Regularized GMM for Time-Varying Models with Applications to Asset Pricing*

Liyuan Cui[†] Guanhao Feng[‡] Yongmiao Hong[§¶]

December 15, 2022

Abstract

We develop a novel regularized GMM (RegGMM) approach to estimating time-varying coefficient models via a ridge fusion penalty with a high-dimensional set of moment conditions. Our RegGMM procedure only requires a mild condition on the total amount of oscillations between consecutive parameter values over the whole sample period, which allows for both abrupt structural breaks and smooth changes. While enjoying a closed-form solution for linear models, RegGMM avoids smoothed nonparametric estimation and implements a global one-step procedure. We establish consistency and derive the convergence rate and limiting distribution of the RegGMM estimator for independent and dependent observations. The simulation study shows its robust finite sample performance over existing methods under various scenarios. When applied to asset pricing modeling, RegGMM provides an alternative solution for estimating the time-varying stochastic discount factor model by utilizing a large cross section and/or many conditioning variables. We apply our method to U.S. equity data from 1972 to 2021. Reflecting the macroeconomic information, our time-varying estimate paths for factor risk prices respond to changing performance for multiple risk factors and summarize potential regime-switching scenarios. By outperforming alternative methods, we document the gains in asset pricing and investment performance from RegGMM for both in-sample and out-of-sample analysis.

Key Words: GMM, ridge fusion penalty, stochastic discount factor, time-varying coefficient model.

*We are grateful to Jesús Fernández-Villaverde (editor) and three anonymous referees for constructive suggestions. We appreciate Fuwei Jiang, Jia Li, Whitney Newey, Aman Ullah, Dacheng Xiu, Jun Yu, Xiaoyan Zhang, and participants at Tsinghua PBC School of Finance, Tulane University, Singapore Management University, 2019 Asian Meeting of the Econometric Society, DUFE Econometrics Conference, Junior Financial Econometric Workshop at Zhejiang University, 16th International Symposium on Econometric Theory and Applications, and 2019 European Meeting of the Econometric Society for their comments and suggestions. We are thankful for the lab support by Jiajun Yuan, Zengguang Zhong, and MOE Key Laboratory of Econometrics at Xiamen University. Jingyi Yao, Yuanzhi Wang, and Jiangshan Yang have provided excellent research assistance. Hong's research is supported by China National Science Foundation through the Fundamental Scientific Center Grant (No. 71988101).

[†]Department of Economics and Finance, City University of Hong Kong. Email: liyuan.cui@cityu.edu.hk.

[‡]Department of Management Sciences, City University of Hong Kong. Email: gavin.feng@cityu.edu.hk.

[§]School of Economics and Management, University of Chinese Academy of Sciences, Center for Forecasting Science and Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: yh20@cornell.edu.

[¶]Corresponding Author.

1 Introduction

There is growing empirical evidence that the structure of economic relationships is changing over time (e.g., Stock and Watson, 1996; Qu and Perron, 2007; Brunnermeier and Nagel, 2008; Hong et al., 2017). Capturing time-varying economic relationships is important for inference and forecasts (e.g., Koop and Korobilis, 2012). Economic theory (e.g., Lucas, 1976) suggests that it is appropriate to view behavioral relationships varying over time, yet it does not indicate how to capture the time variation structurally. There are a few popular approaches to modeling time variation in econometrics. The first is to assume that time-varying parameters are a deterministic function of time (e.g., Robinson, 1989; Orbe et al., 2000; Chen and Hong, 2012; Chen et al., 2018), which is appropriate when the driving factors for time variation are external events, such as policy shifts, technology progress, and preference changes. The second is to specify time variation as a function of observable variables (e.g., Ferson and Harvey, 1991; Nagel and Singleton, 2011), which has desirable interpretability. The third is to model time variation as a latent stochastic process, either stationary or unit root (e.g., Engle, 2002; Dangl and Halling, 2012), which is often used when time-varying parameters are driven by unobservable factors, such as the evolving beliefs of policymakers that lead to policy changes (Cogley and Sargent, 2001).

In a seminal paper, Hansen (1982) proposes the generalized method of moments (GMM), which provides a general framework to unify many estimation and inference procedures, including those for regression with endogeneity. GMM has been widely used to exploit asset pricing models by solving dynamic conditional pricing moment restrictions (Hansen and Singleton, 1982).[1] There has been a growing interest in developing new GMM procedures to accommodate time-varying parameters. For example, Andrews (1993) extends GMM to allow for structural breaks and develops a test to detect unknown breakpoints. Based on Lewbel's (2007) local smoothing GMM estimator, Gospodinov and Otsu (2012) propose estimating time-varying parameters via local kernel smoothing by using a nonparametric conditional moment. Gagliardini et al. (2011) extend the nonparametric method of moments estimation to handle unconditional and conditional moment restrictions. Gagliardini and Ronchetti (2020) employ a local smoothing GMM to characterize a conditional Hansen-Jagannathan (HJ) distance for dynamic pricing errors. Cui et al. (2021) estimate time-varying price-dividend ratios via a global smoothing GMM, whose moment conditions step from Euler equations.

The aforementioned methods make important methodological contributions and significantly extend the scope of application for GMM. However, there are some undesirable features of the existing GMM methods for time-varying models. First, modeling time-varying parameters as a function of observable variables may suffer from the omitted variable problem and the curse of dimensionality issue for smoothed nonparametric methods. An exception is Antoine et al. (2020), which proposes employing a broader choice of conditioning variables to estimate time variation via fixed bandwidth asymptotics. Second, modeling time variation as a deterministic function of the scaled time requires certain smoothness (differentiability) conditions for smooth, nonparametric methods that rule out abrupt structural breaks. Finally, modeling time-varying parameters as a latent stochastic process may suffer from model misspecification. We note that to resolve the curse-of-dimensionality issue, Creal et al. (2018) model time-varying parameters in GMM as a

[1] Hansen and Jagannathan (1997) and Kan and Zhou (1999), provide GMM estimations to the constant parameter stochastic discount factor model. Ghysels (1998) use GMM to estimate the conditional CAPM with time-varying betas.

parametric autoregressive-type updating scheme and provide asymptotic properties for the updating parameters with a special choice of instrumental variables,[2] which is subject to model specification risk.

It has been a long-standing challenge in the literature to avoid model misspecification for time variation and allow for various types of structural changes. This paper proposes a novel regularized GMM (RegGMM) approach to estimating time-varying parameters, which depend on the information set, without assuming its relationships with covariates or time. By utilizing a high-dimensional set of moment conditions with a ridge fusion penalty, RegGMM estimates all realization values of time-varying parameters over the sample period, which avoids specifying the data-generating process (DGP) for time-varying parameters and allows for various types of time variation.[3] Our RegGMM objective function contains a quadratic form of the sample moment vectors that time-varying parameters satisfy and a smoothness condition that regularizes the parameter time-variation magnitude. Built on a mild condition that the total amount of oscillations between consecutive realization values over the sample period is bounded, we promote similarities between the coefficient values over consecutive periods via a ridge fusion penalty, defined as the sum of the squared first differences of unknown parameters over time.

Our proposed regularization scheme is related to the fused LASSO penalty (Tibshirani et al., 2005) and the ridge fusion penalty (Price et al., 2015), which assist in variables selection and precision matrix estimation. The ridge fusion penalty in a high-dimensional moment condition setup helps consistently estimate all time-varying parameter values over the sample period, without specifying the time variation. In particular, it allows for various types of structural changes, including abrupt breaks and smooth changes. The large dimensionality for the moment conditions becomes a " blessing" rather than a "curse" for RegGMM. RegGMM treats all realization values of time-varying parameters over the sample period as a high-dimensional parameter vector, which can be consistently estimated by trading off the overall parameter oscillations and the satisfaction level of moment restrictions. Avoiding directly assuming deterministic or random processes for time-varying parameter values is also related to the analogy with fixed versus random effects in the panel data regression literature. Fernández-Val and Weidner (2018) propose treating the realization values of the individual and time effects as parameters to be estimated without specifying their distributions or relationships with covariates. Our work is among the first to consistently estimate time-varying parameter values without assuming specifications on time variation but by regularizing the smoothness levels of time-varying parameters, which is coherent with the argument by Fu et al. (2022) that main distinctions between deterministic and random processes mainly lie in the smoothness of time-varying parameters.Our asymptotic results complement existing studies that employ a penalization based on a bounded variation of the parameter process (Horenko, 2010), which are particularly useful when time variation changes continuously over time t, rather than the scaled time (t/T).[4]

[2]The instrumental variables include the lagged time-varying parameter value γ_{t-1}, and the lagged updating descent s_{t-1} so that $\gamma_t = (I - B)\omega + B\gamma_{t-1} + As_{t-1}$ where ω, A, B are unknown updating parameter matrices to be estimated. We refer readers to Creal et al. (2018) for details and thank one referee for suggesting this related work.

[3]Section 2 provides detailed discussions on the construction of the required moment conditions when researchers face a set of cross-sectional conditional moment restrictions and many instrumental variables.

[4]The nonstationary time series literature needs to determine time-dependent weights on some stationary models to resemble nonstationary observations. The employed regularizations include the integration of the squared first derivative of the affiliation function for smooth changes (Horenko, 2010), the fused Lasso penalty for abrupt breaks (Marchenko

Our approach has several appealing features. First, we do not have to specify the driving factors and functional forms for time-varying parameters, so it is free of model misspecification. Second, unlike the smoothed nonparametric methods that assume time variation as a deterministic function of time, we avoid differentiability conditions that rule out abrupt structural breaks and the boundary bias problem near the endpoints of the sample period. Third, unlike the smoothed nonparametric methods that assume time variation as a function of observable variables, we avoid the curse of dimensionality problem with many conditioning variables and the omitted variable problem. Fourth, RegGMM is a global one-step procedure that delivers consistent estimates for all realization values of time-varying parameters simultaneously. Moreover, it enjoys an appealing closed-form solution for linear models. However, we need to point out the trade-off between the flexibility of our approach and its lack of interpretability (e.g., sources) for time variation. Therefore, our approach should be a complement to, not a substitute for, the existing estimation methods for time variation.

An important application of RegGMM is asset pricing modeling. There exists a strong demand for empirical studies on the stochastic discount factor (SDF) models with time-varying parameters. Figure 2 illustrates various changes in time-varying performance for multiple well-known risk factors in recent decades that have posed a challenge for empirical researchers. The current literature prefers using conditioning (i.e., explanatory) variables to model time-varying SDF weights, such as fundamental predictors or macroeconomic variables, for better economic interpretation. Still, they might encounter the variable selection and model misspecification difficulty (Nagel and Singleton, 2011). Another empirical modeling challenge is determining the type of time variation in parameters, where RegGMM stands out for its flexible smoothness requirement for time variation.

Our paper also adds to the fast-growing field of high-dimensional model regularization and machine learning in asset pricing. Feng et al. (2020) develop variable selection methods to select risk factors for the SDF model. Gu et al. (2020) use machine learning to perform dimension reduction and predict cross-sectional returns using many predictors. Regularized portfolio optimization problems with a large number of assets and their connection to the SDF are also addressed (Ao et al., 2019; Kozak et al., 2020). Rather than performing shrinkage or dimension reduction, RegGMM benefits from a large cross section of asset returns and/or a long list of instrumental variables, given their role in constructing the required high-dimensional moment restrictions.

