those data are the only signals we can use to adjust the time-varying parameters. We also filter separately for the time-varying parameters $f_t^\mu$ and $f_t^{\log\sigma^2}$, which increases in numerical efficiency and stability of the filter. For our model in equation (1) we have the following result.

**Proposition 1.** *For the observation equation in (1), the scores and information matrices for $f_t^\mu$ and $f_t^{\log\sigma^2}$ in equation (5) reduce to*

$$\begin{aligned}
\nabla_t^\mu &= F_t' S_t' \Sigma_t^{\dagger -1/2} \varepsilon_t^\dagger & \nabla_t^{\log\sigma^2} &= \tfrac{1}{2} F_t' S_t' \left(\varepsilon_t^{\dagger 2} - \iota_{p_t}\right), \\
\mathcal{I}_t^\mu &= F_t' S_t' \Sigma_t^{\dagger -1} S_t F_t, & \mathcal{I}_t^{\log\sigma^2} &= \tfrac{1}{2} F_t' S_t' S_t F_t, \\
s_t^\mu &= \left(F_t' S_t' \Sigma_t^{\dagger -1} S_t F_t\right)^+ F_t' S_t' \Sigma_t^{\dagger -1/2} \varepsilon_t^\dagger, & s_t^{\log\sigma^2} &= \left(F_t' S_t' S_t F_t\right)^+ F_t' S_t' \left(\varepsilon_t^{\dagger 2} - \iota_{p_t}\right),
\end{aligned} \tag{6}$$

*where the squaring $\varepsilon_t^{\dagger 2}$ is done element-wise, and $\varepsilon_t^\dagger = \Sigma_t^{\dagger -1/2} S_t (y_t - \mu_t)$.*

See the appendix for all proofs and derivations.

Both scaled-scores take familiar expressions. For instance, $s_t^\mu$ is a generalized least squares (GLS) improvement using 'regressor matrix' $F_t$ and prediction error $e_t^\mu = y_t - \mu_t$. Similarly, $s_t^{\log\sigma^2}$ is an OLS improvement based on the deviations from squared standardized residuals from their expected values, i.e., $\varepsilon_t^2 - \iota_p$, using again the 'regressor matrix' $F_t$ to combine all the information. We assume $F_t$ to have full column rank, i.e., the design matrix must not be perfectly multicollinear.

Interestingly, the model in (1) can also be written as a VARMA form with vector log GARCH volatility dynamics and possibly time-varying coefficients. To see this, define $\bar{K}_t = \left(F_t' \Sigma_t^{-1} F_t\right)^{-1} F_t' \Sigma_t^{-1}$ and $e_t^\mu = y_t - \mu_t = \Sigma_t^{1/2} \varepsilon_t$, and note that

$$\bar{K}_t y_t = f_t^\mu + \bar{K}_t \Sigma_t^{1/2} \varepsilon_t \quad \Leftrightarrow \quad f_t^\mu = \bar{K}_t \ (y_t - e_t^\mu),$$

such that

$$\begin{aligned}
y_t &= F_t f_t^\mu + \Sigma_t^{1/2} \varepsilon_t = F_t \left(\kappa^\mu + B^\mu f_{t-1}^\mu + A^\mu s_{t-1}^\mu\right) + e_t^\mu \\
&= F_t \kappa^\mu + F_t B^\mu \bar{K}_{t-1} \left(y_{t-1} - e_{t-1}^\mu\right) + F_t A^\mu \bar{K}_{t-1} e_{t-1}^\mu + e_t^\mu \\
&= \underbrace{F_t \kappa^\mu}_{m_t} + \underbrace{F_t B^\mu \bar{K}_{t-1}}_{\Phi_t} y_{t-1} + e_t^\mu \underbrace{- F_t \left(B^\mu - A^\mu\right) \bar{K}_{t-1}}_{\Theta_t} e_{t-1}^\mu.
\end{aligned} \tag{7}$$

The latter can be clearly recognized as a VARMA(1,1) process with possibly time-varying intercept $F_t \kappa^\mu$, time-varying AR matrix $F_t B^\mu \bar{K}_{t-1}$, time-varying MA matrix $-F_t \left(B^\mu - A^\mu\right) \bar{K}_{t-1}$, and time-varying variances as captured by the time-varying conditional covariance matrix $\Sigma_t$. The hierarchical model in equation (1) together with the score-driven dynamics from equation (4) and Proposition 1 can thus describe a wealth of

different dynamic patters in a vector time-series process $y_t$.

## 2.2 Exploiting the hierarchical structure for updating $f_t$

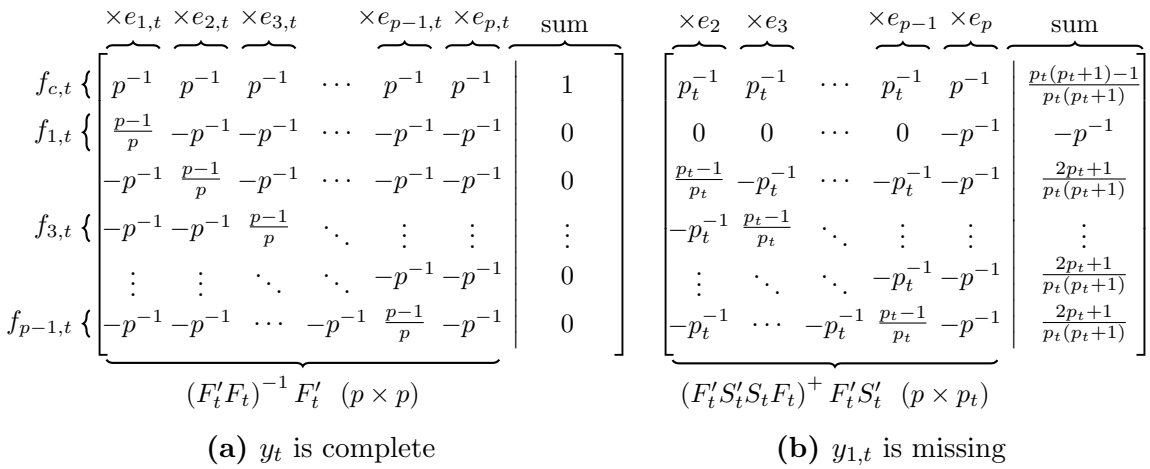
So far, the filter in Proposition 1 accounts for missing values through the presence of the matrix $S_t$. This is particularly important for observation-driven models, where current non-missing observations determine the future value of the parameters. The approach thus far, however, does not yet exploit the hierarchical structure of the model in any way. This hierarchical structure can be used by the researcher to determine which factors should have prioritized access to the available observations when updating the parameters, and which ones should mean-revert given the lack of information at time $t$.

It is clear from the updating mechanism in Eq. (5) that the score is only determined on the basis of the predictive density of the observed data. Nevertheless, even though not all elements of $y_t$ may be observed at time $t$, the score steps typically still induce a change in *all* of the time-varying parameters in the model, including the factors that only affect the observations that are currently missing. This may not always be desirable, and it is here that we can use the hierarchical structure to our advantage.

As an example, consider the $p \times p$ design matrix $F_t$ in (2) for $\Sigma_t = \mathrm{I}_p$. In this simple setting, we can easily re-trace the key features of our approach. The updates for a situation without missing values as well as a situation where one series is missing, are both given as a product of a specific matrix with the available elements of $e_t$. The relevant matrices are given in Figure 2b. The rows of each matrix correspond to the different time-varying factors, as indicated in the figure, while the columns correspond to the elements of $e_{i,t}$.

For the case without missing values (left-hand matrix), the weights of the different prediction errors for updating $f_{c,t}$ clearly sum to one, while those for the idiosyncratic factors $f_{i,t}$ sum to zero. This is in line with the specifications of the factors themselves, where $f_{c,t}$ is common to all series, whereas the $f_{i,t}$ are series-specific and sum to zero. When the first series is missing, however, the weights for $f_{c,t}$ sum to something strictly less than one, while the idiosyncratic factors receive a non-zero weight. This means that relative to the situation without missings, the common factor takes less information from the available observations than before. Instead, part of the information 'leaks' or is transferred to the estimation of the idiosyncratic factors. In such settings it may be useful to let the common factor first exploit all the empirical information in the available

**Figure 2: Score-driven regression weights without and with missing values**




**(a)** $y_t$ is complete **(b)** $y_{1,t}$ is missing

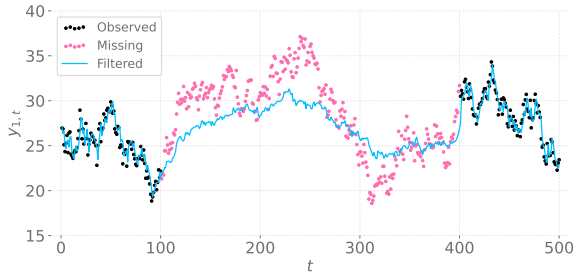
*Notes*: An illustration of the score-driven regression weights, given Proposition 1 and the $p$-dimensional effect-coded design matrix in (2). Figure 2a displays the score-driven OLS weighting matrix when all observations in $y_t$ are complete, whereas Figure 2b displays the weights when only $y_{1t}$ is missing at time $t$ ($p_t = p - 1$). Column values on the right-hand side of the vertical line correspond to the row-wise sum of the weights. The elements $e_t$'s on top of the matrix correspond to the available prediction errors used to obtain the scaled-score estimates $s_t$'s for the corresponding factors $f_t$'s indicated by the curly brackets on the left-hand side of the first matrix. The rows in both matrices correspond to exactly the same factor.

data, while letting the idiosyncratic factor mean-revert. In the context of the example in Figure 2, this means letting $f^{\mu}_{1,t}$ mean-revert and using the information of the observed series $y_{2:p,t}$ to first update the common level $f^{\mu}_{c,t}$.

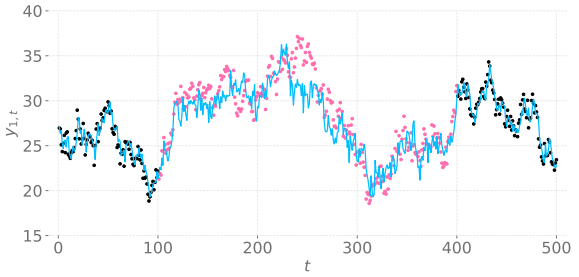
Figure 3 demonstrates this setup for a simulated bivariate time-series system from the same example (i.e., $p = 2$ in Eq. (2)) that features a strong common factor and assumes homoskedastic error variance specification. In this example, $y_{1,t}$ is missing from $t = 100, \ldots, 400$, but $y_{2,t}$ is always observed. When we filter the conditional means of these series using the standard score-driven updates of Creal et al. (2013) from Proposition 1, both $f^{\mu}_{c,t}$ and $f^{\mu}_{1,t}$ (and also $f^{\mu}_{2,t} = -f^{\mu}_{1,t}$) each receive about half of the prediction error $e_{2,t}$ during the period that $y_{1,t}$ is missing. As a result, the filtered (or imputed) mean $\mu_{1,t}$ of $y_{1,t}$ (shown in Subfigure 3a) does not track the missing data very well, even though $\mu_{2,t}$ continues to accurately follow $y_{2,t}$ (see Subfigure 3c). This holds despite the presence of strong common factor in the full system, which one may expect to enable the researcher to obtain a better estimate of the (dominant) level component of $y_{1,t}$. However, if we allow $f^{\mu}_{1,t}$ to mean-revert as we do in this paper and if we allow the common factor to first fully absorb the information in the new observation $y_{2,t}$, the imputed mean for $y_{1,t}$ (displayed in Subfigure 3b) again aligns much more closely with the (missing) observations $y_{1,t}$.

It is here that the modeler has a choice. For instance, if the common factor is much

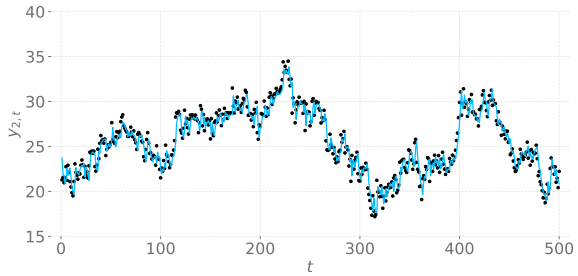
**Figure 3: A bivariate example of filtering with missing values**



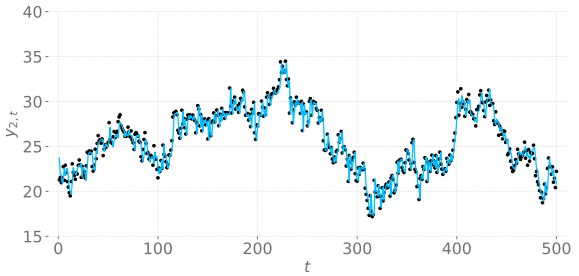

**(a)** $y_{1,t}$ and $f^{\mu}_{1,t}$ using the standard score-driven updates of Proposition 1




**(b)** $y_{1,t}$ and $f^{\mu}_{1,t}$ using the adjusted score-driven updates of Proposition 1′




**(c)** $y_{2,t}$ and $f^{\mu}_{2,t}$ using the standard score-driven updates of Proposition 1




**(d)** $y_{2,t}$ and $f^{\mu}_{2,t}$ using the adjusted score-driven updates of Proposition 1′

*Notes*: this figure presents an example of filtering the mean in a bivariate time-series system with missing values, based on the $p = 2$-dimensional effect-coded (as described in Eq. (2)), using two different score-driven filtering mechanisms. The top panels display the simulated time-series for the first variable $y_{1,t}$, while the bottom panels show those for the second variable $y_{2,t}$. in dots and bottom ones panels display the values for $y_{2,t}$. Observed values are depicted as black dots and missing values as pink dots. In each panels, the solid blue line depicts the filtered mean $\mu_t$ of $y_t$. The time-series in Subfigures 3a and 3c are filtered using the scores from Proposition 1, whereas those in Subfigures 3b and 3d use the scores from Proposition 1′. In the latter case, the idiosyncratic location factor $f^{\mu}_{1,t}$ mean-reverts when $y_{1,t}$ is missing, which also forces the second factor to mean-revert, since $f^{\mu}_{2,t} = -f^{\mu}_{1,t}$. The location model is simulated using the following parameters: $\kappa^{\mu} = [0.25, 0.05]'$, $A^{\mu} = \text{diag}([0.5, 0.25])$, $B^{\mu} = \text{diag}([0.98, 0.95])$, $\Sigma_t = I_2$, and $n = 500$.

larger in size than the idiosyncratic factors, it may be much more important to get the estimate of the common factor correct, possibly at the expense of the idiosyncratic factors. This situation is quite common in many areas of economics and finance. Also in our application later on, the common (level) factor is typically much larger than the subsequent regional, rating, or maturity related factors.

In technical terms, we achieve this objective as follows. Let $K_t \in \mathbb{R}^{r_t \times k}$ be a selection matrix of ones and zeros that selects the elements of $\widetilde{f}_t$ that the researcher chooses to mean-revert at time $t$ in the light of the missing observations. In the example above where $y_{1,t}$ is missing, we would let $f_{1,t}$ mean-revert, resulting in $K_t = (0, 1, 0, 0, 0)$. To ensure that the factors indicated by $K_t$ mean-revert, we impose that the derivatives of

the log predictive density with respect to $K_t f_t$ are zero, i.e., we set

$$K_t \tilde{\nabla}_t^{\mu} = 0, \qquad \tilde{\nabla}_t^{\mu} = \frac{\partial \log p(y_t^{\dagger} \mid f_t, \mathcal{F}_{t-1}; \psi)}{\partial \widetilde{f}_t} = \widetilde{F}_t \nabla_t^{\mu}. \tag{8}$$

Now let $K_{\perp t} \in \mathbb{R}^{(k-r)\times k}$ be a matrix orthogonal to $K_t$, such that $K_t \widetilde{F}_t K_{\perp t}' = 0$. Also define the $k \times k$ projection matrices $M_t = K_{\perp t}'(K_{\perp t} K_{\perp t}')^{-1} K_{\perp t} \in \mathbb{R}^{k\times k}$ and $M_{\perp t} = \widetilde{F}_t' K_t'(K_t \widetilde{F}_t \widetilde{F}_t' K_t')^{-1} K_t \widetilde{F}_t$. Then we rewrite the original score step $\nabla_t^{\mu}$ under the restriction $K_t \tilde{\nabla}_t^{\mu} = K_t \widetilde{F}_t \nabla_t^{\mu} = 0$, as

$$\nabla_t^{\mu} = (M_{\perp t} + M_t)\, \nabla_t^{\mu} = \widetilde{F}_t' K_t'(K_t \widetilde{F}_t \widetilde{F}_t' K_t')^{-1} \left( K_t \widetilde{F}_t \nabla_t^{\mu} \right) + M_t \nabla_t^{\mu} = M_t \nabla_t^{\mu}. \tag{9}$$

We directly see that such a step satisfies the restriction $K_t \widetilde{F}_t \nabla_t^{\mu} = K_t \widetilde{F}_t M_t \nabla_t^{\mu} = 0$ by design. We now obtain the revised Proposition 1′.

**Proposition 1′.** *For the observation equation in (1) and under the restrictions* $K_t \tilde{\nabla}_t^{\mu} = K_t \widetilde{F}_t \nabla_t^{\mu} = 0$ *and* $K_t \tilde{\nabla}_t^{\log\sigma^2} = K_t \widetilde{F}_t \nabla_t^{\log\sigma^2} = 0$, *the scores and information matrices for* $f_t^{\mu}$ *and* $f_t^{\log\sigma^2}$ *in equation (5) reduce to*

$$\begin{aligned}
\nabla_t^{\mu} &= M_t F_t' S_t' \Sigma_t^{\dagger -1/2} \varepsilon_t^{\dagger} & \nabla_t^{\log\sigma^2} &= \tfrac{1}{2} M_t F_t' S_t' \left( \varepsilon_t^{\dagger 2} - \iota_{p_t} \right), \\
\mathcal{I}_t^{\mu} &= M_t F_t' S_t' \Sigma_t^{\dagger -1} S_t F_t M_t, & \mathcal{I}_t^{\log\sigma^2} &= \tfrac{1}{2} M_t F_t' S_t' S_t F_t M_t, \\
s_t^{\mu} &= \left( M_t F_t' S_t' \Sigma_t^{\dagger -1} S_t F_t M_t \right)^{+} M_t F_t' S_t' \Sigma_t^{\dagger -1/2} \varepsilon_t^{\dagger}, & s_t^{\log\sigma^2} &= (M_t F_t' S_t' S_t F_t M_t)^{+} M_t F_t' S_t' \left( \varepsilon_t^{\dagger 2} - \iota_{p_t} \right),
\end{aligned} \tag{10}$$

*where* $K_{\perp t} \in \mathbb{R}^{(k-r)\times k}$ *is a matrix orthogonal to* $K_t$, *such that* $K_t \widetilde{F}_t K_{\perp t}' = 0$, *and* $M_t = K_{\perp t}'(K_{\perp t} K_{\perp t}')^{-1} K_{\perp t} \in \mathbb{R}^{k\times k}$, *and where the squaring* $\varepsilon_t^{\dagger 2}$ *is done element-wise, and* $\varepsilon_t^{\dagger} = \Sigma_t^{\dagger -1/2} S_t (y_t - \mu_t)$.

