Score-Driven Asset Pricing: Predicting Time-Varying Risk Premia Based on Cross-Sectional Model Performance

Dennis Umlandt

July 27, 2022

Abstract

This paper proposes a new parametric approach for estimating linear factor pricing models with dynamic risk premia. Time-varying risk prices and exposures follow an observation-driven updating scheme that reduces the one-step-ahead prediction error from a cross-sectional factor model at the current observation. This agnostic approach is particularly useful in situations where predictors are unknown or of uncertain quality. Updating schemes for elliptically distributed returns are derived and propose cross-sectional regression errors as driving sequence for the parameter dynamics. Estimation and inference are performed by likelihood maximization. A simulation study confirms that the novel method is capable of filtering and predicting substantial risk price movements. The empirical performance of the method is illustrated by an application to a panel of size-sorted equity portfolios.

Keywords: Dynamic Asset Pricing, Generalized Autoregressive Score Models, Time-Varying Risk Premia, Return Predictability

JEL Codes: G12, G17, C58

1 Introduction

Risk premia for holding financial assets vary over time (Campbell and Shiller (1988), Fama and French (1989), Lettau and Ludvigson (2001), Cochrane (2011)). Traditional factor asset pricing models (e.g. Fama and French (1993) and Carhart (1997)) describe these premia with risk prices (lambdas) demanded by investors for each unit of exposure (beta) to a financial or macroeconomic source of risk. Although the predominant methods for testing asset pricing models, such as the two-step regression procedure of Fama and MacBeth (1973) (referred to hereafter as the FMB procedure), rely on constant lambdas and betas, the appropriateness of this assumption has been widely called into question (Jagannathan and Wang (1996), Ghysels (1998)).

Estimation approaches for conditional factor pricing models typically formulate betas (see, e.g., Ferson and Harvey (1999) and Lettau and Ludvigson (2001)) and lambdas (see, e.g., Adrian et al. (2015), Gagliardini et al. (2016), Adrian et al. (2019)) as functions of predictor variables that generate the time dynamics. These approaches are advantageous for testing asset pricing theories that suggest drivers of risk premia. However, beta estimates are found to be especially sensitive to the choice of predictor variables (Harvey (2001)). Moreover, employing inappropriate explanatory instruments or leaving out relevant ones may yield misleading results if one is interested mainly in filtering time-varying risk premia. Empirical research on dynamic asset pricing would therefore benefit from methods that allow for exploring the risk price and exposure dynamics implied solely by the cross-sectional model specification rather than dynamics prescribed by additional external forecast variables. In turn, knowledge of these agnostic dynamics can offer a better understanding of risk premia dynamics in asset pricing theories.

The method presented in this paper allows for estimating dynamic versions of cross-sectional multifactor asset pricing models without specifying predictor variables for driving time dynamics. In every period, the idea is to evaluate current pricing errors with an observation density, implied by the multifactor model, to identify the mispricing that can be attributed to variation in parameters and update them accordingly. To choose the direction and intensity of this parameter updating, I follow the generalized autoregressive score (GAS)[1] approach developed by Creal et al. (2013) and Harvey (2013), who propose updating parameters in econometric models in the direction of the gradient of the log likelihood, the so-called score, evaluated at the current observation.[2] Hence, parameters are pushed in the direction with the steepest increase in the likelihood function pointed out by the gradient.

This paper's main contribution is the introduction of a general framework for dynamic asset pricing models with GAS parameter dynamics. The resulting score-driven likelihood-based asset pricing model (SD-APM) is applicable for every linear (cross-sectional) factor model that can be analyzed using the traditional FMB procedure and generates latent risk price and exposure

[1] Also referred to as score-driven (SD) model or dynamic conditional score (DCS) model.

[2] GAS models have been applied successfully in numerous applications in time series analysis and financial econometrics. See, for example, Harvey and Lange (2017) and Gorgi et al. (2019) for applications in volatility modeling and Oh and Patton (2018) and Bernardi and Catania (2019) for systemic risk applications.

dynamics. Lambda, beta and factor mean series are estimated from asset returns and cross-sectional risk factor data only. The second main contribution is the derivation of optimal model parameter updating schemes with respect to the observational likelihood. They show an intuitive relation to the classical FMB estimates: The updating corrects lambdas and betas with respect to local cross-sectional pricing errors that are produced when employing FMB with constant parameters and thus can be regarded as a local FMB correction. Moreover, the model setup is extended to accommodate tradeable factors as well as general elliptical distributional assumptions for factor and return innovations. Estimation and inference can be carried out according to the maximum likelihood principle.

I perform a Monte Carlo study to investigate the ability of the SD-APM to filter risk price dynamics in a setting with a realistic signal-to-noise ratio and constant betas. The performance of the SD-APM is compared to that of the dynamic asset pricing model developed by Adrian et al. (2015) (referred to hereafter as DAPM), which employs noisy signals of different strengths of the true instrument factor driving the lambda dynamics of the data-generating process. In an asset return panel of 25 assets and 600 (monthly) time observations, a DAPM would need to be informed with a signal containing more than 60% of the information from the true data-generating process to compete with the in-sample pricing performance of the SD-APM. More than 50% of the information is needed to outperform the SD-APM in out-of-sample risk premium forecasting. The results suggest that SD-APM is a valuable method not only in situations where time series predictors are not available but also in situations where the suitability of these predictors is sufficiently uncertain.

An empirical application to a two-factor model of a cross-section of equity portfolios sorted by size shows that the SD-APM produces considerably lower pricing and forecast errors than traditional unconditional benchmark approaches like FMB regressions. Moreover, the SD-APM is able to produce comparable pricing errors to the dynamic regression-based approach of Adrian et al. (2015), which employs external time series predictors. In a setting restricted to time-constant betas, the SD-APM shows greater flexibility by producing considerably lower pricing errors compared to the dynamic benchmark model.

With the SD-APM, this paper contributes a novel methodology to the extensive literature on the estimation of conditional asset pricing models. The closest and most commonly used methods differ as follows. The DAPM of Adrian et al. (2015) is considered the main benchmark for comparisons. Both DAPM and SD-APM rely on the same beta pricing equation but differ in the construction of risk price dynamics (external forecasters vs. recursive observation-driven updating) and their estimation methodology (three-step linear regressions vs. maximum likelihood). Adrian et al. (2019) employ an enhanced version of the DAPM that allows for nonlinear relations between risk prices and forecast variables. Risk prices and exposures that are affine-linear transformations of instruments are also employed in Gagliardini et al. (2016, 2019), with a focus on large unbalanced panels of individual stock returns. The functioning and intentions of these methods are different from those of the method I propose, as their aim is to filter economically meaningful variation in risk prices explained by forecasting factors in regression-based frameworks. Related to these studies, Creal and Kim (2021) study time-varying risk premia explained by external predictors in a unbalanced panel of individual assets based on Bayesian regression trees. In contrast, the study in this paper investigates the portion of risk premium movement that can be learned from observed cross-sectional model pricing errors with a classical non-Bayesian approach.

Early contributions already allow for instrument-free dynamics of risk prices by conducting cross-sectional FMB regressions period by period, such as Fama and MacBeth (1973) and Ferson and Harvey (1991). The approach presented in this paper differs from the traditional one by explicitly modeling an intertemporal relation between lambdas of different time periods that can be fitted and analyzed; in contrast, period-by-period FMB risk price estimates are not explicitly connected over time and are extremely volatile. The risk price updating mechanism in my approach can be understood as an attempt to infer how much of this risk price volatility stems from actual parameter movements and is not caused by estimation errors.

The paper is organized as follows. Section 2 introduces and discusses the proposed score-driven dynamic asset pricing framework as well as parameter updating schemes for elliptically distributed returns. It closes by laying out a strategy for likelihood-based estimation and inference. A Monte Carlo study evaluating the performance of the SD-APM is conducted in Section 3. The empirical application to a equity portfolio cross-sectional model is presented in Section 4. Section 5 concludes.

2 Theoretical Framework

This chapter introduces a framework for score-driven likelihood-based asset pricing models. The model setup is described first, followed by a derivation for optimal parameter-updating schemes in case of elliptical distributions. Before turning to the applications in the following sections, the estimation strategy is explained and discussed.

2.1 Model Setup

The basic model setup is in line with the one presented in Adrian et al. (2015) but differs in some instances, like the specification of conditional factor means as well as risk price and exposure dynamics, which are driven by scores of the observation density.

Let $r_t = (r_t^1, \ldots, r_t^N)^\top$ denote the N-dimensional vector representing the excess returns of N different assets at time $t \in \{0, \ldots, T\}$. The underlying data-generating process is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a filtration $\mathcal{F}_t = \sigma(\{r_t, \ldots, r_0\})$ representing the set of information available at time t. Suppose the risk in the economy is described in terms of K risk factors covered in the state vector f_t that follows

$$f_{t+1} = \mu_t + u_{t+1}, \qquad t = 0, \ldots, T-1, \tag{1}$$

where $u_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_u)$, and μ_t is a conditional mean process adapted to $\mathcal{F}_t$.

Assume the existence of a unique stochastic discount factor (SDF) m_t that prices every asset

$i \in \{1, \dots N\}$ according to

$$\mathbb{E}_t(m_{t+1} r^i_{t+1}) = 0, \tag{2}$$

where $\mathbb{E}_t$ denotes the conditional expectation with respect to time t information $\mathcal{F}_t$. The Euler equation (2) can be used to compute the conditional covariance between the SDF and the asset return as

$$Cov_t(m_{t+1}, r^i_{t+1}) = -\mathbb{E}_t(m_{t+1})\mathbb{E}_t(r^i_{t+1}). \tag{3}$$

Regressing the demeaned return of asset i on the factor innovations u_{t+1} yields an idiosyncratic noise term $e_{i,t+1}$ that is orthogonal to u_{t+1}. Taken together with (3), the return can be decomposed as

$$r^i_{t+1} = \mathbb{E}_t(r^i_{t+1}) + (r^i_{t+1} - \mathbb{E}_t(r^i_{t+1})) \tag{4}$$

$$= -\frac{Cov_t(m_{t+1}, r^i_{t+1})}{\mathbb{E}_t(m_{t+1})} + \beta_{i,t}^\top u_{t+1} + e_{i,t+1}, \tag{5}$$

where $\beta_{i,t} = \Sigma_u^{-1} Cov_t(u_{t+1}, r^i_{t+1})$ denotes the K-dimensional vector of risk exposures. Let the SDF be affine-linear in the economy's risk factor innovations; that is,

$$\frac{m_{t+1} - \mathbb{E}_t(m_{t+1})}{\mathbb{E}_t(m_{t+1})} = -\lambda_t^\top \Sigma_u^{-1} u_{t+1} \tag{6}$$

with time-variant price of risk vector λ_t of dimension K. Plugging the SDF into the return decomposition (5) yields a standard beta representation given by

$$r^i_{t+1} = \lambda_t^\top \Sigma_u^{-1} Cov_t(u_{t+1}, r^i_{t+1}) + \beta_i^\top u_{t+1} + e_{i,t+1} \tag{7}$$

$$= \beta_{i,t}^\top \lambda_t + \beta_{i,t}^\top u_{t+1} + e_{i,t+1}. \tag{8}$$

The decomposition (8) therefore consists of a predictable risk premium $\beta_{i,t}^\top \lambda_t$ that compensates risk exposures, an unpredictable component $\beta_{i,t}^\top u_{t+1}$ depending on risk factor innovations and an asset-specific innovation term $e_{i,t+1}$. We assume $e_t = (e_{1,t}, \dots e_{N,t})^\top \overset{iid}{\sim} \mathcal{N}(0, \Sigma_e)$.[3] Representation (8) may be interpreted as a system of seemingly unrelated regressions (SUR) with time-varying coefficients and identical regressors that can be stacked to

$$r_{t+1} = \beta_t(\lambda_t + u_{t+1}) + e_{t+1}, \tag{9}$$

where $\beta_t = (\beta_{1,t}, \dots, \beta_{N,t})^\top$. What essentially distinguishes the setting proposed here from those of prior studies is the specification of the dynamics of the time-varying $(N+2)K$-dimensional[4] parameter vector $\theta_t = (\mu_t^\top, \lambda_t^\top, vec(\beta_t)^\top)^\top$. Adrian et al. (2015) assume that the conditional mean of f_t is described by a vector autoregressive model of order one, λ_t is affine-linear in a set of predictor variables that have to be specified and derive time-varying betas with a non-parametric

[3] It is later shown that the Gaussianity assumptions for e_t and u_t can be relaxed.

[4] K factor means, K lambdas plus N betas per lambda.

kernel-based approach. In contrast, the approach discussed here will suggest driving the whole set of dynamic parameters by recent observations from the return panel and cross-sectional pricing factors f_t.

2.2 Score-Driven Risk Premia

The GAS model proposed by Creal et al. (2013) provides an opportunity to introduce time variation into general models with specified observation densities. The basic idea is to let the time-varying parameters of a model be updated proportionally to the score of the observation density; that is, the derivative of the logarithmic density with respect to the parameter that should become time-varying. Thus, the parameter vector is pushed in the direction indicated by the gradient. This is the direction in which the update would yield the steepest increase in the observation density. The approach can therefore be understood as a parameter update optimizing the local likelihood in period t.

In the framework described above, asset returns r_t and risk factor realizations f_t are observed and can be evaluated with the conditional observation density $p(r_t, f_t|\mathcal{F}_{t-1}, \theta_{t-1})$, where $\theta_t = (\mu_t^\top, \lambda_t^\top, vec(\beta_t)^\top)^\top$ is the vector of time-varying parameters.[5] Given this specification, the GAS updating scheme for the dynamic vector of risk prices and exposures is then given by:

$$\theta_t = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j \theta_{t-j} \tag{10}$$

with

$$s_t = S_t^{-1} \nabla_t, \qquad \nabla_t = \begin{pmatrix} \nabla_t^\mu \\ \nabla_t^\lambda \\ \nabla_t^\beta \end{pmatrix} = \begin{pmatrix} \frac{\partial \ln p(r_t, f_t|\mathcal{F}_{t-1}, \theta_{t-1})}{\partial \mu_{t-1}} \\ \frac{\partial \ln p(r_t, f_t|\mathcal{F}_{t-1}, \theta_{t-1})}{\partial \lambda_{t-1}} \\ \frac{\partial \ln p(r_t, f_t|\mathcal{F}_{t-1}, \theta_{t-1})}{\partial vec(\beta_{t-1})} \end{pmatrix} \tag{11}$$

and

$$S_t = \begin{pmatrix} \mathbb{E}_{t-1}\left(\nabla_t^\mu \nabla_t^{\mu\top}\right) & \mathbb{E}_{t-1}\left(\nabla_t^\mu \nabla_t^{\lambda\top}\right) & 0 \\ \mathbb{E}_{t-1}\left(\nabla_t^\lambda \nabla_t^{\mu\top}\right) & \mathbb{E}_{t-1}\left(\nabla_t^\mu \nabla_t^{\mu\top}\right) & 0 \\ 0 & 0 & \mathbb{E}_{t-1}\left(\nabla_t^\beta \nabla_t^{\beta\top}\right). \end{pmatrix} \tag{12}$$

We call a model comprising of equations (1), (9), and (10) to (12) the (Gaussian) **S**core-**D**riven **A**sset **P**ricing **M**odel of orders p and q (**SD-APM(p,q)**).

Equation (10) determines the updating mechanism for the time-varying parameter θ_t. Here, ω is an $(N+2)K$-dimensional vector of intercepts, and A_i, B_j are $(N+2)K \times (N+2)K$ matrices of coefficients for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The updating process therefore consists of a constant part, an adjustment due to a innovation sequence s_t, and an autoregressive part. The

[5] Consider that the time-varying parameter in Creal et al. (2013) is denoted with another subindex. Here, the subindex is shifted to harmonize the notation with the baseline conditional asset pricing model of Section 2.1.

centerpiece of score-driven models is the specification of the innovation sequence s_t, that is set proportional to the scores ∇_t^{μ}, ∇_t^{λ} and ∇_t^{β} defined in (11). The score sequences are scaled with the matrix sequence S_t that is chosen as the inverse of a restriction of the Fisher information matrix $\mathbb{E}_{t-1}\left(\nabla_t \nabla_t^{\top}\right)$. This choice has the advantage that the scaling depends directly on the variance of the score.[6] Note that the general Fisher information matrix including all GAS parameters is restricted in the SD-APM such that the cross-information quantities with respect to beta are zero, that is, $\mathbb{E}_{t-1}\left(\nabla_t^{\mu}\nabla_t^{\beta\top}\right) = \mathbb{E}_{t-1}\left(\nabla_t^{\lambda}\nabla_t^{\beta\top}\right) = 0$. This modeling choice is made because it generates simpler and more parsimonious models with less cumbersome derivations. Moreover, the central object in GAS models that carries update information is the score, while the choice of the scaling matrix is to some extent arbitrary. Therefore, it can be assumed that the update is reasonably resilient with respect to the constraint. Results from simulations and the empirical application support this conjecture. Note that the simplification does not mute the potential impact of the lambda (beta) score on the beta (lambda) updating as long as the corresponding entry in the parameter matrix A_i is not zero.