We apply RegGMM to estimate the dynamic SDF model using monthly U.S. equity data from 1972 to 2021 and consider 10 published traded risk factors for the time-varying SDF composition. Because we use macroeconomic conditioning variables, our estimated time-varying factor price paths can reflect market timing information about business cycles, such as treasury bills, inflation, term spread, and default yield. For example, RegGMM produces time-varying risk price estimate paths, which respond to the decaying performance of HML (Value) and the rising performance of SMB (Size) in the 2000s. We find that our time-varying estimates react to the changing performance of multiple risk factors and summarize potential factor regime-switching scenarios. We show that the time-varying SDF model outperforms the constant parameter SDF model in terms of HJ measures, demonstrating the time-varying SDF model's goodness of fit. By outperforming multiple benchmark models, we document the out-of-sample gains for risk-adjusted and model-adjusted

et al., 2018), and the ridge fusion type penalty to de-noise nonstationary time-series observations (Pospíšil et al., 2018). We appreciate one referee for pointing out this intriguing connection.

investment, which suggests that the superior in-sample performance of RegGMM is not due to overfitting.

The rest of this paper proceeds as follows. Section 2 introduces the RegGMM estimation method for time-varying models. Section 2.2 establishes consistency and derives the convergence rate and limiting distribution of the RegGMM estimator. Monte Carlo simulation studies for verifying RegGMM's scope of application are given in section 3. We present an empirical application to a flexible time-varying SDF model and demonstrate its in-sample and out-of-sample performance in section 4. Section 5 concludes the paper, and all technical proofs are provided in Appendix.

2 Regularized GMM Estimation

2.1 Regularized GMM Estimation for Time-Varying Coefficient Models

Given empirical evidence of time-varying dynamic models, we start with the following econometric model:

$$E(u_{t+n}|I_t) = \mathbf{0}_N, \tag{1}$$

where $u_{t+n} = F(X_{t+n}, \gamma_{t+n})$ is a $N \times 1$ vector representing the cross-sectional information with N individual units (e.g., assets), and X_{t+n} contains observable variables. For example, $X_{t+n} = (r'_{t+1}, f'_{t+1})'$ with $n = 1$ in Section 2.3 and $X_{t+n} = (y'_t, x'_t)'$ with $n = 0$ in Section 3. We denote $u_{o,t+n} = F(X_{t+n}, \gamma_{o,t+n})$ for $t = n+1, \cdots, T$. We further denote $z_t \in I_t$ as a $K \times 1$ vector of conditioning variables, which is observed by the modeler.

We consider estimating the unknown time-varying parameters $\gamma_{t+n} = \gamma(I_t)$, which depend on the observations in I_t. Our paper treats the realization values of $\Gamma \equiv \{\gamma_t\}_{t=1}^{T-n}$ as $p(T-n)$ parameters to be estimated, whose dimension grows to infinity as $T \to \infty$. Then, to estimate a high-dimensional vector of parameters $\{\gamma_t\}_{t=1}^{T-n}$, we first transform the conditional moment restriction in (1) into the following unconditional ones by multiplying instrumental variables. Particularly, we consider

$$E[e(U_{t+n}, \gamma_{t+n})] = E[u_{t+n} \otimes \tilde{z}_t] = \mathbf{0}_{\tilde{q}}, \tag{2}$$

where $U_{t+n} = (X'_{t+n}, \tilde{z}'_t)'$, $\otimes$ is the Kronecker product and $\tilde{z}_t$ can be functions of some observable conditioning variables z_t. When N, the number of conditional moments, and K, the number of conditioning variables, are both fixed, without further prior knowledge of time variation, we could consider expanding the conditioning variables, z_t, to a high-dimensional set of instrumental variables by applying a sequence of transformations $\tilde{z}_t = \{\phi(z_t)\}_{t=1}^{\tilde{K}}$ such that the unconditional moments have dimension $\tilde{q} = N\tilde{K}$. The literature has documented substantial benefits of such transformations in different contexts, such as when estimating pricing functions and facilitating model misspecification tests (Hansen and Richard, 1987; Gagliardini and Ronchetti, 2020; Cui et al., 2021).

We propose estimating the realization values Γ in a one-step procedure by utilizing an increasing set of moment conditions from the following regularized optimization problem:

$$\hat{\Gamma} = \arg\min_{\Gamma} \frac{1}{q} \|g_T(\Gamma)\|^2 + \lambda J(\Gamma), \tag{3}$$

where $g_T(\Gamma) \equiv \frac{1}{T-n}\sum_{t=1}^{T-n} e(U_{t+n}, \gamma_{t+n})$ is the sample analogue of equation (2), and $\|\cdot\|$ denotes the Euclidean norm. Because we will use cross-validation (CV) among moment conditions to select the tuning parameter value λ, we denote q as the effective number of moment conditions used in estimation, and we require $q = \tilde{q}(k-1)/k > p(T-n)$ for the k-fold CV.[5] In the rest of this paper, for generality, we will make use of a high-dimensional set of moment restrictions in equation (2) by requiring $q \geq p(T-n) \to \infty$ as the sample size $T \to \infty$, which can be ensured by either $N \to \infty$ or $K \to \infty$, or both, at a suitable rate.

To offset the impact of the divergent dimension of moment restrictions, we normalize the quadratic term by q; otherwise, the total sum of the quadratic sample moment vectors in (3) could be large even if each element is small. The penalized optimization (3) involves a key ingredient $J(\Gamma)$, a ridge fusion penalty, which is defined as the sum of the squared differences of parameters over time:

$$J(\Gamma) = \Gamma' D' D \Gamma = \sum_{t=2}^{T-n} \|\gamma_t - \gamma_{t-1}\|^2, \tag{4}$$

where D is a $p(T-n-1) \times p(T-n)$ matrix such that

$$D_{ij} = \begin{cases} -1 & i = j, \\ 1 & i + p = j, \\ 0 & otherwise. \end{cases} \tag{5}$$

We assume that the total amount of time variation, as measured by $J(\Gamma)$, is bounded, allowing for various structural changes, including abrupt breaks and smooth changes, or a mixture of them. The tuning parameter λ represents the extent of penalization imposed on time variation $J(\Gamma)$. In section 2.2, we explain the selection and evaluation of λ to aid in calibrating the performance of our procedure.

Notably, the ridge fusion regularization on coefficient values over time amounts to shrinking the total consecutive pairwise squared differences of unknown parameters over the sample period. In a special case, if only smooth structural changes exist, in that the true parameter γ_t is continuously differentiable with respect to time or a set of conditioning variables, the ridge fusion penalty is equivalent to penalizing the squared first derivatives of unknown parameters as in Wahba et al. (1995) and Horenko (2010). Under such a scenario, nonparametric local and global smoothing methods are appropriate if econometricians observe the correct set of conditioning variables that drive time variation. However, they suffer from the curse of dimensionality problem if the set of conditioning variables is large. In another important scenario with abrupt structural breaks, existing nonparametric smoothing strategies are no longer applicable due to a loss of estimation consistency. In contrast, RegGMM estimates all time-varying parameter values over the sample without having to specify the DGP for time variation. We allow both abrupt breaks and smooth changes.

We could also use a fused LASSO penalty, $\sum_{t=2}^{T-n} |\gamma_t - \gamma_{t-1}|$, to control for the level of time variation. However, if

[5] We thank one referee for pointing out this aspect. When conducting over-identification tests, such as extending the J-test proposed by Hansen (1982) and the student t test proposed by Dong et al. (2021), we need to regulate the rate at which q will diverge as $T \to \infty$. For high-dimensional moment restriction models with regularization, the over-identification test is an interesting topic, which we leave for future research.

this L_1-type penalty is used, the optimization problem (3) cannot yield a closed-form solution for linear models. More importantly, given the nature of the L_1-type penalty, it is most suitable for abrupt structural breaks because it keeps a few non-zero piece-wise constant parameter values over time while forcing the differences between coefficients in most periods to be zero. It would have little power to capture smooth structural changes over time. Thus, without a certain economically motivated prior for time variation, the ridge fusion regularization in equation (3) appears more suitable for economic studies because it allows for both abrupt breaks and smooth changes as their mixtures.

The intuition behind RegGMM is that the first component in equation (3) provides enriched information for estimating γ_t at each time period. By appropriately penalizing $J(\Gamma)$, we can uncover general time-varying patterns and guarantee a unique global solution given the strict convexity of optimization (3). For linear models, our method delivers an appealing closed-form solution. Another strength of the ridge fusion penalty is that it significantly helps stabilize the resulting estimates because the positive penalty term λ enters into an involved large-dimensional covariance matrix. As a result, the minimum eigenvalue is bounded from below and away from zero.

2.2 Asymptotic Theory

The large sample properties below cover both independent and dependent observations. To simplify our analysis, we use m-dependency to capture the degree of dependency. We first recall the definition of an m-dependent process $\{X_t\}_{t\geq 1}$ and define the distance between two subsets A and B by $d(A,B) := \inf\{|i-j| : i \in A, j \in B\}$. Then, $\{X_t\}_{t\geq 1}$ is an m-dependent process if $\{X_t, t \in A\}$ and $\{X_\tau, \tau \in B\}$ are independent whenever $d(A,B) > m$. We further recall the notation $\Gamma = (\gamma'_{n+1}, \cdots, \gamma'_T)'$ and denote $A(\Gamma) = (T-n)\partial g_T(\Gamma)/\partial\Gamma$ for the rest of the paper. Let $\rho_{\min}(\cdot)$, $\rho_i(\cdot)$ and $\rho_{\max}(\cdot)$ denote the minimum, i-th, and maximum eigenvalues of a matrix, respectively. Further, $\|M\|_F = \sqrt{tr(M'M)}$ denotes the matrix Frobenius norm and $\|M\|_2 = \sqrt{\rho_{\max}(M'M)}$ denotes the matrix L_2-norm.