If there are different design matrices $F_t^{\mu}$ and $F_t^{\log\sigma^2}$, then also the corresponding projection matrices $M_t$ will be different for $\mu$ and $\sigma$.

Our approach for filtering with missing observations is inspired by the so-called 'setting-to-zero' method of Blasques et al. (2021) in the context of univariate score-driven models; see also Creal et al. (2014); Delle Monache et al. (2016); Koopman et al. (2018); Buccheri et al. (2020). In a univariate set-up, the 'setting-to-zero' method puts the score innovation of a factor in the transition equation to zero. Lucas et al. (2016) relate this method to the Expectation-Maximization algorithm as used in the missing data literature, and justify why this approach is reasonable in score-driven models. In our multivariate hierarchical setting, we generalize this approach by allowing the researcher to self-select which derivatives should be set to zero via the matrix $K_t$. The following examples illustrate how Proposition 1′ works out when some of the components of $y_t$ are missing.

**Example 1** (No mean-reverting factors)**.** When missing values at some $t$ are encountered, but there is still sufficient cross-sectional information available to estimate all factors, we have $r = 0$, $K_t$ is empty, and $K_{\perp t}$ is the identity matrix. As a result, $M_t = \mathrm{I}_k$ and Proposition 1′ collapses to Proposition 1. This can happen for instance in cases where multiple elements of $y_t$ identify the same factor. For instance, in case of a single common factor, this factor can still be filtered from the data as long as at least one element of $y_t$ is observed.

**Example 2** (Mean-reverting unrestricted idiosyncratic factors)**.** If we take the example from Figure 2, the first component of $y_t$ is missing. If the common component $f_{c,t}$ is much larger than the idiosyncratic components, it may be preferrable as argued before to let the first idiosyncratic component mean-revert, while using the remaining information to get as accurate an estimate of $f_{c,t}$ as possible. This leads to

$$K_t' = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \widetilde{F}_t' K_t' = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad K_{\perp t}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad M_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The resulting score step follows from Proposition 1′. For the case $\Sigma_t^\dagger = \mathrm{I}$, Figure 4a, then gives the new matrix $(M_t F_t' S_t' S_t F_t M_t)^+ M_t F_t' S_t'$ in front of $\varepsilon_t^\dagger$. This matrix can be compared to the one in Figure 2b. We clearly see that the scaled score step $s_t$ for $f_{1,t}$ equals zero, such that that factor mean-reverts. The common component now has equal weights for each of the three remaining elements of $y_t$ that are observed at time $t$. Also, the weights for $f_{2,t}$ and $f_{3,t}$ again sum to zero, as in the case without missings.

**Example 3** (Mean-reverting unrestricted and restricted idiosyncratic factors)**.** Now we consider the case $r = 2$ in a situation where both $y_{1,t}$ and $y_{4,t}$ are missing. Therefore, we let $f_{1,t}$ and $f_{4,t}$ mean-revert. However, $f_{4,t}$ is effect-coded and therefore a function of the other three (free) idiosyncratic factors. It is therefore not immediate how the information in the remaining series should be spread out over the common factor $f_{c,t}$ and the two

**Figure 4: New score-driven regression weights with missing values**

$$
\begin{array}{c}
\overbrace{\phantom{\times e_{2,t}\times e_{3,t}\times e_{4,t}}}^{\times e_{2,t}\times e_{3,t}\times e_{4,t}} \quad \overbrace{\phantom{\text{sum}}}^{\text{sum}} \\
\begin{array}{l} f_{c,t}\,\{ \\ f_{1,t}\,\{ \\ f_{2,t}\,\{ \\ f_{3,t}\,\{ \end{array}
\left[\begin{array}{ccc|c}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 \\
0 & 0 & 0 & 0 \\
\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 \\
-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0
\end{array}\right]
\end{array}
$$

**(a)** Example 2

$$
\begin{array}{c}
\overbrace{\phantom{\times e_{2,t}\times e_{3,t}}}^{\times e_{2,t}\times e_{3,t}} \quad \overbrace{\phantom{\text{sum}}}^{\text{sum}} \\
\begin{array}{l} f_{c,t}\,\{ \\ f_{1,t}\,\{ \\ f_{2,t}\,\{ \\ f_{3,t}\,\{ \end{array}
\left[\begin{array}{cc|c}
\frac{1}{2} & \frac{1}{2} & 1 \\
0 & 0 & 0 \\
\frac{1}{2} & -\frac{1}{2} & 0 \\
-\frac{1}{2} & \frac{1}{2} & 0
\end{array}\right]
\end{array}
$$

**(b)** Example 3

*Notes*: An illustration of the application of $M_t$ through Proposition 1′, related to the examples given in the main text. It presents the matrix $(M_t F_t' S_t' S_t F_t M_t)^+ M_t F_t' S_t'$ in front of $\varepsilon_t^\dagger$ in Proposition 1′ for $\Sigma_t^\dagger = \mathrm{I}$.

idiosyncratic factors $f_{2,t}$ and $f_{3,t}$. Using Proposition 1′, we obtain

$$
S_t = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad K_t = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad K_t \widetilde{F}_t = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & -1 \end{bmatrix},
$$

$$
K_{\perp t}' = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad M_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix},
$$

and the resulting matrix $(M_t F_t' S_t' S_t F_t M_t)^+ M_t F_t' S_t'$ is given in Figure 4b. We see that indeed only $f_{c,t}$, $f_{2,t}$ and $f_{3,t}$ possibly make a non-zero step. Moreover, again the weights for the common factor sum to one, while those for the idiosyncratic factors sum to zero. The common factor thus takes all the information from the remaining observations. The rest of the information is used to filter the factors for which some direct information is available, while $f_{1,t}$ and $f_{4,t}$ revert to their long term means.

**Example 4** (Mean-reverting common factors)**.** As a final example, we consider the case where the idiosyncratic factors are large, while the common factor is small in magnitude. Consider the same setting as in Example 3, with $y_{1,t}$ and $y_{4,t}$ missing, but where we now choose to let the (in this case smaller) common factor $f_{c,t}$ mean-revert, as well as one of

the idiosyncratic factors, e.g. $f_{1,t}$. We obtain

$$S_t = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad K_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad K_t\widetilde{F}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \end{bmatrix},$$

$$K'_{\perp\, t} = \begin{bmatrix} 0 & 0 \\ \sqrt{\frac{1}{3}} & 0 \\ -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{6}} & \sqrt{\frac{1}{2}} \end{bmatrix}, \quad M_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}, \qquad [\star] = \begin{bmatrix} 0 & 0 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{11}$$

where the last matrix $[\star] = (M_t F_t' S_t' S_t F_t M_t)^+ M_t F_t' S_t'$ is comparable to that in Figure 4b. Now the score of the common factor $f_{c,t}$ in (11) puts the step for the first element of $f_t$, i.e., $f_{c,t}$, to zero such that it mean reverts. The rest of the entries in the last matrix in Eq. (11) ensure also $f_{4,t}$ mean reverts. This is done by transferring the information in $e^{\mu}_{2,t}$ and $e^{\mu}_{3,t}$ into $f_{2,t}$ and $f_{3,t}$, while $f_{1,t}$ is modified by the mean reversion of $f_{4,t}$ and the identifying restriction that the idiosyncratic factors should always sum to zero. This illustrates how different choices regarding the mean reversion of the time-varying factors in the model lead in a consistent way to the available information in the data at time $t$ to spread out over the non-mean-reverting factors. Later in the paper we illustrate how these choices lead to a useful imputation mechanism in a hierarchical factor model setting and to adequate imputation uncertainty bands.

The above examples illustrate that our methodology from Proposition 1′ can thus be used to let either idiosyncratic or common factors mean-revert, depending on their relative importance in the hierarchical set-up. The choice of which factors to prioritize is fully under control of the modeler through the specification of the matrix $K_t$. The approach accommodates customized versions of the 'setting-to-zero' method in a multivariate, hierarchical set-up. Our approach remains valid for other popular ways to encode dummy variables, such as the plain vanilla dummy encoder,[2] deviation/difference encoding, repeated effect-coding, and design matrices including exogenous regressors.

[2]When $F_t$ is a pure dummy coded matrix, the scaled-score of factors without any contributions are put automatically to zero by just using $S_t$ (and pseudo-inverses).

**Algorithm 1:** In-sample forecast bands

Initialize score-driven filters with some initial values $f_1^{\mu,(s)}, f_1^{\log\sigma^2,(s)} \in \mathbb{R}^k$.

**for** $s = 1, \ldots, S$ **do**

  Simulate $\widehat{\psi}^{(s)}$ using its asymptotic distribution

  **for** $t = 1, \ldots, n$ **do**

    **if** $y_t$ *is complete* **then**

      Compute $s_t^{\mu}$ and $s_t^{\log\sigma^2}$ using (6);

    **else if** $y_t$ *is partially observed* **then**

      Simulate $S_t^c y_t$ from $\mathrm{N}\left(S_t^c \mu_t^{(s)}, S_t^c \Sigma_t^{(s)} S_t^{c\prime}\right)$ and combine this with the observations $y_t^{\dagger}$ into the vector $y_t^{(s)}$. Use $y_t^{(s)}$ to update to $f_{t+1}^{\mu,(s)}$ and $f_{t+1}^{\log\sigma^2,(s)}$ using (10).

    **else**

      Simulate $y_t^{(s)}$ from $\mathrm{N}\left(\mu_t^{(s)}, \Sigma_t^{(s)}\right)$ and use it to update to $f_{t+1}^{\mu,(s)}$ and $f_{t+1}^{\log\sigma^2,(s)}$ using (10).

    **end**

  **end**

**end**

Use $\left\{f_t^{\mu,(s)}\right\}_{s=1}^{S}$ and $\left\{f_t^{\log\sigma^2,(s)}\right\}_{s=1}^{S}$ for $t = 1, \ldots, n$ to describe the in-sample uncertainty of the factors.

## 2.3 Imputing missing observations

The above treatment of the score through Eq. (10) in Proposition 1′ is adequate for obtaining point forecasts of $f_t^{\mu}$ and $f_t^{\log\sigma^2}$ and thus of the missing components in $y_t$. For interval forecasts, however, we resort to simulation-based techniques. A simulation algorithm for pure out-of-sample forecasting for score-driven models without missing value problems was introduced in Blasques et al. (2016). Here, we generalize the approach to the in-sample data imputation problem.

Let $S_t^c \in \mathbb{R}^{(p-p_t)\times p}$ denote a selection matrix such that $S_t^c y_t$ contains the *missing* elements of $y_t$ at time $t$, just as $S_t y_t$ contains the *non-missing* or observed elements. For imputing missing data and their uncertainty over the entire sample period, we consider the first time point $t$ where we have missing data. At time $t$, we have $S_t^c y_t | \mathcal{F}_{t-1} \sim \mathrm{N}(S_t^c \mu_t^{(s)}, S_t^c \Sigma_t^{(s)} S_t^{c\prime})$ with $\mu_t^{(s)} = \mu_t$ and $\Sigma_t^{(s)} = \Sigma_t$, such that we directly obtain imputation values and prediction intervals.[3] To propagate the uncertainty due to missing values

[3]The current conditioning scheme builds on the diagonal structure of $\Sigma_t^{(s)}$ and the fact that $S_t^c$ is a

forward, we simulate values for $S_t^c y_t$ from the above normal distribution. These are combined with the observed elements $y_t^\dagger$ into a simulated data point $y_t^{(s)}$, which is then used to obtain simulated updates of the time-varying parameters $\mu_{t+1}^{(s)}$ and $\Sigma_{t+1}^{(s)}$. With these simulated updated parameters, we proceed to time $t+1$, and so on, until the end of the sample. The process can be repeated many times, and pointwise simulation bands can be constructed for $S_t^c y_t$ and for the underlying time-varying parameters for all $t$. The algorithm is summarized as Algorithm 1. As in Blasques et al. (2016), we can extend the approach above to also include parameter uncertainty. We do so by embedding the above simulation algorithm into a second loop, where we simulate over the vector of static parameters using its asymptotic distribution. This step is described in the outer loop in Algorithm 1.

## 2.4 Filtering equations for score-driven updates and smoothing

The score-driven recursion in (4) lets the parameter vector $f_t$ evolve based on past observations only. The model specification in (1) is thereby observation-driven model. This implies that all time-varying parameters are one-step-ahead perfectly predictable and that, under correct specification, these time-varying parameters are completely revealed by past observations and the values of the static parameters. A better estimate of the time-varying parameter might be obtainable if we can use a filter based on both past and contemporaneous (and possibly even future) observations, i.e., filtering and smoothing as opposed to prediction could possibly provide a more accurate estimate of the time-varying parameter, particularly in the presence of missing data.

Buccheri et al. (2021) recently proposed a methodology for score-driven transition equations that uses more than just past data to estimate time-varying parameters. The propose both a score-driven filter that uses past and current observations and a score-driven smoother that uses all past, present, and future observations. Their approach builds on an analogy between the score-driven model and a linear Gaussian state-space model. In particular, a set of score-driven updates and smoothing recursions is derived through a general representation of the Kalman filter and Kalman smoother, respectively. The proposed recursions are approximate, but their numerical results show that their approach can be equally powerful as exact filtering techniques; see Buccheri et al. (2021)

selection matrix. In more general cases, we can sample from $S_t^c y_t | \mathcal{F}_{t-1}, y_t^\dagger$ with conditional mean $S_t^c \mu_t^{(s)} + S_t^c \Sigma_t^{(s)} S_t' \left(S_t \Sigma_t^{(s)} S_t'\right)^{-1} \left(y_t^\dagger - S_t \mu_t^{(s)}\right)$ and conditional variance $S_t^c \Sigma_t^{(s)} S_t^{c\prime} - S_t^c \Sigma_t^{(s)} S_t' \left(S_t \Sigma_t^{(s)} S_t'\right)^{-1} S_t \Sigma_t^{(s)} S_t^{c\prime}$.

for more details.

Filtering with the score-driven update and smoothing equations is straightforward as these recursions do not require additional elements than the ones already derived for the predictive filter. The score-drive contemporaneous filter recursion for $f_t$ is given by

$$f_{t|t} = f_t + B^{-1} A s_t, \tag{12}$$

for $t = 1, \ldots, n$. The score-driven smoother is given by

$$\begin{aligned} r_{t-1} &= s_t (B - A)' r_t \\ f_{t|n} &= f_t + B^{-1} A r_{t-1}, \end{aligned} \tag{13}$$

where $r_n = 0$, and $t = n, \ldots, 1$. Just like for the predictive filter in (4), updated and smoothed estimates of the location and log-variance factors, i.e., $f^{\mu}_{t|t}$, $f^{\mu}_{t|n}$, $f^{\log \sigma^2}_{t|t}$, and $f^{\log \sigma^2}_{t|n}$, are obtained by running the above recursions.

## 2.5 Estimation

The filtered parameters $f_t$ are a function of the data. This allows us to estimate the score-driven model's static parameter vector $\psi$ by maximum likelihood. In the presence of missing data, however, likelihood evaluations require careful attention. When no observations are available, the log-likelihood collapses to zero; see Blasques et al. (2020). If at least one observation in $y_t$ is not missing, i.e., $p_t \geq 1$, then $\ell_t$ is computed via the marginal likelihood function $\log p\left(y_t^{\dagger} \mid f_t, \mathcal{F}_{t-1}; \psi\right)$. Formally, the ML criterion function for multivariate score-driven models with missing observations is defined as

$$\begin{aligned} \widehat{\psi} &\coloneqq \arg\max_{\psi} \sum_{t=1}^{n} \ell_t(\psi), \\ \ell_t(\psi) &= \begin{cases} \log p\left(y_t^{\dagger} \mid f_t, \mathcal{F}_{t-1}; \psi\right), & \text{if } p_t \geq 1, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{14}$$

where $y_t^{\dagger} = S_t y_t$, for $t = 1, \ldots, n$. Blasques et al. (2020) refer to the log-likelihood in (14) as the pseudo log-likelihood function. During ML estimation, the regression weight matrix $M_t F_t' S_t' S_t F_t M_t$ in (10) for updating the log-variance factors can be pre-computed as they do not depend on any of the unknown parameters. They can therefore be fixed during optimization. This significantly reduces the run-time of the log-variance filter as taking (pseudo)-inverses at each $t$ and across each optimization iteration is avoided. Similar, further increases in numerical efficiency and stability are reached when one opts to filter the location factors using OLS rather than GLS type updates, i.e., by replacing

$\Sigma_t^\dagger$ by a unit matrix in the expression for $s_t^\mu$ in (10).