An apparent alternative to specifying score-driven risk parameters would be to model s_t in (10) as Gaussian i.i.d. innovation to achieve so-called parameter-driven dynamics. This would resemble a variant of the popular Kalman filter if either λ_t or β_t is constant. The crucial difference, however, is that s_t would be another independent source of randomness, whereas the mechanism in (11) relates s_t to the innovations e_t and u_t that already exist in the model. The mechanism therefore captures systematic variation in idiosyncratic and factor innovations, signaled by current observations, to generate updates in risk exposures and prices. This allows time-varying parameters in the SD-APM to be perfectly predictable given past information. The latter feature is particular appealing in an asset pricing framework, as risk prices and exposures are assumed to be pre-determined in most asset pricing theories. Moreover, the observation-driven structure of parameter dynamics facilitates the likelihood evaluation considerably.

Proposition 1. Let f_t and r_t be the factors and the returns of an Gaussian SD-APM(p,q), respectively. The scaled score series for the dynamic parameter updating is then given by

$$s_t = \begin{pmatrix} s_t^{\mu} \\ s_t^{\lambda} \\ s_t^{\beta} \end{pmatrix} = \begin{pmatrix} f_t - \mu_{t-1} \\ (\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^{\top}\Sigma_e^{-1}r_t - \lambda_{t-1} \\ \left((\lambda_{t-1}\lambda_{t-1}^{\top} + \Sigma_u)^{-1}(\lambda_{t-1} + u_t)\right) \otimes e_t \end{pmatrix} \tag{13}$$

Proof. See Appendix A.1.

Intuitively, Proposition 1 suggests that the factor mean should be updated according to the deviation of the factor realization to the current factor mean belief. The scaled scores with respect to λ, denoted as s_t^{λ} in (13), can be regarded as generalized least squares (GLS) regression errors

[6] Other popular scalings proposed in Creal et al. (2013) use the identity matrix or the Cholesky factor of the inverse Fisher information. The latter choice may be attractive for achieving a unit variance of the scaled score process s_t.

from cross-sectionally regressing r_t on β_{t-1}. The driving mechanism therefore intuitively corrects local deviations in the cross-sectional fit. This observation can be related to the widely used FMB regressions, in which constant risk prices are estimated with a cross-sectional regression of the average portfolio returns $\overline{r}$ on their (costant) betas; that is, the risk price estimate is given by $(\beta^\top \Sigma_e^{-1}\beta)^{-1}\beta^\top \Sigma_e^{-1}\overline{r}$. The correction step proposed by (10) in conjunction with (13) for the lambda essentially enforces a drift to this average prescription. Finding that λ_{t-1} is greater than $(\beta_{t-1}^\top \Sigma_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^\top \Sigma_e^{-1} r_t$ would indicate that the current risk price is too high. The resulting s_t^λ would therefore be negative and would downsize the risk price for the next period. The coefficient matrices $A_1, \ldots A_p$ reveal how much of the recent cross-sectional regression errors can potentially be attributed to a change in risk prices and not to factor or idiosyncratic innovations.

The discussed updating mechanism underlines the difference between the proposed score-driven method and the time-varying lambda framework of Fama and MacBeth (1973), which would choose λ_t in each period to minimize the cross-sectional regression error, hence choosing λ_t such that $s_t = 0$. This updating is nested in the SD-APM but comes with the drawback that time-varying lambdas become unrealistically volatile because cross-sections are fitted independently period by period and do not draw information from connections between lambdas of different time periods. The constant lambda framework of Fama and MacBeth (1973) averages these lambda series to achieve risk price estimates. Its model setting is recovered if innovations are uncorrelated and $A_i = B_j = 0$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. Although the maximum likelihood estimator of the SD-APM prescribes a cross-sectional regression based on betas for estimating risk prices in the time-invariant setting, it differs from the FMB estimator, as betas and lambdas are estimated simultaneously and and not in a two-step procedure.

Also notable are the similarities to the popular generalized conditional heteroscedastictiy (GARCH) time series models developed by Engle (1982) and Bollerslev (1986). If the time-varying parameter in (10) was the conditional variance of a zero-mean time series ε_t (that is, $\theta_t = \sigma_t^2$) and the score was the squared current observation (that is, $s_t = \varepsilon_t^2$) the famous GARCH(p,q) updating equation would be obtained.[7] This updating process is quite intuitive because, in a setting with a constant variance parameter, this would be estimated via the mean of squared observations. The appropriate statistic of new incoming information ε_t^2 in relation to the past observation therefore indicates whether it is likely that the variance has increased (if ε_t^2 is high relative to the current variance estimate) or decreased (if ε_t^2 is relatively low). SD-APM works according to the same principle, as $(\beta_{t-1}^\top \Sigma_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^\top \Sigma_e^{-1} r_t$ summarizes the relevant incoming information in period t for estimating risk prices.

The score-driven updating sequence for the β, that is s_t^β in (13), follows a similar intuition as the lambda updating. It is the local error from regressing the return r_t on the stochastic regressor $\lambda_{t-1} + u_t$. This becomes plausible when acknowledging from equation (9) that β_{t-1} can be interpreted as the corresponding regressor. Moreover, in section 2.4.3. it is shown that this updating relates to the well-known first-stage time series regressions on factors if the latter are

[7] Creal et al. (2013) argue that the Gaussian GARCH(1,1) model is indeed a GAS model.

assumed to be traded.

2.3 Estimation and Inference

One major advantage of the SD-APM is that its likelihood function can be evaluated directly. The set of static parameter vectors and matrices to be estimated is given by $\omega, A_1, \dots A_p, B_1, \dots B_q, \mu, \Phi, \Sigma_e$ and Σ_u. This collection of parameters, stacked in the parameter vector ϑ, needs to be chosen to maximize the (conditional) log-likelihood function given by

$$\mathcal{L} = \sum_{t=1}^{T} \ln p(r_t, f_t | \mathcal{F}_{t-1}, \theta_{t-1}, \vartheta) \tag{14}$$

$$= \sum_{t=1}^{T} \ln \left(p(r_t | f_t, \mathcal{F}_{t-1}, \theta_{t-1}, \vartheta) p(f_t | \mathcal{F}_{t-1}, \theta_{t-1}, \vartheta) \right) \tag{15}$$

$$= -T(N+K) \ln \pi - \frac{T}{2} \ln |\Sigma_e| - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} e_t^\top \Sigma_e^{-1} e_t - \frac{1}{2} \sum_{t=1}^{T} u_t^\top \Sigma_u^{-1} u_t. \tag{16}$$

Besides the problems associated with the enormous number of parameters, closed-form solutions of the ML estimators are not available due to the strong dependencies of the parameters on each other. This makes it necessary to employ numerical optimization procedures to optimize (16).

The number of parameters can be crucially reduced when assuming cross-sectionally homoscedastic and independent errors e_t. This additionally facilitates the derivation of the risk price updating scheme in (13), which henceforth simplifies to $s_t^\lambda = (\beta^\top \beta)^{-1} \beta^\top r_t - \lambda_{t-1}$. Because of the missing occurrence of Σ_e in the updating scheme, the maximum likelihood estimator of the covariance matrix is given by $\hat{\Sigma}_e = \frac{1}{T} \sum_{t=1}^{T} e_t e_t^\top$. This approach may weight innovations in the updating scheme incorrectly, especially if idiosyncratic errors are highly correlated. However, one can construct a correction by running a second maximum likelihood estimation with a prespecified dispersion matrix estimated from a first estimation stage with homoscedastic and independent errors e_t in the spirit of feasible GLS estimation approaches. Another promising approach is to compute a prior covariance matrix estimate $\hat{\Sigma}_e$ based on pricing errors from Fama and MacBeth (1973). The results turn out to be close to the two-stage approach, as the variation of time-varying parameters is relatively low compared to the error variances.

Inference is conducted based on the inverse Hessian of the log-likelihood evaluated at the optimum, as suggested by Creal et al. (2013) for GAS models. If ϑ stacks all the static parameters of the model, standard asymptotic theory for ML estimators establishes asymptotic normality under some regularity conditions, in particular

$$\sqrt{T} \left(\hat{\vartheta} - \vartheta \right) \xrightarrow{d} \mathcal{N} \left(0, \mathcal{I}^{-1}(\vartheta) \right), \tag{17}$$

with Fisher information matrix $\mathcal{I}(\vartheta) := -\mathbb{E} \left(\partial^2 l_t / \partial \vartheta \partial \vartheta^\top \right)$, where l_t is the log-likelihood contribution of the i-th observation evaluated at ϑ. Blasques et al. (2022) provide general conditions for consistency and asymptotic normality of the ML estimator for correctly specified univariate time series models with a single time-varying parameter. However, a formal proof of this theoretical result for the multivariate SD-APM is beyond the scope of this paper. Because the use of a Gaussian SD-APM is a potential source of model misspecification, it could be advisable to rely on quasi-ML standard errors that can be derived with

$$\sqrt{T}\left(\hat{\vartheta}-\vartheta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}^{-1}(\vartheta)\mathcal{J}(\vartheta)\mathcal{I}^{-1}(\vartheta)\right), \tag{18}$$

where $\mathcal{J}(\vartheta) := \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\left(\partial\mathcal{L}(\vartheta)/\partial\vartheta\right)\left(\partial\mathcal{L}(\vartheta)/\partial\vartheta\right)^\top\right)$. Standard errors in the following empirical application are derived by numerically differentiating the score function with a finite difference approximation.

2.4 Alternative Specifications and Extensions

The following sections present some alternative specifications and extensions of the baseline Gaussian SD-APM.

2.4.1 Traded Factors

Many asset pricing models include factors that are themselves returns and are therefore traded. A well-known example is the CAPM. The pricing equation (8) would then also apply to these factors. An immediate implication is then that factor means are equal, that is, $\mu_t = \mathbb{E}\left(f_{t+1}\right) = \lambda_t$. Imposing this identity in the SD-APM framework causes the factor model to become

$$f_{t+1} = \lambda_t + u_{t+1} \tag{19}$$

and the beta representation to become

$$r_{t+1} = \beta_t(\lambda_t + u_{t+1}) + e_{t+1} = \beta_t f_{t+1} + e_{t+1}. \tag{20}$$

Hence, assets are priced with an regression model, where the risk factors in f_t serve as regressors and risk prices enter the model as conditional factor means. Score-driven updating schemes for this restricted model can be derived almost analogously to the unrestricted case. A crucial difference is that Fisher information does not have to be restricted in order to get feasible updating schemes because $\mathbb{E}_{t-1}\left(\nabla_t^\lambda \nabla_t^{\beta\top}\right) = 0$ in this setting. Proposition 2 provides the scaled score series for the parameter updating.

Proposition 2. Let f_t and r_t be the traded factors and returns of an Gaussian SD-APM(p,q), respectively. The scaled score series for the risk price and exposure updating is then given by

$$s_t = \begin{pmatrix} s_t^\lambda \\ s_t^\beta \end{pmatrix} = \begin{pmatrix} f_t - \lambda_{t-1} \\ \left((\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u)^{-1} f_t\right) \otimes e_t \end{pmatrix}. \tag{21}$$

Proof. See Appendix A.2.

It is striking that the dynamics of risk prices are no longer driven by the cross-sectional regression errors but exclusively by deviations from the current factor realization. Therefore, the score approach proposes to use only the information from the role of risk prices as factor means. This seems sensible in theory, as other returns in (20) are composed of the factor realization as signal of the time-varying factor mean plus idiosyncratic noise and therefore do not carry additional information. Thus, it is quite reasonable that the score approach suggests the use of the signal f_t only. However, this can lead to problems in practical applications, as many asset pricing factors have an extremely low signal-to-noise ratio. This may imply that the deviation from the current realization is only a weak signal for the direction of the parameter update. This could explain the relatively weak performance of this model specification in the following simulations and application, although the factors themselves are returns.

In terms of beta updating, we see that the sequence s_t^β is identical to the case without the traded factor assumption, given the identity (19). It can still be interpreted as an error from regressing current returns on the stochastic regressor f_t.

2.4.2 Elliptically Distributed Returns and Factors

In order to compute the score ∇_t, assumptions on the distributions of the innovations u_t and e_t have to be made. The basic SD-APM developed so far assumes that these innovations are normally distributed, which means that the returns and factors are also assumed to be (conditionally) normally distributed. This could be problematic because empirical distributions of returns often show departures from the normal distribution. In particular, the tails of the return distribution tend to be heavier than in the Gaussian case, which is known as excess kurtosis. To allow for additional flexibility in fitting the data while keeping the framework tractable, I consider the distribution families from the general class of elliptical distributions[8] for an alternative SD-APM specification. The general n-dimensional density of such distributions can be formulated as

$$p(x) = |\Omega|^{-\frac{1}{2}} \psi \left((x-\mu)^\top \Omega^{-1} (x-\mu) \right), \tag{22}$$

with mean $\mu \in \mathbb{R}^n$, dispersion matrix $\Omega \in \mathbb{R}^{n \times n}$, and a characteristic density generator $\psi : \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$. Denote with $X \sim \mathcal{E}_n(\mu, \Omega, \psi)$ that a random variable X possesses a density given by (22). The elliptical class includes not only the normal distribution, as can be seen by choosing the generator to be $\psi(x) = (2\pi)^{-N/2} \exp(-x/2)$, but also many other distributions that account for excess kurtosis and are commonly used to fit financial returns, such as the Laplace and the Student's t-distribution.

We define SD-APM(p,q) in which the innovations are not (necessarily) Gaussian but follow a common elliptical distribution as Elliptical SD-APM(p,q). In particular, we require that

[8] See Fang et al. (1990) or Chapter 6 of Embrechts et al. (2015) for a comprehensive treatment of elliptical distributions.

$(u_t^\top, e_t^\top)^\top \sim \mathcal{E}_{N+K}(0, \Omega, \psi)$ with

$$\Omega = \begin{pmatrix} \Omega_u & 0 \\ 0 & \Omega_e \end{pmatrix} \tag{23}$$

and a characteristic density generator ψ. Note that u_t and e_t are still orthogonal as required in the baseline model but not necessarily independent as in the Gaussian case. Proposition 3 provides the implied factor mean and risk price score sequences for the elliptical model specification.

Proposition 3. Let $x_t = (f_t^\top, r_t^\top)^\top$ be the stacked factors and returns of an Elliptical SD-APM(p,q) with density generator ψ. The scaled score series for the factor mean and risk price updating is then given by

$$\begin{pmatrix} s_t^\mu \\ s_t^\lambda \end{pmatrix} = C(\|\tilde{x}_t\|^2, \psi) \begin{pmatrix} f_t - \mu_{t-1} \\ (\beta_{t-1}^\top \Omega_e^{-1} \beta_{t-1})^{-1} \beta_{t-1}^\top \Omega_e^{-1} r_t - \lambda_{t-1} \end{pmatrix} \tag{24}$$

with

$$C(\|\tilde{x}_t\|^2, \psi) = \frac{-(N+K)\frac{\psi'(\|\tilde{x}_t\|^2)}{\psi(\|\tilde{x}_t\|^2)}}{2\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2 \left(\frac{\psi'(\|\tilde{x}_t\|^2)}{\psi(\|\tilde{x}_t\|^2)}\right)^2\right)} \tag{25}$$

and $\|\tilde{x}_t\|^2 = (x_t - \mathbb{E}_{t-1}(x_t))^\top \Omega_x^{-1} (x_t - \mathbb{E}_{t-1}(x_t))$.

Proof. See Appendix A.3.