Assumption 1. *$E(u_{t+n}|I_t) = \mathbf{0}_N$ has a unique zero at $\gamma_{o,t+n}$, which depends on the information set I_t for each t. $\Gamma_o = (\gamma'_{o,n+1}, \cdots, \gamma'_{o,T})' \in S \equiv S_0^{T-n}$ with S_0 being a finite-dimensional compact parameter space.*

Assumption 2. *(i) Let $\{e(U_{t+n}, \gamma_{t+n})\}$ and $\{U_{t+n}\}$ be stationary m-dependent processes with $m/T \to 0$ as $T \to \infty$; (ii) For some $\eta > 0$, a positive constant c_1 exists that $E|e_j(U_{t+n}, \gamma_{o,t+n})|^{2+\eta} \leq c_1 < \infty$ for $1 \leq t \leq T-n$ and $1 \leq j \leq q$.*

Assumption 3. *(i) For any $\epsilon > 0$, there exists a sufficiently small constant $\nu_{T,\epsilon} > 0$ such that $\inf_{\Gamma\in S, \|\Gamma-\Gamma_o\|>\epsilon} \frac{1}{q}\|E[g_T(\Gamma)]\|^2 > \nu_{T,\epsilon}$; (ii) $\max\{\lambda J_o, (1+m)/T\} = o(\nu_{T,\epsilon})$ with $J_o = \Gamma'_o D' D \Gamma_o < \infty$.*

Assumption 4. *For $\tilde{\Gamma}, \bar{\Gamma} \in S$, a measurable positive function $B_1(U_t)$ exists that $\|e_j(U_{t+n}, \tilde{\gamma}_{t+n}) - e_j(U_{t+n}, \bar{\gamma}_{t+n})\| \leq B_1(U_t)\|\tilde{\Gamma} - \bar{\Gamma}\|$, where $E[B_1^2(U_t)] < \infty$ for $1 \leq t \leq T-n$ and $1 \leq j \leq q$.*

Assumption 5. *For Γ in a neighborhood of Γ_o, positive constants c_2 and c_3 exist that $\rho_{\min}\{E\frac{\partial e_j(U_{t+n},\gamma_{t+n})}{\partial \gamma_{t+n}} E\frac{\partial e_j(U_{t+n},\gamma_{t+n})}{\partial \gamma'_{t+n}}\} \geq c_2 > 0$ and $\rho_{\max}\{E\frac{\partial e_j(U_{t+n},\gamma_{t+n})}{\partial \gamma_{t+n}} E\frac{\partial e_j(U_{t+n},\gamma_{t+n})}{\partial \gamma'_{t+n}}\} \leq c_3 < \infty$ for $1 \leq t \leq T-n$ and $1 \leq j \leq q$.*

Assumption 6. *(i) There exist positive constants c_4 and c_5 such that $\rho_{\min}\left\{\frac{1}{q}Ee(U_{t+n}, \gamma_{o,t+n})e(U_{t+n}, \gamma_{o,t+n})'\right\} \geq c_4 > 0$ and $\rho_{\max}\left\{\frac{1}{q}Ee(U_{t+n}, \gamma_{o,t+n})e(U_{t+n}, \gamma_{o,t+n})'\right\} \leq c_5 < \infty$ for all $1 \leq t \leq T-n$; (ii) For $\tilde{\Gamma}, \bar{\Gamma} \in S$, a measurable positive* *function* $B_2(U_t)$ *exists that* $\|\partial e_j(U_{t+n}, \tilde{\gamma}_{t+n})/\partial\gamma_{t+n} - \partial e_j(U_{t+n}, \bar{\gamma}_{t+n})/\partial\gamma_{t+n}\| \leq B_2(U_t)\|\tilde{\Gamma} - \bar{\Gamma}\|$, *where* $E[B_2^2(U_t)] < \infty$ *for* $1 \leq j \leq q$ *and* $1 \leq t \leq T - n$.

Assumption 1 is a standard condition in the GMM literature, which implies that the time-varying parameter vector Γ_o can be uniquely identified through the moment conditions. In this paper, $p(T-n)$, the number of parameters to be estimated, and q, the number of moment conditions, are both large, with $q \geq p(T-n)$ for identification purposes. The compactness condition in Assumption 1 avoids some ill-posedness issues encountered in the literature. It holds when the finite-dimensional parameter space S is closed and bounded.[6] Alternatively, we could consider a compact subset of $R^{p(T-n)}$ that $S_c = \{\Gamma \in R^{p(T-n)} : \|\Gamma\| \leq C_T\}$, where $C_T = o(T/\sqrt{1+m})$ is some positive number that diverges with T. The proof is collected in Appendix B.[7]

Assumption 2 imposes a mild condition on the serial dependency of the dataset over time, where m is allowed to increase with the sample size T as long as $m/T \to 0$. Such a blocking technique on moment conditions is also used in Chang et al. (2015) when the dimensions of both moment restrictions and parameters of interest grow with the sample size T. Like mixing conditions, m-dependency is also commonly used to regulate serial dependency in time series analysis (Hansen and Singleton, 1982; Rao and Sreehari, 2016). For instance, a special case of an m-dependent process is an i.i.d. process where $m = 0$. In another example, a first-order moving average (MA) process with i.i.d. innovations is a stationary m-dependent process with $m = 1$. In nonparametric statistics, one often uses m-dependence to test for independence or lack of correlation. For example, a random indicator variable $W_n = \mathbb{1}\{X_t > X_{t+1}\}$ with $\mathbb{1}(\cdot)$ being the indicator function can be used to test the dependent structure in X_t's, which is also m-dependent (Islak, 2013). Inoue and Shintani (2006) point out that, given the Wold decomposition theorem, a stationary process admits a moving average (MA) representation with possibly infinite order, which can be approximated by an m-dependent process with m growing to infinity with T. We conjecture that similar results from our paper could be established under mixing conditions. However, assuming m-dependence simplifies our analysis, we can establish an explicit relationship that describes how serial dependency may affect the convergence rate. We note that the m-dependence assumption could be further weakened to be a strictly stationary process, and we refer readers to Hong and Lee (2005) for such an extension in a different context.[8]

Assumption 3 (i) is a generalized condition on global identification. Note that we need to scale down the squared norm due to the diverging dimension of moment restrictions. This relaxed condition can also be seen in Dong et al.

[6]We refer readers to Chen (2007) for a detailed discussion.

[7]To obtain time-varying parameter values, existing econometric literature often further assumes some parametric specification on time variation, such as $\gamma_{t+n} = \gamma(t/T)$ or $\gamma_{t+n} = \gamma(\omega_t)$ for some observable $\omega_t \in I_t$ (Chen and Hong, 2012; Dong et al., 2021). Creal et al. (2018) also consider such a conditional moment restriction with time-varying parameters, whose realization values depend on the past data in I_t. However, in this paper, we avoid specifying the DGP for time variation, which is allowed to display various types. Econometricians and empirical researchers may be more interested in discovering the path of the time-varying parameter values.

[8]Specifically, Assumption 2 (i) could be weakened such that for each sufficiently large m, there exists a strictly stationary process $\{e_{m,t+n}\}$ that is measurable with respect to the sigma field generated by $\{e(U_{t+n-1}, \gamma_{t+n-1})', \cdots, e(U_{t+n-m}, \gamma_{t+n-m})'\}'$ and satisfy that as $m \to \infty$, $\{e_{m,t+n}\}$ is independent of $\{e(U_{t+n-m-1}, \gamma_{t+n-m-1})', e(U_{t+n-m-2}, \gamma_{t+n-m-2})', \cdots\}'$, and $E|e_j(U_{t+n}, \gamma_{t+n}) - e_{j,m,t+n}|^2 \leq Cm^{-\zeta}$ for some $\zeta \geq 1$ and all $1 \leq j \leq q$ and $1 \leq t \leq T - n$.

(2021). A stronger version of this condition has been used in the literature with $\nu_{T,\epsilon} = \nu > 0$ (Ai and Chen (2003)). As argued by Chen and Pouzo (2012), given that $(S, \|\cdot\|)$ is compact, the condition in Assumption 3 (ii) can be reduced to $\max\{\lambda J_o, (1+m)/T\} \to 0$. Assumption 3 (ii) also imposes restrictions on the ridge fusion penalty tuning parameter λ and its relationship with q via the normalized squared norm. Similar to Cui et al. (2021), we can let the penalty term play a key role in smoothing when deriving large sample properties. We show in Theorem 1 that when λ satisfies Assumption 3, $\hat{\Gamma}$, which minimizes the RegGMM criterion over a compact space, is consistent. In Theorem 3, we show that Assumption 3 is also required to ensure the asymptotic unbiasedness nature of the estimated values of time-varying parameters.

Assumption 4 is a Lipschitz condition, which is widely used in the GMM literature (Han and Phillips, 2006). It is essential for establishing uniform convergence because it relates to stochastic equicontinuity (Newey, 1991). Assumption 4 also ensures the applicability of our result to time-varying nonlinear GMM models. Dong et al. (2021) use this condition through global smoothing estimation with many moment conditions.

Assumption 5 is to regulate the eigenvalues of moment conditions. Such conditions are employed and justified in Chang et al. (2015) when maximizing a high-dimensional generalized empirical likelihood, which can also be framed into a GMM setup with many parameters and moments. Assumption 6 facilitates the limit theory by imposing an additional Lipschitz condition on the first-order derivative of moments with respect to unknown parameters.

Theorem 1 (Consistency)**.** *Suppose Assumptions 1-4 hold. Let $\hat{\Gamma}$ minimizes RegGMM in (3). If further $(1+m)/T = O(\lambda)$, we have $\max_{n+1\le t\le T} \|\hat{\gamma}_t - \gamma_{o,t}\| = o_p(1)$ as $T \to \infty$.*

Theorem 1 implies the consistency of $\hat{\Gamma}$ that minimizes the RegGMM criterion. It can be achieved by choosing a positive but small penalization parameter that satisfies $(1+m)/T = O(\lambda)$ and $\max\{\lambda J_o, (1+m)/T\} = o(1)$ or a smaller order $o(\nu_{T,\epsilon})$ with $\nu_{T,\epsilon} \to 0$. Intuitively, if λ tends to infinity, it leads to time-invariant parameter estimates with probability tending to 1. Similarly, if λJ_o tends to a finite strictly positive constant, $\hat{\Gamma}$ will not converge in probability to Γ_o. If $\lambda = 0$, it leads to overfitting of time variation. Theorem 1 implies the consistency of the RegGMM estimator for both dependent and independent observations. Note that for estimation consistency and identification purposes, q must go to infinity at a rate at least as fast as T on one hand and satisfy Assumption 3 on the other hand. In the next theorem, we narrow the neighborhood around Γ_o to drive an estimation error upper bound, which requires additional moment assumptions on pricing errors and their first-order derivatives when applied to asset pricing modeling.

Theorem 2 (Convergence Rate)**.** *Suppose Assumptions 1-5 hold and $(1+m)/T \le \lambda J_o \le \frac{1}{2}(1+m)\log T/T$. Then, we have $\max_{n+1\le t\le T} \|\hat{\gamma}_t - \gamma_{o,t}\| = O_p\big(\sqrt{(1+m)\log T/T}\big)$ as $T \to \infty$ and $q \to \infty$.*

Theorem 2 provides an error bound for the RegGMM estimator in a shrinking neighborhood of Γ_o. This covers all time points in the sample period, including the boundary regions near the endpoints. We note that the estimation error bound involves a factor of $\log T$, which is common to recent studies that involve high dimensionality in the machine learning literature. The reason $\log T$ emerges in our proof is that we use the Markov inequality when obtaining a probability bound for the maximum of a sequence of random variables (Huang et al., 2008). Also, the serial dependency m affects the error bound because it affects how fast sample analogues could converge in probability to their population moments (Romano and Wolf, 2000). Since $\log T$ grows to infinity at a rate slower than T^{ϵ} for any small constant $\epsilon > 0$, $\sqrt{\log T/T}$ converges to zero at a rate slower than $\sqrt{1/T^{1+\epsilon}}$. Therefore, the rate obtained in Theorem 2 is slightly slower than the parametric root-T rate if m is fixed or grows at a logarithmic rate as $T \to \infty$. Hence, our method provides desirable estimation accuracy while avoiding the restrictive specification of time variation. In particular, the consistency of the RegGMM estimator does not require γ_t to be continuously differentiable with respect to time or conditioning variables—an otherwise indispensable condition for smoothed nonparametric estimation. Instead, the condition required by our approach is more general and covers a broader range of applications.