To carry out the ML estimation, one requires to first initialize the score-driven filter recursion in (4) at a fixed point $f_1^\mu, f_1^{\log\sigma^2} \in \mathbb{R}^k$. A typical choice for the initialization of observation-driven models is the unconditional mean of the filter itself, namely $f_1 = (\mathrm{I} - B)^{-1}\kappa$. Using the unconditional mean of the score-driven recursion as initial values does introduce vulnerability to outliers in the beginning of the filter. To mitigate this issue, one may fix initial values at some estimates inspired by the sample moment of the data. For example, consider the $p \times 1$ vector of sample averages over $\tau$ consecutive observations $\overline{y}_{1:\tau} = \left(\overline{y}_{1,1:\tau}, \ldots, \overline{y}_{p,1:\tau}\right)'$, where $\overline{y}_{i,1:\tau} = \tau^{-1}\sum_{t=1}^{\tau} y_{i,t}$, for $i = 1, \ldots, p$. This vector of averages can be used to analytically estimate the first values of the location factors, $\widehat{f_1^\mu}$, by solving

$$\widehat{f_1^\mu} = F_1^{-1}\overline{y}_{1:\tau},$$

for any (pseudo-)invertable factor loading matrix $F_1$. By setting $\tau$ to a small number, say $\tau = 10$, one can ensure that the initial values yields a prediction $\widehat{\mu}_1 = F_1\widehat{f_1^\mu}$ relatively close to $y_1$. Similarly, we can define a $p \times 1$ vector of prediction error variances over the first $\tau$ observations: $\widehat{s}^2_{1:\tau} = \left(\widehat{s}^2_{1,1:\tau}, \ldots, \widehat{s}^2_{p,1:\tau}\right)'$, where $\widehat{s}^2_{i,1:\tau} = \tau^{-1}\sum_{t=1}^{\tau}\left(y_{i,t} - \widehat{\mu}_{i,1}\right)^2$, which we can use to solve for the log-variance factors' initial value $\widehat{f_1^{\log\sigma^2}} = F_1^{-1}\log\left(\widehat{s}^2_{1:\tau}\right)$. Note that the above initialization procedure can also be carried out when there are missing values among the first $\tau$ observations. In such cases, one uses the number of available observations only to compute the means $\overline{y}_{i,1:\tau}$ and variances $\widehat{s}^2_{i,1:\tau}$. Factors that cannot be estimated in this way due to (a sequence) of missing observations in the beginning of the time-series can be simply initialized with their unconditional mean or by increasing the value of $\tau$.

Sample moments may also be used to reduce the number of static parameters by employing a mean and/or variance targeting approach. A targeting approach for score-driven models is established by setting the unconditional mean of the filter equal to its corresponding sample moment, which avoids numerical estimation of the conditional intercept $\kappa$. This approach can be easily implemented if the score-driven factors are one-to-one related with the time-varying moments. In our multi-factor location-scale model, this relationship includes the dependence on a (potentially) time-varying matrix $F_t$. For the location factors, the mean-targeting approach thereforetakes the following form

$$\kappa^\mu := \left(\mathrm{I}_k - B^\mu\right)\overline{F}^{+}\overline{y},$$

where $\overline{F} = n^{-1}\sum_{t=1}^{n} F_t$ is an (pseudo-)invertible matrix and $\overline{y} = \overline{y}_{1:n}$. A similar expres-

sion for $\kappa^{\log \sigma^2}$ can be obtained by replacing $\overline{y}$ with $\log(\widehat{s}^2_{1:n})$. However, our experience shows that variance-targeting is not recommended as $\widehat{s}^2$ includes both the variance of the noise $\varepsilon_t$ as well the variation of the mean $\mu_t$, whereis $\sigma_t$ should only reflect the former, not the latter. Only in cases where the mean $\mu_t$ is static, variance targeting seems applicable, e.g., in pure volatility models. In other cases, variance targeting in the way above almost invariably induces inconsistent estimation results.

# 3 Monte-Carlo Evidence

## 3.1 Design

In our Monte Carlo (MC) experiment we investigate the methodology's performance when forecasting missing values in structured data. We consider a $p = 4$ dimensional setting and two distinct scenarios for the missing values, namely MAR (missing-at-random) and NMAR (not-MAR). In the NMAR setting, we consider long patches of missing values at the start, middle, and end of the sample. Such patterns resemble the gaps observed empirically in Section 4. The length of the three patches of missings is set to 20% of the sample size each, such that only 40% of the data is observed. We consecutively introduce these missing values in only the first time-series, the first two series, the first three, and finally all four series. In this way, it becomes increasingly challenging to back out any information on the common and series-specific components via the factor structure.

In the MAR experiment, we use a standard i.i.d. Bernoulli indicator to put $y_{i,t}$ to missing with probability $\pi_{i,t}$. We consider $\pi_{i,t} \in \{25\%, 50\%, 75\%\}$ and again introduce the missing values in only the first series (scenario 1) up to all four series (scenario 4). A visual impression of the extent of the missing data problems is given in Figures B.1–B.4 in AppendixB. Out of the 16 scenarios in total, the challenges range from mild (25% missing at random in only one series) to severe (75% missing in all four series, either at random or in patches).

The DGP for each of the 16 scenarios is the score-driven model from Section 2 with one common factor, three free idiosyncratic factors, and one effect-coded idiosyncratic

factor. The corresponding factor loading matrix in the DGP is

$$F = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}. \tag{15}$$

The static parameters used in the DGP reflect the values found for the empirical data in Section 4 and are given by $\kappa^{\mu} = 0 \times \iota_4$, $A^{\mu} = \text{diag}\,(0.1, 0.025, 0.025, 0.025)$, $B^{\mu} = \text{diag}\,(0.98, 0.96, 0.96, 0.96)$, and $\kappa^{\log\sigma^2} := \kappa^{\mu}$, $A^{\log\sigma^2} := A^{\mu}$, $B^{\log\sigma^2} := B^{\mu}$. The data $y_t$ then include a strong location and scale common factor structure, combined with persistent idiosyncratic dynamics of a substantially smaller magnitude.

After simulating from this DGP and introducing the missing values, we fit the mean-variance GAS model to each of the 16 missingness scenarios, utilizing the scaled-scores in Proposition 1′. We perform $M = 500$ Monte-Carlo simulations and report the model's mean absolute error (MAE) for the in-sample forecasts of the mean and variance factors. We also compute MAEs for relevant functions thereof, such as tail risk measures like value-at-risk (VaR) and expected shortfall (ES).

To investigate the accuracy of the in-sample forecast error bands for missing values, we use Algorithm 1 including the parameter uncertainty outer loop, using $S = 500$ simulated parameter paths. From these, we construct pointwise forecast error bands for each simulation and for each missing value. We compute the coverage rate of these bands, i.e., how often the band contains the true missing value, where the latter is known in the simulation context. Coverage is computed across $M = 500$ Monte-Carlo simulations over an (evenly spread) grid of time points and we compare the simulated coverage percentage to its nominal level.

## 3.2 Results

The MAE results for the mean and variance factors and tail risk measures are provided in Tables 1 and 2, respectively. Supplementary results are provided in Appendix B. Panel A in Table 1 reports the MAE for the mean factors. The left-most column (labeled 0) serves as the benchmark and corresponds to the scenario without missing values. The next panel of four columns ($y_{1,t}$ to $y_{1:4,t}$) present the results for the four scenarios where patches of missings are present in only the first, the first two, the first three, or in all four series. We implemented shading in each of the panels to indicate when one or more

**Table 1: In-sample forecast performance for mean and log-variance factors**

| *Mechanism*: | | Missing Patches (NMAR) | | | | Randomly Missing (MAR) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missing %*: | | 60% | | | | 25% | | | | 50% | | | | 75% | | | |
| | 0 | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ |
| **Panel A: In-sample MAE for mean factors** | | | | | | | | | | | | | | | | | |
| $f_c$ | 0.014 | 0.059 | 0.096 | 0.152 | 0.216 | 0.028 | 0.039 | 0.049 | 0.057 | 0.039 | 0.056 | 0.074 | 0.086 | 0.049 | 0.073 | 0.100 | 0.118 |
| $f_1$ | 0.027 | 0.072 | 0.072 | 0.072 | 0.072 | 0.038 | 0.039 | 0.039 | 0.041 | 0.049 | 0.050 | 0.052 | 0.055 | 0.061 | 0.062 | 0.064 | 0.069 |
| $f_2$ | 0.026 | 0.035 | 0.071 | 0.071 | 0.071 | 0.027 | 0.038 | 0.040 | 0.041 | 0.029 | 0.050 | 0.051 | 0.054 | 0.032 | 0.061 | 0.063 | 0.068 |
| $f_3$ | 0.026 | 0.035 | 0.047 | 0.072 | 0.072 | 0.027 | 0.029 | 0.040 | 0.041 | 0.029 | 0.033 | 0.051 | 0.055 | 0.031 | 0.038 | 0.063 | 0.068 |
| $f_4$ | 0.033 | 0.038 | 0.048 | 0.072 | 0.073 | 0.035 | 0.037 | 0.039 | 0.047 | 0.037 | 0.040 | 0.045 | 0.060 | 0.037 | 0.044 | 0.055 | 0.074 |
| **Panel B: Coverage of 95% forecast error bands for mean factors** | | | | | | | | | | | | | | | | | |
| $f_c$ | 94.384 | 94.582 | 93.899 | 92.291 | 96.267 | 93.641 | 92.843 | 91.304 | 90.796 | 93.545 | 92.792 | 91.626 | 94.591 | 95.971 | 96.412 | 96.116 | 99.345 |
| $f_1$ | 97.541 | 97.453 | 97.782 | 98.054 | 98.016 | 97.723 | 97.864 | 98.080 | 98.383 | 98.225 | 98.519 | 98.999 | 99.467 | 99.258 | 99.568 | 99.702 | 99.910 |
| $f_2$ | 97.428 | 97.296 | 97.973 | 98.244 | 98.220 | 98.148 | 98.026 | 98.153 | 98.523 | 97.882 | 98.612 | 98.943 | 99.493 | 98.599 | 99.519 | 99.699 | 99.900 |
| $f_3$ | 97.414 | 97.384 | 96.998 | 98.090 | 98.086 | 97.872 | 97.961 | 98.066 | 98.447 | 98.007 | 98.064 | 98.948 | 99.438 | 98.651 | 98.862 | 99.665 | 99.901 |
| $f_4$ | 98.234 | 99.453 | 99.580 | 99.421 | 99.623 | 98.733 | 98.932 | 99.182 | 99.245 | 99.284 | 99.588 | 99.670 | 99.738 | 99.735 | 99.828 | 99.859 | 99.933 |
| **Panel C: In-sample MAE for log-variance factors** | | | | | | | | | | | | | | | | | |
| $f_c$ | 0.023 | 0.083 | 0.134 | 0.212 | 0.301 | 0.041 | 0.056 | 0.069 | 0.081 | 0.055 | 0.080 | 0.106 | 0.125 | 0.069 | 0.104 | 0.145 | 0.174 |
| $f_1$ | 0.037 | 0.101 | 0.101 | 0.101 | 0.102 | 0.054 | 0.054 | 0.056 | 0.058 | 0.069 | 0.070 | 0.072 | 0.077 | 0.086 | 0.087 | 0.089 | 0.095 |
| $f_2$ | 0.038 | 0.050 | 0.102 | 0.102 | 0.103 | 0.039 | 0.055 | 0.056 | 0.058 | 0.042 | 0.070 | 0.072 | 0.077 | 0.045 | 0.086 | 0.089 | 0.094 |
| $f_3$ | 0.037 | 0.049 | 0.066 | 0.101 | 0.102 | 0.040 | 0.042 | 0.056 | 0.058 | 0.043 | 0.048 | 0.073 | 0.077 | 0.045 | 0.055 | 0.089 | 0.095 |
| $f_4$ | 0.048 | 0.055 | 0.068 | 0.101 | 0.103 | 0.050 | 0.052 | 0.055 | 0.066 | 0.052 | 0.057 | 0.064 | 0.084 | 0.054 | 0.063 | 0.079 | 0.104 |

*Notes*: We explain Panel A, containing the MAE of the mean factors. The other panels are structured similarly. The first column contains MAE results without missing values (row-wise) for the common factor $f_{c,t}$ and the four idiosyncratic factors $f_{i,t}$, $i = 1, \ldots, 4$. The first panel contains the results for the NMAR set-up, with 60% of missing values in three patches (20% at the start, 20% in the middle, 20% at the end) in the first series only ($y_{1,t}$ column), the first two series ($y_{1:2,t}$ column), up to in all four series ($y_{1:4,t}$ column). The remaining three panels give similar results, but for the MAR case with $\pi = 25\%, 50\%, 75\%$ of missing values in one up to four of the series. Panel B gives the results for the coverage bands for the mean factors' in-sample forecast error bands. Panel C gives the MAE results for the log-variance factors. Results are based on $M = 500$ Monte-Carlo simulations. The forecast-error bands are based on another $S = 500$ parameter path simulations with each Monte-Carlo simulation. More results are found in Appendix B.

of the informative series for that factor are missing.

The occurrence of patches of missings affects the MAE in a predictable way. If only the first series has patches of missings ($y_{1,t}$ column), the forecast quality of the first idiosyncratic factor deteriorates substantially. Also the forecast quality of the common factor $f_{c,t}$ goes down, but much less so. Note that the forecast accuracy of $f_{1,t}$ does not really deteriorate further if more series start to have the same long patches of missing values. This is due to the fact that the key information to identify this factor, namely $y_{1,t}$, was already missing before. This contrasts with $f_{c,t}$, the MAE of which increases further if more series have (overlapping) patches of missings. We also see that as more series exhibit patches of missings, the MAE of each of the idiosyncratic components ($f_{i,t}$) for that series ($y_{1:i,t}$) goes up in turn as well.

To understand these outcomes, consider the setting where $y_{1,t}$ is not observed for some time, while all the other series are. In that case, the idiosyncratic factor $f_{1,t}$ cannot be

estimated at that point in time. Its scaled-score is then zero through our choice of the matrix $M_t$, such that the filtered factor mean reverts. Given that the other series are available, however, the common component $f_{c,t}$ can still be filtered from the data, albeit with a higher error. As more time-series (e.g., $y_{2,t}$) start having long stretches of missings at the same time as $y_{1,t}$, the MAE of $f_{1,t}$ remains unchanged as it was already mean reverting, while the error of $f_{2,t}$ increases. In addition, the MAE of $f_{c,t}$ also increases as less data is available to filter out the common level component.

Comparing the MAE for patches of missings (NMAR) versus randomly missing data (MAR), we see that patches result in substantially higher MAEs, except if the probability of missings is extreme (75%). Even then, though, the common component's ($f_{c,t}$) MAE is only about half the size of that of the MAR setting. This is due to the fact that the common factor model averages both in the cross-section and time-series direction, and can therefore exploit any data point to obtain information on the common level $f_{c,t}$ and possibly its variance. The MAR setting profits from this (with high probability), whereas the NMAR setting does not, as the missing patches are in the same locations for all series in our set-up.

Panel B in Table 1 reports the coverage rates of the simulation-based in-sample forecast error bands for the factors. We use a 95% confidence level. The first column shows the performance of the bands in case there are no missing values. We see that the common factor's coverage rate lies close to the nominal level of 95%, while the bands for the idiosyncratic factors are slightly conservative. The pattern persists if more and more missings are added to the data. For the NMAR case, all idiosyncratic coverage rates are somewhat too high at 97.5%–99.5%. The common factor, by contrast, has too low coverage rates, going down from 93.6% to 90.7% as more and more series have common patches of missings. This makes sense, as it then becomes increasingly difficult to track the $f_{c,t}$ component and its volatility. For MAR missing data patterns, we again find that the bands for the idiosyncratic factors are conservative with too high coverage ratios, while the bands for the common factor have coverage ratios that range between 90%–99%. This is likely due to the fact that under the MAR assumption, common components can be tracked well over time, even with (potentially) multiple values in the cross-section missing incidentally.

The results in Panel C confirm the pattern of Panel A for the log-variance factors. Also the results for coverage rates (not shown) are very much in line with those for the level factor.

**Table 2: In-sample forecast performance for tail risk measures**

| *Mechanism*: | | Missing Patches (NMAR) | | | | Randomly Missing (MAR) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missing %*: | | 60% | | | | 25% | | | | 50% | | | | 75% | | | |
| **Series** | 0 | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ |
| **Panel A: MAE of in-sample forecasts for VaR with $\alpha = 5\%$** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.048 | 0.196 | 0.227 | 0.291 | 0.353 | 0.091 | 0.096 | 0.103 | 0.109 | 0.126 | 0.137 | 0.150 | 0.160 | 0.162 | 0.179 | 0.204 | 0.217 |
| $y_2$ | 0.049 | 0.075 | 0.226 | 0.290 | 0.353 | 0.055 | 0.097 | 0.103 | 0.110 | 0.061 | 0.137 | 0.151 | 0.160 | 0.067 | 0.177 | 0.202 | 0.216 |
| $y_3$ | 0.048 | 0.074 | 0.103 | 0.289 | 0.353 | 0.055 | 0.062 | 0.103 | 0.109 | 0.062 | 0.075 | 0.150 | 0.160 | 0.068 | 0.087 | 0.204 | 0.218 |
| $y_4$ | 0.058 | 0.078 | 0.103 | 0.149 | 0.350 | 0.065 | 0.071 | 0.078 | 0.115 | 0.071 | 0.083 | 0.097 | 0.165 | 0.073 | 0.091 | 0.117 | 0.222 |
| **Panel B: MAE of in-sample forecasts for VaR with $\alpha = 1\%$** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.061 | 0.246 | 0.286 | 0.366 | 0.443 | 0.115 | 0.122 | 0.130 | 0.138 | 0.159 | 0.172 | 0.190 | 0.202 | 0.205 | 0.226 | 0.258 | 0.274 |
| $y_2$ | 0.062 | 0.095 | 0.286 | 0.366 | 0.444 | 0.070 | 0.122 | 0.130 | 0.139 | 0.078 | 0.173 | 0.190 | 0.202 | 0.085 | 0.224 | 0.256 | 0.273 |
| $y_3$ | 0.061 | 0.093 | 0.130 | 0.364 | 0.445 | 0.070 | 0.078 | 0.130 | 0.138 | 0.079 | 0.095 | 0.190 | 0.202 | 0.086 | 0.110 | 0.258 | 0.276 |
| $y_4$ | 0.074 | 0.099 | 0.130 | 0.188 | 0.440 | 0.082 | 0.090 | 0.099 | 0.145 | 0.089 | 0.105 | 0.123 | 0.208 | 0.093 | 0.115 | 0.148 | 0.281 |
| **Panel C: MAE of in-sample forecasts for ES with $\alpha = 1\%$** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.068 | 0.273 | 0.317 | 0.406 | 0.491 | 0.127 | 0.135 | 0.144 | 0.153 | 0.177 | 0.191 | 0.210 | 0.224 | 0.227 | 0.251 | 0.286 | 0.305 |
| $y_2$ | 0.069 | 0.105 | 0.317 | 0.407 | 0.492 | 0.077 | 0.136 | 0.145 | 0.154 | 0.086 | 0.192 | 0.211 | 0.225 | 0.095 | 0.249 | 0.284 | 0.303 |
| $y_3$ | 0.068 | 0.104 | 0.144 | 0.404 | 0.493 | 0.078 | 0.087 | 0.144 | 0.154 | 0.087 | 0.105 | 0.211 | 0.224 | 0.095 | 0.122 | 0.286 | 0.306 |
| $y_4$ | 0.082 | 0.110 | 0.145 | 0.209 | 0.488 | 0.091 | 0.100 | 0.110 | 0.161 | 0.099 | 0.116 | 0.137 | 0.231 | 0.103 | 0.128 | 0.165 | 0.312 |

*Notes*: We explain Panel A, containing the MAE of the 5% Value-at-Risk (VaR). The other panels are structured similarly. The first column contains MAE results without missing values (row-wise) for each of the series $y_{i,t}$, $i = 1, \ldots, 4$. The first panel contains the results for the NMAR set-up, with 60% of missing values in three patches (20% at the start, 20% in the middle, 20% at the end) in the first series only ($y_{1,t}$ column), the first two series ($y_{1:2,t}$ column), up to in all four series ($y_{1:4,t}$ column). The remaining three panels give similar results, but for the MAR case with $\pi = 25\%, 50\%, 75\%$ of missing values in one up to four of the series. Panels B and C give the results for the 1% VaR and Expected Shortfall (ES), respectively. Results are based on $M = 500$ Monte-Carlo simulations. More results are found in Appendix B.