A derivation of the beta updating sequence based on the elliptical model is not feasible because the betas enter the covariance matrix of x_t. This makes the derivation of Fisher information cumbersome in the multivariate setting and highly dependent on the shape of the particular distribution. A straightforward way to include time-varying betas in the elliptical model is to replace the Fisher information with the identity matrix that yields the beta score given by

$$s_t^\beta = -\frac{\psi'(\|\tilde{x}_t\|^2)}{2\psi(\|\tilde{x}_t\|^2)} \left((\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u)^{-1} f_t\right) \otimes e_t. \tag{26}$$

Another straightforward solution would be to numerically compute the Fisher information matrix with respect to β. With regard to the factor mean and risk price updating in Proposition 3, it is striking that the particular choice of the elliptical return distribution does not alter the direction of both the mu and the lambda score and therefore also does not alter the direction of the factor mean and risk price parameter updating. Only the scalar function C depends on the distributional shape represented by the density generator ψ. Updating steps are therefore scaled differently according to the specified distribution. The following corollary collects particular scaling functions for normal and the student-t distribution.

Corollary 1. *The scaling in* (25) *is given by*

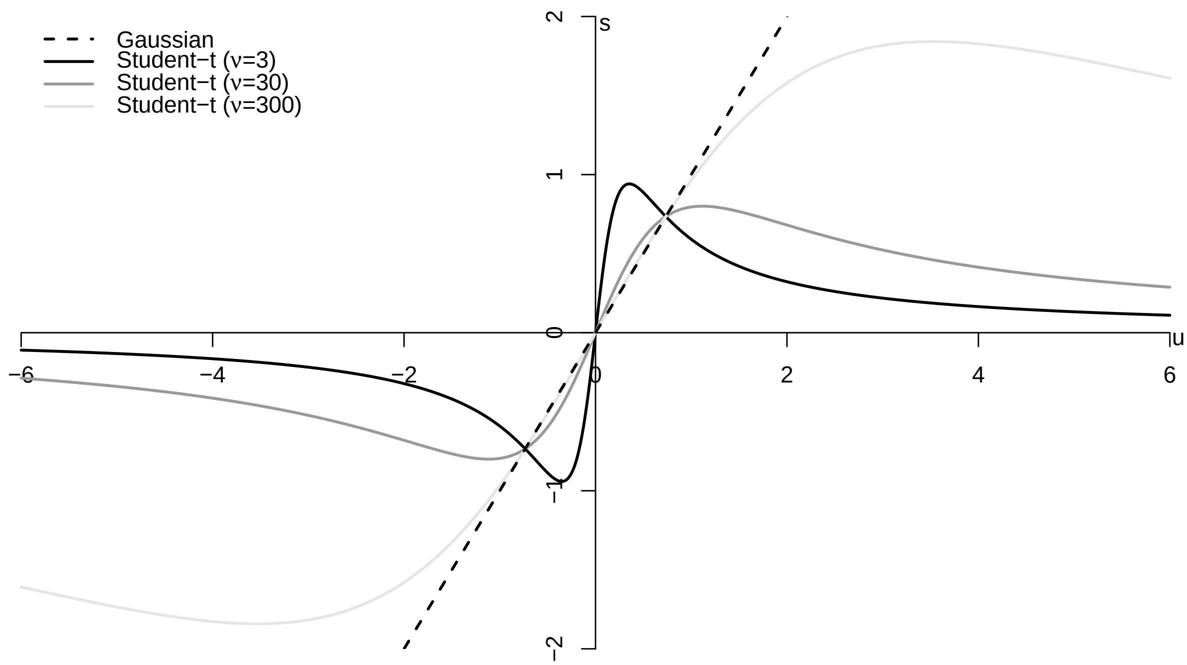


Figure 1: **Factor Mean Score** The figure shows the factor mean score s_t^{μ} as a function of $u_t = f_t - \mu_{t-1}$ for different elliptical distributions. The number of assets and factors is chosen as $N = 10$ and $K = 1$, respectively. Covariance matrices are fitted on a return panel of portfolios sorted by size an a market return factor.

(i) $C(\|\tilde{x}_t\|^2, \psi) = 1$ *if* $(u_t^\top, e_t^\top)^\top \overset{iid}{\sim} \mathcal{N}(0, \Omega)$.

(ii) $C(\|\tilde{x}_t\|^2, \psi) = \frac{\nu+N+K+2}{\nu+\|\tilde{x}_t\|^2}$ *if* $(u_t^\top, e_t^\top)^\top \overset{iid}{\sim} t_\nu(0, \Omega)$.

Proof. See Appendix A.4.

Part (i) of Corollary 1 just confirms that Proposition 3 is in line with Proposition 1. Hence, a particularly simple updating scheme is achieved when assuming normally distributed innovations because the scaling is equal to unity in this case. The derivation of the scaling is more cumbersome for the majority of elliptical distributions, but Proposition 1 indicates that the cross-sectional regression errors should be informative for driving the parameter dynamics irrespective of the true elliptical distribution. Setting the scaling C to unity can therefore crucially improve tractability, while the updating still works in the correct direction. This reasoning could also be used to justify the use of the Gaussian beta updating. However, the gained tractability comes at the cost of incorrectly weighting the magnitude of observed model errors on parameter updates, if the data-generating process is not Gaussian.

The effect of the scaling becomes clear when considering the Student's t-updating sequences presented in Corollary 1(ii). Figure 1 shows the factor mean score as a function of a demeaned factor observation $u_t = f_t - \mu_t$ for a SD-APM with different elliptical distributions. Covariance matrices are fitted on a return panel of portfolios sorted by size and a market return factor that are also used in the following empirical application. The Gaussian score is the 45° line, and new observations therefore impact the updating according to their magnitude. Regarding the Student-t distribution, the score shows a steeper ascent for observations close to the mean but approaches zero for $|u_t| \to \infty$. Hence, the score gives more weight to observations close to the mean and less weight to those that crucially deviate from the mean. This down-weighting is a result of dividing the observation by $\|\tilde{x}_t\|^2$ in the Student-t scaling function C in Corollary 1(ii). The updating scheme therefore takes the more pronounced tails of the return and factor distributions into account and dampens the impact of outliers. This is also in line with the observation that the effect is less pronounced for a higher degrees of freedom parameter ν. The discussion for the factor mean can be applied analogously to the risk price score s_t^{λ}. Hence, using the Student-t distribution makes the score-driven parameter dynamics more robust to outliers in the data. This latter feature has also been documented and discussed for score models of location (Creal et al. (2014), Harvey and Luati (2014)) and scale (Creal et al. (2011)).

3 Monte Carlo Study

In this section, the performance of the constant-beta SD-APM is evaluated with a Monte Carlo study on forecasting risk prices and excess returns. The central question is whether the SD-APM is able to filter risk price movements without knowing the driving forces and how well it can compete with models making use of information about the predictable component. The simulation setting assumes time-constant betas for two reasons. First, the estimation with time-varying betas is highly computationally demanding in panels with $N > 10$ because the introduction of every additional asset requires taking account of K additional time-varying beta parameters. This would render reliable Monte Carlo results with a reasonable number of replications infeasible, at least for larger panels. Moreover, empirical studies such as Braun et al. (1995) and Ghysels (1998) document that changes in betas, in contrast to lambdas, are rather slow, if they occur at all. However, time-varying betas are allowed for in the following empirical application, and Appendix B includes Monte Carlo results for a particular cross-section with N=10, showing that the SD-APM is capable of filtering beta dynamics as well.

3.1 Data-Generating Process

Assume there is one (cross-sectional) risk factor f_t and a forecasting factors z_t such that

$$r_{t+1} = \beta(\lambda_0 + \Lambda_1 z_t) + \beta u_{t+1} + e_{t+1}, \tag{27}$$

where u_{t+1} is the innovation to the risk factors f_{t+1} from the factor model (1). The pricing equation employed for simulations is chosen in line with the DAPM modeling approach of Adrian et al. (2015) with exactly one risk factor f_t, where the dynamic risk price is an affine-linear function of the forecasting factor z_t, i.e., $\lambda_t = \lambda_0 + \Lambda_1 z_t$.[9] This poses a challenge for the SD-APM procedure,

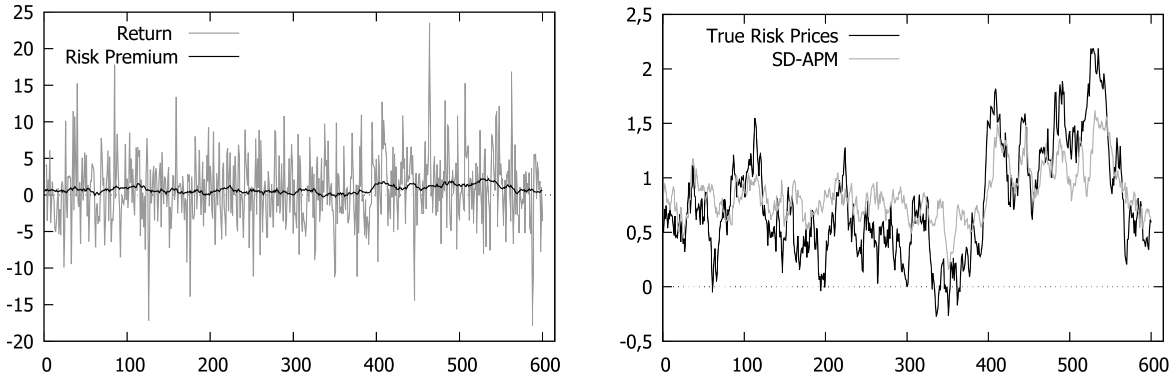
[9] Adrian et al. (2015) explicitly include the possibility that factors can be both risk and forecasting factors simultaneously. In the simulation study, I abstain from doing so for simplicity.

which has to prove itself within the framework of the competing DAPM approach. In order to generate realistic returns, the DGP is calibrated with a cross-section of 10 portfolios sorted by size from the Kenneth French data library. This cross-section is also used in the following empirical application.

To get a candidate process for the forecasting factor, a DAPM with three forecasting factors is fitted on the size portfolio cross-section. The first two forecasting factors under consideration are the 10-year treasury yield as well as the term spread, computed as the difference between the yields of the 10-year treasury note and the three-month treasury bill. Both series are obtained from the H.15 statistical release of the Board of Governors of the Federal Reserve System. The third forecasting factor is the dividend yield of the $S\&P$-500 index. All three series have been reported to predict stock returns in the past[10] and are also used in the following empirical application. For simplicity, the three predictors are combined to form one forecast process by computing the linear combination of the three weighted by coefficients from the fitted DAPM. The forecasting factor series shows a high degree of persistence, with an estimated AR(1) coefficient of 0.9796. Hence, the dynamics are well described by an AR(1) process given by $z_{t+1} = 0.5 + 0.98(z_t - 0.5) + u_{z,t+1}$ with $u_{z,t+1} \overset{iid}{\sim} \mathcal{N}\left(0, 0.137^2\right)$ that is used for simulation in the following. The process always starts with its unconditional mean, i.e., $z_0 = 0.5$. Based on using the combined forecasting factor processes in a one-factor DAPM of the size portfolio cross-section, the risk price parameters are calibrated with $\lambda_0 = 0$ and $\Lambda_1 = 1$. This means that $\lambda_t = z_t$ for every t.

If the risk factor is tradeable, its conditional mean should relate to the risk prices $\lambda_t = z_t$. The risk factor is therefore simulated from $f_t = 0.5z_t + u_t$, where the coefficient is chosen based on regressing the Small-Minus-Big (SMB) factor from the Kenneth French data library on the combined forecasting factor. Returns are derived from the beta representation according to (27), with N being the number of assets in the panel and T being the number of time observations. The betas are equidistantly spread over the interval $[0.6, 1.5]$, which is the range commonly observed for portfolio exposures in models with only one factor. Factor innovations u_t and idiosyncratic errors e_{it} are drawn from a common zero-mean Student t-distribution with covariance matrix according to (23). Motivated by the results in the empirical application below, the degrees of freedom parameter is set to $\nu = 6$. Covariances are set to $\Sigma_u = 18$ and $\Sigma_e = 4I_N$ in order to match the variance of the size-sorted portfolio returns in the empirical application. Note that this calibration yields a realistically low signal-to-noise ratio. In particular, the λ_t process to be filtered has a variance of about 0.48, which is very low relative to the variances of the factor and the idiosyncratic innovations. For illustration, Figure 2(a) shows a simulated return of an asset with a risk factor exposure of 1 from a panel with $N = 10$ and $T = 600$ together with its conditional risk premium given by $\beta\lambda_{t-1}$. The rather small magnitude of the risk premium fits the observation that stock returns show little if any predictability.

[10] See the data section of the empirical application for more information.



(a) Simulated Return Series

(b) Simulated Risk Price Series

Figure 2: **Simulated Return and Lambda.** The figure shows simulated excess returns together with the conditional expectation $\beta\lambda_{t-1}$(panel (a)) of an asset with $\beta = 1$ from one draw of a panel with $N = 10$ assets and $T = 600$ observations. Panel (b) shows the associated lambda series together with a SD-APM(1,1) estimated series.

3.2 Models and Benchmarks

The simulated data are used to evaluate the performance of a (Gaussian) SD-APM(1,1) with score-driven factor means and risk prices but time-constant betas. Two further variants of the SD-APM are considered. The first is the specification discussed in section 2.4.1 assuming the pricing factor f_t is perfectly traded. A second alternative model specification assumes that the innovation terms u_t and e_t are commonly t-distributed, as discussed in section 2.4.2, and is referred to as t-SD-APM.

An empirical analysis with the SD-APM requires a careful choice of starting values for the optimization routine and potentially additional parameter restrictions to ensure convergence of the optimization and to reach a global maximum. This is particular challenging in a simulation study in which such an estimation needs to be done for thousands of simulated panels. Because an individual model specification for every Monte Carlo draw is infeasible, some adequate parameter restrictions are imposed on matrices A and B that would also likely be imposed in an actual empirical application. A rather stable SD-APM specification can be reached by restricting the factor mean equation to an purely autoregressive process. Moreover, this levels the playing field with the benchmark models that treat factor returns with an autoregression, such that improvements in pricing and forecast errors can merely be attributed to an improved estimation of risk price dynamics. In addition, it is worthwhile to restrict the risk price updating such that the coefficient with respect to s_t^{μ} is the negative of the one with respect to s_t^{λ}. This allows further adjusting risk premium dynamics according to innovations in the factor time series model.

A risk price series from a simulated panel of $N = 10$ assets and $T = 600$ observations, together with the risk prices filtered with a SD-APM(1,1), is shown in Figure 2(b). The SD-APM seems to correctly anticipate updating directions and turning points but with some delay, as information in the forecast factor becomes available via pricing errors in the subsequent time period.

I consider the unconditional FMB regression estimates to be the first benchmark method for investigating the gain from introducing dynamics into risk prices. The DAPM of Adrian et al. (2015) is considered as dynamic benchmark model. It models risk prices as affine-linear functions of risk price forecasters as in (27) and is estimated with a three-pass regression approach. Because the part of the data-generating process concerning the risk price dynamics completely follows the same specification, the DAPM estimator has a trivially high information advantage over the other approaches when using the correct forecaster z_t as the explanatory variable. To address this issue, the DAPM is estimated with a diffuse signal of the true forecast factor realization, ranging from pure noise to the true signal. The DAPM is therefore estimated with a signal $\tilde{z}_t^{\kappa}$ drawn from

$$\tilde{z}_t^{\kappa} = \kappa z_t + (1-\kappa)\epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_z^2), \tag{28}$$

where $\kappa \in [0, 1]$. For $\kappa = 0$, the signal exhibits no information about the true factors, and for $\kappa = 1$ the DAPM exploits the true forecasting factor series. The results will of course depend heavily on the variance of the noise term ϵ_t. For better comparability, I set this variance parameter to be equal to the variance of the true predictor process z_t. The $\tilde{z}_t^{\kappa}$ is not used in SD-APM estimation because this model does not make use of any external forecaster information in z_t.

3.3 Simulation Results

This section presents the results of the estimated SD-APM specifications and benchmark models. The central question is whether the SD-APM is able to filter risk premia movements without knowing the driving forces and how well it can compete with models making use of information about the predictable component.