Theorem 3 (Asymptotic Normality). *Suppose Assumptions 1-6, and the conditions in Theorem 2 hold. Consider $a = c\mathcal{I} \otimes a_p$ with $\mathcal{I}$ being a $(T-n)\times 1$ vector of ones, $a_p \in R^p$ with $\|a\|^2 = 1$. Let $\bar{H} = \frac{1}{q^2} a_p' D_T V D_T' a_p$ with $D_T = \frac{1}{T-n}\sum_{t=n+1}^{T} E\big[\frac{\partial e(U_{t+n},\gamma_{o,t+n})}{\partial \gamma_{t+n}}\big]$ and $V = var[\frac{1}{\sqrt{T-n}}\sum_{t=1}^{T-n} e(U_{t+n},\gamma_{o,t+n})]$. For $\sqrt{1+m}\log T = o(\sqrt{T})$, we have*

$$\bar{H}^{-1/2}\sqrt{T-n}a'\Big(\big[\frac{1}{q(T-n)^2}A'(\hat{\Gamma})A(\hat{\Gamma}) + \lambda D'D\big]\Big)(\hat{\Gamma} - \Gamma_o) \xrightarrow{d} N(0,1).$$

Theorem 3 recognizes a standard formula for the asymptotic variance of the RegGMM estimator with a correction due to the penalization for the amplitude of parameter variations between consecutive periods. The Hessian matrix, $\mathcal{H}(\hat{\Gamma},\hat{\Gamma}) = \frac{\partial^2}{\partial\Gamma\partial\Gamma'}\frac{1}{q}\|g_T(\hat{\Gamma})\|^2 + \lambda D'D$, contains two parts. The first component is the Hessian function of an unpenalized GMM loss function as in a high-dimensional moment conditional framework that $\frac{\partial^2}{\partial\Gamma\partial\Gamma'}\frac{1}{q}\|g_T(\hat{\Gamma})\|^2 = \frac{1}{q}A'(\hat{\Gamma})A(\hat{\Gamma}) + \Delta(\Gamma,\Gamma)$, where $\Delta(\Gamma,\Gamma) = \frac{1}{q}g_T(\Gamma)\frac{\partial}{\partial\Gamma}A(\Gamma)$. It has been shown that $\frac{1}{q}A'(\hat{\Gamma})A(\hat{\Gamma})$ is almost surely positive definite and $\|\Delta(\Gamma,\Gamma)\|_F = o_p(1)$ for $\Gamma \in S$ (Dong et al., 2021). The second component $\lambda D'D$ is due to the ridge fusion penalty that $\frac{\partial^2}{\partial\Gamma\partial\Gamma'}J(\Gamma) = \lambda D'D$. We note that such an additional term associated with the penalty and the differencing matrix D also appears in the covariance estimator of Li and Ruppert (2008) and Cui et al. (2021), who propose estimating unknown functions in linear regression and high-dimensional GMM by a ridge fusion penalty on coefficient values of B-splines basis functions. We note that Theorem 3 involves a bias term, $\bar{H}^{-1/2}\sqrt{T-n}a'\lambda D'D\Gamma_o$, which is proven to be a small order term and thus assists with a zero mean in Theorem 3. Such a result could be interpreted as an "undersmoothing" condition.[9] Theorem 3 involves a sandwich formula for computing the covariance of our penalized estimator, which coincides with the sandwich expression by the local quadratic approximation method (Fan and Li, 2001).[10] The sandwich formula in Theorem 3 reflects a bias-variance trade-off.

In practice, we would need a consistent estimator for $\bar{H}$, which involves the unknown matrix V in the limiting normal distribution as in Theorem 3. We first define the sample autocovariance function

$$\hat{\Omega}(j) = \begin{bmatrix} \frac{1}{T-n}\sum_{t=j+1}^{T-n} e(U_{t+n},\hat{\gamma}_{t+n})e(U_{t+n-j},\hat{\gamma}_{t+n-j})' & \text{for } j \geq 0 \\ \frac{1}{T-n}\sum_{t=-j+1}^{T-n} e(U_{t+n+j},\hat{\gamma}_{t+n+j})e(U_{t+n},\hat{\gamma}_{t+n})' & \text{for } j < 0 \end{bmatrix}.$$

If m is known, a consistent variance estimator of $\bar{H}$ can be obtained by defining $\hat{V} = \sum_{j=-m}^{m}\hat{\Omega}(j)$. However, in practice,

[9] There is a similar result in Dong et al. (2021), who require the approximation errors to be of a small order term.
[10] Zou (2006) has a similar sandwich form in the adaptive LASSO estimator by iteratively computing the involved ridge regression.

the value of serial dependency m is generally unknown, which is analogous to the case when mixing conditions are used to regulate serial dependency where the α-or β-mixing values are also generally unknown. In such a case, we can employ the well-known class of heteroskedasticity and autocorrelation (HAC) estimators as in Newey and West (1987) and Andrews (1991) without having to estimate or select the practical value of m or mixing values (Inoue, 2006; Ma et al., 2021). Specifically, we consider the class of HAC estimators for V as follows:

$$\hat{V} = \sum_{j=-(T-n-1)}^{T-n-1} k(j/b_T)\hat{\Omega}(j),$$

where b_T is a smoothing parameter that grows with the sample size T and $k(\cdot)$ is a real-valued kernel function, which generally declines as j increases, with $k(0) = 1$.[11] A suitable choice of $k(\cdot)$, such as the Bartlett kernel $k(u) = 1 - |u|$ for $|u| \le 1$ and 0 otherwise, ensures the semi-positive definiteness property of $\hat{V}$ in finite samples. In the present context, the involved smoothing parameter b_T grows to infinity with the sample size T at a slower rate, whose practical value can be determined by the procedure in Inoue (2006) among others.[12]

We then consider the following estimator of $\bar{H}$:

$$\hat{H} = \frac{1}{q^2} a_p' \hat{D}_T \hat{V} \hat{D}_T a_p,$$

where we save notations by denoting $\hat{D}_T \equiv \frac{1}{T-n} \sum_{t=1}^{T-n} \partial e(U_{t+n}, \hat{\gamma}_{t+n})/\partial \gamma_{t+n}$.

Theorem 4 (Asymptotic Variance Estimation)**.** *Suppose Assumptions 1-6 hold. There exists a positive constant c_6 such that $E|e_j(U_{t+n}, \gamma_{o,t+n})|^4 < c_6 < \infty$ for all $1 \le j \le q$ and $1 \le t \le T-n$. Suppose $\max\{m, b_T\} \max_{n+1 \le t \le T} \|\hat{\gamma}_t - \gamma_{o,t}\| = o_p(1)$ and $b_T\sqrt{(1+m)/T} = o(1)$. Then we have $|\hat{H} - \bar{H}| \xrightarrow{p} 0$ as $T \to \infty$.*

Remarks on Asymptotic Theory. The asymptotic analysis will facilitate the construction of the confidence interval for the time-varying parameters and their combinations. The conditions in Theorem 4 are a natural extension of those in Newey and West (1987), which establishes the consistency of the HAC covariance estimator when both the numbers of parameters and moment conditions are fixed.

It is well known that the asymptotic theory of the HAC covariance estimator and related test statistics do not perform well in finite samples. The literature has documented that the bootstrap can help enhance the finite sample performance compared with the critical values based on first-order asymptotic theory in GMM (Hall and Horowitz, 1996). Based on Theorems 3 and 4, we provide detailed descriptions of a bootstrap procedure by Kato (2011) and Inoue and Shintani (2006) for conducting inference in Appendix C.2.

[11]The weights $k(j/b_T)$ satisfy $|k(j/b_T)| \le C$ for finite constant C and $\lim_{b_T \to \infty} k(j/b_T) = 1$ for each j. Andrews (1991) generalizes the class of kernels that satisfy $\{k(\cdot) : R \to [-1,1] | k(0) = 1, k(x) = k(-x), \int_{-\infty}^{\infty} k^2(x)dx < \infty, k(\cdot)$ is continuous at 0 and at all but a finite number of other points$\}$. Examples of kernels include the truncated, Bartlett, Parzen, Tukey-Hanning, and Quadratic spectral functions. We refer readers to Andrews (1991) for details.

[12]The main idea is to approximate the serially dependent moment conditions with a MA representation of possibly large order. Such a strategy is also used in Ng and Perron (1995). Other methods include minimizing the mean squared error of the HAC covariance matrix estimator in (Andrews, 1991), which excludes truncated and trapezoidal kernels.

Estimation efficiency may be improved by including a $q \times q$ weighting matrix $\hat{W}$ that may depend on the sample data, in which case the objective function can be replaced by $\hat{\Gamma} = \arg\min_{\Gamma} \frac{1}{q} \|g_T(\Gamma)\hat{W}^{-1/2}\|^2 + \lambda J(\Gamma)$. Similar to Dong et al. (2021), who study a GMM procedure with global smoothing by choosing $W = I$, for ease of representation, in the main context of this paper, we explore the RegGMM estimator by choosing $\hat{W}$ as the identity matrix but provide a detailed analysis of the consistency of RegGMM under a general weight matrix in Appendix B. Efficiency gain could be further achieved by considering a one-step continuously updated estimator (CUE), which is studied by Newey and Smith (2004) under a conventional GMM framework without regularizations.

We do not have to specify the DGP for time variation in our setup, but we need a large $q \geq p(T-n)$. However, if we impose the additional condition that time variation exhibits certain smoothness over the scaled time (e.g., $\gamma_t = \gamma(t/T)$ for some unknown twice differentiable function $\gamma(\cdot)$), we conjecture that q could be fixed. In such a case, we could use local linear smoothing and formulate a local objective function at each fixed time period $t \in [n+1, T]$:

$$\min_{\beta \in R^{2p}} \|\frac{1}{T} \sum_{s=t+n-[Th]}^{t+n+[Th]} K_{s,t+n} e(U_s, \gamma_s)\|^2 + \lambda \sum_{s=t+n-[Th]+1}^{t+n+[Th]} \|\gamma_s - \gamma_{s-1}\|^2,$$

where $\gamma_s = \alpha_0 + \alpha_1[s-(t+n)]/T$, $\beta = (\alpha_0', \alpha_1')'$, $K_{s,t+n} = \frac{1}{h} K(\frac{s-(t+n)}{Th})$, the kernel $K(\cdot) : [-1,1] \in R^+$ is a prespecified symmetric probability density, and $h = h(T)$ is a bandwidth with $h \to 0$ and $Th \to \infty$ as $T \to \infty$. We conjecture that using the ridge fusion penalty could help alleviate the key smoothing role that the bandwidth h has been playing in local smoothing estimation.

Selection of Penalty Tuning Parameter. RegGMM involves one penalty tuning parameter, λ, which affects its finite sample performance. Andrews (1999) and Inoue (2006) propose studying the GMM estimation efficiency and asymptotic refinement by sampling the cross-sectional moment conditions. Empirically, in an application to asset pricing modeling, a desirable SDF price test assets accurately while being less sensitive to different choices of instrumental variables (Nagel and Singleton, 2011; Gagliardini and Ronchetti, 2020; Antoine et al., 2020). Hence, obtaining a balance between first-order estimation accuracy and second-order stability across different moments is desirable. Hence, we select λ using cross-validation (CV) by sampling from moment conditions.