Table 2 presents the MAE performance of tail risk measures under increasing missing value problems. This is particularly relevant for practice, where such risk measures still have to be reported to regulators, even though the data can be highly incomplete, similar to the current simulation set-up. Given we know the DGP as well as the model, we can analytically compute the Value-at-Risk (VaR) and Expected Shortfal (ES) for both the DGP and the model, and thus compute the MAE of these quantities. Note that the rows in the table now relate to the VaR and ES for each series rather than for the factors $f_{c,t}$ and $f_{i,t}$. The benchmark setting without missings (column 0) indicates MAE in the range of 0.05 to 0.08, with higher numbers for higher confidence levels (smaller tail probabilities $\alpha$), and slightly higher numbers for ES than for VaR. Unlike Table 1, we see that the MAE keeps increasing for all risk measures as more series contain missing values. This holds for both the MAR and NMAR case, though the MAE in the MAR case is about 40% smaller than in the NMAR case. The reason for this is that more series with missings increase the estimation error of the common factor. As both the common

and idiosyncratic factors play a role in the determination of VaR and ES, the MAE of these keeps increasing across columns. Understandably the biggest jump for each series $y_{i,t}$, however, is made if that series starts exhibiting missing values itself, so from column $y_{1:i-1,t}$ to $y_{1:i,t}$.

# 4 Empirically forecasting missing values for sparse CDS curves

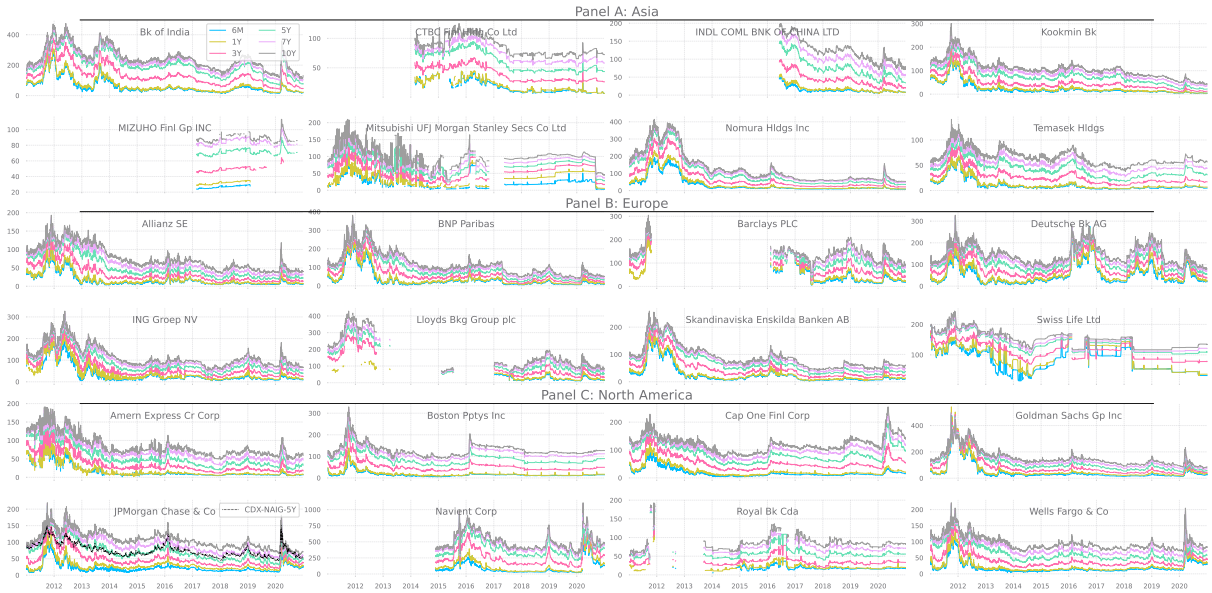
## 4.1 Data

In this section, we apply our missing value forecasting methodology for treating missing values in credit default swap (CDS) spread curves. CDS spreads play an important role in the financial industry and in supervision, as they reflect the markets' perceptions on the credit quality of large financial institutions, and therefore of the financial system itself. They can be regarded as the insurance premium one would pay for protection against default of that institution. CDSs are traded on financial markets at different maturities. Typically, CDS spreads are not always observed for every institution and for every maturity at every time point: CDSs are not always issued and/or traded for every underlying institution and maturity combination. This poses severe challenges for both supervisors and the institutions under their mandate. We investigate whether our methodology succeeds in capturing the salient features of these data and uncover missing values that we introduce into these empirical series.

We gather daily CDS spread data for different maturities for 24 financial institutions from Markit$^{\text{TM}}$. The institutions are equally distributed across the regions Asia (8), Europe (8), and North America (8). For each institution, we use six different CDS maturities, ranging from 6 months to 10 years. The sample spans a 10 year period from January 2, 2011 to December 31, 2020. This leads to a total of $p = 144$ time-series with $n = 2609$ days of observations. Figure 5 shows the observed CDS data.

The credit curves among all financial institutions show a similar pattern over time, especially during crises. For example, all series peak around the 2012 credit crunch and the more recent Covid-19 crisis. Note that the subplot of JPMorgan Chase & Co (JPM) also includes the North American CDX index for investment-grade companies with a 5Y tenor/maturity (CDXNAIG-5Y). We use this series later on as a potential proxy for the

**Figure 5: CDS term structures of financial institutions**




*Notes*: each panel of this figure shows 8 historical CDS term structures belonging to a specific region: Asia (top), Europe (middle), and North America (bottom). CDS maturities range from 6M–10Y. Spreads are measured in basis points (bps = 0.01%). The subplot for JPMorgan Chase & Co also includes (black dotted) the time-series of the North American CDS index of investment grade with a 5Y tenor (CDXNAIG-5Y).

common factor $f_{c,t}$.

Figure 5 also clearly shows that the data is far from complete. Many missing values occur, and the patterns of missings is generally not random, but rather in patches. Large gaps occur at the beginning and the middle of the sample. Table 3 reports for each institution the percentage of missing observations per maturity. Missing value percentages range between zero (the empty cells) to high percentages in the eighties. Challenges for missing value forecasting are thus substantial in this application. We note that the lower end of the curves (up to maturities of 3 years) have more missing observations than CDS series for higher tenors. The reason for this is that longer maturity CDSs are more liquid.

Table 3 also provides an overview of the rating frequencies for each firm over the 10 year sample period. We see that most institutions have an A rating, followed by banks with a BBB and AA rating. Only one bank has an AAA, respectively a B rating in the sample, implying that missing value forecasting will be harder for these series if a rating related factor plays a role.

**Table 3: Missing observations and rating decomposition per CDS name**

| Institution | Missing Values (%) | | | | | | Rating (% of time) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 6M | 1Y | 3Y | 5Y | 7Y | 10Y | AAA | AA | A | BBB | BB |
| **Panel A: Asia** | | | | | | | | | | | |
| Bk of India | 0.038 | | | | 0.077 | 0.077 | | | | 95.4 | 4.6 |
| CTBC Finl Hldg Co Ltd | 42.545 | 36.374 | 31.430 | 31.430 | 31.430 | 31.430 | | | 48.1 | 51.9 | |
| INDL COML BNK OF CHINA LTD | 54.274 | 54.274 | 54.274 | 54.274 | 54.274 | 54.274 | | 0.1 | 99.9 | | |
| Kookmin Bk | 0.23 | | | | | | | 0.1 | 99.9 | | |
| MIZUHO Finl Gp INC | 81.219 | 81.219 | 76.121 | 67.190 | 67.190 | 71.867 | | 0.1 | 99.9 | | |
| Mitsubishi UFJ Morgan Stanley Secs Co Ltd | 24.607 | 12.265 | 8.816 | 8.049 | 14.795 | 10.310 | | 0.8 | 99.2 | | |
| Nomura Hldgs Inc | | | | | | | | | 38.1 | 61.9 | |
| Temasek Hldgs | 0.652 | | | | | | 100 | | | | |
| **Panel B: Europe** | | | | | | | | | | | |
| Allianz SE | | | | | | | 0.1 | 99.9 | | | |
| BNP Paribas | | | | | | | | 11.0 | 89.0 | | |
| Barclays PLC | 58.068 | 49.828 | 48.524 | 45.343 | 45.381 | 45.381 | | 6.401 | 44.845 | 48.754 | |
| Deutsche Bk AG | | | | | | | | 1.4 | 72.8 | 25.8 | |
| ING Groep NV | 15.102 | 3.795 | 0.345 | 0.077 | 0.077 | 0.153 | | | 100 | | |
| Lloyds Bkg Group plc | 60.636 | 55.424 | 44.845 | 37.792 | 38.865 | 40.82 | | 22.3 | 77.7 | | |
| Skandinaviska Enskilda Banken AB | 0.077 | 0.077 | | | 0.498 | 0.498 | 0.1 | 38.2 | 61.7 | | |
| Swiss Life Ltd | 5.788 | 9.467 | 0.038 | 0.038 | 0.038 | 0.038 | | 0.1 | 90.9 | 9.0 | |
| **Panel C: North America** | | | | | | | | | | | |
| Amern Express Cr Corp | 16.098 | 3.680 | 0.613 | | | | | 0.1 | 99.9 | | |
| Boston Pptys Inc | 1.035 | 0.115 | | | | | | | 15.3 | 84.7 | |
| Cap One Finl Corp | 0.038 | 0.038 | 0.038 | 0.038 | 0.038 | 0.153 | | | 0.1 | 99.9 | |
| Goldman Sachs Gp Inc | 0.077 | 0.038 | | | | | | | 99.9 | 0.1 | |
| JPMorgan Chase & Co | 0.038 | 0.038 | | | | | | 8.1 | 91.9 | | |
| Navient Corp | 39.095 | 39.057 | 38.904 | 38.904 | 39.785 | 39.785 | | | | | 100 |
| Royal Bk Cda | 38.367 | 21.349 | 18.666 | 17.631 | 22.231 | 22.422 | 0.1 | 99.4 | 0.5 | | |
| Wells Fargo & Co | 0.077 | 0.038 | | | | | | 8.7 | 91.3 | | |

*Notes*: this table reports the percentage of missing values for each CDS curves (over the 10 year period: January 2, 2011 - December 31, 2020) for each of the 24 financial institutions in our data set. The historical rating frequencies for each firm are also reported. Empty table cells indicate that the percentage of missing values is zero.

## 4.2 Single name CDS curve imputation

In our first application, we evaluate the quality of our missing value forecasting methodology to CDS curves of one firm only. For this, we select the upper-end of the JPM CDS curve (3Y-10Y), because it has no missing values. We thus have a four-dimensional system, each maturity corresponding to a time series $y_{i,t}$, similar to Section 3. As in our earlier Monte Carlo experiment, we generate artificial missing data by putting specific (isolated or patches of) JPM observations to missing using either a MAR or NMAR scheme. The objective is to assess the performance of both the point and interval forecasts of the missing values.

We add four features to the current empirical analysis compared to the earlier simulation design from Section 3. First, next to the location-scale model set-up with one common-factor $f_{c,t}$ and four idiosyncratic factors $f_{i,t}$ as in Eq. 15, we also consider a linear Gaussian state-space model with the exact same design matrix for location factors.

A multi-factor state-space framework, in combination with Kalman filter, offers is often used as device for missing value imputation, see for instance Jungbacker et al. (2011), Stock and Watson (2016) and Durbin and Koopman (2012). The state-space model serves as a natural benchmark for our observation-driven approach. We refer to the classical linear Gaussian state-space with independently time-varying location (state) factors as the `SS-l` model. Additionally, we consider a sore-driven model with just time-varying mean factors (`GAS-l`) to study the added value of including dynamic variances in the imputation model (`GAS-l-s`). Second, we also consider the aforementioned GAS location and location-scale model with a static equi-correlation covariance matrix. CDS tenors display strong comovements across the term-structure in levels, but historically also exhibit a strong correlation structure between their returns (between 85%-95%). The GAS models with an equi-correlation covarinace matrix are denoted as `GAS-l-ec` and `GAS-l-s-ec`.

Third, we also consider a setting where we add the North American CDX index for investment grade companies at a 5 year maturity to the four JPM time-series. The index is liquidly traded and would typically be a prime candidate in practice to anchor the common component of the CDS curves. We never set the index values to missing, such that there is always at least one observation in the cross-section in this setting. We consider this set-up for the GAS location-scale model and refer to this specification as the `GAS-l-s-CDX` model. As a result, it is always possible to estimate the time-varying common location and scale factors.

Fourth, we cross-validate the result across the order in which missing values are added to the four JPM series. This gives a total of $4! = 24$ different orders for adding missing values to the four series. The patterns for generating MAR and NMAR missing data are the same as in in Section 3. For a particular scenario, the data that is (randomly) assigned to be missing is not used for estimating the model parameters. The MAE is then computed between the observed (but unused) value $y_t$ (in basis point (bp) levels) and its point forecast $\widehat{\mu}_t$. We only measure the MAE and the accuracy of the 95% forecast-error bands over the observations that were assigned to be missing. The bands for the state-space model are generated based on the sum of the predicted state-covariance by the Kalman filter and diagonal measurement variance matrices. The bands for the sore-driven models are simulated whilst accounting for both parameter and innovation uncertainty based on $S = 500$ MC paths, as done in Section 3.

Tables 4 and 5 report the MAE and coverage results, respectively, as the average over $4! = 24$ cross-validation combinations that introduce the missing values to the four series

**Table 4: In-sample point forecast accuracy (MAFE) for CDS JPM term structure**

| *Mechanism*: | | Missing Patches (NMAR) | | | | | Randomly Missing (MAR) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missing %*: | | 60% | | | | | 25% | | | | 50% | | | | 75% | | | |
| | | # time-series in system with missing values | | | | | | | | | | | | | | | | |
| Model | Tenor | 0 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| SS-l | 3Y | 1.641 | 8.067 | 7.981 | 9.559 | 19.341 | 1.726 | 1.786 | 1.714 | 1.709 | 1.951 | 1.847 | 1.833 | 1.864 | 2.259 | 2.042 | 2.059 | 2.344 |
| | 5Y | 2.072 | 6.097 | 6.395 | 8.510 | 23.198 | 2.113 | 2.040 | 2.083 | 2.099 | 2.171 | 2.183 | 2.168 | 2.215 | 2.240 | 2.214 | 2.251 | 2.642 |
| | 7Y | 2.323 | 4.668 | 5.370 | 7.012 | 22.973 | 2.336 | 2.326 | 2.305 | 2.337 | 2.249 | 2.359 | 2.352 | 2.439 | 2.363 | 2.404 | 2.444 | 2.870 |
| | 10Y | 2.620 | 11.143 | 7.809 | 9.104 | 23.693 | 2.578 | 2.605 | 2.569 | 2.599 | 2.587 | 2.613 | 2.611 | 2.689 | 2.670 | 2.687 | 2.698 | 3.127 |
| GAS-l | 3Y | 1.652 | 9.525 | 9.418 | 10.045 | 19.165 | 1.705 | 1.729 | 1.714 | 1.716 | 1.831 | 1.825 | 1.851 | 1.901 | 2.092 | 2.095 | 2.100 | 2.456 |
| | 5Y | 2.052 | 4.726 | 5.534 | 7.091 | 22.618 | 2.012 | 2.078 | 2.066 | 2.092 | 2.125 | 2.118 | 2.152 | 2.222 | 2.189 | 2.226 | 2.288 | 2.678 |
| | 7Y | 2.326 | 4.350 | 5.361 | 6.307 | 22.262 | 2.483 | 2.371 | 2.305 | 2.331 | 2.368 | 2.337 | 2.399 | 2.453 | 2.395 | 2.400 | 2.462 | 2.879 |
| | 10Y | 2.546 | 11.733 | 9.967 | 8.972 | 22.677 | 2.586 | 2.551 | 2.569 | 2.556 | 2.649 | 2.582 | 2.631 | 2.694 | 2.643 | 2.658 | 2.687 | 3.131 |
| GAS-l-ec | 3Y | 1.829 | 10.471 | 10.334 | 10.688 | 19.101 | 1.870 | 1.832 | 1.832 | 1.825 | 1.968 | 1.955 | 1.938 | 1.963 | 2.129 | 2.122 | 2.106 | 2.471 |
| | 5Y | 2.220 | 5.436 | 5.821 | 7.716 | 22.459 | 2.208 | 2.178 | 2.153 | 2.170 | 2.198 | 2.247 | 2.232 | 2.261 | 2.304 | 2.280 | 2.297 | 2.689 |
| | 7Y | 2.471 | 4.615 | 5.498 | 7.628 | 22.158 | 2.409 | 2.451 | 2.382 | 2.389 | 2.465 | 2.425 | 2.415 | 2.472 | 2.466 | 2.463 | 2.475 | 2.876 |
| | 10Y | 2.693 | 7.364 | 7.954 | 8.823 | 22.640 | 2.670 | 2.652 | 2.658 | 2.624 | 2.665 | 2.674 | 2.684 | 2.713 | 2.704 | 2.713 | 2.726 | 3.142 |
| GAS-l-s | 3Y | 1.648 | 8.438 | 8.530 | 8.796 | 19.038 | 1.667 | 1.762 | 1.719 | 1.753 | 1.824 | 1.831 | 1.829 | 1.890 | 2.064 | 2.065 | 2.089 | 2.426 |
| | 5Y | 2.048 | 5.795 | 5.847 | 7.653 | 22.712 | 2.143 | 2.130 | 2.066 | 2.124 | 2.110 | 2.134 | 2.124 | 2.215 | 2.190 | 2.206 | 2.259 | 2.667 |
| | 7Y | 2.329 | 4.043 | 4.722 | 6.228 | 22.436 | 2.345 | 2.361 | 2.341 | 2.372 | 2.342 | 2.326 | 2.363 | 2.453 | 2.370 | 2.407 | 2.471 | 2.901 |
| | 10Y | 2.544 | 9.072 | 9.371 | 8.230 | 22.782 | 2.524 | 2.646 | 2.529 | 2.602 | 2.568 | 2.610 | 2.562 | 2.678 | 2.593 | 2.627 | 2.700 | 3.137 |
| GAS-l-s-ec | 3Y | 2.013 | 8.537 | 8.386 | 8.474 | 19.094 | 2.004 | 2.037 | 1.939 | 1.910 | 1.978 | 1.948 | 1.909 | 1.931 | 2.194 | 2.143 | 2.111 | 2.447 |
| | 5Y | 2.524 | 4.493 | 4.551 | 6.022 | 22.648 | 2.555 | 2.551 | 2.401 | 2.345 | 2.511 | 2.436 | 2.263 | 2.271 | 2.568 | 2.374 | 2.304 | 2.683 |
| | 7Y | 2.824 | 4.582 | 4.651 | 5.884 | 22.382 | 2.823 | 2.737 | 2.602 | 2.579 | 2.758 | 2.555 | 2.475 | 2.490 | 2.711 | 2.539 | 2.484 | 2.884 |
| | 10Y | 3.063 | 8.159 | 7.738 | 7.977 | 22.689 | 3.027 | 3.090 | 2.918 | 2.815 | 2.959 | 2.831 | 2.701 | 2.729 | 2.918 | 2.786 | 2.727 | 3.120 |
| GAS-l-s-CDX | 3Y | 1.615 | 7.469 | 7.562 | 8.306 | 13.652 | 1.675 | 1.671 | 1.656 | 1.649 | 1.728 | 1.714 | 1.735 | 1.788 | 1.938 | 1.942 | 1.991 | 2.171 |
| | 5Y | 2.005 | 6.591 | 7.139 | 8.350 | 12.119 | 2.033 | 1.988 | 2.016 | 2.025 | 2.048 | 2.091 | 2.080 | 2.107 | 2.134 | 2.175 | 2.262 | 2.440 |
| | 7Y | 2.310 | 5.046 | 5.606 | 6.948 | 10.882 | 2.308 | 2.313 | 2.293 | 2.314 | 2.335 | 2.334 | 2.357 | 2.410 | 2.406 | 2.463 | 2.517 | 2.714 |
| | 10Y | 2.562 | 8.361 | 8.537 | 9.038 | 11.202 | 2.637 | 2.520 | 2.575 | 2.577 | 2.606 | 2.628 | 2.648 | 2.662 | 2.709 | 2.747 | 2.813 | 2.982 |