3.3.1 In-Sample Pricing and Forecasting

The first metric for evaluating the in-sample performance is the root mean squared pricing error (RMSE), which is computed as

$$RMSE_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{i,t}^2} \tag{29}$$

for asset i. It reflects the average return share of asset i that cannot be explained by the (cross-sectional) factor model. Table 2(a) shows the RMSEs averaged across N assets and $S = 2500$ replications. The number of time observations is given by T. The starting value of the risk price λ_0 is initialized with the unconditional regression estimate, and the first 50 time observations are discarded before computing pricing errors in order to mitigate the influence of the starting value. Not surprisingly, the SD-APM shows lower pricing errors than the static FMB benchmark across all panels. Comparing the SD-APM to the DAPM reveals that the latter always performs better if $\kappa > 0.7$, or put differently, if more than 70% of the information in signal $\tilde{z}_t$ results from the true predictor. The breakeven κ at which the two models perform on par, slightly varies across

Table 1: Pricing and Forecast Error Comparison

The table shows the average root mean squared pricing errors ($RMSE$) and root mean squared forecast errors ($RMSFE$) of a Gaussian SD-APM(1,1), a SD-APM(1,1) with traded factor and a t-SD-APM(1,1) with t-distributed innovations as well as competing benchmark models. The eight simulated panels have different numbers of assets N, time observations T and are replicated 2500 times each. Benchmarks are a constant risk price specification estimated fitted with the unconditional approach of Fama and MacBeth (1973) (FMB) and the regression-based dynamic asset pricing model of Adrian et al. (2015) (DAPM). The share of information from the correct forecast variable made available to the DAPM is denoted with κ.

		N=10			N=25			N=100		
		T= 300	600	1200	T=300	600	1200	T=300	600	1200
(a) Average RMSE										
SD-APM		2.016	2.014	2.013	2.012	2.012	2.010	2.011	2.008	2.007
SD-APM	(traded)	2.041	2.048	2.049	2.038	2.047	2.050	2.040	2.045	2.050
t-SD-APM		2.017	2.015	2.013	2.015	2.013	2.010	2.014	2.010	2.008
FMB		2.020	2.027	2.032	2.018	2.027	2.032	2.019	2.026	2.031
DAPM	$\kappa = 0.0$	2.030	2.033	2.035	2.029	2.033	2.034	2.029	2.032	2.034
	0.1	2.029	2.033	2.034	2.028	2.033	2.034	2.029	2.031	2.034
	0.2	2.028	2.031	2.033	2.028	2.031	2.033	2.028	2.030	2.032
	0.3	2.026	2.028	2.030	2.026	2.029	2.030	2.026	2.027	2.029
	0.4	2.023	2.024	2.025	2.022	2.024	2.025	2.023	2.023	2.024
	0.5	2.019	2.019	2.019	2.018	2.019	2.019	2.019	2.018	2.018
	0.6	2.014	2.013	2.013	2.014	2.014	2.013	2.015	2.013	2.012
	0.7	2.011	2.009	2.008	2.010	2.009	2.008	2.011	2.008	2.007
	0.8	2.008	2.006	2.005	2.007	2.006	2.005	2.008	2.005	2.004
	0.9	2.006	2.004	2.003	2.005	2.004	2.003	2.007	2.003	2.002
	1.0	2.006	2.004	2.003	2.005	2.004	2.003	2.006	2.003	2.002
(b) Average RMSFE										
SD-APM		4.916	4.921	4.932	4.904	4.919	4.928	4.909	4.911	4.916
SD-APM	(traded)	4.921	4.941	4.955	4.908	4.943	4.956	4.918	4.933	4.948
t-SD-APM		4.915	4.920	4.930	4.902	4.921	4.928	4.914	4.910	4.917
FMB		4.932	4.943	4.957	4.923	4.944	4.956	4.930	4.938	4.947
DAPM	$\kappa = 0.0$	4.925	4.940	4.955	4.917	4.941	4.955	4.923	4.935	4.946
	0.1	4.925	4.939	4.955	4.916	4.940	4.954	4.923	4.934	4.945
	0.2	4.923	4.937	4.953	4.915	4.938	4.952	4.921	4.932	4.943
	0.3	4.920	4.933	4.948	4.912	4.934	4.947	4.918	4.928	4.938
	0.4	4.915	4.926	4.941	4.907	4.928	4.940	4.914	4.922	4.931
	0.5	4.909	4.918	4.932	4.901	4.920	4.931	4.908	4.914	4.922
	0.6	4.902	4.910	4.922	4.894	4.912	4.922	4.901	4.905	4.912
	0.7	4.895	4.903	4.915	4.887	4.905	4.914	4.895	4.898	4.905
	0.8	4.891	4.898	4.910	4.883	4.900	4.909	4.891	4.894	4.900
	0.9	4.888	4.895	4.908	4.880	4.898	4.907	4.888	4.891	4.897
	1.0	4.888	4.895	4.907	4.879	4.897	4.906	4.888	4.891	4.897

panels between 0.5 and 0.7. In particular, pricing error differences between the two dynamic models indicate that the breakeven kappa becomes more favorable for the SD-APM with increasing N.

Comparing the pricing errors of the SD-APM variants, we notice two findings. First, the pricing errors of the SD-APM crucially increase when assuming that f_t is a traded factor. This is surprising, as the conditional factor mean and risk prices follow the same dynamics, although the factor is not perfectly priced. The second finding is that the t-SD-APM specification assuming that innovations are commonly t-distributed, as in the true DGP, performs equally well as the baseline Gaussian SD-APM. If anything, the pricing errors of the t-SD-APM appear slightly higher.

The second metric for evaluating the in-sample performance is the root mean squared forecast error (RMSFE), which is computed as

$$RMSFE_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_{i,t} - \hat{\beta}_i\hat{\lambda}_{t-1}\right)^2}. \tag{30}$$

for asset i. Whereas the pricing errors in RMSE evaluate the the overall pricing performance of the model, including the fitted factor residuals $\hat{u}_t$, the RMSFE can be used to evaluate how well the model captures the one-period-ahead prediction equation $\mathbb{E}_t(r_t) = \beta\lambda_t$ in-sample. Put differently, the RMSFE measures the adequacy of the in-sample filtered risk premium series, which is given by $\beta\lambda_t$. Notably, the RMSFEs shown in panel (b) of Table 2 slightly increase with sample length T, which is likely a result of the extremely persistent lambda dynamics. Because of the latter, large departures from the mean level, which are hard to capture by the models, are more likely in longer samples. We see that the relative performance of the SD-APM compared to DAPM is weaker when comparing forecast errors in panel (b) instead of pricing errors in (a). The breakeven κ in case of forecast errors is located between 0.3 and 0.6, and it tends to be higher with an increasing number of assets N. However, the improved performance of the SD-APM over the FMB approach is even more pronounced for forecast errors. The comparison of the forecast errors of the SD-APM variants resembles the results of the corresponding pricing error comparisons.

In conclusion, the in-sample pricing and forecast error comparisons show that the SD-APM is able to partly filter risk premium variation in a realistically noisy panel setting. As expected, knowledge of the true time series predictor z_t causes the regression-based DAPM approach to be the most promising one. However, the situation is different if we realistically assume that the predictor is noisy and that only fifty percent of the variation results from the true predictor. Then, the SD-APM, which completely abstains from using information from z_t, can keep up with the DAPM in a typical return panel of $N = 10$ assets and $T = 600$ monthly observations. This suggests that the use of the SD-APM is worthwhile not only in situations where no time series predictors are available but also when they are sufficiently uncertain. The baseline Gaussian SD-APM with no traded factor assumption appears to be a particular worthwhile specification, as it produces considerably lower pricing and forecasting errors than its traded version. In particular, it appears that the factor innovations are too noisy to serve as an adequate forcing variable for risk price dynamics. Moreover, it performs similar to the t-SD-APM that captures the correct distribution of the innovations in the DGP, but it is less parsimonious because of the additional degrees of freedom parameter. The similar pricing and forecasting performance likely stems from the risk price being dependent on the first two moments of the innovations and not the ones of higher order, which are

Table 2: Risk Premium Forecast Error Comparison

The table shows the average root mean squared pricing error ($RMSE$) of a Gaussian SD-APM(1,1), a SD-APM(1,1) with traded factor and a t-SD-APM(1,1) with t-distributed innovations as well as competing benchmark models. The eight simulated panels have different numbers of assets N, time observations T and are replicated 2 500 times each. Benchmarks are a constant risk price specification estimated fitted with the unconditional approach of Fama and MacBeth (1973) (FMB) and the regression-based dynamic asset pricing model of Adrian et al. (2015) (DAPM). The share of information from the correct forecast variable made available to the DAPM is denoted with κ.

		N=10			N=25			N=100		
		T= 300	600	1200	T=300	600	1200	T=300	600	1200
SD-APM		0.566	0.523	0.526	0.523	0.506	0.484	0.503	0.466	0.457
SD-APM	(traded)	0.816	0.781	0.784	0.820	0.768	0.770	0.815	0.777	0.766
t-SD-APM		0.553	0.514	0.518	0.523	0.499	0.479	0.536	0.474	0.461
FMB		0.727	0.726	0.735	0.720	0.738	0.734	0.723	0.733	0.742
DAPM	$\kappa = 0.0$	0.777	0.752	0.745	0.764	0.769	0.745	0.771	0.753	0.751
	0.1	0.773	0.747	0.741	0.761	0.765	0.740	0.766	0.748	0.744
	0.2	0.759	0.730	0.725	0.748	0.748	0.723	0.753	0.732	0.725
	0.3	0.730	0.696	0.689	0.723	0.712	0.686	0.726	0.698	0.688
	0.4	0.682	0.639	0.628	0.682	0.653	0.625	0.681	0.643	0.628
	0.5	0.617	0.561	0.542	0.625	0.571	0.540	0.621	0.566	0.545
	0.6	0.546	0.472	0.441	0.558	0.477	0.441	0.553	0.476	0.447
	0.7	0.483	0.386	0.340	0.494	0.389	0.342	0.488	0.388	0.347
	0.8	0.440	0.319	0.257	0.445	0.323	0.258	0.440	0.318	0.262
	0.9	0.418	0.281	0.204	0.418	0.287	0.204	0.412	0.279	0.207
	1.0	0.412	0.270	0.189	0.409	0.276	0.187	0.401	0.268	0.187

more precisely captured by the Student-t distribution.

3.3.2 Out-of-Sample Risk Premium Forecasting

Because $\mathbb{E}_T(r_{T+1}) = \beta\lambda_T$, predicting stock market returns one period ahead requires an accurate estimate of β and λ_T. The performance of the SD-APM is therefore evaluated with respect to its ability to predict the one-period-ahead risk premium $\beta\lambda_T$ out-of-sample. The specifications of the SD-APM are the same as in the in-sample analysis, and the DAPM and FMB regressions are again considered as benchmarks.

The predictive accuracy is evaluated with the out-of-sample risk premium error computed as

$$\sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\beta_i(s)\lambda_T(s) - \hat{\beta}_i(s)\hat{\lambda}_T(s)\right)^2} \tag{31}$$

for asset i. Note that, unlike in the in-sample analysis, the errors are now averaged only across the Monte Carlo draws s. Hence, the only error to be used for the evaluations is the one made in the last period T, which is the relevant one for forecasting $T+1$ returns.

Table 2 provides out-of-sample risk premium forecast errors averaged across assets. Comparing the errors of the FMB and the SD-APM approaches suggests that the SD-APM provides an improved out-of-sample risk premium prediction by capturing risk price variation. Moreover, the risk premium prediction errors of the SD-APM generally decrease with the number of assets N and time observations T. The breakeven kappa at which the SD-APM and the diffused DAPM perform on par lies again between 0.6 and 0.7. Hence, we can generally confirm that the in-sample results also apply out-of-sample as well. Moreover, the relative performance of the SD-APM compared to the DAPM is slightly better for samples with T=300. Hence, the results suggest that the SD-DAPM suffers less from a low number of time observations than the DAPM. One observation that might explain this is that the latter regression-based approach tends to estimate the regression intercept λ_0 in (27) with a large standard error in short samples. A poor estimate of λ_0 would then crucially weaken the prediction, even if an informative predictor $\tilde{z}_T$ is provided.

Comparing the three SD-APM specifications with each other, we see first that the traded factor version performs even worse than the FMB benchmark assuming static risk prices. This supports the impression gained from the in-sample results that factor innovations are too noisy to predict risk price dynamics, and one might be better off using cross-sectional errors instead. Another observation, which fits the in-sample results, is that the t-SD-APM and the Gaussian specification perform similarly well. However, the latter performs slightly better with a high panel size N, whereas the t-SD-APM produces lower prediction errors for N equal to 10 or 25.

4 Empirical Application

The Monte Carlo study presented in the previous section provides evidence that the SD-APM performs well in cross-sections simulated from correctly specified models. The following empirical application intends to shed light on whether and to what extent the SD-APM can detect predictable components of risk factors from actual equity return data.

4.1 Data

Test assets are ten size-sorted portfolios based on US equities. Monthly series have been obtained from Kenneth French's online library and span the period from January 1964 to June 2021. Hence, we work with a return panel with dimensions $N = 10$ and $T = 691$. Two risk factors are considered to price this cross-section of test assets:

$$f_t = \begin{pmatrix} MKT_t \\ SMB_t \end{pmatrix}, \tag{32}$$

where MKT is the excess return on the value-weighted equity market portfolio, and SMB is the small minus big portfolio return from Fama and French (1993).

Additional forecasting factors are used in the regression-based DAPM approach of Adrian et al. (2015), which serves as the main dynamic benchmark. These factors are the same that are used for calibrating the DGP in the simulation study, hence we have

$$z_t = \begin{pmatrix} TSY10_t \\ TERM_t \\ DY_t \end{pmatrix} \tag{33}$$

where TSY10 is the 10-year treasury yield and TERM is the term spread, computed as the difference between the yields of the 10-year treasury note and the three-month treasury bill. Both series are obtained from the H.15 statistical release of the Board of Governors of the Federal Reserve System. The third forecasting factor is the dividend yield of the S&P 500 index. All three series are also used as forecasters in Adrian et al. (2015). More evidence on equity return predictability from these factors can be found in Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), and Campbell and Thompson (2008) for long-run treasury yields and Campbell and Shiller (1988), Fama and French (1989), Campbell and Thompson (2008), and Cochrane (2008) for the term structure and dividend yields.

4.2 Empirical Model Specifications

A dynamic version of the cross-sectional model in line with the baseline SD-APM framework of Section 2 yields the following pricing equation:

$$r^i_{t+1} = \beta^{MKT}_{i,t}(\lambda^{MKT}_t + u^{MKT}_{t+1}) + \beta^{SMB}_{i,t}(\lambda^{SMB}_t + u^{SMB}_{t+1}) + e_{i,t+1} \tag{34}$$

for test assets $i = 1, \ldots 10$. The main specification to be considered is an SD-APM(1,1) with Gaussian innovations. Non-diagonal elements in parameter matrices A and B in the updating scheme (10) are set to zero. The diagonalization mutes the impact of scaled scores and factor innovations on the parameter updating of the other factor. However, excluding those effects in the present application does not crucially impair the model performance but rather yields a much more parsimonious model. The resulting updating equations, parameterized along the long-run values $\overline{\mu}, \overline{\lambda}$, and $\overline{\beta}$, are given by

$$\mu^j_t = \overline{\mu}^j + a^\mu_j s^\mu_{j,t} + b^\mu_j \left(\mu^j_{t-1} - \overline{\mu}^j\right) \tag{35}$$

$$\lambda^j_t = \overline{\lambda}^j + a^\lambda_j s^\lambda_{j,t} + b^\lambda_j \left(\lambda^j_{t-1} - \overline{\lambda}^j\right) \tag{36}$$

$$vec\left(\beta^j_{i,t}\right) = vec\left(\overline{\beta}^j_i\right) + a^\beta_{i,j} s^\beta_{i,j,t} + b^\beta_{i,j} vec\left(\beta^j_{i,t-1} - \overline{\beta}^j_i\right) \tag{37}$$

with scalar parameters $\overline{\mu}^j, \overline{\lambda}^j, \overline{\beta}^j, a^\mu_j, a^\lambda_j, a^\beta_{i,j}, b^\mu_j, b^\lambda_j$ and $b^\beta_{i,j}$, where $i = 1, \ldots, 10$ and $j = MKT, SMB$.

Two established benchmark specifications are considered to investigate whether the SD-APM approach can actually improve on filtering risk premia and explain return variation in cross-section and time simultaneously. The first benchmark is the constant risk price specification underlying classical Fama and MacBeth (1973) regressions. In line with Adrian et al. (2015), estimated innovations $\hat{u}_t$ from a VAR(1) model including the abovementioned factors are provided as pricing factors in order to account for the significant persistence of the pricing factors. Betas in this benchmark are either constant over time or time-varying in a moving window of 60 months. The second benchmark is a DAPM that explains risk price variations with the forecasting factors described above. This yields a regression equation given by

$$\lambda_t^j = \lambda_0 + \Lambda_1^{j,TSY10} TSY10_t + \Lambda_1^{j,TERM} \Delta TERM_t + \Lambda_1^{j,\Delta DY} \Delta DY_t \tag{38}$$

for each of the two cross-sectional risk factors $j = MKT, SMB$. Time-varying betas in the DAPM are estimated with a Gaussian kernel regression approach, including a data-driven bandwidth choice following Ang and Kristensen (2012). Estimation and inference for the DAPM is conducted as described in Adrian et al. (2015), and I refer to them for more detail.