Specifically, for k-fold CV, $(k-1)$ folds are used for parameter estimation, and the hold-out fold is used to validate the model. Borrowing the notations from Andrews (1999), for each $\lambda \in \mathcal{M}$, for v-th fold, we obtain $\hat{\Gamma}_\lambda^{(v)} = \arg\min_{\Gamma \in S} \frac{1}{q} \|g_T^{(-v)}(\Gamma)\|^2 + \lambda \Gamma D D' \Gamma'$, where $g_T^{(-v)}(\Gamma) = \frac{1}{T-n} \sum_{t=1}^{T-n} I_{-v} e(U_{t+n}, \gamma_{t+n})$ and I_{-v} is a $q \times q$ diagonal matrix whose i-th diagonal value is 0 if the i-th moment condition is included in the v-fold, and 1 otherwise. Then, the tuning parameter can be obtained by minimizing the average predictive error in the testing samples (CV error): $\lambda^* = \arg\min_{\lambda \in \mathcal{M}} \sum_{v=1}^{k} \|g_T^{(v)}(\hat{\Gamma}_\lambda^{(v)})\|^2$, where $g_T^{(v)}(\hat{\Gamma}^{(v)})$ collects the predictive pricing errors in the testing moments. To reduce variation in the assessment due to the sample splitting issue, we perform the 5-fold CV 10 times in our applications and report the average selected tuning parameter. We will assess the validity of a selected tuning parameter by performing a test on whether the practically chosen tuning parameter is optimal, which can produce the least cross-validation error. Practically, we conduct this hypothesis testing procedure and obtain the p-value via the CV with confidence (CVC) method from Lei (2020). We explain the p-value construction details in Appendix C.1.

2.3 Dynamic Stochastic Discount Factor Models.

Figure 2 shows smooth and abrupt changes in time-varying performance for multiple well-known risk factors during recent decades, which challenge empirical asset pricing studies using constant parameter models. We now illustrate the use of RegGMM in the conditional asset pricing literature for the SDF model. In this subsection, we consider a special setup, where the pricing error u_{t+n} is a linear function of time-varying parameters. In this case, the RegGMM estimator enjoys a closed-form solution for the entire set of unknown time-varying parameters over the sample period.

In the absence of arbitrage, a time-varying SDF, m_{t+1}, exists such that for any traded asset i with an excess return at time t of r_t, we have the conditional moment equation:

$$E[m_{t+1}r_{t+1}|I_t] = \mathbf{0}_N, \tag{6}$$

where r_{t+1} is an $N \times 1$ vector of excess returns on N assets. A beta pricing model can be cast in the SDF framework by specifying the SDF as a linear function of f_{t+1}, where f_{t+1} is a $p \times 1$ vector of observable risk factors:

$$m_{t+1} = 1 - \gamma'_{t+1} f_{t+1}, \tag{7}$$

where γ_t is a $p \times 1$ vector of time-varying SDF loadings or risk price parameters. We can plug equation (7) into equation (6) to estimate the SDF loadings. In the following notation, $r_{t+1}f'_{t+1}$ is an $N \times p$ matrix of explanatory variables and

$$\mathbf{0}_N = E[m_{t+1}r_{t+1}|I_t] = E[(1 - \gamma'_{t+1}f_{t+1})r_{t+1}|I_t] = E[r_{t+1} - (r_{t+1}f'_{t+1})\gamma_{t+1}|I_t].$$

The interpretation for γ_{t+1} is important from the economic perspective. Cochrane (1996) assumes $E(f_{t+1}|I_t) = 0_p$, and then, $E(r_{t+1}f'_{t+1}|I_t)$ is the conditional covariance matrix between asset returns and risk factors, and γ_{t+1} is the corresponding conditional factor risk price. When $E(f_{t+1}|I_t) \neq 0_p$ and $E(r_{t+1}f'_{t+1}|I_t)$ is the conditional second moment, the factor risk price interpretation also applies in Cochrane (2009). As discussed in Feng et al. (2020), the risk prices of factors differ from their risk premia. Risk premia are supposed to be positive to compensate for risk-taking behavior. However, risk prices can be negative due to their combination with other factor exposures when constructing the SDF model. The closed-form solution for the unconditional model exists when γ_{t+1} is constant.

In our setup, the time-varying parameter is not assumed to be a deterministic function of time or conditioning variables. As discussed in Cochrane (2009), conditional modeling requires instrumental variables for the GMM estimation. To estimate $\Gamma = \{\gamma_{t+1}\}_{t=1}^{T}$, we transform equation (6) from a conditional SDF representation to an unconditional one by scaling instrumental variables, which can be functions of some conditioning variables. For the choice of conditioning variables in the empirical literature, econometricians usually consider those that do not correlate with future pricing errors but might be weakly correlated with future returns, such as lagged returns, predictable firm or portfolio characteristics, and macroeconomic or aggregate predictors. For example, Welch and Goyal (2008) document inflation, treasury bill return, various aggregate market characteristics, yields for TERM, and default factors as conditioning variables. Suppose the instrumental vector $\tilde{z}_t \in R^{\tilde{K}}$ is uncorrelated with the future pricing error u_{t+1}. Then we have

$$\begin{aligned}\mathbf{0}_{\tilde{q}} &= E\{(1-\gamma'_{t+1}f_{t+1})r_{t+1}\otimes\tilde{z}_t\} \quad (8)\\ &= E(u_{t+1}\otimes\tilde{z}_t), \quad (9)\end{aligned}$$

where $u_{t+1} = m_{t+1}r_{t+1} = (1-\gamma'_{t+1}f_{t+1})r_{t+1}$ is the $N \times 1$ vector of pricing errors, which is also the ex-post discounted return in asset pricing.

To implement the proposed procedure, we introduce the following equivalent matrix expression for the optimization problem in equation (3):

$$\hat{\Gamma} = \arg\min_{\Gamma} \frac{1}{(T-1)^2 q}\|A\Gamma - B\|^2 + \lambda J(\Gamma), \quad (10)$$

where $q = \tilde{q}(k-1)/k$ is the effective number of moment conditions in the k-fold CV procedure, $B = \sum_{t=1}^{T-1} B_t$ with $B_t = r_{t+1}\otimes z_t$ being a $q\times 1$ vector, and $A = (A_1,\cdots,A_{T-1})$ is a $q\times p(T-1)$ matrix with $A_t = (r_{t+1}\otimes z_t)\otimes f'_{t+1}$ being a $q\times p$ matrix for $t = 1,\cdots,T-1$. Therefore, the proposed RegGMM estimator enjoys a closed-form solution:

$$\hat{\Gamma} = [A'A + \lambda q(T-1)^2 D'D]^{-1}A'B, \quad (11)$$

where D is as specified in equation (5) with $n = 1$.

Practitioners often encounter a large pool of conditioning variables that affect return prediction or asset pricing performance. In this paper, dimensionality becomes a blessing because it assists in estimating $p(T-1)$ parameter values. Rather than shrinking or selecting variables from a large pool of conditioning variables, RegGMM utilizes the rich information from a high-dimensional set of moment restrictions to determine the time-varying SDF weights. Hence, it can help handle large-dimensional data and avoid degenerated model performance due to misspecification issues.

RegGMM could also be used to test parametric conditional SDF specifications.[13] For example, one could consider a model specification $m_{t+1} = 1-(\theta' x_t + \gamma_t)' f_{t+1}$, where θ is a matrix of time-invariant parameters and x_t is a vector of conditioning variables that drive the time-varying SDF weights. Nagel and Singleton (2011) consider this linear functional form in testing a conditional SDF model, and Roussanov (2014) propose a nonparametric cross-sectional regression that is robust to functional form but still needs to specify conditioning variables. In our RegGMM setup, one could estimate the path of γ_t, and a constant (or zero if x_t includes an intercept) path subject to sampling variation would suggest linear modeling of risk premia through x_t is adequate. This would provide a direct test for a parametric time-varying SDF model that assumes time-varying SDF loadings as a linear function of a prespecified set of priced risk factors $m_{t+1} = 1-(\theta' x_t)' f_{t+1}$, as is discussed in Cochrane (2009).

[13] We thank one referee for suggesting this interesting extension for future work.

3 Simulation Study

We now conduct comprehensive simulation studies to examine the finite sample performance of RegGMM. We test two different DGPs for robustness. In DGP 1, we model time variation as a deterministic function of time, exhibiting both abrupt breaks and smooth changes. In DGP 2, we focus on the capability of RegGMM to estimate parameter values when time variation is a function of a relatively large set of conditioning variables. For DGP 2, smoothed nonparametric estimators, such as global and local smoothing methods, encounter the curse of dimensionality problems. We also provide performance comparisons by reporting estimation results from existing parametric and smoothed nonparametric methods. Specifically, we conduct simulation exercises in the following standard panel regression setting:

$$y_{i,t} = \gamma_{1,t} + \gamma_{2,t}x_{i,t} + u_{i,t},\ E(u_{i,t}|z_t) = 0 \text{ for } t = 1, \cdots, T \text{ and } i = 1, \cdots, N,$$

where $\gamma_t = (\gamma_{1,t}, \gamma_{2,t})'$ is the unknown time-varying parameter vector, and $x_t \sim N(0, I_N)$ and is orthogonal to the $N \times 1$ conditional pricing error vector u_t, which may have serial dependency $u_{i,t} = \rho_1\epsilon_{i,t-1} + \epsilon_{i,t}$ with $\epsilon_{i,t} \sim i.i.d.N(0, 0.1)$ and $N = 100$. The $K \times 1$ vector z_t with $K = 10$ is a set of conditioning variables that are orthogonal to u_t and follows $z_t = \rho_2 z_{t-1} + \nu_t$ with $\nu_{i,t} \sim i.i.d.N(0, 1)$. This is a set of base instruments to generate a large set of instrumental variables for RegGMM. We examine the finite sample performance of RegGMM under different degrees of serial dependency by altering the values of $\rho_1 \in [0, 0.95]$ and $\rho_2 \in [0.5, 0.95]$. When $\rho_1 = 0$, the unconditional moment functions have serial dependency $m = 1$ with weak or strong linear dependence. When $\rho_1 \neq 0$, the unconditional moments exhibit serial dependency as a mixture of moving average and autoregressive processes, which can be approximated by an $MA(m)$ sequence with m increasing with the sample size (e.g., Inoue, 2006).

We consider two sample sizes, $T = 120$ and 360 with an out-of-sample period of 60 observations. The choice of T covers sample sizes typically encountered for monthly and daily data in the empirical asset pricing literature. To construct the set of instrumental variables, we consider a sequence of orthogonal basis functions in the Hilbert space $L^2[0, 1]$ with $\Psi(z) = \{\psi_j(z)\}_{j=1}^r$, where $\psi_0(z) = 1$ and $\psi_j(z) = \sqrt{2}cos(\pi jz)$ for $j > 1$. Let the instrumental variables $\tilde{z}_t = (1, \Psi(\bar{x}_t)', \Psi(\bar{x}_{t-1})', \Psi(\bar{x}_{t-2})', \Psi(z_{1,t})', \Psi(z_{1,t-1})', \Psi(z_{1,t-2})', \cdots, \Psi(z_{l,t})', \Psi(z_{l,t-1})', \Psi(z_{l,t-2})')'$, where $\bar{x}_t = \frac{1}{N}\sum_{i=1}^N x_{i,t}$. For each T, we let $r = 3$ or 9 and so have $\tilde{K} = 133, 331$ respectively, which correspond to two choices of $\tilde{q} = N\tilde{K} = 13300, 33100$ given $N = 100$. We then have the following unconditional moment restrictions:

$$E[(y_t - \gamma_{1,t} - \gamma_{2,t}x_t) \otimes \tilde{z}_t] = \mathbf{0}_{\tilde{q}},$$

where $n = 0$ in this example and we select the tuning parameter λ via the repeated 5-fold CV and thus effectively use $q = 4\tilde{q}/5 > pT$ moment conditions for RegGMM.