*Notes*: this table reports the empirical cross validation-based Monte Carlo imputation performance results for various forecasting models on JPM's CDS time-series (tenors run from 3Y till 10Y). The performance results are reported for different missing data patterns (i.e., combinations of missing data mechanisms and number of sparse time-series in the system), given three different specifications of the factor loading matrix. For each missingness scenario, the MAE in basis points (bps) between the observed (but unused) value $y_t$ (in basis point (bp) levels) and its point forecast $\widehat{\mu}_t$ is computed at sparse entries only. The `SS-l` label refers to a linear Gaussian state-space `SS` model with just time-varying location `l` factors only. Similarly, score-driven models including dynamic location (and scale) factors `GAS-l(-s)`. GAS models including a static equi-correlation include the term '`-ec`' in their label. The `GAS-l-s-CDX` model refers to a time-series system in which the 5Y CDX is also included as an additional instrument. The hierarchical design matrix is the same for all models, and always includes one common-factor $f_{c,t}$ and one idiosyncratic factor $f_{i,t}$ for each time-series $y_{i,t}$ in the system, as in Eq. 15. The estimation period is January 2, 2011 - December 31, 2020 ($n = 2609$).

in different orders. The first column (labeled 0) is again the benchmark scenario without any missing values and low MAE in the order of 1.5bp to 2.5bp. Dynamic multivariate Gaussian models thus fits the data well if there are no missing values, regardless the choice of parameter-driven vs. observation-driven framework, covariance matrix $\Sigma_t$ or design of the factor loading matrix $F$. The MAE for `SS-l` model in the top rows of Table 4 gradually increases across the columns per missing value design. This is as expected, because factors (and thus the data) are less accurately tracked as number of missing values within a cross-section increase. The effect is particularly bad for patches

**Table 5: In-sample coverage of forecast bands for CDS JPM term structure**

| *Mechanism*: | | Missing Patches (NMAR) | | | | | Randomly Missing (MAR) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missing %*: | | 60% | | | | | 25% | | | | 50% | | | | 75% | | | |
| | | # time-series in system with missing values | | | | | | | | | | | | | | | | |
| Model | Tenor | 0 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| SS-l | 3Y | 96.646 | 87.045 | 92.500 | 95.669 | 51.284 | 96.958 | 96.357 | 96.685 | 96.562 | 97.035 | 96.830 | 96.860 | 96.530 | 96.762 | 96.707 | 96.827 | 96.568 |
| | 5Y | 95.056 | 86.968 | 86.853 | 91.964 | 48.256 | 95.450 | 95.360 | 95.373 | 94.868 | 95.795 | 94.817 | 95.190 | 95.041 | 95.688 | 94.564 | 95.140 | 95.288 |
| | 7Y | 93.965 | 94.097 | 91.632 | 92.871 | 48.754 | 93.686 | 93.852 | 94.300 | 94.025 | 94.185 | 93.520 | 94.018 | 94.175 | 93.788 | 93.465 | 94.135 | 94.572 |
| | 10Y | 93.219 | 85.588 | 92.922 | 94.187 | 48.831 | 93.354 | 93.188 | 93.678 | 93.488 | 93.916 | 93.494 | 93.852 | 93.881 | 93.797 | 93.533 | 94.055 | 94.297 |
| GAS-l | 3Y | 95.362 | 84.354 | 84.237 | 86.790 | 88.351 | 90.516 | 90.222 | 90.423 | 90.523 | 91.973 | 91.309 | 91.172 | 91.651 | 93.899 | 93.473 | 93.286 | 94.621 |
| | 5Y | 95.516 | 68.209 | 68.422 | 61.983 | 78.470 | 89.852 | 87.871 | 87.764 | 87.781 | 90.146 | 90.184 | 89.617 | 90.270 | 93.448 | 92.003 | 91.965 | 93.328 |
| | 7Y | 95.171 | 74.048 | 74.319 | 70.758 | 80.674 | 85.097 | 85.442 | 86.273 | 85.941 | 88.216 | 88.433 | 88.037 | 88.545 | 92.655 | 92.293 | 92.044 | 93.298 |
| | 10Y | 95.401 | 95.948 | 97.208 | 88.779 | 95.950 | 84.714 | 84.535 | 85.003 | 85.155 | 87.027 | 87.685 | 87.781 | 88.270 | 91.658 | 92.123 | 92.797 | 94.035 |
| GAS-l-ec | 3Y | 95.477 | 97.213 | 97.516 | 96.114 | 84.703 | 93.226 | 93.903 | 93.857 | 93.744 | 93.954 | 93.986 | 93.827 | 94.015 | 95.169 | 94.866 | 94.569 | 95.433 |
| | 5Y | 94.059 | 98.086 | 98.107 | 96.937 | 82.982 | 91.334 | 91.846 | 91.752 | 91.705 | 92.472 | 92.069 | 92.178 | 92.769 | 93.090 | 93.452 | 93.848 | 94.954 |
| | 7Y | 93.369 | 98.415 | 98.362 | 98.018 | 83.703 | 89.775 | 90.031 | 90.363 | 90.472 | 90.849 | 90.976 | 91.074 | 91.715 | 92.323 | 92.698 | 93.200 | 94.949 |
| | 10Y | 93.292 | 98.000 | 98.224 | 97.951 | 88.582 | 88.548 | 88.625 | 88.931 | 89.360 | 89.213 | 89.986 | 89.924 | 91.037 | 92.169 | 92.757 | 93.274 | 94.817 |
| GAS-l-s | 3Y | 95.669 | 98.022 | 93.874 | 87.971 | 85.601 | 97.418 | 97.009 | 97.367 | 95.565 | 98.543 | 98.000 | 97.371 | 99.038 | 99.455 | 96.690 | 99.690 | 97.171 |
| | 5Y | 94.864 | 97.618 | 79.270 | 81.997 | 78.257 | 96.677 | 96.472 | 96.924 | 95.412 | 97.482 | 97.674 | 97.912 | 98.438 | 97.733 | 98.999 | 99.412 | 96.822 |
| | 7Y | 94.941 | 95.714 | 83.285 | 72.850 | 81.820 | 94.530 | 95.437 | 95.348 | 94.204 | 95.987 | 96.696 | 95.590 | 98.153 | 98.364 | 96.242 | 99.321 | 96.485 |
| | 10Y | 94.787 | 97.224 | 94.740 | 96.696 | 95.448 | 96.754 | 96.421 | 97.069 | 95.942 | 97.572 | 97.795 | 96.685 | 98.524 | 99.156 | 99.676 | 99.719 | 96.947 |
| GAS-l-s-ec | 3Y | 93.714 | 99.936 | 99.936 | 99.883 | 78.965 | 99.284 | 99.604 | 99.855 | 99.930 | 99.923 | 99.987 | 99.983 | 99.987 | 99.974 | 99.983 | 99.980 | 93.024 |
| | 5Y | 91.989 | 100.000 | 100.000 | 99.986 | 78.451 | 99.080 | 99.642 | 99.864 | 99.917 | 99.859 | 99.981 | 99.987 | 99.990 | 99.983 | 99.983 | 99.994 | 93.096 |
| | 7Y | 91.376 | 99.745 | 99.750 | 99.649 | 81.653 | 98.313 | 99.463 | 99.770 | 99.898 | 99.796 | 99.962 | 99.983 | 99.997 | 100.000 | 100.000 | 99.997 | 93.043 |
| | 10Y | 91.606 | 99.468 | 99.553 | 99.170 | 88.444 | 98.569 | 99.565 | 99.864 | 99.904 | 99.898 | 99.987 | 99.983 | 99.997 | 100.000 | 100.000 | 100.000 | 93.134 |
| GAS-l-s-CDX | 3Y | 95.631 | 81.164 | 79.488 | 78.877 | 75.731 | 97.137 | 97.022 | 97.231 | 97.342 | 98.837 | 99.054 | 99.037 | 98.882 | 99.625 | 99.868 | 94.209 | 95.199 |
| | 5Y | 95.056 | 95.607 | 95.044 | 88.616 | 82.746 | 97.009 | 96.856 | 96.788 | 97.220 | 98.185 | 98.760 | 98.730 | 98.690 | 99.327 | 99.813 | 94.200 | 95.005 |
| | 7Y | 94.327 | 95.193 | 90.491 | 77.714 | 67.233 | 96.472 | 96.766 | 96.779 | 96.766 | 97.610 | 98.025 | 98.100 | 98.441 | 99.753 | 99.663 | 93.842 | 94.964 |
| | 10Y | 94.902 | 97.883 | 94.682 | 91.392 | 76.143 | 94.734 | 95.450 | 95.569 | 96.102 | 95.642 | 97.003 | 97.661 | 98.268 | 97.751 | 98.786 | 99.338 | 94.545 |

*Notes*: this table reports the empirical cross validation-based Monte Carlo performance of the forecast error bands for various forecasting models on JPM's CDS time-series (tenors run from 3Y till 10Y). The coverage ratios (%) of the imputed values are reported for the 95% confidence level. The bands for the state-space model are generated based on the sum of the predicted state-covariance by the Kalman filter and diagonal measurement variance matrices. The bands for the sore-driven models are simulated whilst accounting for both parameter and innovation uncertainty based on $S = 500$ MC paths, as done in Section 3. See caption of Table 4 for more details.

of missing values (NMAR), where the MAE increases to about its tenfold. For the MAR scenario, the MAE roughly doubles if 75% of the data are missing. In fact, the MAR MAEs remain highly robust and only increase slowly, if at all. This also holds for the GAS models. The performance of score-driven models is similar in terms of in-sample point forecasts to the state-space model, regardless the choice of the covariance matrix $\Sigma_t$ (e.g., with or without dynamic (co)variances). This not only supports the findings in Koopman et al. (2016) indicating that parameter-and observation-driven models have similar point forecast accuracy. the robust forecast performance across state-space and score-driven models, but also the decay in dealing with missing.

However, in the last rows of this table, we see the MAE performance if there is a liquid anchoring time series that is always observed. In that case, the MAE is only about half the size of that of the analysis without the CDX index in the NMAR setting, even when all four series show the same patches of missings. The index thus helps to anchor

the common component, but does so quite imperfectly. As soon as one of the series shows no patches of missings, the performance of the effect coded $F$ with or without the CDX index is about the same. The specific JPM components vis-á-vis the index thus play a non-negligible role. As long as there are some observations left in the cross-section, the common level can be determined with some accuracy by pooling cross-sectional and time-series properties of the data. Thus, the common factor provides a good anchoring value for imputing the missing value. Besides, the robust empirical point forecast performance between SS and GAS also highlights the adequacy of our proposed treatment for filtering with missing values in hierarchical score-driven models.

Moreover, the effect on the coverage ratios of the forecast error-bands seems to differ per imputation model and missingness scenario, as shown in Table 5. The `SS-l` model's coverage ratios are fairly close to the 95% confidence level when at least 1 observation in the cross-section is present for NMAR scenarios. When all observations are NMAR, the bands of `SS-l` model are too tight. On the other hand, when all four series show the same patches of missing, the score-driven density forecasts are much closer to the 95% confidence level. This clearly highlights effectiveness of density forecasts of the proposed simulation algorithm in extreme missingness scenarios. The density forecasts seem to further improve if a correlations are also fitted in $\Sigma_t$, even though the point forecast accuracy remains unaffected between `GAS-l` and `GAS-l-ec` models. The simulation bands are in generally closer to 95% across all NMAR scenarios when either correlations and dynamic volatilities are modeled. At the same time, these bands can also quickly become too conservative as the influence of of parameter uncertainty increases when the number of static parameters. On the contrary, the analytical bands of the Kalman filter perform consistently across all NMAR scenarios. Note that prediction intervals for NMAR scenarios based on just the fitted score-driven means and variances perform on par with the state-space model, but is not shown here. All in all, the point forecasts and simulation bands of the score-driven location-scale models perform well enough to be used in practice for large scale imputation, as we do in the following subsection.

## 4.3 Global missing value imputation

The previous results for one firm suggest that additional sources of information can significantly boost the imputation accuracy our dynamic factor model, especially when $y_t$ has patches of missing values. In our second application, we therefore extend the

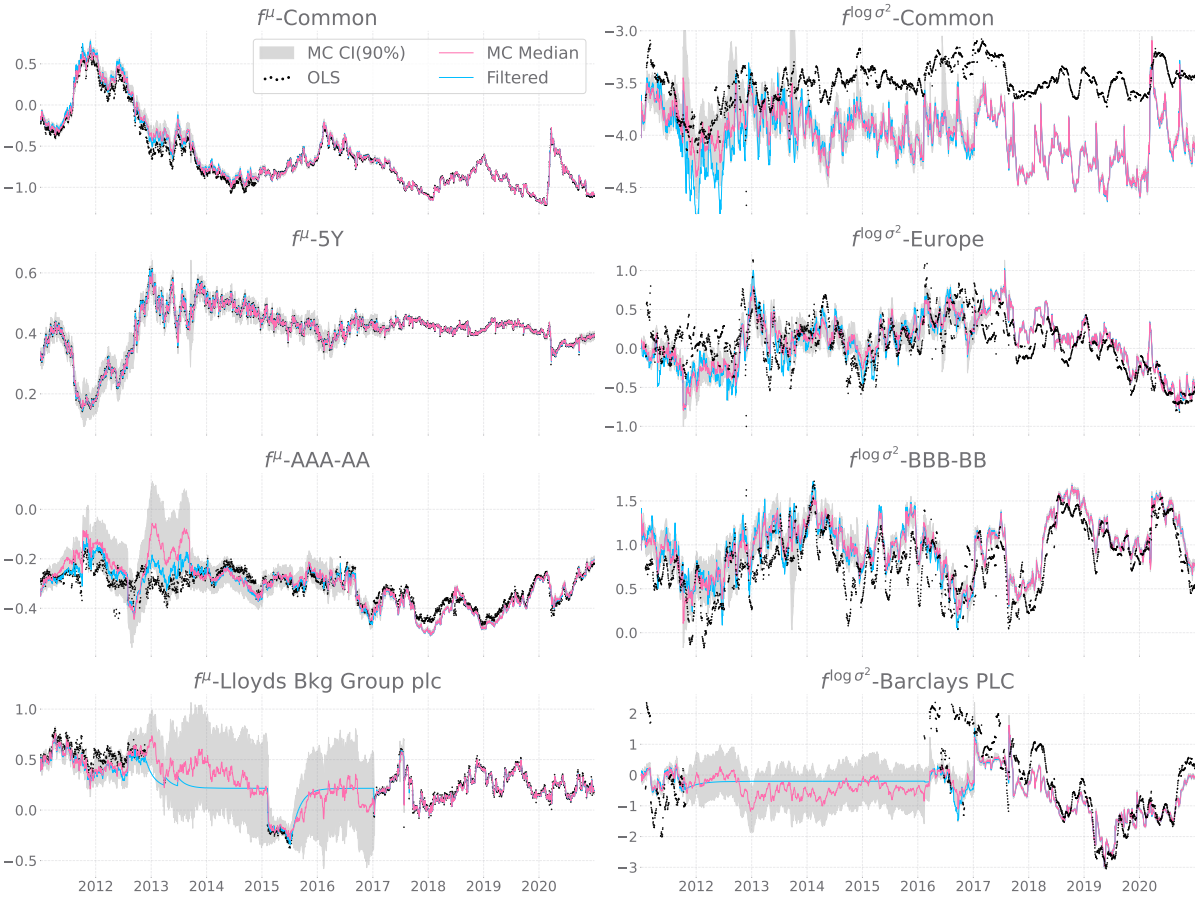
previous model to a setting with different firms in different geographical regions. The aim is to show how the model can be applied for imputing large gaps in high-dimensional panels by utilizing cross-sectional and time-series information in a structured way. In particular, we are interested in what hierarchical structure is most suited to construct synthetic CDS series, i.e., a CDS series for a firm–maturity combination for which no data is actively traded. For this analysis, we consider a number of factor specifications of increasing complexity for our full sample of $p = 144$ CDS time-series.