4.3 Empirical Results

In the following, the results of the empirical analysis will be discussed along different lines. First, we examine parameter estimates of factor risk exposures and risk prices. Afterwards, we compare the SD-APM with several benchmarks with respect to pricing and forecasting errors.

4.3.1 Factor Risk Exposures

Table 3 shows estimates of factor risk exposure of the 10 size-sorted equity portfolios. The first block of five columns contains estimated parameters for the market factor (MKT), whereas the second block contains those of the small-minus-big factor (SMB). Parameter estimates of the beta updating scheme from a SD-APM(1,1) with time-varying betas are presented in the first three columns of each block. The fourth and fifth columns of each block contain risk exposure estimates from a constant beta SD-APM(1,1) and first-stage Fama and MacBeth (1973) regressions, respectively.

First, we can observe that the risk exposure estimates from the constant beta SD-APM specification are very close to those estimated by FMB. This holds particularly true for market betas. Moreover, the unconditional betas $\overline{\beta}$ in the SD-APM specification with time-varying betas are quite close to the unconditional FMB estimates, especially those with respect to MKT. In line with the literature on the size effect, we find that all 10 size portfolios have exposure around unity to MKT, whereas these loadings show a wedge with respect to SMB. We find significant time-variation for all risk exposures, as indicated by a^β being significantly different from zero. The estimated values for b^β reveal that the time-varying SD-APM betas quite persistently fluctuate around the unconditional risk exposure. Notable exceptions are the size2 and size10 exposures to SMB, which show negative values b^β that are also considerably smaller than for the other exposures.

The filtered SD-APM betas can be compared to those estimated with the non-parametric approach of the time-varying beta DAPM. Figure 3 shows the corresponding plots of some selected exposures together with betas estimated from rolling 60-month window Fama and MacBeth (1973) first-stage regressions. Exposures implied by the SD-APM show a movement that is much larger

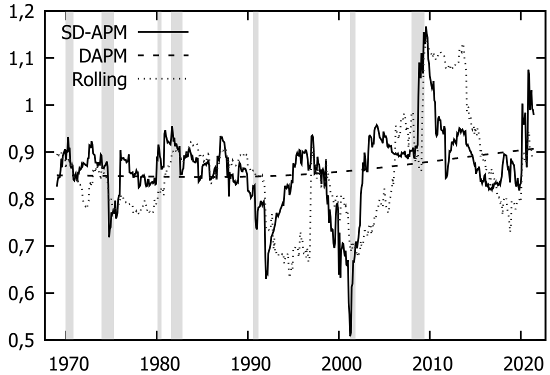
Table 3: Risk Exposure Estimates

This table shows estimates of factor risk exposure parameters for the market factor (MKT) and the small-minus-big factor (SMB). It reports results from a SD-APM(1,1) model with time-varying betas and a specification with constant betas as well risk exposure estimates from first stage Fama and MacBeth (1973) (FMB) regressions. Standard errors are shown in parentheses. Test assets are 10 value-weighted equity portfolios sorted on size with monthly returns denoted in percentages. The sample period is 1964:01 - 2021:06.

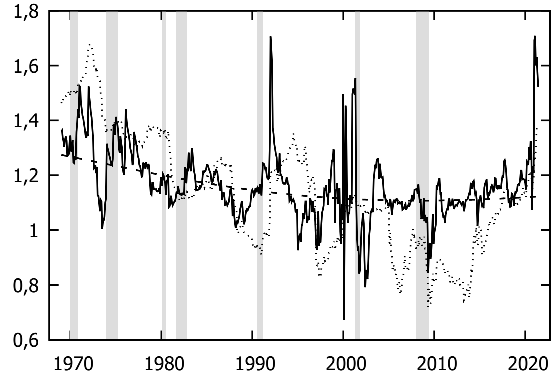
	MKT					SMB				
	SD-APM (t.-v. beta)			SD-APM	FMB	SD-APM (t.-v. beta)			SD-APM	FMB
	$\overline{\beta}$	a^{β}	b^{β}	β	β	$\overline{\beta}$	a^{β}	b^{β}	β	β
size1	0.872	0.046	0.964	0.875	0.868	1.161	0.070	0.902	1.205	1.177
	(0.038)	(0.011)	(0.020)	(0.017)	(0.023)	(0.040)	(0.003)	(0.041)	(0.025)	(0.034)
size2	1.001	0.023	0.983	0.997	1.001	1.052	-0.112	-0.184	1.087	1.088
	0.026	0.009	0.019	0.012	0.017	0.006	0.016	0.121	0.017	0.025
size3	1.025	0.031	0.911	1.021	1.022	0.920	0.047	0.861	0.908	0.911
	0.013	0.010	0.053	0.010	0.015	0.018	0.017	0.050	0.015	0.021
size4	1.015	0.044	0.755	1.013	1.015	0.850	0.069	0.841	0.809	0.812
	0.011	0.010	0.087	0.010	0.014	0.019	0.003	0.059	0.014	0.021
size5	1.035	0.027	0.888	1.029	1.031	0.712	0.075	0.869	0.672	0.678
	0.012	0.009	0.070	0.010	0.013	0.022	0.003	0.049	0.015	0.020
size6	1.021	0.034	0.900	1.020	1.020	0.530	0.057	0.793	0.482	0.484
	0.014	0.010	0.062	0.011	0.014	0.019	0.005	0.089	0.016	0.021
size7	1.051	0.042	0.729	1.048	1.048	0.390	0.027	0.958	0.366	0.370
	0.012	0.012	0.107	0.011	0.013	0.023	0.005	0.017	0.016	0.019
size8	1.047	0.025	0.700	1.038	1.039	0.244	0.019	0.984	0.251	0.258
	0.011	0.011	0.147	0.010	0.012	0.027	0.005	0.011	0.015	0.018
size9	1.002	0.015	0.965	1.000	1.001	0.057	0.021	-0.570	0.061	0.063
	0.013	0.006	0.020	0.009	0.011	0.011	0.007	0.190	0.014	0.016
size10	0.981	0.044	0.858	0.983	0.983	-0.192	0.050	1.000	-0.266	-0.266
	0.005	0.007	0.040	0.004	0.007	0.027	0.003	0.002	0.006	0.010

than those implied by the DAPM, which are extremely smooth. In fact, the smooth non-parametric estimates could be a kind of long-term trend of the score-driven beta, best seen for size1-SMB in panel (b). The variations of the SD-APM betas resemble those of the rolling regression estimates but appear less persistent. Rather, one observes exposure peaks in the shaded NBER recessions, which tend to quickly fall and revert back to a medium level. An exception is the size10-SMB beta in panel (f), which moves very persistently like the corresponding rolling beta.

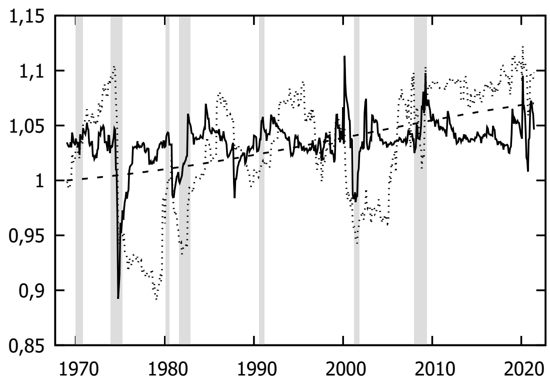
In summary, with few exceptions, betas fluctuate significantly and persistently over time. However, it is also clear from Figure 3 that the range in which the betas fluctuate is quite narrow in most cases.



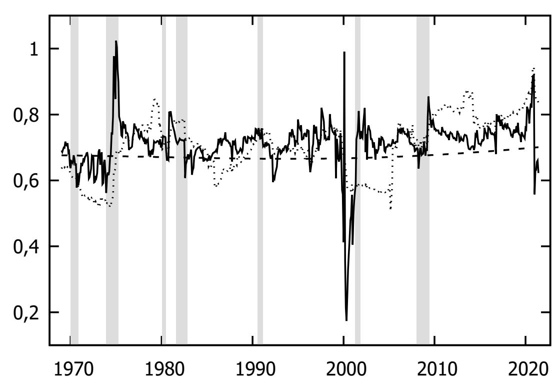
(a) size1 and MKT



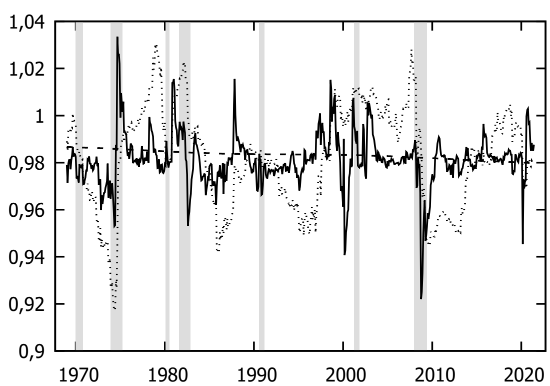
(b) size1 and SMB



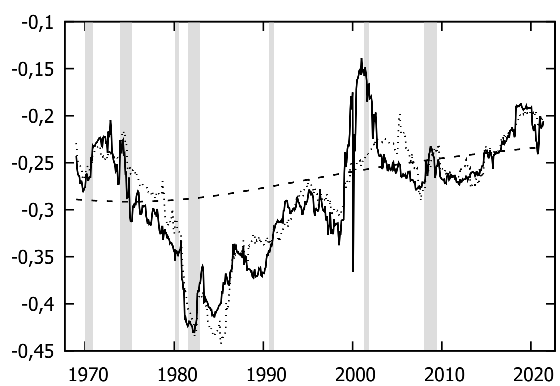
(c) size5 and MKT



(d) size5 and SMB



(e) size10 and MKT



(f) size10 and SMB

Figure 3: **Time-Varying Beta Estimates.** This figure shows estimated factor risk exposures for the market factor (MKT) and the small-minus-big factor (SMB). It reports results from a SD-APM(1,1) and a DAPM from Adrian et al. (2015). Rolling refers to betas estimated on a rolling 60-month window using Fama and MacBeth (1973) first stage regressions. Test assets are value-weighted equity decile portfolios sorted on size with monthly returns denoted in percentages. Shaded areas refer to NBER recessions. The sample period is 1964:01 - 2021:06.

Table 4: Price of Risk Estimates

This table shows estimates of risk price parameters for the market factor (MKT) and the small-minus-big factor (SMB). The first two columns show results from a constant-beta SD-APM(1,1). Following two rows provide estimates from a dynamic asset pricing model (DAPM) in line with Adrian et al. (2015). Forecasting factors are those discussed in Section 4.1. Columns 5 and 6 provide FMB regression results of a constant risk price specification. Results from time-varying beta specifications of the SD-APM, the DAPM and 60-month rolling FMB regressions are provided in the columns on the right. Standard errors are shown in parentheses. SD-APM standard errors are derived from the numerically computed Fisher Information as described in Section 2.3. Errors for the DAPM estimates are adjusted for cross-asset correlation in the residuals and for estimation error of the time-series betas. FMB standard errors include a Shanken (1992) correction. Test assets are 10 value-weighted equity portfolios sorted on size with monthly returns denoted in percentages. The sample period is 1964:01 - 2021:06.

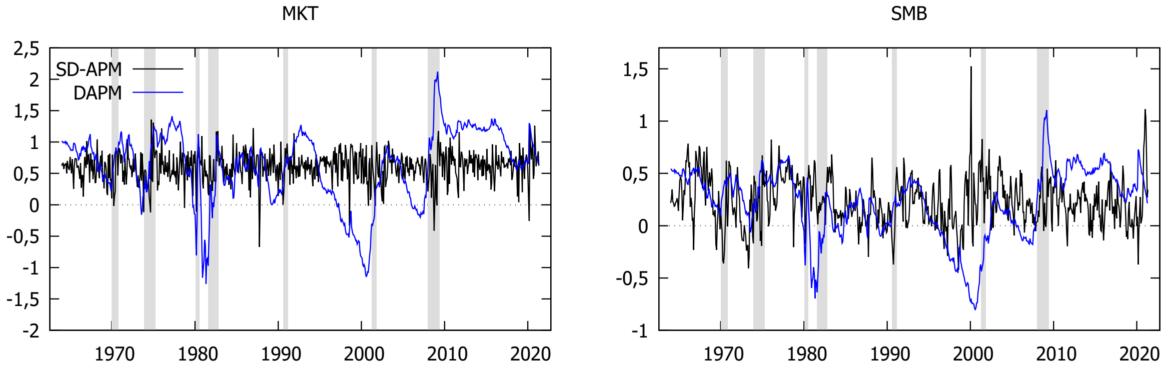
	Constant betas						Time-varying betas					
	SD-APM		DAPM		FMB		SD-APM		DAPM		FMB	
	MKT	SMB	MKT	SMB	MKT	SMB	MKT	SMB	MKT	SMB	MKT	SMB
$\overline{\lambda}$	0.600	0.232	0.621	0.217	0.621	0.217	0.773	0.104	0.633	0.257	0.603	0.111
	(0.181)	(0.142)	(0.185)	(0.158)	(0.172)	(0.125)	(0.915)	(0.309)	(0.525)	(0.373)	(0.239)	(0.165)
a^{λ}	0.054	0.055					0.097	0.044				
	(0.024)	(0.022)					(0.028)	(0.011)				
b^{λ}	0.151	0.681					0.310	0.750				
	(0.300)	(0.163)					(0.163)	(0.108)				
TSY10			-0.228	-0.148					-0.196	-0.125		
			(0.084)	(0.063)					(0.087)	(0.061)		
TERM			0.262	0.090					0.256	0.111		
			(0.138)	(0.101)					(0.143)	(0.100)		
DY			1.553	1.021					1.218	0.918		
			(0.629)	(0.482)					(0.636)	(0.464)		

4.3.2 Factor Risk Prices

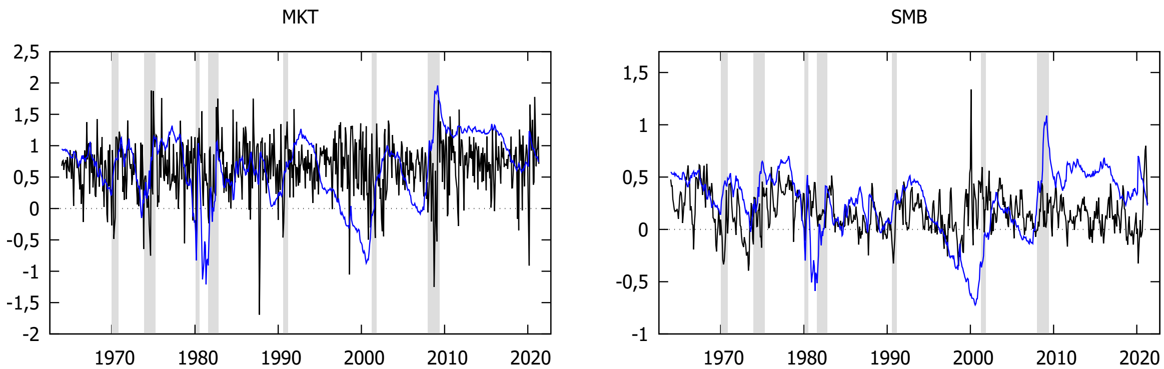
We now come to the factor risk price dynamics estimated by the SD-APM. Table 4 provides estimates of the risk price parameters from the SD-APM(1,1) and benchmark models. The first six columns show results of model specifications with betas assumed to be constant over time. The unconditional risk prices $\overline{\lambda}$ are reported in the first row. We see that the unconditional estimates of the two dynamic models are quite close to the FMB estimates. In addition, their standard errors, shown in parentheses, are only slightly higher than those of the FMB estimates including a Shanken (1992) correction for the pre-estimation bias from first-stage time series regressions. The estimates for a^{λ} reveal how risk prices react in response to the scaled score, which is the contemporaneous cross-sectional regression error. They are estimated as 0.054 and 0.055 for the MKT and SMB factors respectively, and are statistically different from zero. This reveals that the risk prices of both pricing factors significantly vary over time and that contemporaneous cross-sectional regression errors are informative for inferring the time-variation of risk prices. The coefficients $\Lambda_1^{TSY10}, \Lambda_1^{TERM}$, and Λ_1^{DY} indicate how risk prices react in the benchmark DAPM based on equation (38). The reported estimates show, in line with Adrian et al. (2015), that the three forecasting factors are capable of predicting risk price movements. This is valid for the TERM factor to a limited extent, as the corresponding coefficient is only significant in the MKT regression.