We consider two classes DGPs to generate γ_t. In DGP 1, we consider the time-varying parameter as a deterministic function of time but exhibits multiple structural breaks and different degrees of smoothness. In DGP 2, we allow time variation as a function of 10 conditioning variables z_t, where different conditioning variables drive time variation in different subsampling periods.

DGP 1 [Time-varying parameters as a deterministic function of time]:

$$\gamma_{1,t} = \begin{cases} 2t|sin(4\pi t/T)|/T, & t = 1, \cdots, T/2, \\ G_1(10t/T, 2, 7), & t = T/2+1, \cdots, T \end{cases}$$

and

$$\gamma_{2,t} = \begin{cases} 6(3t/T)^5 - 5(3t/T)^4 + 8(3t/T)^3 - 7(3t/T)^2 + 3t/T, & t = 1, \cdots, T/3, \\ 3\cos(6\pi t/T), & t = T/3+1, \cdots, 2T/3, \\ 9t^2|\sin(9\pi t/T)|/T^2, & t = 2T/3+1, \cdots, T \end{cases}$$

where $G_1(z, k, a) = \{1 + \exp[-k(z-a)]\}^{-1}$.

DGP 2 [Time-varying parameters as a function of conditioning variables]:

$$\gamma_{2,t} = log\left(1 + \left|\sum_{j=1}^{l} \theta_{t,j} z_{t,j}\right|\right) + \sum_{j=1}^{l} \eta_{t,j} \sin z_{t,j} \quad t = 1, 2, ..., T$$

where

$$\theta_{t,j} = \begin{cases} \theta_j \mathbb{I}(j \leq 5), & t = 1, \cdots, T/2, \\ \theta_j \mathbb{I}(j > 5), & t = T/2+1, \cdots, T, \end{cases} \quad \eta_{t,j} = \begin{cases} \eta_j \mathbb{I}(j \leq 5), & t = 1, \cdots, T/2, \\ \eta_j \mathbb{I}(j > 5), & t = T/2+1, \cdots, T, \end{cases}$$

with $\theta_j \sim U(0,1), \eta_j \sim U(0,1)$ and $\gamma_{1,t}$ is as in DGP 1.

We compare RegGMM with Hansen's (1982) constant parameter GMM (CGMM), and various alternative time-varying estimation methods, including local kernel smoothing GMM (LGMM) (Lewbel, 2007), global series smoothing GMM (GGMM) (Dong et al., 2021), and affine GMM (AGMM) (Nagel and Singleton, 2011).[14] For out-of-sample comparison, following Lettau and Pelger (2020), we first estimate parameter values up to time t with a rolling window of T observations and obtain the one-step ahead forecast for the parameter's value at $t+1$ using the estimates at time t.[15] Such an out-of-sample approach, which assumes similarities between the consecutive parameter values, is widely used in econometric and finance studies.[16] We also provide a roll window update for the out-of-sample evaluation in the

[14]We demonstrate the use of these time-varying methods using DGP 1 that $\gamma_t = \gamma(t/T)$, where we need to scale t by the sample size T for estimation consistency (Chen and Hong, 2012). (a) The LGMM estimates are obtained by for each $t \in [Th, T-Th]$, $(\hat{\gamma}_{1,t}, \hat{\gamma}_{2,t}) = \arg\min_{\gamma \in R} \frac{1}{T} \sum_{s=t-Th}^{s=t+Th} K_{st}[(y_{i,t} - \gamma_1(t/T) - \gamma_2(t/T)x_{i,t}) \otimes \tilde{z}_t]^2$; for $t \in [1, Th) \bigcup (T-Th, T]$, we obtain a pseudo data at time t by reflecting the observations from the time period $[t-Th, t]$ (see Chen and Hong (2012)). $K_{st} = \frac{1}{h} K(\frac{s-t}{Th})$ and $K(u)$ is the Epanechnikov kernel function $K(u) = \frac{3}{4}(1-u^2)$ if $|u| \leq 1$ and 0 if $|u| > 1$. The bandwidth $h = cT^{-1/5}$ with $c = 1/\sqrt{12}$ as suggested by Chen and Hong (2012). (b) The GGMM results are based on the series estimation with Chebyshev polynomials $\{\psi(\cdot)\}_{j=1}^{g}$, where g is the order of series expansion and is determined by AIC as suggested in Dong et al. (2021). For all $t \in [1, T]$ and $a \in R^g$ and $b \in R^g$, $\gamma_1(t/T) = \psi'(t/T)a$ and $\gamma_2(t/T) = \psi'(t/T)b$, where $(\hat{a}, \hat{b}) = \arg\min_{a,b} \frac{1}{T} \sum_{t=1}^{T} [(y_{i,t} - \psi'(t/T)a - \psi'(t/T)bx_{i,t}) \otimes \tilde{z}_t]^2$. (c) AGMM assumes time variation as a linear function of the conditioning variables that $\gamma_t = a + b(t/T)$ in DGP 1 and $\gamma_t = z_t'\theta$ in DGP 2. (d) For LGMM, we also have tried $c = 1$ and the one based on CV as in Li and Yang (2011) with the uniform and Daniel kernels. For GGMM, we also tried the Fourier and Spline series. RegGMM demonstrates superior performance over LGMM and GGMM with different kernels, bandwidths, and basis functions.

[15]When conducting out-of-sample studies using local smoothing methods, we construct a pseudo data at time t by reflecting the observations from the period $[t-Th, t]$.

[16]For example, when evaluating latent factor models, Lettau and Pelger (2020) use the factor loadings estimated from the previous T observations on the testing sample. When evaluating covariance matrix estimators, Aït-Sahalia and Xiu

empirical study.

We assess estimation accuracy based on the following performance metrics, including the average absolute estimation error (Avg.Abs.Err.) $\frac{1}{T}\sum_{t=1}^{T}\|\hat{\gamma}_t-\gamma_t\|$, the maximum absolute estimation error (Max.Abs.Err) $\max_{t\in[1,T]}\|\hat{\gamma}_t-\gamma_t\|$, and the average squared moment errors (Avg.Mom.Err.) $\frac{1}{NT}\sum_{t=1}^{T}\hat{u}_t'\hat{u}_t$, where $T=120$ or 360 for in-sample studies and $T=60$ for out-of-sample comparisons. For fair comparisons, we implement CGMM, LGMM, GGMM, and AGMM estimation under the two augmented unconditional moments respectively.

Table 1 summarizes the finite sample results for RegGMM and other methods under DGP 1. CGMM restricts parameters to constant and encounters misspecification for time variation. Although AGMM allows for time variation in parameter values, it also suffers from severe misspecification because it fits parameter values using a straight line. LGMM and GGMM significantly improve estimation accuracy upon CGMM and AGMM. However, they are inferior to RegGMM for all studied norms, even though time-variation satisfies the twice-continuous differentiability requirement in all subsampling periods before and after structural breaks. The reason is that RegGMM can cater to abrupt, smooth, or mixed types of structural changes and thus produces a more accurate estimation than smoothed nonparametric estimation over the whole sample, including the boundary regions. The out-of-sample results in Table 2 also confirm the superior performance of RegGMM, which is robust to different choices of moment conditions, sample sizes, and degrees of serial dependency. We show that even though we assist the existing estimation methods by providing the correct set of conditioning variables, RegGMM still strictly dominates for both the in-sample and out-of-sample results.

Table 3 assesses estimation accuracy when time variation is a function of conditioning variables. We feed CGMM, AGMM, LGMM, and GGMM with the correct set of conditioning variables so that the correct specification for time variation is maintained for these methods. Compared with DGP 1, except for the difficulty in capturing the unknown functional form, DGP 2 raises another serious concern about the curse of dimensionality for smoothed nonparametric methods. In particular, GGMM is unfavorable because of a clear bias-variance trade-off: an increasing order of series expansion reduces the approximation errors but leads to less precise or even inconsistent estimates. Not surprisingly, compared with CGMM and AGMM, GGMM does not help improve estimation accuracy for most cases. We also observe deteriorated performance for LGMM, mainly due to the boundary bias issue and the curse of dimensionality problem.[17] In contrast, RegGMM attains the best estimation accuracy because it circumvents high dimensionality issues by directly estimating time-varying parameter realizations with the assistance of increasingly enriched information due to its ability to strike a balance between model complexity and goodness of fit. In Table 4, we find that the curse of dimensionality and misspecification issues further escalate the challenges faced by all the alternative methods. Still, RegGMM provides a desirable resolution and considerably enhances out-of-sample performance in most cases.

We obtain the RegGMM estimates for each scenario based on the data-driven tuning parameter value λ described

(2017) directly adopt the estimated covariance matrices using data from the previous month for the portfolio rebalancing.

[17]Both local constant and local linear GMM estimators are proven consistent, but the latter doubles the number of parameters to be estimated compared with the first one. When there is a large number of conditioning variables, the number of effective observations that fall in the neighborhood of the realization value of each conditioning variable decreases. Thus, we report the results based on the local constant LGMM based on fixed bandwidth because the fixed-bandwidth asymptotics is helpful when there are many conditioning variables (Antoine et al., 2020)

in section 2.2. For robustness check, we also consider two other choices of tuning parameter, namely, $\lambda = 0$ and $\lambda = 1$, which correspond to overfitting and underfitting for parameter values, respectively. We evaluate the appropriateness of a given tuning parameter by calculating its associated cross-validation error and p-value.[18] A p-value smaller than a certain significance level α indicates the rejection of the optimal hypothesis for a chosen tuning parameter. Table 5 summarizes the tuning parameter performance. We note that the choice of $\lambda = 1$, which results in underfitting, is inappropriate due to the significant CV errors and close-to-zero p-values for all DGPs considered. When $\lambda = 0$, zero penalties are imposed on time variation, and we sometimes fail to reject its appropriateness when the sample size is small. The reason is that fewer parameters need to be estimated when T is small, therefore more moment conditions can aid in accurate estimation. When T is large, the choice of $\lambda = 0$ does not balance the goodness of fit and the magnitude of time variation. Therefore, we need a data-driven tuning parameter value. For each studied DGP with various degrees of serial dependency, the 5-fold repeated CV criterion generates the optimal tuning parameter values with the least CV error and p-values close to 1. The penalty values in DGP 2 are generally smaller than those in DGP 1 because DGP 2 exhibits a more significant amount of time variation and thus requires a smaller penalty. The appropriateness of the chosen tuning parameter λ is robust to different sample sizes and numbers of moment conditions.

Finally, Figure 1 plots the histograms of the studentized RegGMM's time-varying estimates for $\frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{2,t}$ from bootstrap samples with $(\rho_1, \rho_2) = (0.95, 0.95)$. The asymptotic theory suggests that the studentized estimates should follow an $N(0, 1)$ distribution. We document that enabling serial dependency still offers a good approximation in finite samples. The histograms reasonably match the $N(0, 1)$ distribution, which verifies our inference results.

4 Empirical Applications to Asset Pricing Models

4.1 Data

In this section, we examine the performance of the time-varying SDF model based on the U.S. equity monthly data from January 1972 to December 2021. We implement RegGMM using the training sample from 1972 to 2011 to study the in-sample goodness of fit of the model. We consider the recent decade for the out-of-sample study.

The empirical literature primarily uses characteristics-sorted or managed portfolios for testing asset pricing models because these portfolios have stable factor loadings and reflect various risk exposures. We follow the data construction in Feng et al. (2021) and take 61 firm characteristics from 6 major categories: momentum, value, investment, profitability, frictions (or size), and intangibles. To evaluate the RegGMM performance, we try different test assets for different cross-section sizes. First, we use monthly bivariate-sorted 3×2 value-weighted portfolios between size and other characteristics ($3 \times 2 \times 60 = 360$), which considers the effect for small caps. Second, we consider monthly univariate-sorted 5×1 value-weighted portfolios ($10 \times 1 \times 61 = 610$).