Table 6 summarizes the different specifications we consider. Note that the global panel of CDS curves contains substantial missing data patches; see Figure 5 and Table 3. To impute these missing values, some common factor structure is needed. In the simplest set-up, we only impose once common factor (Common) for all 144 time-series. We then subsequently augment this specification by adding factors for the different tenors (or maturities), regions, rating classes, underlying firms, and finally tenor-rating combinations. This yields a total of 6 models of increasing complexity. Adding further factors did not substantially increase the predictive performance of the models. The increase in performance of a one-factor model with MAE of about 46 basis points, to a model with all factors and and MAE of about 10 basis points, is however substantial. The last two models have a 3 level hierarchy: a common factor (level 1), followed by tenor, region, rating, and possibly tenor-rating factors (level 2), and firm-specific factors (level 3).

To limit the number of free static parameters that need to be estimated, we proceed as follows. First, we use the approach of Subsection 2.5 to target $\kappa^\mu$. Second, we use single scalar pooled persistence parameters $B^\mu$ and $B^{\log \sigma^2}$, respectively. Each of the factors can thus move along their own dynamic path, but they do so with the same persistence. Given the high persistence of all factors, this does not come at a substantial loss in fit. The parameters $A^\mu$ and $A^{\log \sigma^2}$ are still different for every factor. After a preliminary analysis, we also reduce the dimension of $F_t$ restricting the tenor-rating interactions to the interactions between all rating categories and the lowest three tenors. Finally, we pool the rating categories AAA and AA into a single rating class AAA/AA. The categories BBB and BB are also pooled into a single rating class (BBB/BB). The final two models then have a total of 74 (62 free, 12 effect coded) and 92 (74 free, 18 effect coded) factors, respectively. As reported in Table 6, their corresponding number of parameters is 103 and 127, respectively, which is perfectly doable given the size of the sample.

Figure 6 shows an example of eight fitted location and log-variance factors and their MC forecast error bands. Both the filtered estimate (blue) and the MC median (pink) are

**Figure 6: Example of estimated and simulated factors by largest hierarchical density model**




*Notes*: left hand-side of this plot shows 4 filtered and simulated location factors by the hierarchical density model with a total of 92 factors Similarly, the right hand-side depicts an example of 4 log-variance factors. The shaded are highlights the 90% forecast errors bands (including innovation and parameter uncertainty) based on $S = 500$ MC paths. The fitted filtered factors are plotted in blue and their respective MC median is shown in pink. Each subplot also includes a proxy based on a daily application of the OLS model. The log-variance proxies are obtained by applying by a 21-day rolling simple linear regression on the log of squared residuals, obtained from the daily application of the OLS model for location factors.

**Table 6: Estimation results for dynamic high-dimensional hierarchical density models**

| Hierarchical Density Model Specification | Panel A: Dynamic Factors & Static Parameters | | | | Panel B: In-Sample Performance | | | |
|---|---|---|---|---|---|---|---|---|
| | Effect Coded Factors | # Free Factors ($2 \times k$; location + scale) | # Total Factors | # Param. | Log Lik. | Predictive (MAE) | Update (MAE) | Smoother (MAE) |
| Common | | $2 \times 1 = 2$ | 2 | 5 | -397,552 | 46.382 | 46.358 | 46.358 |
| Common, Tenor | 10Y | $2(1 + 5) = 12$ | 14 | 22 | -258,612 | 32.942 | 32.904 | 32.905 |
| Common, Tenor, Region | 10Y, EU | $2(1 + 5 + 2) = 16$ | 20 | 30 | -235,669 | 32.029 | 31.971 | 31.972 |
| Common, Tenor, Region, Rating | 10Y, EU, A | $2(1 + 5 + 2 + 2) = 20$ | 26 | 38 | -162,145 | 28.619 | 28.541 | 28.542 |
| Common, Tenor, Region, Rating, Firm | 10Y, EU, A, Kookmin, JPMorgen, Scandinavian | $2(1 + 5 + 2 + 2 + 21) = 62$ | 74 | 103 | 178,982 | 11.890 | 11.562 | 11.621 |
| Common, Tenor, Region, Rating, Firm, Tenor-Rating | 10Y, EU, A, Kookmin, JPMorgan, Scandinavian, 6M-A, 1Y-A, 3Y-A | $2(1+5+2+2+21+9) = 74$ | 92 | 127 | 195,579 | 11.195 | 10.854 | 10.925 |

*Notes*: the table provides an overview of the estimation results of the fitted hierarchical dynamic density models on the high-dimensional panel of $p = 144$ CDS time-series displayed in Figure 5. The first column of the table reports the specification of the factor loading matrices. The first column in Panel A lists the effect-coded factors of the included categorical variables in the corresponding $F_t$ $(p \times k)$. Remaining columns in this Panel report, the sum of total free time-varying factors location and log-variance factors, followed by the number of total factors. The numbers in this column are the sum of the number of free factors plus two times the number of effect coded factors in the first column of Panel A. The number of parameters indicates the size of the unknown parameter set $\psi$. Panel B reports the in-sample performance statics, i.e., the Log Likelihood of the estimated model and the MAE of the fitted values in bps. The MAE accuracy is taken as the average of in-sample prediction errors across all observed values in the CDS panel data. The MAE is reported based on the conditional estimates of three different GAS filters, i.e., for $f_{t|t-1}$ (predictive filter), $f_{t|t}$ (update filter) and $f_{t|n}$ (smoother) of Buccheri et al. (2021). The estimation period is January 2, 2011 - December 31, 2020.

depicted. Full results for all factors can be found in Figures C.1 and C.2 in Appendix C We find similar factor patterns by a daily application of the (rolling) OLS model(s). Interestingly, the forecast bands show highly dynamic patterns for factors that cannot be estimated due to missing observations. For example, the factors for Lloyds Bank and Barclays show a long patches of missing values (see Figure 5). Over these missing periods, the fitted values corresponds to forecasts that converge to the unconditional mean of the factors, whereas the simulation based median dynamically evolves around unconditional mean. Once information arrives that allows for the estimation of this factor, the factor itself as well as its forecast error band change substantially.

The implied median of the forecast bands is in generally different from the filtered estimate at cross-sections with many missing values. This difference potentially arises from a change in the daily hierarchical structure used simulations vs MLE. The simulation

is run with a complete design matrix $F_t$, whereas the model is estimated based on the subset $S_tF_t$. Compared to $S_tF_t$, the complete matrix $F_t$ also includes ratings dummies at missing entries. The ratings were backwards filled at the missing entries to obtain a complete design matrix at each time point. As a result, a different hierarchical structure is used for simulation, and thus yielding a different weighting scheme for the factors.

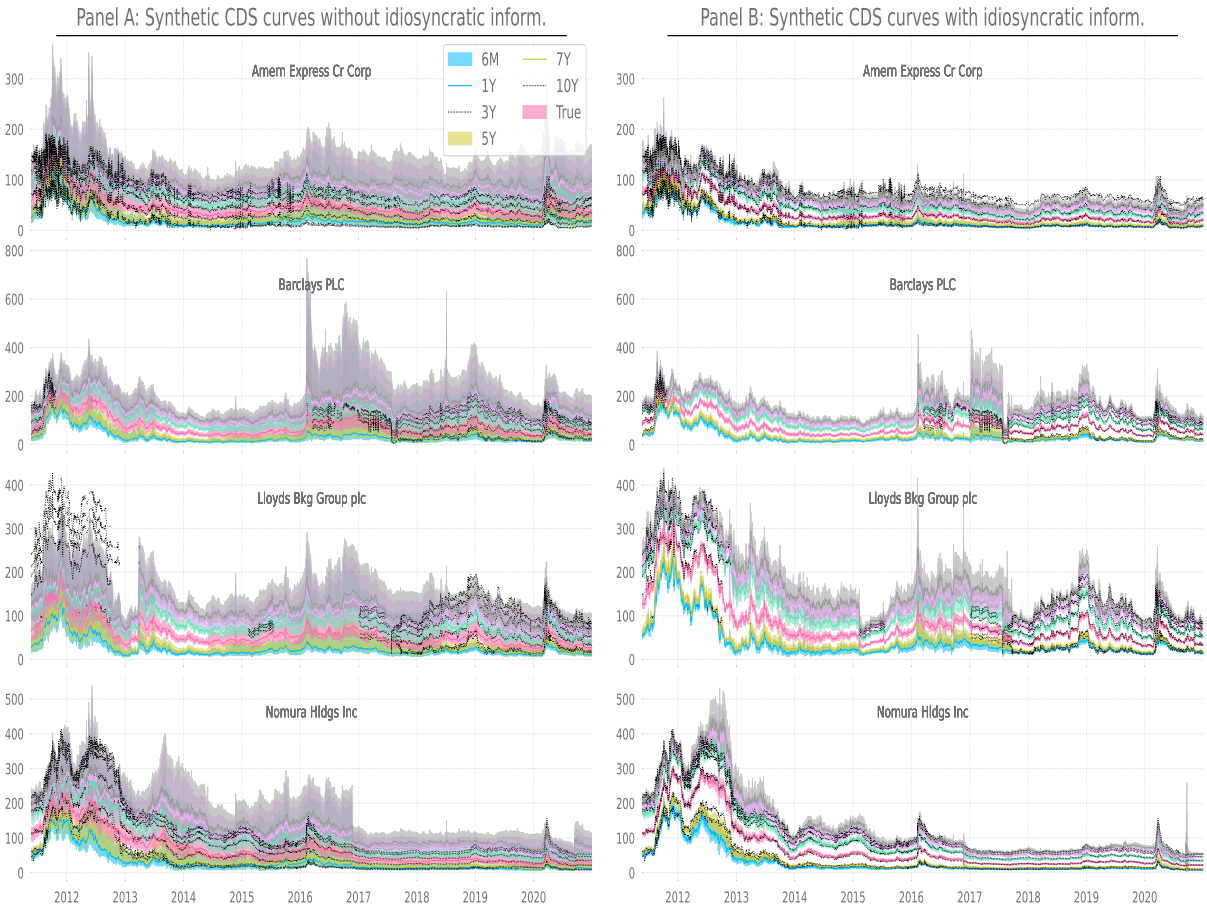
To better understand the adequacy of our largest imputation model, we use its conditional filtered factors to construct synthetic credit spreads. For this we visually compare the synthetic spreads with and without the idiosyncratic information. A possible realization of a synthetic spread at time point $t$, say $\widehat{y}_{i,t}$, is obtained as follows: $\widehat{y}_{i,t} = \exp\left(\widehat{\mu}_{i,t} + \widehat{\sigma}_{i,t}\widehat{\varepsilon}_{i,t}\right)$, for $i = 1, \ldots, p = 144$ and $t = 1, \ldots, n = 2609$. To obtain an approximate density for the credit spreads values, we propagate also the remaining residuals of the system in the formula, i.e,

$$H_t\left(\widehat{y}_{i,t}\right) = \Pr\left(\widehat{y}_{i,t} \leq x\right) = p^{-1} \sum_{j=1}^{p} \mathbb{1}\left(\exp(\widehat{\mu}_{i,t} + \widehat{\sigma}_{i,t}\widehat{\varepsilon}_{j,t}) \leq x\right),$$

where $H_t\left(\widehat{y}_{i,t}\right)$ is an approximate empirical cumulative density function for predicting $y_{i,t}$ on observation level and $\widehat{\mu}_{i,t}$ and $\widehat{\sigma}_{i,t}$ are pre-computed and fixed moment estimates for $y_{i,t}$. Imputing observations on the levels provides a more relativistic estimate for the missing observations because empirical innovations are also considered. When considering the case in which we do not make use of any idiosyncratic information for the prediction of the 6 tenors $y_t$ of the $j$-th firm, we do not include the firm fixed-effects factor $f^{\mu}_{j,t}$ and $f^{\log\sigma^2}_{j,t}$ in the computation of and $\widehat{y}_t$'s and leave out all 6 local-in-time residuals that belong to this firm. That is, only the common factors are used in this case and the sum will run for up to $p = 144 - 6 = 138$ residuals to compute $H_t\left(\widehat{y}_{i,t}\right)$. Note that for certain time points less than $p = 144$ residuals are available because the CDS panel data is in presence of missing observations.

Figure 7 provides four examples of synthetic CDS curves computed with and without the use of idiosyncratic information. The synthetic curves are computed for the firms: American Express Credit Corporation (AXP), Barclays PLC (BARC), Lloyds Banking Group Plc (LYG) and Nomura Holdings Inc (NMR) The two panels of the figure highlight that without the use of idiosyncratic information much larger errors are made. At many time points, the synthetic curves are either over predicting or under-predicting the true data points. The idiosyncratic firm factors are highly informative, even when the respective curve consists of large block missing gaps. For example, the CDS curves of BARC and LYG include many missing entries, and the overall level of these curves is

**Figure 7: Examples of synthetic credit curves**




*Notes*: This figure displays the synthetic CDS term structures for four different companies. Panel A plots the curves that are computed without the use of idiosyncratic, i.e., with common factors only. The conditional filtered location and scale factors are obtained from the largest hierarchical density model specified in Table 6. The solid colored lines correspond to the median of the approximate spread density and the shaded areas represent the 95% empirical confidence intervals, for which the model's residuals were used. The true observed curves for the corresponding firms are plotted in black dots. Panel B displays the synthetic curves for the same firms in Panel A but is computed with the use of idiosyncratic factors.

**Table 7: Coverage rates of synthetic curve bands at 95% confidence level**

| | Tenor | | | | | |
|---|---|---|---|---|---|---|
| Region | 6M | 1Y | 3Y | 5Y | 7Y | 10Y |
| **Panel A: Coverage without idiosyncratic information (%)** | | | | | | |
| Asia | 89.723 | 89.440 | 88.818 | 92.545 | 96.474 | 97.710 |
| North America | 97.882 | 97.403 | 97.456 | 97.312 | 97.456 | 97.307 |
| Europe | 95.937 | 93.015 | 91.764 | 85.823 | 82.738 | 82.474 |
| **Panel B: Coverage with idiosyncratic information (%)** | | | | | | |
| Asia | 95.415 | 95.353 | 92.971 | 94.136 | 95.108 | 92.722 |
| North America | 93.714 | 93.748 | 92.914 | 91.395 | 95.578 | 95.348 |
| Europe | 95.027 | 95.132 | 91.889 | 93.628 | 97.135 | 95.295 |

*Notes*: the table reports the in-sample coverage rates of the empirical 95% confidence bands for the true CDS data points present in the panel of $p = 144$ time-series. The first panel reports the coverage rates of the empirical bands for which no idiosyncratic information was used. The coverage accuracy is further divided per region-tenor combination. Similarly, Panel B reports coverage accuracy of empirical bands for which idiosyncratic was used.

significantly shifted by their idiosyncratic factors and the predictions are much closer to observed values. Most of the time, the true spreads data points are within the empirical bands with a 95% confidence level.

The coverage rates based on $H_t\left(\widehat{y}_{i,t}\right)$ with and without the use of idiosyncratic information are reported in Table 7. The empirical bands for which no idiosyncratic information was utilized perform moderately across all regions and tenors. In contrast, the confidence intervals implied by $H_t\left(\widehat{y}_{i,t}\right)$ that does include idiosyncratic dynamics are much accurate and are very close to the nominal level. As a result, it is also much more likelier that the imputed densities $H_t\left(\widehat{y}_{i,t}\right)$ provide a better description of the unobserved values $y_{i,t}$. Hence, our largest time-varying hierarchical density model serves as an adequate choice for imputing the missing values present in the high-dimensional panel data of CDS time-series.

# 5 Conclusion

We have introduced a hierarchical dynamic factor location-scale model for forecasting missing values and their associated imputation uncertainty for large multivariate time-series systems. The model exploits both the cross-sectional and time-series properties of the data while allowing for very general patterns of missing data. Estimation of the model's static parameters could be performed using standard maximum likelihood

methods given the observation-driven structure of the model. At the same time, the observation-driven structure poses challenges to the formation of intuitive observation weights to construct estimates of the time-varying parameters. We showed how to overcome these by mapping the observations onto the proper orthogonal (sub)-basis.

The method worked well in a pure simulation environment and also compared well against a state-space benchmark in an empirical setting. Both isolated, independent missing values as well as long stretches of missings could be dealt with by the method. When applied to empirical data on panels of sparse credit default swap (CDS) term structures, the GAS model demonstrated strong scalability to high dimensions. Our empirical results indicate the significant importance of incorporating firm-specific location-scale factor for imputation, as hierarchical models incorporating such factor produced more accurate risk quantiles for the panel compared to models without these idiosyncratic components.

# References


Abadir, K. (2005). *Matrix algebra*. Cambridge New York: Cambridge University Press.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Bai, J. and S. Ng (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 1–18.

BCBS (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. *Basel Committee on Banking Supervision*.

Beutner, E. A., Y. Lin, and A. Lucas (2023). Consistency, distributional convergence, and optimality of score-driven filters. Technical report, Tinbergen Institute Discussion Paper, 23-051/III.

Blasques, F., C. Francq, and S. Laurent (2020). A new class of robust observation-driven models. WorkingPaper 20-073/III, Tinbergen Instituut.

Blasques, F., P. Gorgi, and S. Koopman (2021). Missing observations in observation-driven time series models. *Journal of Econometrics 221*(2), 542–568.

Blasques, F., S. J. Koopman, K. Łasak, and A. Lucas (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting 32*(3), 875–887.