An examination of the time-varying beta specifications, the results of which are shown in columns 7 to 10 of Table 4, reveals that unconditional risk prices estimated by the SD-APM deviate more strongly from those obtained with the classical FMB approach. This particularly holds for the MKT risk price. Moreover, standard errors for the unconditional risk prices increase greatly. Consider that the time-varying beta specification requires two additional parameters to be estimated for each beta. If we leave out the entries in the covariance matrices, this makes it necessary to estimate $3K + 3NK = 66$ parameters instead of $3K + 3NK = 26$ in the constant beta specification. Hence, the increased number of parameters to be estimated may explain the increased uncertainty in estimating $\overline{\lambda}$. With respect to the a^{λ} estimates, we see that this coefficient becomes larger for MKT and slightly smaller for SMB, while both remain statistically significant. The b^{λ} estimates reveal that both risk price processes become more persistent when allowing for time-varying betas. The estimates of coefficients $\Lambda_1^{TSY10}, \Lambda_1^{TERM}$, and Λ_1^{DY} from the benchmark DAPM are quite similar in both the time-varying and the time-constant beta setting. However, we also see a clear increase of the corresponding standard errors of the unconditional risk price estimate.

We now turn to a graphical comparison of filtered factor risk prices that are provided in Figure 4. The upper-left panel shows market risk prices from specifications with constant betas. The most striking observation is that SD-APM-implied risk prices are much less persistent than the DAPM trajectories that would be predicted by forecasting factors. Moreover, the two risk price processes differ considerably from each other, which is also evidenced by an empirical correlation of 0.086. If anything, one can recognize some connection in the early 70s. The SD-APM-implied risk premia do not show the commonly observed pattern of generally increased risk prices in recessions, which are indicated by the shaded areas in Figure 4. What can be found instead is a pattern where the market risk price tends to go down at the beginning and increases towards then end of a recession. This is best seen in the global financial crisis period around 2008. This down and up behavior is even more pronounced in the risk prices filtered from the SD-APM with time-varying betas that is shown in the lower-left panel. Whereas the corresponding DAPM trajectory is almost unchanged, SD-APM-implied market risk prices are considerably more volatile based on time-varying beta specifications. It is also noteworthy that the dynamics in the SD-APM react in particular to financial market-specific shocks. Consider, for example, the striking drop in October 1987, when a massive unexpected crash, Black Monday, occurred. This is also a result of the observation-driven construction of the SD-APM. While the DAPM explains fluctuations in risk prices through instrument variables, the SD-APM dynamics are driven by deviations between the observed returns and portfolios and those predicted by the assumed factor model. In the case of an extreme event such as Black Monday, it must then be quantified how much the observation



(a) Constant Beta Specifications



(b) Time-Varying Beta Specifications

Figure 4: **Time-Varying Risk Prices.** This figure shows estimated factor risk prices for the market factor (MKT) and the small-minus-big factor (SMB). It reports results from a SD-APM(1,1) and a DAPM from Adrian et al. (2015). Reults in panel (a) are based on constant beta specifications whereas those in panel (b) are based on time-varying betas. Test assets are value-weighted equity decile portfolios sorted on size with monthly returns denoted in percentages. Shaded areas refer to NBER recessions. The sample period is 1964:01 - 2021:06.

results from a change in the parameters and how much merely results from an extreme realization of the innovation terms. However, it cannot be conclusively clarified here without doubt whether the drop on Black Monday resulted from a correction of future return expectations, which would be reflected by a lower lambda, or is due to the way the SD-APM is constructed.

The estimated SMB risk price paths are shown in the plots on the left of Figure 4. As expected from the parameter estimates, the SMB risk price dynamics are much more persistent and track the DAPM-implied SMB risk prices more closely until the second half of the 1990s. However, in the last 25 years of the sample, the series implied by the two different methods differ considerably. The DAPM proposes low lambdas at the turn of the millennium and increased lambdas during the global financial crisis. Neither of these movements is captured by the SD-APM. Instead, the SD-APM-implied SMB risk price shows a pronounced peak just before the burst of the dotcom bubble and considerably lower SMB risk prices during and after the global financial crisis. The latter is in line with the SMB factor realizations showing no reaction to the global financial crisis, whereas the regression-based benchmark would imply higher SMB premia during the financial crisis. It could be that the SD-APM is not able to capture the increased post-crisis SMB premia. However, another explanation for the differences in estimated risk premia could be that the relations to the predictor variables are unstable and thus the DAPM implied risk premia are misleading. An observation supporting this view is that DAPM-implied premia to MKT and SMB share very similar dynamics, although returns and SD-APM premia differ considerably across factors.

4.3.3 Pricing and Forecasting Error Comparisons

In order to evaluate the fit of the SD-APM, we compare pricing and forecasting errors from different SD-APM specifications as well as from DAPM and FMB benchmarks, which are shown in Table 5. The first four columns show the results of the Gaussian SD-APM specifications, which differ in whether the betas are constant or vary over time and whether the pricing factors are traded or not. Columns 5 and 6 show results for constant beta SD-APMs with t-distributed residuals in specifications with either non-traded or traded factors. Column 7 refers to a two-step score-driven filter in which time-varying betas are estimated in a first stage based on score-driven time-varying coefficient model. Time-varying risk prices and factor means are then estimated in a second stage with an SD-APM based on betas from the first stage. [11] This approach can be understood as computational alternative to estimate the full SD-APM in Column 1. Columns 8 and 9 show results from DAPM specifications using the forecasting variables described in section 4.1 with betas being either constant or time-varying. The last two columns show errors from constant lambda specifications estimated with Fama and MacBeth (1973) regressions, where betas are constant or pre-estimated with a rolling window of 60 months. Errors from fitting the first 60 observations are discarded in order to make the results comparable with the rolling model for which these are not available. In addition, this exclusion serves as a burn-in period to mitigate the impact of starting values in the score-driven models.

With respect to the pricing performance, we recognize that all dynamic models perform considerably better than the unconditional FMB specification, which shows an average RMSE of 1.565. The best-performing model with respect to the average RMSE is the regression-based DAPM with time-varying lambdas and betas. This is not surprising, as this model uses predictor information in z_t that is not used by the score-driven and constant model specifications. More surprising is that the (Gaussian) time-varying beta SD-APM specifications have average RMSEs of 1.159 and 1.160, which are quite close to the one of the DAPM benchmark that is given by 1.108. If we estimate the time-varying beta SD-APM with the computationally less demanding but also less efficient two-stage approach 2SF, we still get an average RMSE of 1.177. Hence, in terms of pricing errors, the score-driven models can crucially make up for not knowing the predictors by evaluating contemporaneous model errors. If we consider models that assume constant betas, we even see that the (Gaussian) SD-APM with an RMSE of 1.194 performs much better than the DAPM benchmark with an RMSE of 1.352. This could be due to the greater flexibility of the SD-APM to accom-

[11] More details on this benchmark are provided in Supplementary Appendix D.3.

Table 5: Root mean squared pricing and forecasting errors

The table shows root means squared pricing errors (RMSE) and rooot mean squared forecast errors (RMSFE) of different asset pricing model specifications. Pricing factors are the market factor (MKT) and the small-minus-big factor (SMB). The first four columns show the results of the Gaussian SD-APM specifications, which differ in whether the betas are constant or vary over time and whether the pricing factors are traded or not. Columns 5 and 6 show results for constant beta SD-APMs with t-distributed residuals in specification with non-traded and traded factors. Columns 7 and 8 show results from DAPM specifications using the forecasting variables described in section 4.1 and having betas either constant or time-varying. The last two columns show errors from constant lambda specifications estimated with Fama and MacBeth (1973) regressions where betas are constant or pre-estimated on a rolling window of 60 months. Test assets are 10 value-weighted equity portfolios sorted on size with monthly returns denoted in percentages. The sample period is 1964:01 - 2021:06.

	SD-APM				t-SD-APM		2SF	DAPM		FMB	
λ_t	x	x	x	x	x	x	x	x	x		
β_t	x	x					x	x		x	
traded		x		x		x					
(a) RMSE											
size1	1.921	1.932	1.974	1.996	1.990	1.995	1.974	1.702	2.406	2.467	2.590
size2	1.353	1.369	1.369	1.377	1.372	1.378	1.350	1.259	1.703	1.829	1.896
size3	1.057	1.064	1.127	1.130	1.129	1.133	1.103	1.046	1.369	1.552	1.634
size4	1.033	1.031	1.122	1.120	1.124	1.125	1.071	1.064	1.326	1.519	1.597
size5	1.083	1.074	1.145	1.140	1.141	1.141	1.123	1.060	1.251	1.417	1.500
size6	1.208	1.203	1.262	1.260	1.265	1.269	1.230	1.186	1.335	1.496	1.583
size7	1.212	1.209	1.217	1.218	1.215	1.219	1.211	1.159	1.296	1.409	1.479
size8	1.167	1.165	1.153	1.154	1.151	1.154	1.152	1.106	1.191	1.321	1.362
size9	1.071	1.072	1.066	1.066	1.064	1.067	1.068	1.032	1.098	1.211	1.250
size10	0.486	0.487	0.502	0.501	0.501	0.501	0.490	0.466	0.544	0.735	0.754
Average	1.159	1.160	1.194	1.196	1.195	1.198	1.177	1.108	1.352	1.496	1.565
(b) RMSFE											
size1	6.340	6.325	6.367	6.427	6.390	6.398	6.381	6.371	6.371	6.444	6.437
size2	6.510	6.500	6.525	6.564	6.557	6.558	6.539	6.513	6.512	6.563	6.560
size3	6.166	6.165	6.173	6.201	6.213	6.211	6.186	6.135	6.134	6.199	6.194
size4	5.936	5.937	5.940	5.965	5.981	5.978	5.955	5.898	5.897	5.964	5.960
size5	5.778	5.782	5.778	5.797	5.816	5.810	5.788	5.735	5.733	5.793	5.790
size6	5.423	5.443	5.420	5.437	5.459	5.451	5.431	5.369	5.367	5.437	5.432
size7	5.363	5.387	5.360	5.378	5.395	5.387	5.370	5.328	5.329	5.379	5.375
size8	5.161	5.193	5.154	5.166	5.184	5.174	5.163	5.121	5.121	5.167	5.162
size9	4.752	4.788	4.745	4.754	4.771	4.762	4.751	4.715	4.716	4.756	4.753
size10	4.371	4.417	4.357	4.355	4.366	4.363	4.359	4.328	4.327	4.363	4.359
Average	5.580	5.594	5.582	5.604	5.613	5.609	5.592	5.551	5.551	5.606	5.602

modate different patterns of risk price dynamics compared to DAPM risk prices, whose dynamics are tied to the predictor variables. In addition, we also find that taking time-varying lambdas into account within the SD-APM results in a greater improvement of pricing performance than taking beta dynamics into account as well. This is different in the case of the DAPM, where the additional inclusion of beta dynamics leads to a considerable improvement in pricing performance. A potential reason for this could be that the non-parametrically estimated DAPM beta dynamics cover some lambda dynamics that cannot be explained by the employed risk price predictors.

Comparing the different SD-APM specifications, we see that modeling residuals with a multivariate t-distribution does not improve the pricing performance compared to the Gaussian specifications, although the degrees of freedom parameter is estimated with 6.23 (6.19 with traded factors). Hence, the gains from a better fit of the residual distribution do not seem to outweigh the additional uncertainty from introducing the additional degrees of freedom parameter. This seems reasonable considering that risk premia are mainly determined by the dynamics of the first two moments of the distribution and not by the higher order moments, which are better captured by the Student-t distribution.

Panel (b) of Table 5 shows the RMSFEs, which reveal how well the conditional risk premium $\beta_t \lambda_t$ forecasts next period returns r_t. These errors are generally higher than pricing errors because the risk premium makes up a rather small share of the overall return. Moreover, the differences between the models become reasonably smaller. However, we see that the DAPM benchmarks perform best with an RMSFE of 5.51 each. This is 0.092 less than the unconditional FMB benchmark. The best-performing SD-APM specification is the one with time-varying betas and no traded factor assumption that has an RMSFE of 5.580 which is 0.022 less than the unconditional FMB benchmark. Hence, 24 percent of the forecasting improvement from using external predictors in a DAPM can also be achieved by an SD-APM that makes no use of the predictor information. Moreover, we see that including time-varying betas does not have a major impact on the forecast performance. Hence, the improved forecast performance merely stems from risk price dynamics and not from changing exposures. Another remarkable finding is that the SD-APM specifications, which enforce pricing factors to be tradeable perform considerably worse than those specifications without such enforcement. In particular, the RMSFE of the constant beta SD-APM with traded factors is higher than the one of the unconditional FMB specification. This observation can be explained by the fact that the updating scheme of the traded factor version only uses factor innovations as drivers. Consideration of the cross-sectional relationship and thus of the other returns is completely omitted. This suffices to keep pricing errors small, but the information of the return panel for forecasting the risk premium remains unused. In addition, the consideration t-distributed innovations worsen the forecasting performance as well, and is slightly in line with the simulation results. A possible explanation is that the additional scaling function C in the updating (25) over-dampens the impact of extreme observations, which seem to be particularly informative.

5 Conclusions

This paper has introduced an empirical dynamic asset pricing framework that allows for time-varying lambdas and betas, which are unobserved processes filtered from the cross-section of asset returns and the asset pricing model's factor structure, in line with the more general GAS model developed by Creal et al. (2013). It is applicable to a wide range of linear factor models in the finance literature. A main advantage is that no external predictors are required to describe the time dynamics of risk prices or exposures. A simulation study provides evidence that the method is capable of filtering substantial risk price movements from a model with correctly specified cross-sectional factors under a realistic signal-to-noise ratio. Moreover, it can compete with a dynamic estimation approach taking signals of true time series drivers into account. The results point to a non-negligible source of possible misspecification in the time series model that can be evaluated and possibly circumvented by utilizing the SD-APM framework.

Updating schemes for the SD-APM class with elliptically distributed innovations have been derived. It turns out that the risk price updating direction within this class is unaffected by the distributional assumptions but the magnitude of the parameter movement depends on the shape of the corresponding probability density function. The presented SD-APM(1,1) specification with constant betas and normally distributed innovation terms is a particularly tractable model with respect to the complexity and computational burden for estimation. Its use has been illustrated by an empirical application filtering an adequate series of equity risk premia, revealing movements that are not shown by the typically extremely persistent economic forecasting factors.

Appendix

A Proofs

Vector-by-vector derivatives are denoted in denominator layout such that $\frac{\partial v}{\partial w}$ is a $n \times m$-matrix given a n-dimensional vector v and m-dimensional vector w. The vec operator stacks the columns of a $n \times m$-matrix X underneath the other such that $vec(X)$ is a $n \cdot m$-dimensional vector. The n-dimensional identity matrix is denoted with I_n and the Kronecker product with $\otimes$.

Before dealing with the main technical results of the paper, I provide two lemmas collecting some helpful properties of spherically distributed random vectors. A n-dimensional random vector x is spherically distributed if $x = rs$, where s is uniformly distributed on the (n-1)-dimensional unit sphere and r is a non-negative random number that is independent of s. Moreover, let $\|x\|$ denote the euclidean norm of a vector x.

Lemma 1. *Suppose the n-dimensional random vector x follows a spherical distribution, then:*

(i) $\|x\|$ *and* $\frac{x}{\|x\|}$ *are independent.*

(ii) $\mathbb{E}\left(\frac{x}{\|x\|}\frac{x^\top}{\|x\|}\right) = \frac{1}{n} I_n$

Proof. See proofs of Theorem 2.3 and Theorem 2.7 in Fang et al. (1990).

Lemma 2. *Suppose that $Z \sim F(n, \nu)$, then:*

$$\mathbb{E}\left(\frac{Z}{\left(\frac{\nu}{n} + Z\right)^2}\right) = \frac{n^2}{(n+\nu+2)(n+\nu)} \tag{A.1}$$

Proof. See Supplementary Material C.1.