We also try different factor combinations for robustness checks. Because of the out-of-sample research design for

[18] See Section 2.2 for the definition of the cross-validation error. The p-value is based on the null hypothesis that a given tuning parameter is optimal. The construction of the p-value is provided in Appendix C.1.

investment performance, only traded factors are considered. We consider 5 factors of Fama and French (2015): excess market factor, small-minus-big (SMB), high-minus-low (HML), robust-minus-weak (RMW), and conservative-minus-aggressive (CMA). For an extended model of 10 factors, we further add momentum (MOM, winner-minus-loser), short-term reversal (STR), long-term reversal (LTR), betting-against-beta (BAB), and quality-minus-junk (QMJ).

Finally, we have two lists of instrumental variables for further robustness checks. We follow Welch and Goyal (2008) and include 10 equity predictors for conditioning variables, including treasury bill return, inflation, yields for TERM and default factors, and various aggregate equity market characteristics (earnings-to-price ratio, stock variance, net equity issues, dividend yields, leverage, and liquidity). We use these 10 conditioning variables and create 110 instrumental variables from the one-month lagged predictors, quadratic terms, and interactions. We also create two-month and three-month lagged versions and produced 330 instrumental variables.

4.2 Asset Pricing Performance

Hansen-Jagannathan Measure. The literature has adopted multiple performance criteria to evaluate asset pricing models. To evaluate the goodness of fit of an SDF model, the HJ distance (Hansen and Jagannathan, 1997) is widely used for GMM estimation.Recent literature adopts multiple R^2 measures for model performance to evaluate a linear beta pricing model. For example, Feng et al. (2021) consider total R^2 and cross-sectional R^2 to evaluate the time-series and cross-sectional goodness of fit. We follow these criteria and create an HJ-R^2, which assesses the moment condition fitness and reflects an SDF model's relative asset pricing ability over a benchmark one. Given a model M_A with $m_t = 1 - \gamma_t' f_t$ and $E(m_t r_t) = 0_{N\times 1}$, the $N \times 1$ vector of cross-sectional pricing errors is

$$e(M_A) = \frac{1}{T}\sum_{t=1}^{T} m_t r_t.$$

Then, the aggregated cross-sectional pricing errors with equal weight to each asset follow

$$Q(M_A) = e(M_A)' e(M_A).$$

We can do the same calculation for the constant parameter SDF model. Then, we define the HJ-R^2 as

$$\text{HJ-}R^2 = 1 - \frac{Q(M_A)}{Q(M_B)}. \quad (12)$$

A higher HJ-R^2 value indicates better asset pricing goodness of fit, and a positive HJ-R^2 value indicates that model M_A outperforms the benchmark model M_B. In the empirical analysis, we use the standard Capital Asset Pricing Model (CAPM) as M_B, which assumes that the market factor is used to approximate the SDF. We estimate the corresponding constant parameter SDF model in our training sample for this calculation.

Table 6 shows the improvement of the goodness of fit by allowing time-varying factor risk price estimates over the constant parameter GMM model. We provide results for different test assets (bivariate- and univariate-sorted portfolios), different factor models (10 and 5 factors), and different numbers of instrumental variables (330 and 110). We also calculate the performance values for alternative time-varying methods, including local GMM, global GMM, and affine GMM. For in-sample goodness of fit, RegGMM generates higher HJ-R^2 values than all alternative methods. Allowing time-varying modeling (local, global, and affine GMM) does not always help because the constant parameter GMM is more robust than other methods. Alternative time-varying estimation methods might still suffer from model misspecification errors, but RegGMM performs better because it avoids specifying the DGP for time variation. Furthermore, we do not find that a larger number of instrumental variables has better goodness of fit. The reasons for this may be that the predictive signals are mainly from the one-month lagged conditioning variables in Welch and Goyal (2008) and that there is no additional information from two- and three-month lagged ones.

Time-Varying Risk Price. Figure 2 plots the year-by-year performance (annualized returns) of those 10 risk factors. Though all factors achieve highly positive gains in the overall sample, they show unsynchronized time-varying performance in different periods. For example, during the 2008 financial crisis, the market and Betting-against-Beta (BAB) factors dramatically dropped by 44%, but other fundamental factors such as RMW (Profitability) and QMJ (Quality) delivered 20% and 30% gains, respectively. Even for the famous Fama-French factors, one can also find the relatively weak performance of CMA in the 1980s, RMW in the 1990s, and HML in the 2000s, respectively. These unsynchronized performances provide strong evidence for the time-varying risk premia or prices of factors. We must understand the importance of their time-varying decomposition in forming the SDF.

The main advantage of RegGMM is that it provides the changing paths for time-varying parameters regardless of the sources or forms of time variation. Thanks to the macroeconomic conditioning variables used in RegGMM, our time-varying factor price paths can reflect market timing information about the business cycle, such as treasury bills, inflation, term spread, and default yield. We can illustrate how risk price estimates of multiple factors change smoothly over time [19], showing the individual factor time-varying importance of composing the SDF. We plot time-varying factor risk price estimates in Figure 3 and 4.

In addition to fixing the last values for the out-of-sample strategy, we have also provided a rolling window updating scheme to show the robustness. To maintain the smoothness for the time variation, we plug the previously estimated parameters into the fusion penalty and use the same penalization level. Therefore, for γ_t in a new period, its estimation only uses information up to this period. More importantly, the previous estimation before this period is never revised with the updated information. We show the rolling window update for γ_t in Figure 3 and 4 for the recent 10 years. Therefore, the out-of-sample γ_t forecasts are also time-varying and capture information updates. For example, the increasing weights for the market factor reflect the bull market in the recent decade.

Finally, the asymptotic variance and confidence intervals are another output of RegGMM, which helps identify the significance of factor time-varying importance. We construct the 95% confidence intervals to evaluate the usefulness and regime-switching of a factor for the time-varying SDF construction. Two examples are provided with graded 95% confidence intervals in Figures 3 and 4. In the Fama-French 5-factor model, the factor risk price for CMA (Investment) has

[19] Researchers usually find abrupt changes or jumps for portfolio or factor returns in the daily or higher frequency data, yet we study monthly data in this paper.

decreased since its slow decline in the 1980s and became insignificant after 1997. For the 10-factor SDF, we also find that Short-Term Reversal (STR) has become insignificant since 1996 due to its previous weak performance. These decreasing factor risk price estimation results are robust for both bivariate- and univariate-sorted portfolios.

4.3 Investment Performance

Risk-Adjusted Performance. Figure 5 plots in-sample and out-of-sample absolute investment gains. In theory, a well-fitted SDF of traded factors is supposed to be the tangency portfolio as a portfolio. We normalize the time-varying factor risk price estimates (SDF loadings) as portfolio weights and assess their investment performance over alternative methods. The trade-off for avoiding model misspecification risk faced by alternative methods (local, global, and affine GMM) is the lack of an explicit specification of driving factors for time-varying parameters. However, we could set the risk price estimate for the most recent estimate from the time-varying parameter path, which follows the out-of-sample implementation in the simulation study. Given the slow-moving path of time-varying factor risk price, the most recent estimates should be helpful in subsequent periods. For a fair comparison between RegGMM and other methods, we fix the portfolio weights for the entire out-of-sample period in the recent decade. Our RegGMM SDF strategies outperform most alternative methods in both in-sample and out-of-sample periods in Figure 5.

As the tangency portfolio, the SDF should provide the highest Sharpe ratio for risk-adjusted performance. To demonstrate its performance in fitting the time-varying risk price, we provide annualized Sharpe ratio numbers in Panel A of Table 7. Our RegGMM SDF time-varying strategies deliver the highest in-sample Sharpe ratios for monthly rebalanced portfolios, even higher than the mean-variance efficient portfolio. This strong evidence shows that time-varying SDF portfolio optimization benefits from additional investment information, possibly due to the macroeconomic conditioning variables used in RegGMM. By responding to the time-varying performance of factors in different periods, our RegGMM SDF portfolio weights show corresponding reactions in the SDF loadings. For factors that deliver robust performance in the training sample, such as QMJ, RegGMM provides robust estimates for their SDF loadings.

The positive in-sample results do not necessarily guarantee similar out-of-sample performance. Overfitting risks exist for regularization methods with the use of a high-dimensional set of macroeconomic conditioning variables. Therefore, an out-of-sample evaluation is necessary to demonstrate the performance robustness of the SDF model, and it can be conducted using the SDF strategy's investment performance. This is why we mainly consider traded factors for constructing the time-varying SDF in the empirical analysis. RegGMM allows for both traded and nontraded factors (like the consumption factor). For the out-of-sample performance in the recent decade, both the 5- and 10-factor RegGMM SDF strategies outperform the market factor with higher Sharpe ratio values. These out-of-sample results further demonstrate the advantages of allowing flexible time variation in the SDF model by RegGMM. For a robustness check, the rolling-window updated RegGMM SDF strategies deliver slightly higher performance than the static ones.

Evaluating downside risk measures associated with trading the SDF or the tangency portfolio is also worthwhile.

We follow Gu et al. (2020) and define the maximum drawdown for any overlapping one-year period as

$$MDD = max_{0\leq t_1\leq t_2\leq T}(Y_{t_1} - Y_{t_2}), \text{ s.t. } |t_2 - t_1| < 12, \tag{13}$$

where Y_{t_1} and Y_{t_2} refer to the cumulative log return from month 0 to t_1 and t_2, and the duration between t_1 and t_2 is no longer than 12 months. The one-year overlapping maximum drawdown reflects the investment tail risk better than the overall goodness of fit. Results are provided in Panel B of Table 7. This downside risk measure reveals the source of the high Sharpe ratio of our RegGMM SDF strategies: low maximum drawdown values. Although the market factor achieves the highest cumulative return in the most recent decade (Figure 5), its maximum drawdown value is the highest because the market factor differs from all the other long-short factors that hedge market exposure. Although the Fama-French 5-factor RegGMM SDF strategies produce lower Sharpe ratios than the 10-factor case, they provide lower maximum drawdown values.

Model-Adjusted Performance. As the tangency portfolio, our flexible time-varying SDF provides more investment information than existing asset pricing models, such as CAPM and the Fama-French 3-factor model. To demonstrate the advantage of estimating the time-varying risk price by RegGMM, we provide robust and highly positive results for unexplained monthly alphas and their significance in Table 8. Following the above implementation, we regress the in-sample and out-of-sample RegGMM SDF strategies over the benchmark models and report the unexplained intercepts. We also show comparison results for other methods.

We include the constant parameter GMM, the mean-variance efficient portfolio, and the equal-weighted portfolio, which are commonly used comparison benchmarks for investigating the investment performance of SDF. We show robust results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets, 5- and 10-factor models, and 110 and 330 instrumental variables. First, we find our 5- and 10-factor RegGMM SDF strategies deliver the highest in-sample alphas against CAPM and the Fama-French 3-factor model. All RegGMM SDF strategies have economically and statistically significant monthly alphas of more than 0.35% for a 40-year sample in Table 8. For a robustness check, the rolling-window updated RegGMM SDF strategies deliver higher and more significant investment performance than the static ones, demonstrating the advantages of the time-varying models.