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.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2020). A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. *Journal of Business & Economic Statistics*, 1–17.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2021). Filtering and smoothing with score-driven models.

Cahan, E., J. Bai, and S. Ng (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions.

Chan, J. C., A. Poon, and D. Zhu (2023). High-dimensional conditionally gaussian state space models with missing data. *Journal of Econometrics 236*(1), 105468.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

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

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics 96*(5), 898–915.

Delle Monache, D., I. Petrella, and F. Venditti (2016). Adaptive state space models with applications to the business cycle and financial stress.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

EBA (2015). EBA Report on CVA under Article 456 (2) of Regulation (EU) No 575/2013. *European Banking Authority*.

Foroni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting 30*(3), 554–568.

Green, A. (2015). *XVA: Credit, Funding and Capital Valuation Adjustments*. Wiley.

Gregory, J. (2020). *The xVA Challenge: Counterparty Risk, Funding, Collateral, Capital and Initial Margin*. Wiley.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press.

Jin, S., K. Miao, and L. Su (2021). On factor models with random missing: Em estimation, inference, and cross validation. *Journal of Econometrics 222*(1), 745–777.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*(8), 1358–1368.

Koopman, S. J., R. Lit, A. Lucas, and A. Opschoor (2018). Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics 33*(7), 966–985.

Koopman, S. J., A. Lucas, and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics 98*(1), 97–110.

Lucas, A., A. Opschoor, and J. Schaumburg (2016). Accounting for missing values in score-driven time-varying parameter models. *Economics Letters 148*, 96–98.

Morini, M. and A. Prampolini (2011). Risky funding: A unified framework for counterparty and liquidity charges. *Risk Magazine*.

Pallavicini, A., D. Perini, and D. Brigo (2011). Funding valuation adjustment: a consistent framework including CVA, DVA, collateral, netting rules and re-hypothecation.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*(3), 386–398.

Stock, J. and M. Watson (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Handbook of Macroeconomics, pp. 415–525. Elsevier.

van der Merwe, C., D. Heyman, and T. de Wet (2018). Approximating risk-free curves in sparse data environments. *Finance Research Letters 26*, 112–118.

Xiong, R. and M. Pelger (2023). Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics 233*(1), 271–301.

# A Derivation of Scaled-Scores

Assume $S_t \coloneqq \mathrm{I}$ at time $t$, then the log of the conditional density in (1) is

$$\log p\left(y_t \mid f_t, \mathcal{F}_{t-1}; \psi\right) = -\frac{1}{2}\left[p \log 2\pi + \log|\Sigma_t| + \left(y_t - \mu_t\right)' \Sigma_t \left(y_t - \mu_t\right)\right],$$

with $p$-dimensional mean and variance vectors $\mu_t = F_t f_t^{\mu}$ and $\sigma_t^2 = \exp\left(F_t f_t^{\log\sigma^2}\right)$, respectively, and $p \times p$ diagonal variance matrix $\Sigma_t = \mathrm{diag}(\sigma_t^2)$, for $t = 1, \ldots, n$.

***Proof of Proposition 1:*** The score and information matrix of the location factors are straightforward to derive:

$$\nabla_t^{\mu} = \frac{\log p\left(y_t \mid f_t, \mathcal{F}_{t-1}; \psi\right)}{\partial f_t^{\mu}} = F_t' \Sigma_t^{-1}\left(y_t - \mu_t\right)$$
$$\mathcal{I}_t^{\mu} = -\mathbb{E}_{t-1}\left[\frac{\partial^2 \log p\left(y_t \mid f_t, \mathcal{F}_{t-1}; \psi\right)}{\partial f_t^{\mu} \partial f_t^{\mu\prime}}\right] = F_t' \Sigma_t^{-1} F_t$$

From basic matrix calculus results it follows that for any triangular matrix $\Sigma_t = J_t J_t'$, its log determinant can be written as

$$\log|\Sigma_t| = \log|J_t J_t'| = \log|J_t|^2 = 2\log \Pi_{i=1}^{p} J_{ii,t} = 2\sum_{i=1}^{p} \log J_{ii,t},$$

see for instance Abadir (2005) for this and other useful results. Since $\Sigma_t$ is diagonal, we can simply define $J_t \coloneqq \Sigma_t^{\frac{1}{2}} = \mathrm{diag}(\sigma_t)$. Hence, we have

$$\log|\Sigma_t| = 2\sum_{i=1}^{p} \log \sigma_{i,t} = 2\sum_{i=1}^{p} \log \exp\left(\frac{1}{2} F_t f_t^{\log\sigma^2}\right)_i = \sum_{i=1}^{p}\left(F_t f_t^{\log\sigma^2}\right)_i = \mathrm{tr}(\mathrm{diag}(\log\sigma_t)),$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator. By using this matrix result, the score and information matrix of the log-variance factors take the following expressions:

$$\begin{aligned}
\nabla_t^{\log\sigma^2} &= \frac{1}{2} F_t' \left[\Sigma_t\left(y_t - \mu_t\right) \circ \left(y_t - \mu_t\right) - \iota_p\right], \\
&= \frac{1}{2} F_t' \left(\varepsilon_t^2 - \iota_p\right), \\
\mathcal{I}_t^{\log\sigma^2} &= \frac{1}{2} F_t' F_t,
\end{aligned}$$

where $\circ$ denotes the Hadamard's element wise matrix multiplication operator. Note that we have made use of the fact that squared mean prediction errors divided by their variance are equal to standardized normal innovations, i.e.,

$$\Sigma_t\left(y_t - \mu_t\right) \circ \left(y_t - \mu_t\right) = \Sigma_t\left(y_t - \mu_t\right)^2 = \varepsilon_t^2$$

The expressions in Proposition 1 are based on the marginal density $\log p\left(y_t^{\dagger} \mid f_t, \mathcal{F}_{t-1}; \psi\right)$, which are obtained by merely replacing $F_t$ with $S_t F_t$. This completes the proof. ■

# B Additional Monte Carlo simulation results

This appendix reports additional results of our Monte Carlo study in Section 3. Each subplot in the Figures B.1-B.4 corresponds to a specific *missingness scenario* of the Monte Carlo study based on a time-series system of $p = 4$ time-series. The missing data mechanism is exactly the same for all series in each of the respective figures, but is introduced stages. For example, Figures B.1 displays four unique scenarios in which up to $p = 4$ time-series are simultaneously subject to the block missing pattern. Subfigure B.1a corresponds to scenario in which only $y_{1,t}$ contain block missing observations, whereas Subfigure B.1b illustrates the scenario in which both $y_{1,t}$ and $y_{2,t}$ have the exact same block missing entries, etc. Similarly, the subplots in the Figures B.2, B.3 and B.4 illustrate the randomly missing scenarios for different values of $\pi_{i,t}$.

**Figure B.1: 60% block missing scenarios**

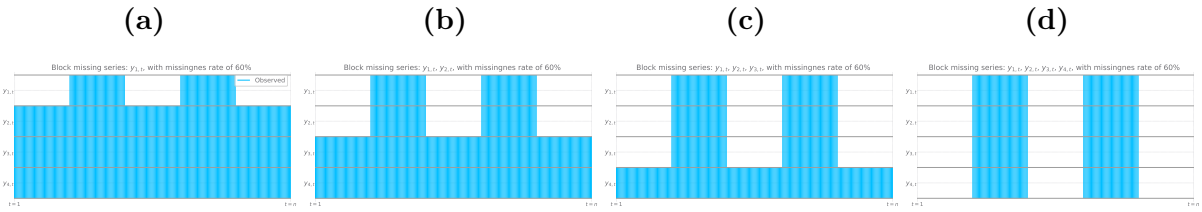



**Figure B.2: $\pi_{i,t} = 25\%$ randomly missing scenarios**

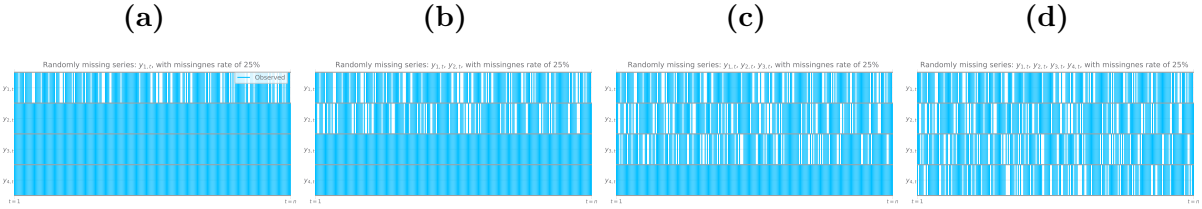



**Figure B.3: $\pi_{i,t} = 50\%$ randomly missing scenarios**

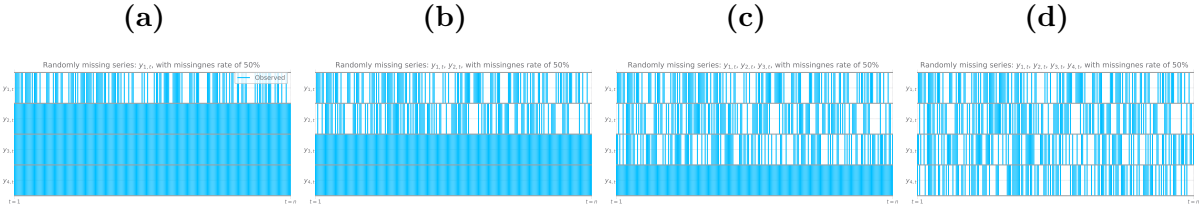
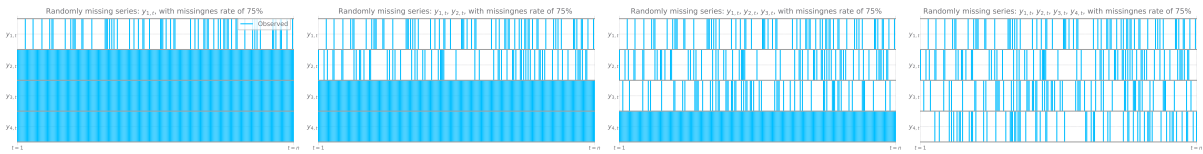


**Figure B.4:** $\pi_{i,t} = 75\%$ **randomly missing scenarios**

**(a)** **(b)** **(c)** **(d)**



*Notes*: Each subplot in the Figures B.1, B.2, B.3 and B.4 illustrates a potential missingnes scenario among a panel data consisting of $p = 4$ time-series. A non-missing time-series observation is marked by a vertical blue stripe over time for $t = 1, \ldots, n$. Figures B.1 displays four unique scenarios in which up to $p = 4$ time-series are simultaneously subject to the block missing pattern. For example, Subfigure B.1a corresponds to scenario in which only $y_{1,t}$ contain block missing observations, whereas Subfigure B.1b illustrates the scenario in which both $y_{1,t}$ and $y_{2,t}$ have the exact same block missing entries, etc. Similarly, the subplots in the Figures B.2, B.3 and B.4 illustrate the randomly missing scenarios for different values of $\pi_{i,t}$.

Additional experiments highlight that the forecast performance results obtained on factor level, also hold for functions of the location and log-variance factors. For example, panels A-D in Table B.1 reports these results for all time-varying mean and volatility predictions at sparse locations. The MAFE errors and percentage coverage rates are reported for the exact same missing data mechanism and scenario combinations in Table **??**. The last panel in Table B.1 reports the Pearson's correlation between the forecast errors of mean and variance pairs (in percentages), i.e., $\text{Corr}\left(\mu_i - \widehat{\mu}_i, \sigma_i^2 - \widehat{\sigma}_i^2\right)$, for $i = 1, \ldots, p = 4$.

We also measure the forecast error and accuracy of the bands for the model implied risk measures, given the true and predicted paths of the time-varying means and variances for the four series in our system series. We consider the measures: value-at-risk (VaR) and expected shortfall (ES) at several confidence levels. The MAFE metric for the mode implied VaR and ES at conventional $1 - \alpha\%$ confidence levels is reported in Table 2. The coverage rates of the forecast bands for these time-varying risk measures is reported in Table B.2. Furthermore, we have also gathered the precision of these risk measures based on the accuracy of the predicted hit variables. All hit variables take value one if the the time-series variables exceed their VaR/ES threshold over time, and are zero otherwise. The forecast error of the hit-variables is then measured as the MAFE between the true and predicted hit-variables. Since the hit-variable is a binary variable, the MAFE quantifies the precision because it is now naturally bounded between 0 and 1. Table B.3 reports the MAFE precision rates for each time-series, among all $1 - \alpha$ confidence levels.

**Table B.1: In-sample forecast perfomance results on mean-variance level**

| *Missing Data Mechanism*: | | Block Missing | | | | Randomly Missing | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missingnes Rate*: | | 60% | | | | 25% | | | | 50% | | | | 75% | | | |
| *Missing Scenario Figure*: | | B.1a | B.1b | B.1c | B.1d | B.2a | B.2b | B.2c | B.2d | B.3a | B.3b | B.3c | B.3d | B.4a | B.4b | B.4c | B.4d |
| **Moments** | 0 | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ |
| **Panel A: MAFE of in-sample forecasts for mean equation** | | | | | | | | | | | | | | | | | |
| $\mu_1$ | 0.030 | 0.126 | 0.147 | 0.187 | 0.229 | 0.059 | 0.062 | 0.066 | 0.070 | 0.082 | 0.088 | 0.097 | 0.102 | 0.104 | 0.114 | 0.130 | 0.137 |
| $\mu_2$ | 0.030 | 0.047 | 0.145 | 0.186 | 0.226 | 0.034 | 0.062 | 0.066 | 0.070 | 0.039 | 0.088 | 0.097 | 0.102 | 0.043 | 0.113 | 0.128 | 0.136 |
| $\mu_3$ | 0.030 | 0.047 | 0.066 | 0.187 | 0.228 | 0.034 | 0.039 | 0.066 | 0.070 | 0.039 | 0.047 | 0.096 | 0.102 | 0.043 | 0.055 | 0.129 | 0.137 |
| $\mu_4$ | 0.036 | 0.049 | 0.066 | 0.093 | 0.227 | 0.041 | 0.045 | 0.049 | 0.074 | 0.045 | 0.052 | 0.062 | 0.105 | 0.047 | 0.058 | 0.074 | 0.140 |
| **Panel B: MAFE of in-sample forecasts for variance equation** | | | | | | | | | | | | | | | | | |
| $\sigma_1$ | 0.022 | 0.091 | 0.106 | 0.135 | 0.163 | 0.043 | 0.045 | 0.048 | 0.051 | 0.059 | 0.064 | 0.070 | 0.075 | 0.076 | 0.084 | 0.096 | 0.102 |
| $\sigma_2$ | 0.023 | 0.035 | 0.106 | 0.136 | 0.163 | 0.026 | 0.045 | 0.048 | 0.052 | 0.029 | 0.064 | 0.071 | 0.075 | 0.032 | 0.084 | 0.096 | 0.102 |
| $\sigma_3$ | 0.023 | 0.035 | 0.048 | 0.135 | 0.164 | 0.026 | 0.029 | 0.048 | 0.052 | 0.029 | 0.035 | 0.071 | 0.075 | 0.032 | 0.041 | 0.096 | 0.102 |
| $\sigma_4$ | 0.028 | 0.037 | 0.049 | 0.070 | 0.163 | 0.030 | 0.033 | 0.037 | 0.054 | 0.033 | 0.039 | 0.046 | 0.077 | 0.035 | 0.043 | 0.055 | 0.105 |
| **Panel C: Coverage of mean equation at 95% confidence level (%)** | | | | | | | | | | | | | | | | | |
| $\mu_1$ | 96.937 | 95.921 | 95.723 | 94.855 | 97.121 | 95.972 | 95.643 | 94.868 | 94.485 | 96.191 | 96.141 | 95.708 | 97.283 | 97.903 | 98.371 | 98.285 | 99.769 |
| $\mu_2$ | 96.783 | 95.317 | 95.878 | 94.917 | 97.133 | 96.551 | 95.734 | 95.025 | 94.773 | 95.525 | 96.154 | 95.587 | 97.283 | 96.774 | 98.334 | 98.309 | 99.757 |
| $\mu_3$ | 96.801 | 95.322 | 94.507 | 94.826 | 96.985 | 96.197 | 95.339 | 94.811 | 94.680 | 95.676 | 94.294 | 95.671 | 97.261 | 96.775 | 96.552 | 98.263 | 99.722 |
| $\mu_4$ | 97.990 | 98.670 | 98.408 | 97.239 | 98.155 | 97.723 | 97.544 | 97.262 | 96.883 | 98.053 | 98.203 | 97.925 | 98.644 | 99.350 | 99.471 | 99.449 | 99.891 |
| **Panel D: Coverage of variance equation at 95% confidence level (%)** | | | | | | | | | | | | | | | | | |
| $\sigma_1$ | 96.932 | 96.098 | 95.607 | 94.121 | 96.618 | 95.599 | 95.068 | 93.921 | 93.343 | 96.143 | 95.432 | 94.472 | 96.306 | 97.717 | 97.795 | 97.369 | 99.395 |
| $\sigma_2$ | 96.597 | 94.787 | 95.535 | 94.042 | 96.582 | 96.019 | 95.117 | 93.899 | 93.380 | 95.497 | 95.387 | 94.419 | 96.210 | 96.575 | 97.674 | 97.400 | 99.369 |
| $\sigma_3$ | 97.089 | 95.017 | 93.613 | 93.955 | 96.499 | 95.776 | 94.927 | 93.970 | 93.444 | 95.663 | 93.679 | 94.356 | 96.170 | 96.559 | 96.065 | 97.535 | 99.435 |
| $\sigma_4$ | 97.621 | 98.406 | 98.049 | 96.378 | 97.869 | 97.524 | 97.404 | 96.798 | 96.278 | 97.976 | 97.878 | 97.196 | 98.140 | 99.082 | 99.156 | 99.063 | 99.728 |
| **Panel E: Forecast error correlation between mean and variance pairs (%)** | | | | | | | | | | | | | | | | | |
| $\rho_{\mu_1,\sigma_1^2}$ | -0.951 | 0.616 | 0.891 | 0.796 | 1.474 | -0.311 | -0.162 | -0.343 | 0.298 | -0.023 | -0.414 | 0.276 | 0.321 | 0.180 | 0.308 | 0.899 | 0.132 |
| $\rho_{\mu_2,\sigma_2^2}$ | -1.013 | -0.504 | 1.339 | 0.772 | 0.948 | -0.442 | 0.350 | -0.288 | 0.100 | 0.401 | -0.572 | 0.088 | -0.336 | -1.094 | -0.020 | 0.418 | -0.271 |
| $\rho_{\mu_3,\sigma_3^2}$ | 1.415 | 1.711 | 1.476 | 0.385 | 0.801 | 1.543 | 2.012 | 1.057 | 0.854 | -0.108 | -0.330 | 0.487 | 0.097 | -1.586 | -0.883 | 0.068 | -0.804 |
| $\rho_{\mu_4,\sigma_4^2}$ | -0.396 | 0.395 | 0.895 | 0.366 | 0.803 | 0.859 | 0.833 | 1.045 | 0.181 | -0.234 | -0.641 | 0.282 | -0.431 | 0.635 | 0.513 | 0.441 | -0.562 |

*Notes*: The first two panels of this table report the MAFE of the mean and volatility for each time-series. The next two panels report the coverage rates of the forecast bands corresponding to a confidence level of 95% for the mean and volatility. The last panel reports the Pearson's correlation coefficient (in %) between the forecast errors of mean and variance pairs. For more details on the scenarios and missing data mechanisms, see caption of Table **??**.