A.1 Proof of Proposition 1

Proof of Proposition 1. The common observational density of f_t and r_t can be derived with the given density functions. We find

$$p(r_t, f_t|\mathcal{F}_{t-1}, \theta_t) = p(r_t|f_t, \mathcal{F}_{t-1}, \theta_t)p(f_t|\mathcal{F}_{t-1}, \theta_t) \tag{A.2}$$

$$= (2\pi)^{-\frac{N}{2}} |\Sigma_e|^{-\frac{1}{2}} \exp\left\{-\frac{e_t^\top \Sigma_e^{-1} e_t}{2}\right\} (2\pi)^{-\frac{K}{2}} |\Sigma_u|^{-\frac{1}{2}} \exp\left\{-\frac{u_t^\top \Sigma_u^{-1} u_t}{2}\right\} \tag{A.3}$$

and the conditional log-likelihood

$$l_t = \ln p(r_t, f_t|\mathcal{F}_{t-1}, \theta_t) = -(N+K)\ln\pi - \frac{1}{2}\ln|\Sigma_e| - \frac{1}{2}\ln|\Sigma_u| - \frac{1}{2}e_t^\top \Sigma_e^{-1} e_t - \frac{1}{2}u_t^\top \Sigma_u^{-1} u_t. \tag{A.4}$$

The score with respect to the factor mean μ can be derived as:

$$\nabla_t^\mu = \frac{\partial l_t}{\partial \mu_{t-1}} = -\frac{\partial e_t}{\mu_{t-1}} \cdot \Sigma_e^{-1} e_t - \frac{\partial u_t}{\mu_{t-1}} \cdot \Sigma_u^{-1} u_t = -\beta_{t-1}^\top \Sigma_e^{-1} e_t + \Sigma_u^{-1} u_t \tag{A.5}$$

and the corresponding block entry of the Fisher information as

$$\mathcal{I}_t^\mu = \mathbb{E}_{t-1}\left(\nabla_t^\mu \nabla_t^{\mu\top}\right) = \beta_{t-1}^\top \Sigma_e^{-1} \mathbb{E}_{t-1}\left(e_t e_t^\top\right) \Sigma_e^{-1} \beta_{t-1} + \Sigma_u^{-1} \mathbb{E}_{t-1}\left(u_t u_t^\top\right) \Sigma_u^{-1} \tag{A.6}$$

$$= \beta_{t-1}^\top \Sigma_e^{-1} \beta_{t-1} + \Sigma_u^{-1}. \tag{A.7}$$

The score with respect to the risk price λ can be derived as:

$$\nabla_t^\lambda = \frac{\partial l_t}{\partial \lambda_{t-1}} = -\frac{\partial e_t}{\lambda_{t-1}} \cdot \Sigma_e^{-1} e_t = \beta_{t-1}^\top \Sigma_e^{-1} e_t \tag{A.8}$$

and

$$\mathcal{I}_t^\lambda = \mathbb{E}_{t-1}\left(\nabla_t^\lambda \nabla_t^{\lambda\top}\right) = \beta_{t-1}^\top \Sigma_e^{-1} \mathbb{E}_{t-1}\left(e_t e_t^\top\right) \Sigma_e^{-1} \beta_{t-1} = \beta_{t-1}^\top \Sigma_e^{-1} \beta_{t-1}. \tag{A.9}$$

For the Fischer Information part concerning factor means and risk prices we find:

$$\mathcal{I}_t^{\mu,\lambda} = \mathbb{E}_{t-1}\left(\nabla_t^\mu \nabla_t^{\lambda\top}\right) = -\beta_{t-1}^\top \Sigma_e^{-1} \mathbb{E}_{t-1}\left(e_t e_t^\top\right) \Sigma_e^{-1} \beta_{t-1} + \Sigma_u^{-1} \mathbb{E}_{t-1}\left(u_t e_t^\top\right) \Sigma_e^{-1} \beta_{t-1} \tag{A.10}$$

$$= -\beta_{t-1}^\top \Sigma_e^{-1} \beta_{t-1}. \tag{A.11}$$

For the risk exposure updating, we derive

$$
\begin{aligned}
\nabla_t^{\beta} &= \frac{\partial l_t}{\partial vec(\beta_{t-1})} = -\frac{\partial e_t}{\partial vec(\beta_{t-1})}\Sigma_e^{-1}e_t && \text{(A.12)}\\
&= \frac{\partial}{\partial vec(\beta_{t-1})}\left[vec(\beta_{t-1}(\lambda_{t-1}+u_t))\right]\cdot\Sigma_e^{-1}e_t && \text{(A.13)}\\
&= \frac{\partial}{\partial vec(\beta_{t-1})}\left[((\lambda_{t-1}+u_t)^{\top}\otimes I_N)vec(\beta_{t-1})\right]\cdot\Sigma_e^{-1}e_t && \text{(A.14)}\\
&= ((\lambda_{t-1}+u_t)\otimes I_N)\Sigma_e^{-1}e_t && \text{(A.15)}\\
&= ((\lambda_{t-1}+u_t)\otimes\Sigma_e^{-1})e_t. && \text{(A.16)}
\end{aligned}
$$

The corresponding Fisher information is given by

$$
\begin{aligned}
\mathcal{I}_t^{\beta} &= \mathbb{E}_{t-1}\left(\nabla_t^{\beta}\nabla_t^{\beta\top}\right) && \text{(A.17)}\\
&= \mathbb{E}_{t-1}\left(((\lambda_{t-1}+u_t)\otimes\Sigma_e^{-1})\mathbb{E}_{t-1}\left(e_te_t^{\top}|u_t\right)((\lambda_{t-1}+u_t)^{\top}\otimes\Sigma_e^{-1})\right) && \text{(A.18)}\\
&= \mathbb{E}_{t-1}\left(((\lambda_{t-1}+u_t)\otimes\Sigma_e^{-1})(1\otimes\Sigma_e)((\lambda_{t-1}+u_t)^{\top}\otimes\Sigma_e^{-1})\right) && \text{(A.19)}\\
&= (\lambda_{t-1}\lambda_{t-1}^{\top}+\Sigma_u)\otimes\Sigma_e^{-1}. && \text{(A.20)}
\end{aligned}
$$

The scaling matrix can therefore be derived as

$$
\begin{aligned}
S_t &= \begin{pmatrix} \beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1}+\Sigma_u^{-1} & -\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1} & 0\\ -\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1} & \beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1} & 0\\ 0 & 0 & (\lambda_{t-1}\lambda_{t-1}^{\top}+\Sigma_u)\otimes\Sigma_e^{-1}\end{pmatrix}^{-1} && \text{(A.21)}\\
&= \begin{pmatrix} \Sigma_u & \Sigma_u & 0\\ \Sigma_u & (\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1})^{-1}+\Sigma_u & 0\\ 0 & 0 & (\lambda_{t-1}\lambda_{t-1}^{\top}+\Sigma_u)^{-1}\otimes\Sigma_e\end{pmatrix} && \text{(A.22)}
\end{aligned}
$$

which can be used to derive the driving sequences:

$$
\begin{aligned}
s_t &= \begin{pmatrix} s_t^{\mu}\\ s_t^{\lambda}\\ s_t^{\beta}\end{pmatrix} = S_t\begin{pmatrix}\nabla_t^{\mu}\\ \nabla_t^{\lambda}\\ \nabla_t^{\beta}\end{pmatrix} && \text{(A.23)}\\
&= \begin{pmatrix} -\Sigma_u\beta_{t-1}^{\top}\Sigma_e^{-1}e_t+\Sigma_u\Sigma_u^{-1}u_t+\Sigma_u\beta_{t-1}^{\top}\Sigma_e^{-1}e_t\\ -\Sigma_u\beta_{t-1}^{\top}\Sigma_e^{-1}e_t+\Sigma_u\Sigma_u^{-1}u_t+(\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^{\top}\Sigma_e^{-1}e_t+\Sigma_u\beta_{t-1}^{\top}\Sigma_e^{-1}e_t\\ \left((\lambda_{t-1}\lambda_{t-1}^{\top}+\Sigma_u)^{-1}\otimes\Sigma_e\right)((\lambda_{t-1}+u_t)\otimes\Sigma_e^{-1})e_t\end{pmatrix} && \text{(A.24)}\\
&= \begin{pmatrix} r_t-\mu_{t-1}\\ (\beta_{t-1}^{\top}\Sigma_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^{\top}\Sigma_e^{-1}r_t-\lambda_{t-1}\\ \left((\lambda_{t-1}\lambda_{t-1}^{\top}+\Sigma_u)^{-1}(\lambda_{t-1}+u_t)\right)\otimes e_t\end{pmatrix} && \text{(A.25)}
\end{aligned}
$$

□

A.2 Proof of Proposition 2

Proof of Proposition 2. In case of traded factor, the observation density of u_t and e_t can be derived as follows

$$p(r_t, f_t|\mathcal{F}_{t-1}, \theta_t) = (2\pi)^{-\frac{N}{2}} |\Omega_e|^{-\frac{1}{2}} \exp\left\{-\frac{e_t^\top \Sigma_e^{-1} e_t}{2}\right\} (2\pi)^{-\frac{K}{2}} |\Sigma_u|^{-\frac{1}{2}} \exp\left\{-\frac{u_t^\top \Sigma_u^{-1} u_t}{2}\right\} \quad \text{(A.26)}$$

The conditional log-likelihood of the observation density can then be written as

$$\begin{aligned} l_t = \ln p(r_t, f_t|\mathcal{F}_{t-1}, \theta_t) =& -(N+K)\ln\pi - \frac{1}{2}\ln|\Sigma_e| - \frac{1}{2}\ln|\Sigma_u| - \frac{1}{2}(r_t - \beta_{t-1}f_t)^\top \Sigma_e^{-1}(r_t - \beta_{t-1}f_t) \\ & - \frac{1}{2}(f_t - \lambda_{t-1})^\top \Sigma_u^{-1}(f_t - \lambda_{t-1}). \end{aligned} \quad \text{(A.27)}$$

The score with respect to the risk price λ can be derived as:

$$\nabla_t^\lambda = \frac{\partial l_t}{\partial \lambda_{t-1}} = \Sigma_u^{-1}(f_t - \lambda_{t-1}) \quad \text{(A.28)}$$

and the corresponding Fisher information as

$$\mathcal{I}_t^\lambda = \mathbb{E}_{t-1}\left(\nabla_t^\lambda \nabla_t^{\lambda\top}\right) = \Sigma_u^{-1}. \quad \text{(A.29)}$$

The driving sequence for the risk price updating can therefore derived as

$$s_t^\lambda = (\mathcal{I}_t^\lambda)^{-1}\nabla_t^\lambda = f_t - \lambda_{t-1}. \quad \text{(A.30)}$$

For the risk exposure updating, we derive

$$\nabla_t^\beta = \frac{\partial l_t}{\partial vec(\beta_{t-1})} = (f_t \otimes I_N)\Sigma_e^{-1}(r_t - \beta_{t-1}f_t) = (f_t \otimes \Sigma_e^{-1})(r_t - \beta_{t-1}f_t) \quad \text{(A.31)}$$

The corresponding Fisher information is given by

$$\begin{aligned} \mathcal{I}_t^\beta = -\mathbb{E}_{t-1}\left(\frac{\partial}{\partial vec(\beta_{t-1})}\nabla_t^\beta\right) &= \mathbb{E}_{t-1}\left((f_t \otimes I_N)(f_t^\top \otimes \Sigma_e^{-1})\right) = \mathbb{E}_{t-1}\left(f_t f_t^\top \otimes \Sigma_e^{-1}\right) && \text{(A.32)} \\ &= \left(\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u\right) \otimes \Sigma_e^{-1} && \text{(A.33)} \end{aligned}$$

such that the driving sequence for the beta updating can be derived as

$$\begin{aligned} s_t^\beta = (\mathcal{I}_t^\beta)^{-1}\nabla_t^\beta &= \left(\left(\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u\right)^{-1} \otimes \Sigma_e\right)(f_t \otimes \Sigma_e^{-1})(r_t - \beta_{t-1}f_t) && \text{(A.34)} \\ &= \left(\left(\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u\right)^{-1} f_t \otimes I_N\right)(r_t - \beta_{t-1}f_t) = vec\left(e_t f_t^\top (\lambda_{t-1}\lambda_{t-1}^\top + \Sigma_u)^{-1}\right) && \text{(A.35)} \end{aligned}$$

□

A.3 Proof of Proposition 3

Proof of Proposition 2. Because of Theroem 2.16 in Fang et al. (1990), it holds $x_t := (f_t^\top, r_t^\top)^\top \sim \mathcal{E}_{N+K}(\mu_x, \Omega_x, \psi)$ with

$$\mu_x = \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1}\lambda_{t-1} \end{pmatrix} \quad \text{and} \quad \Omega_x = \begin{pmatrix} \Omega_u & \Omega_u \beta_{t-1}^\top \\ \beta_{t-1}\Omega_u & \beta_{t-1}\Omega_u\beta_{t-1}^\top + \Omega_e \end{pmatrix}. \tag{A.36}$$

The log observation density can then be derived as

$$\ln p(x_t|\mathcal{F}_{t-1}, \theta) = -\frac{1}{2}\ln|\Omega_x| + \ln\psi\left((x_t - \mu_x)^\top \Omega_x^{-1}(x_t - \mu_x)\right) \tag{A.37}$$

$$= -\frac{1}{2}\ln|\Omega_x| + \ln\psi\left(\|\tilde{x}_t\|^2\right). \tag{A.38}$$

The scores for the factor mean and risk price updating are given by

$$\nabla_t^\delta = -2\frac{\partial \ln p(x_t|\mathcal{F}_{t-1}, \theta)}{\partial \delta_{t-1}} = -2\frac{\psi'\left(\|\tilde{x}_t\|^2\right)}{\psi\left(\|\tilde{x}_t\|^2\right)} D\Omega_x^{-1}(x_t - \mu_x) \tag{A.39}$$

with $\delta_t = (\mu_t^\top, \lambda_t^\top)^\top$ and $D = \frac{\partial \mu_x}{\partial \delta_{t-1}}$. Define $\tilde{x}_t = \Omega_x^{-1/2}(x_t - \mu_x)$ with $\Omega_x^{1/2}$ being the Cholesky factor of Ω_x. Remark that $\tilde{x}_t$ is spherically distributed and therefore fulfills the conditions for applying Lemma 1. The Fisher information matrix can then be computed as

$$\mathcal{I}_t^\delta = \mathbb{E}_{t-1}\left(\nabla_t^\delta \nabla_t^{\delta\top}\right) = 4\mathbb{E}_{t-1}\left(\left(\frac{\psi'\left(\|\tilde{x}_t\|^2\right)}{\psi\left(\|\tilde{x}_t\|^2\right)}\right)^2 D\Omega_x^{-1}(x_t - \mu_x)(x_t - \mu_x)^\top \Omega_x^{-1} D^\top\right) \tag{A.40}$$

$$= 4\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2 \left(\frac{\psi'\left(\|\tilde{x}_t\|^2\right)}{\psi\left(\|\tilde{x}_t\|^2\right)}\right)^2 D\Omega_x^{-1/2}\frac{\tilde{x}_t}{\|\tilde{x}_t\|}\frac{\tilde{x}_t^\top}{\|\tilde{x}_t\|}\Omega_x^{-1/2}D^\top\right) \tag{A.41}$$

$$= 4\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2 \left(\frac{\psi'\left(\|\tilde{x}_t\|^2\right)}{\psi\left(\|\tilde{x}_t\|^2\right)}\right)^2\right) D\Omega_x^{-1/2}\mathbb{E}_{t-1}\left(\frac{\tilde{x}_t}{\|\tilde{x}_t\|}\frac{\tilde{x}_t^\top}{\|\tilde{x}_t\|}\right)\Omega_x^{-1/2}D^\top \tag{A.42}$$

$$= \frac{4}{N+K}\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2 \left(\frac{\psi'\left(\|\tilde{x}_t\|^2\right)}{\psi\left(\|\tilde{x}_t\|^2\right)}\right)^2\right) D\Omega_x^{-1}D^\top \tag{A.43}$$