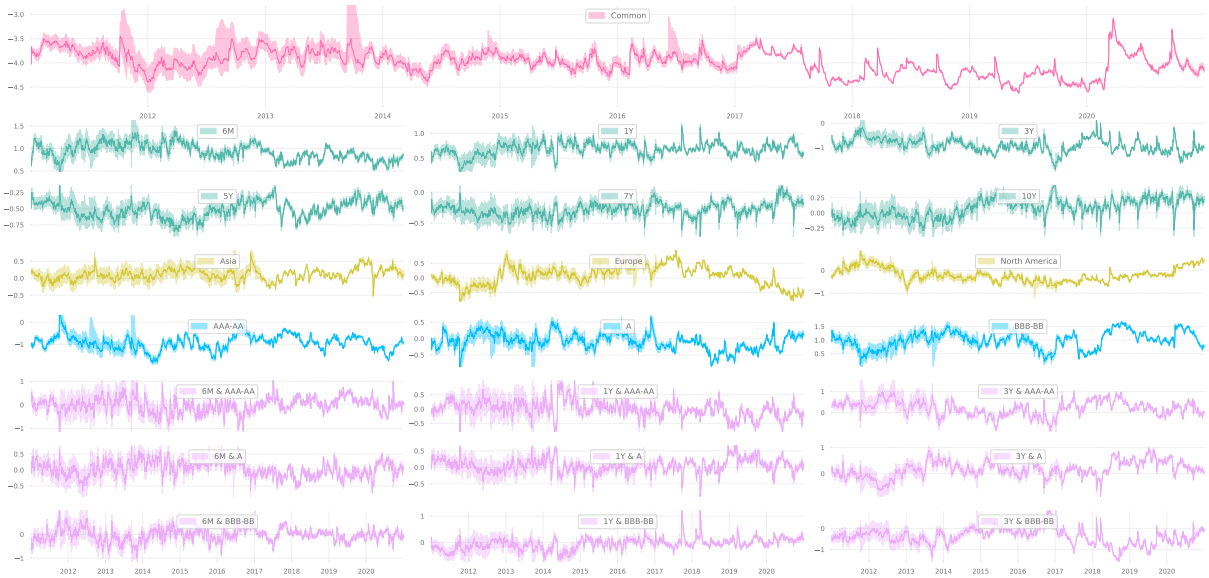
# C Additional empirical results

**Figure C.1: Filtered common location-scale factors**

**(a)** Mean factors




**(b)** Log-variance factors




*Notes*: the first panel of this figure displays the common CDS mean factors, given the parameter estimates of the largest specification of the dynamic density model in Table 6. The filtered estimates (solid) are computed as the median of predictive GAS filter ($f_t$) paths, given $S = 500$ simulations. The 95% forecast bands that inherit both innovation and parameter uncertainty are also plotted (shaded areas). Each factor category is highlighted with a different color, e.g., the regional factors are plotted in yellow and the interaction terms in purple. Similarly, the second panel depicts the time-varying common log-variance factors. Rolling-window estimates for the mean factors are displayed in black dots.

## Table B.2: In-sample accuracy of forecast bands for tail risk measures

| *Missing Data Mechanism*: | | Block Missing | | | | Randomly Missing | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missingnes Rate*: | | 60% | | | | 25% | | | | 50% | | | | 75% | | | |
| *Missing Scenario Figure*: | | B.1a | B.1b | B.1c | B.1d | B.2a | B.2b | B.2c | B.2d | B.3a | B.3b | B.3c | B.3d | B.4a | B.4b | B.4c | B.4d |
| **Series** | 0 | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ |
| **Panel A: Coverage of VaR with $\alpha = 5\%$ at 95% confidence level** (%) | | | | | | | | | | | | | | | | | |
| $y_1$ | 96.749 | 97.916 | 97.512 | 96.679 | 98.180 | 95.972 | 95.599 | 94.719 | 94.185 | 96.561 | 96.111 | 95.267 | 96.851 | 98.249 | 98.322 | 97.982 | 99.596 |
| $y_2$ | 96.705 | 96.314 | 97.546 | 96.586 | 98.139 | 96.118 | 95.531 | 94.567 | 94.058 | 95.792 | 96.056 | 95.234 | 96.858 | 96.829 | 98.287 | 97.986 | 99.579 |
| $y_3$ | 96.749 | 96.465 | 95.872 | 96.499 | 98.183 | 96.018 | 95.214 | 94.522 | 94.072 | 95.765 | 94.332 | 95.356 | 96.938 | 96.677 | 96.293 | 98.049 | 99.589 |
| $y_4$ | 97.735 | 98.933 | 98.925 | 98.210 | 98.846 | 97.727 | 97.666 | 97.289 | 96.793 | 98.163 | 98.245 | 97.862 | 98.519 | 99.292 | 99.390 | 99.329 | 99.810 |
| **Panel B: Coverage of VaR with $\alpha = 1\%$ at 95% confidence level** (%) | | | | | | | | | | | | | | | | | |
| $y_1$ | 96.765 | 97.842 | 97.406 | 96.509 | 98.102 | 95.873 | 95.487 | 94.523 | 93.954 | 96.471 | 95.938 | 95.019 | 96.666 | 98.154 | 98.187 | 97.800 | 99.529 |
| $y_2$ | 96.657 | 96.205 | 97.410 | 96.424 | 98.048 | 96.029 | 95.381 | 94.366 | 93.810 | 95.735 | 95.867 | 95.014 | 96.675 | 96.757 | 98.145 | 97.820 | 99.513 |
| $y_3$ | 96.836 | 96.399 | 95.716 | 96.359 | 98.094 | 95.944 | 95.101 | 94.333 | 93.824 | 95.690 | 94.146 | 95.108 | 96.725 | 96.598 | 96.176 | 97.909 | 99.545 |
| $y_4$ | 97.665 | 98.870 | 98.822 | 98.016 | 98.786 | 97.639 | 97.582 | 97.114 | 96.609 | 98.121 | 98.151 | 97.670 | 98.387 | 99.228 | 99.317 | 99.249 | 99.788 |
| **Panel C: Coverage of ES with $\alpha = 2.5\%$ at 95% confidence level** (%) | | | | | | | | | | | | | | | | | |
| $y_1$ | 96.765 | 97.839 | 97.404 | 96.505 | 98.100 | 95.871 | 95.484 | 94.520 | 93.949 | 96.470 | 95.937 | 95.014 | 96.664 | 98.153 | 98.184 | 97.797 | 99.528 |
| $y_2$ | 96.657 | 96.203 | 97.408 | 96.419 | 98.046 | 96.026 | 95.379 | 94.364 | 93.805 | 95.732 | 95.865 | 95.010 | 96.672 | 96.755 | 98.142 | 97.818 | 99.512 |
| $y_3$ | 96.838 | 96.399 | 95.715 | 96.355 | 98.092 | 95.944 | 95.100 | 94.329 | 93.821 | 95.691 | 94.144 | 95.104 | 96.721 | 96.596 | 96.174 | 97.908 | 99.544 |
| $y_4$ | 97.665 | 98.869 | 98.820 | 98.013 | 98.786 | 97.640 | 97.582 | 97.111 | 96.605 | 98.119 | 98.149 | 97.666 | 98.385 | 99.228 | 99.316 | 99.248 | 99.787 |
| **Panel D: Coverage of ES with $\alpha = 1\%$ at 95% confidence level** (%) | | | | | | | | | | | | | | | | | |
| $y_1$ | 96.766 | 97.804 | 97.350 | 96.420 | 98.060 | 95.827 | 95.437 | 94.437 | 93.870 | 96.420 | 95.865 | 94.919 | 96.600 | 98.098 | 98.131 | 97.731 | 99.509 |
| $y_2$ | 96.643 | 96.153 | 97.348 | 96.335 | 98.002 | 96.009 | 95.328 | 94.274 | 93.707 | 95.700 | 95.787 | 94.913 | 96.599 | 96.729 | 98.080 | 97.754 | 99.488 |
| $y_3$ | 96.874 | 96.367 | 95.650 | 96.297 | 98.053 | 95.916 | 95.063 | 94.268 | 93.738 | 95.668 | 94.093 | 95.003 | 96.643 | 96.578 | 96.146 | 97.855 | 99.526 |
| $y_4$ | 97.647 | 98.838 | 98.773 | 97.935 | 98.764 | 97.616 | 97.560 | 97.037 | 96.530 | 98.086 | 98.104 | 97.591 | 98.344 | 99.204 | 99.291 | 99.216 | 99.780 |

*Notes*: The panels of this table report the coverage rates of the forecast bands for the time-varying VaR and ES at several confidence levels for each time-series in the system. For more details on the scenarios and missing data mechanisms, see caption of Table **??**.

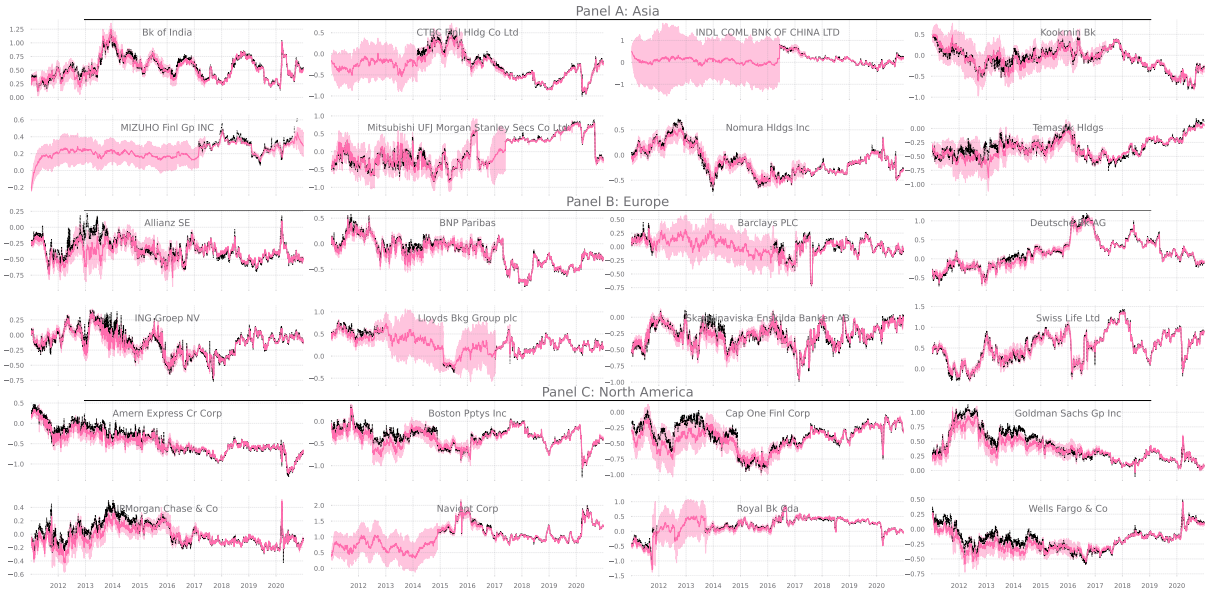
## Table B.3: In-sample forecast precision of tail risk measures

| *Missing Data Mechanism*: | | Block Missing | | | | Randomly Missing | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Missingnes Rate*: | | 60% | | | | 25% | | | | 50% | | | | 75% | | | |
| *Missing Scenario Figure*: | | B.1a | B.1b | B.1c | B.1d | B.2a | B.2b | B.2c | B.2d | B.3a | B.3b | B.3c | B.3d | B.4a | B.4b | B.4c | B.4d |
| **Moments** | 0 | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ | $y_{1,t}$ | $y_{1:2,t}$ | $y_{1:3,t}$ | $y_{1:4,t}$ |
| **Panel A: MAFE of in-sample forecasts for VaR hit-variable with $\alpha = 5\%$ (%)** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.478 | 4.955 | 4.955 | 4.955 | 4.955 | 4.998 | 1.970 | 2.000 | 2.043 | 5.060 | 3.189 | 3.277 | 3.333 | 5.029 | 4.209 | 4.269 | 4.310 |
| $y_2$ | 0.489 | 0.760 | 4.938 | 4.938 | 4.938 | 0.542 | 4.956 | 2.015 | 2.030 | 0.625 | 5.005 | 3.264 | 3.267 | 0.675 | 5.026 | 4.260 | 4.297 |
| $y_3$ | 0.475 | 0.750 | 1.044 | 5.015 | 5.015 | 0.551 | 0.601 | 4.951 | 2.039 | 0.630 | 0.766 | 4.979 | 3.322 | 0.708 | 0.910 | 5.037 | 4.305 |
| $y_4$ | 0.589 | 0.770 | 1.026 | 1.466 | 4.981 | 0.672 | 0.723 | 0.768 | 5.027 | 0.710 | 0.819 | 0.975 | 4.970 | 0.737 | 0.915 | 1.170 | 4.990 |
| **Panel B: MAFE of in-sample forecasts for VaR hit-variable with $\alpha = 1\%$ (%)** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.161 | 0.985 | 0.985 | 0.985 | 0.985 | 0.980 | 0.482 | 0.491 | 0.521 | 1.015 | 0.721 | 0.763 | 0.770 | 1.030 | 0.919 | 0.942 | 0.945 |
| $y_2$ | 0.158 | 0.233 | 0.997 | 0.997 | 0.997 | 0.177 | 0.969 | 0.484 | 0.523 | 0.209 | 1.003 | 0.752 | 0.755 | 0.224 | 1.006 | 0.914 | 0.930 |
| $y_3$ | 0.154 | 0.239 | 0.338 | 1.016 | 1.016 | 0.187 | 0.188 | 0.989 | 0.519 | 0.207 | 0.261 | 1.004 | 0.769 | 0.221 | 0.292 | 1.009 | 0.930 |
| $y_4$ | 0.192 | 0.265 | 0.341 | 0.489 | 0.997 | 0.207 | 0.236 | 0.241 | 0.992 | 0.228 | 0.276 | 0.321 | 1.004 | 0.238 | 0.301 | 0.378 | 0.986 |
| **Panel C: MAFE of in-sample forecasts for ES hit-variable with $\alpha = 2.5\%$ (%)** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.158 | 0.955 | 0.955 | 0.955 | 0.955 | 0.955 | 0.468 | 0.481 | 0.515 | 0.985 | 0.706 | 0.741 | 0.749 | 0.998 | 0.893 | 0.915 | 0.920 |
| $y_2$ | 0.155 | 0.228 | 0.970 | 0.970 | 0.970 | 0.163 | 0.944 | 0.476 | 0.514 | 0.205 | 0.972 | 0.737 | 0.735 | 0.219 | 0.975 | 0.887 | 0.902 |
| $y_3$ | 0.149 | 0.238 | 0.331 | 0.986 | 0.986 | 0.185 | 0.191 | 0.960 | 0.505 | 0.203 | 0.256 | 0.972 | 0.749 | 0.218 | 0.287 | 0.977 | 0.903 |
| $y_4$ | 0.189 | 0.254 | 0.335 | 0.480 | 0.965 | 0.209 | 0.236 | 0.240 | 0.960 | 0.225 | 0.270 | 0.309 | 0.975 | 0.235 | 0.298 | 0.375 | 0.958 |
| **Panel D: MAFE of in-sample forecasts for ES hit-variable with $\alpha = 1\%$ (%)** | | | | | | | | | | | | | | | | | |
| $y_1$ | 0.077 | 0.373 | 0.373 | 0.373 | 0.373 | 0.379 | 0.206 | 0.226 | 0.222 | 0.384 | 0.295 | 0.318 | 0.315 | 0.399 | 0.367 | 0.379 | 0.383 |
| $y_2$ | 0.078 | 0.119 | 0.385 | 0.385 | 0.385 | 0.090 | 0.390 | 0.217 | 0.229 | 0.100 | 0.390 | 0.325 | 0.322 | 0.105 | 0.387 | 0.366 | 0.375 |
| $y_3$ | 0.074 | 0.121 | 0.169 | 0.394 | 0.394 | 0.099 | 0.097 | 0.386 | 0.221 | 0.091 | 0.110 | 0.372 | 0.308 | 0.105 | 0.141 | 0.389 | 0.383 |
| $y_4$ | 0.090 | 0.120 | 0.164 | 0.241 | 0.375 | 0.099 | 0.101 | 0.120 | 0.368 | 0.112 | 0.137 | 0.152 | 0.392 | 0.117 | 0.144 | 0.182 | 0.382 |

*Notes*: The panels of this table report the precision of hit-variables for the time-varying VaR and ES at several confidence levels for each time-series in the system. For more details on the scenarios and missing data mechanisms, see caption of Table **??**.

**Figure C.2: Filtered idiosyncratic location-scale factors**

**(a)** Mean factors




**(b)** Log-variance factors




*Notes*: the first plot if this panel displays the filtered conditional idiosyncratic (fixed-effect) median-based location factors (solid pink). Each panel of the first panel plots 8 factors, namely one for each of the financial firms within the respective region (i.e., for Asia, North America and Europe). The subplots also plot the simulation-based 95% confidence bands. Similarly, the second plot of this figure displays the time-varying idiosyncratic log-variance factors. For more details, see caption of Figure C.1.