The third and fourth equality result from Lemma 1(i) and (ii), respectively. Because of

$$\Omega_x^{-1} = \begin{pmatrix} \Omega_u^{-1} + \beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1} & -\beta_{t-1}^\top\Omega_e^{-1} \\ -\Omega_e^{-1}\beta_{t-1} & \Omega_e^{-1} \end{pmatrix} \quad \text{and} \quad D = \frac{\partial \mu_x}{\partial \delta_{t-1}} = \begin{pmatrix} I_K & 0 \\ 0 & \beta^\top \end{pmatrix}, \tag{A.44}$$

we get

$$s_t^\delta = (\mathcal{I}_t^\delta)^{-1}\nabla_t^\delta = \frac{-(N+K)\frac{\psi'(\|\tilde{x}_t\|^2)}{\psi(\|\tilde{x}_t\|^2)}}{2\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2\left(\frac{\psi'(\|\tilde{x}_t\|^2)}{\psi(\|\tilde{x}_t\|^2)}\right)^2\right)}(D\Omega_x^{-1}D^\top)^{-1}D\Omega_x^{-1}(x_t-\mu_x) \tag{A.45}$$

$$= C(\|\tilde{x}_t\|^2,\psi)\begin{pmatrix}\Omega_u^{-1}+\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1} & -\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1}\\ -\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1} & \beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1}\end{pmatrix}^{-1}$$

$$\cdot\begin{pmatrix}(\Omega_u^{-1}+\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1})(f_t-\mu_{t-1})-\beta_{t-1}^\top\Omega_e^{-1}(r_t-\beta_{t-1}\lambda_{t-1})\\ \beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1}(f_t-\mu_{t-1})-\beta_{t-1}^\top\Omega_e^{-1}(r_t-\beta_{t-1}\lambda_{t-1})\end{pmatrix} \tag{A.46}$$

$$= C(\|\tilde{x}_t\|^2,\psi)\begin{pmatrix}\Omega_u & \Omega_u\\ \Omega_u & (\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1})^{-1}+\Omega_u\end{pmatrix}\begin{pmatrix}(\Omega_u^{-1}+\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1})(f_t-\mu_{t-1})-\beta_{t-1}^\top\Omega_e^{-1}(r_t-\beta_{t-1}\lambda_{t-1})\\ -\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1}(f_t-\mu_{t-1})+\beta_{t-1}^\top\Omega_e^{-1}(r_t-\beta_{t-1}\lambda_{t-1})\end{pmatrix} \tag{A.47}$$

$$= C(\|\tilde{x}_t\|^2,\psi)\begin{pmatrix}f_t-\mu_{t-1}\\ (\beta_{t-1}^\top\Omega_e^{-1}\beta_{t-1})^{-1}\beta_{t-1}^\top\Omega_e^{-1}r_t-\lambda_{t-1}\end{pmatrix}. \tag{A.48}$$

□

A.4 Proof of Corollary 1

(i) The density generator of a (N+K)-dimensional normal distribution is given by $\psi(x) = (2\pi)^{N/2}e^{-x/2}$. By observing that ψ solves the differential equation $\psi' = -\frac{1}{2}\psi$, we find that $-2\psi'(\|\tilde{x}_t\|)/\psi(\|\tilde{x}_t\|) = 1$ and therefore

$$C(\|\tilde{x}_t\|,\psi) = \left(\frac{4}{N+K}\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2\left(\frac{-\psi'(\|\tilde{x}_t\|)/2}{\psi(\|\tilde{x}_t\|)}\right)^2\right)\right)^{-1} \tag{A.49}$$

$$= \left(\frac{1}{N+K}\mathbb{E}_{t-1}\left(\|\tilde{x}_t\|^2\right)\right)^{-1} = \left(\frac{1}{N+K}\sum_{i=1}^{N+K}\mathbb{E}_{t-1}\left(\tilde{x}_{it}^\top\tilde{x}_{it}\right)\right)^{-1} = 1 \tag{A.50}$$

because $\tilde{x}_t \sim N(0, I_{N+K})$.

(ii) The density generator ψ of the (N+K)-dimensional Student's t-distribution is therefore given by

$$\psi(x) = \frac{\Gamma\left(\frac{\nu+N+K}{2}\right)}{(\nu\pi)^{\frac{N+K}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x}{\nu}\right)^{-\frac{\nu+N+K}{2}} \tag{A.51}$$

and can be used to compute

$$-\frac{\psi'(\|\tilde{x}_t\|^2)}{\psi(\|\tilde{x}_t\|^2)} = -\frac{-\frac{\nu+N+K}{2}\left(1+\frac{\|\tilde{x}_t\|^2}{\nu}\right)^{-\frac{\nu+N+K}{2}-1}\frac{1}{\nu}}{\left(1+\frac{\|\tilde{x}_t\|^2}{\nu}\right)^{-\frac{\nu+N+K}{2}}} = \frac{1}{2}\frac{\nu+N+K}{\nu+\|\tilde{x}_t\|^2}. \tag{A.52}$$

The scaling can be reformulated as

$$C(\|\tilde{x}_t\|, \psi) = \frac{(N+K)^2}{(\nu+N+K)(\nu+\|\tilde{x}_t\|^2)}\left(\mathbb{E}_{t-1}\left(\frac{Z}{\left(\frac{\nu}{N+K}+Z\right)^2}\right)\right)^{-1}, \tag{A.53}$$

with $Z := \|\tilde{x}_t\|^2/(N+K)$. Since $\|\tilde{x}_t\|^2$ is a sum of $N+K$ squared centered t-distributed random numbers with degrees-of-freedom-parameter ν, we know that $Z \sim F(N+K, \nu_e)$ (see, for example, p.22 in Fang et al. (1990)). Applying Lemma 2 yields

$$C(\|\tilde{x}_t\|, \psi) = \frac{\nu_e + N + K + 2}{\nu + \|\tilde{x}_t\|^2}. \tag{A.54}$$

□

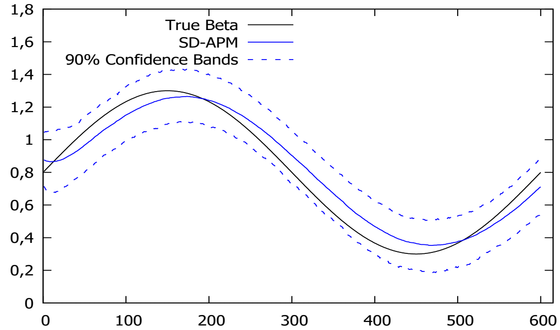
B Time-Varying Beta Simulation Results

Table 6: Dynamic Beta Processes

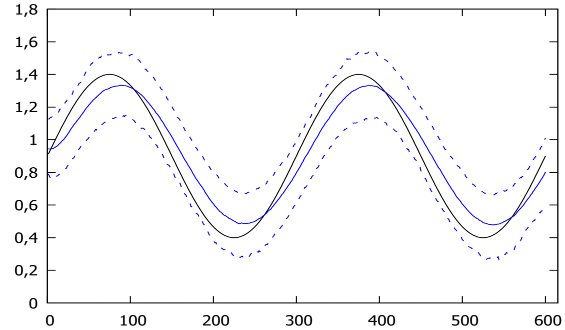
$\beta_{3,t}$	Slow Cycle	$0.8 + 0.5 \cdot \sin(2\pi t/T)$
$\beta_{4,t}$	Fast Cycle	$0.9 + 0.5 \cdot \sin(4\pi t/T)$
$\beta_{5,t}$	Constant	1
$\beta_{6,t}$	Linear Trend	$1.1 + 0.5t/T$
$\beta_{7,t}$	Break	$1.2 + 0.5 \cdot 1(t < T/2) - 0.5 \cdot 1(t \geq T/2)$
$\beta_{8,t}$	Several Breaks	$1.3 + 0.5 \cdot 1(t < T/4 \wedge 2T/4 \leq t < 3T/4)$ $-0.5 \cdot 1(T/4 \leq t < 2T/4 \wedge t \geq 3T/4)$

This section provides simulation evidence that the SD-APM is able to adequately filter time-varying betas in addition to time-varying risk prices and factor means. The data generating process (DGP) ist the same as discussed in section 3.1, but with some of the betas being varying in time. Table 6 shows the different specifications for β_3 to β_8. The remaining betas are constant at their previously calibrated level. The time-varying beta simulation exercise is conducted on a panel with N=10 assets and T=600, as this choice reflects most adequately the size of the data set in the empirical stock return application. A (Gaussian) SD-APM model with the same specification as described in section 3.2 but with score-driven beta dynamics is estimated on S=2500 Monte Carlo replications. Coefficient matrices A^β and B^β are restricted to be diagonal and score-driven betas are initialize at the constant FMB estimate.

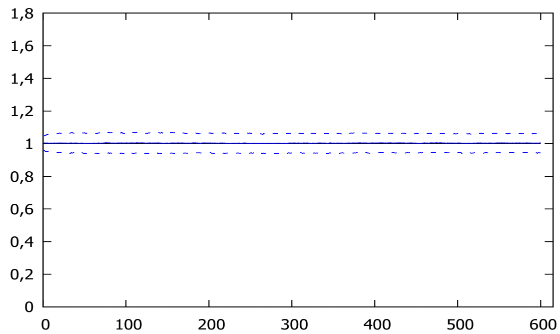
Figure 5 shows beta estimates from the SD-APM averaged across S=2500 Monte Carlo replications alongside 90% confidence bands and the true trajectory. Based on the average estimates and the rather narrow confidence bands, we see that the SD-APM is quite successful in tracking the different beta dynamics. As characteristic for observation-based models, the SD-APM estimates the beta dynamics with a small delay, best seen in the cycles in panels (a) and (b). Panel (c)



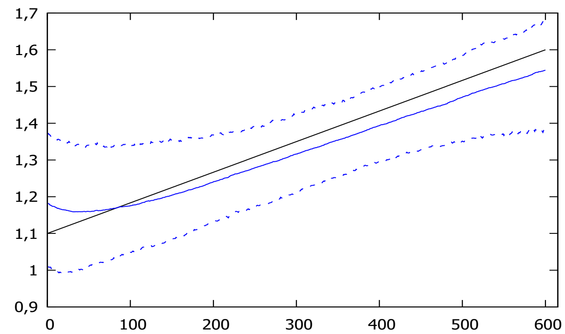
(a) Slow Cycle



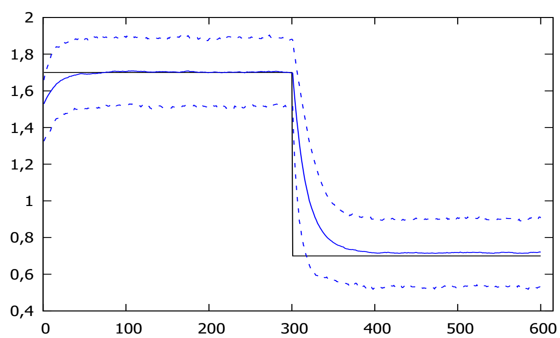
(b) Fast Cycle



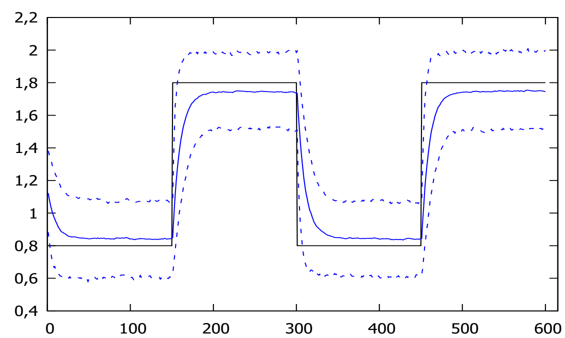
(c) Constant



(d) Linear Trend



(e) Break



(f) Several Breaks

Figure 5: **Time-Varying Beta Estimates.** This figure shows estimated factor risk exposures from a (Gaussian) SD-APM specification with time-varying betas.

reveals that the SD-APM can also nicely adapt to situations in which the exposure parameter is indeed constant over time. Linear trends can be captured as seen in panel (d), but the average estimate slightly drifts apart for higher t. Moreover, breaks are nicely anticipated by the SD-APM, as can be seen in in panel (e). The adjustment period after the break takes roughly 80 periods. The performance weakens with the occurrence of several breaks in panel (f), but the SD-APM still correctly detects the breaks and adjusts the parameter accordingly.

In conclusion, we can tell that the SD-APM is able to track different styles of time-variation in betas. This provides reasonable confidence that the results from the constant beta simulation study hold true in a more general setting including time-varying betas.

References

Adrian, T., Crump, R.K., Moench, E., 2015. Regression-based Estimation of Dynamic Asset Pricing Models. Journal of Financial Economics 118, 211–244.

Adrian, T., Crump, R.K., Vogt, E., 2019. Nonlinearity and Flight-to-Safety in the Risk-Return Trade-Off for Stocks and Bonds. Journal of Finance 74, 1931–1973.

Ang, A., Kristensen, D., 2012. Testing Conditional Factor Models. Journal of Financial Economics 106, 132–156.

Bernardi, M., Catania, L., 2019. Switching Generalized Autoregressive Score Copula Models with Application to Systemic Risk. Journal of Applied Econometrics 34, 43–65.

Blasques, F., van Brummelen, J., Koopman, S.J., Lucas, A., 2022. Maximum likelihood estimation for score-driven models. Journal of Econometrics 227, 325–346.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307–327.

Braun, P.A., Nelson, D.B., Sunier, A.M., 1995. Good News, Bad News, Volatility, and Betas. Journal of Finance 50, 1575–1603.

Campbell, J.Y., 1987. Stock Returns and the Term Structure. Journal of Financial Economics 18, 373–399.

Campbell, J.Y., Shiller, R.J., 1988. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Thompson, S.B., 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509–1531.

Carhart, M.M., 1997. On Persistence in Mutual Fund Performance. Journal of Finance 52, 57–82.

Cochrane, J.H., 2008. The Dog That Did Not Bark: A Defense of Return Predictability. Review of Financial Studies 21, 1533–1575.

Cochrane, J.H., 2011. Discount Rates. Journal of Finance 66, 1047–1108.

Creal, D., Kim, J., 2021. Empirical Asset Pricing with Bayesian Regression Trees. Working Paper University of Notre Dame, Department of Economics.

Creal, D., Koopman, S.J., Lucas, A., 2011. A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations. Journal of Business & Economic Statistics 29, 552–563.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized Autoregressive Score Models with Applications. Journal of Applied Econometrics 28, 777–795.

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

Embrechts, P., Frey, R., McNeil, A.J., 2015. Quantitative Risk Management: Concepts, Techniques and Tools (Revised Ed.). Princeton University Press.

Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50, 987–1007.

Fama, E.F., French, K.R., 1989. Business Conditions and Expected Returns on Stocks and Bonds. Journal of Financial Economics 25, 23–49.

Fama, E.F., French, K.R., 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., MacBeth, J.D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607–636.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall/CRC, London and New York.

Ferson, W.E., Harvey, C.R., 1991. The Variation of Economic Risk Premiums. Journal of Political Economy 99, 385–415.

Ferson, W.E., Harvey, C.R., 1999. Conditioning Variables and the Cross Section of Stock Returns. Journal of Finance 54, 1325–1360.

Gagliardini, P., Ossola, E., Scaillet, O., 2016. Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets. Econometrica 84, 985–1046.

Gagliardini, P., Ossola, E., Scaillet, O., 2019. Estimation of Large Dimensional Conditional Factor Models in Finance. Swiss Finance Institute Research Paper No. 19-46.

Ghysels, E., 1998. On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt? Journal of Finance 53, 549–573.

Gorgi, P., Hansen, P.R., Janus, P., Koopman, S.J., 2019. Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model. Journal of Financial Econometrics 17, 1–32.

Harvey, A., Lange, R.J., 2017. Volatility Modeling with a Generalized t Distribution. Journal of Time Series Analysis 38, 175–190.

Harvey, A., Luati, A., 2014. Filtering With Heavy Tails. Journal of the American Statistical Association 109, 1112–1122.

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

Harvey, C.R., 2001. The Specification of Conditional Expectations. Journal of Empirical Finance 8, 573–637.

Jagannathan, R., Wang, Z., 1996. The Conditional CAPM and the Cross-Section of Expected Returns. Journal of Finance 51, 3–53.

Keim, D., Stambaugh, R., 1986. Predicting Returns in the Stock and Bond Markets. Journal of Financial Economics 17, 357–390.

Lettau, M., Ludvigson, S., 2001. Consumption, Aggregate Wealth, and Expected Stock Returns. Journal of Finance 56, 815–849.

Oh, D.H., Patton, A.J., 2018. Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads. Journal of Business & Economic Statistics 36, 181–195.

Shanken, J., 1992. On the Estimation of Beta-Pricing Models. Review of Financial Studies 5, 1–33.