Thanks to the use of macroeconomic conditioning variables with flexible time variation assumptions by RegGMM, the additional investment information is driven by the factor timing ability given by the dynamic SDF model. For example, in Figures 2 and 3, our time-varying risk price estimates respond to the decaying performance of HML and the rising performance of SMB in the 2000s. For the 10-factor RegGMM SDF strategies in Figure 4, we also find that the slowly decreasing RegGMM estimates capture the decreasing performances for MOM and STR. The changes in our time-varying SDF loadings help capture the time-varying risk premia for risk factors in various regime-switching scenarios.

Second, although the market factor dominates in the most recent decade in Figure 2, our 10-factor RegGMM strategies robustly provide significantly high alphas of about 0.55% against CAPM and the Fama-French 3-factor model in Table 8. By contrast, most investment strategies rarely show positive and significant results. We see that, though the

Fama-French 3 factors are included in these portfolios, any portfolio with nested factors does not easily outperform the benchmark. The constant GMM SDF strategies are helpful, but our RegGMM SDF strategies provide better results because of the use of a high-dimensional set of macroeconomic variables with flexible assumptions on time variation.

5 Conclusion

By exploiting a high-dimensional set of moment restrictions, we propose a ridge fusion-based regularized GMM (RegGMM) estimation method for flexible time-varying coefficient models. RegGMM consistently estimates time variation over the entire sample period without performing shrinkage, selection, or smoothed nonparametric estimation. Existing smoothed nonparametric estimation methods assume that time-varying parameters are functions of time or observable conditioning variables and require differentiability conditions on time-varying parameters that rule out abrupt structural breaks. They do not perform well near the endpoints of the sample period, due to the well-known boundary bias problem. They also suffer from the curse of dimensionality problem when the set of conditioning variables is large. To the best of our knowledge, our approach is among the first with the capability of consistently estimating flexible structural changes, which can be driven by time, observable or unobservable factors, or a mixture of them. Moreover, by introducing a ridge fusion regularization on the total amount of time variation in the whole sample, our method allows for abrupt, smooth, and dual-type structural changes. It has a lower computational cost than existing methods because it is a global one-step procedure and enjoys an appealing closed-form solution for linear models. We establish consistency and derive the convergence rate and asymptotic distribution of the proposed RegGMM estimator.

In applications to asset pricing, RegGMM offers an alternative solution for estimating time-varying stochastic discount factor models. For U.S. equities from the past five decades, our time-varying estimates for factor risk price respond to the changing performance of multiple risk factors and summarize potential factor regime-switching scenarios. We document improved asset pricing performance of the flexible time-varying SDF model over the constant parameter SDF model in terms of HJ measures. By outperforming multiple benchmark models, we demonstrate the gains for risk-adjusted and model-adjusted investment performance of the time-varying SDF model estimated by RegGMM for both in-sample and out-of-sample analysis. The positive out-of-sample investment gains demonstrate the advantages of RegGMM in capturing time variation, and our positive in-sample results are not due to overfitting. As Hansen (2001) points out, structural changes can happen in the short or long run in various forms. Therefore, avoiding restrictive specifications for time variation makes our proposed method free of misspecification errors and the curse of dimensionality issues.

Our approach offers a new direction for addressing some methodological constraints of existing smoothed nonparametric methods. For example, it is possible to introduce the sparse L_1-norm on time-varying parameters in our penalty function to study regime-switching of factor model selection. One could extend our method to address situations where time-varying parameters depend on cross-section units in fixed- and large-dimensional panel regression models. One could also employ our method to estimate large-dimensional dynamic covariance matrices by promoting similarities in covariance matrices over consecutive periods. All of these could be pursued in future work.

References

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Aït-Sahalia, Y. and D. Xiu (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics 201*(2), 384–399.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 821–856.

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica 67*(3), 543–563.

Antoine, B., K. Proulx, and E. Renault (2020). Pseudo-true sdfs in conditional asset pricing models. *Journal of Financial Econometrics 18*(4), 656–714.

Ao, M., L. Yingying, and X. Zheng (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies 32*(7), 2890–2919.

Brunnermeier, M. K. and S. Nagel (2008). Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals. *American Economic Review 98*(3), 713–36.

Chang, J., S. X. Chen, and X. Chen (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *Journal of Econometrics 185*(1), 283–304.

Chen, B. and Y. Hong (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica 80*(3), 1157–1183.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics 6*, 5549–5632.

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*(1), 277–321.

Chen, X. B., J. Gao, D. Li, and P. Silvapulle (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business & Economic Statistics 36*(1), 88–100.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy 104*(3), 572–621.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Cogley, T. and T. J. Sargent (2001). Evolving post-world war ii us inflation dynamics. *NBER Macroeconomics Annual 16*, 331–373.

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2018). Generalized autoregressive method of moments.

Cui, L., Y. Hong, and Y. Li (2021). Solving euler equations via two-stage nonparametric penalized splines. *Journal of Econometrics 222*(2), 1024–1056.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Dong, C., J. Gao, and O. Linton (2021). High dimensional semiparametric moment restriction models. *Journal of Econometrics*.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American statistical Association 96*(456), 1348–1360.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance 75*(3), 1327–1370.

Feng, G., N. Polson, and J. Xu (2021). Deep learning in characteristics-sorted factor models. Technical report, City University of Hong Kong.

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics 10*, 109–138.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy 99*(2), 385–415.

Fu, Z., Y. Hong, L. Su, and X. Wang (2022). Specification tests for time-varying coefficient models. *Journal of Econometrics*.

Gagliardini, P., C. Gourieroux, and E. Renault (2011). Efficient derivative pricing by the extended method of moments. *Econometrica 79*(4), 1181–1232.

Gagliardini, P. and D. Ronchetti (2020). Comparing asset pricing models by the conditional hansen-jagannathan distance. *Journal of Financial Econometrics 18*(2), 333–394.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance 53*(2), 549–573.

Gospodinov, N. and T. Otsu (2012). Local gmm estimation of time series models with conditional moment restrictions. *Journal of Econometrics 170*(2), 476–490.

Götze, F. and H. R. Künsch (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics 24*(5), 1914–1933.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Hall, P. and J. L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 891–916.

Han, C. and P. C. B. Phillips (2006). Gmm with many moment conditions. *Econometrica 74*(1), 147–192.

Hansen, B. E. (2001, December). The new econometrics of structural change: Dating breaks in u.s. labour productivity. *Journal of Economic Perspectives 15*(4), 117–128.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.

Hansen, L. P. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance 52*(2), 557–590.

Hansen, L. P. and S. F. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 587–613.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica 50*(5), 1269–1286.

Hong, Y. and Y.-J. Lee (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *Review of Economic Studies 72*(2), 499–541.

Hong, Y., X. Wang, and S. Wang (2017). Testing strict stationarity with applications to macroeconomic time series. *International Economic Review 58*(4), 1227–1277.

Horenko, I. (2010). Finite element approach to clustering of multidimensional time series. *SIAM Journal on Scientific Computing 32*(1), 62–83.

Huang, J., J. L. Horowitz, and S. Ma (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics 36*(2), 587–613.

Inoue, A. (2006). A bootstrap approach to moment selection. *Econometrics Journal 9*(1), 48–75.

Inoue, A. and M. Shintani (2006). Bootstrapping gmm estimators for time series. *Journal of Econometrics 133*(2), 531–555.

Islak, U. (2013). Asymptotic normality of random sums of m-dependent random variables.

Kan, R. and G. Zhou (1999). A critique of the stochastic discount factor methodology. *Journal of Finance 54*(4), 1221–1248.

Kato, K. (2011). A note on moment convergence of bootstrap m-estimators. *Statistics & Decisions 28*(1), 51–61.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Lei, J. (2020). Cross-validation with confidence. *Journal of the American Statistical Association 115*(532), 1978–1997.

Lettau, M. and M. Pelger (2020). Estimating latent asset-pricing factors. *Journal of Econometrics 218*(1), 1–31.

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters 94*(1), 124–128.

Li, Y. and D. Ruppert (2008). On the asymptotics of penalized splines. *Biometrika 95*(2), 415–436.

Li, Y. and L. Yang (2011). Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance 18*(5), 972–992.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy 1*, 19–46.

Ma, S., O. Linton, and J. Gao (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics 222*(1), 295–323.

Marchenko, G., P. Gagliardini, and I. Horenko (2018). Towards a computationally tractable maximum entropy principle for nonstationary financial time series. *SIAM Journal on Financial Mathematics 9*(4), 1249–1285.

Nagel, S. and K. J. Singleton (2011). Estimation and evaluation of conditional asset pricing models. *Journal of Finance 66*(3), 873–909.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 1161–1167.

Newey, W. K. and R. J. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica 72*(1), 219–255.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Ng, S. and P. Perron (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*(429), 268–281.

Orbe, S., E. Ferreira, and J. Rodríguez-póo (2000). A nonparametric method to estimate time varying coefficients under seasonal constraints. *Journal of Nonparametric Statistics 12*(6), 779–806.

Pospíšil, L., P. Gagliardini, W. Sawyer, and I. Horenko (2018). On a scalable nonparametric denoising of time series signals. *Communications in Applied Mathematics and Computational Science 13*(1), 107–138.

Price, B. S., C. J. Geyer, and A. J. Rothman (2015). Ridge fusion in statistical learning. *Journal of Computational and Graphical Statistics 24*(2), 439–454.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

Rao, B. P. and M. Sreehari (2016). On the order of approximation in the random central limit theorem for m-dependent random variables. *Probability and Mathematical Statistics 36*(1), 47–57.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253–264. Springer.

Romano, J. P. and M. Wolf (2000). A more general central limit theorem for m-dependent random variables with unbounded m. *Statistics & Probability Letters 47*(2), 115–124.

Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section of stock returns. *Journal of Financial Economics 111*(2), 352–380.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(1), 91–108.

Wahba, G., Y. Wang, C. Gu, R. Klein, and B. Klein (1995). Smoothing spline anova for exponential families, with application to the wisconsin epidemiological study of diabetic retinopathy. *Annals of Statistics 23*(6), 1865–1895.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.

Table 1: Simulation in-sample Performance: Time Variation as a Function of Time

(ρ_1, ρ_2)	Criteria	RegGMM	CGMM	LGMM	GGMM	AGMM	RegGMM	CGMM	LGMM	GGMM	AGMM
				q=13300					q=33100		
						T=120					
	Avg.Abs.Err.	**0.02**	1.25	0.19	0.18	1.25	**0.02**	1.25	0.19	0.18	1.24
$(0, 0.5)$	Max.Abs.Err.	**0.09**	3.11	0.75	0.70	3.14	**0.05**	3.12	0.74	0.69	3.17
	Avg.Mom.Err.	**0.01**	2.61	0.07	0.06	2.61	**0.01**	2.61	0.07	0.06	2.61
	Avg.Abs.Err.	**0.02**	1.26	0.19	0.18	1.26	**0.02**	1.26	0.19	0.18	1.26
$(0, 0.95)$	Max.Abs.Err.	**0.09**	3.15	0.75	0.69	3.18	**0.04**	3.14	0.75	0.68	3.19
	Avg.Mom.Err.	**0.01**	2.61	0.07	0.06	2.61	**0.01**	2.61	0.07	0.06	2.61
	Avg.Abs.Err.	**0.04**	1.25	0.19	0.18	1.25	**0.02**	1.25	0.19	0.15	1.24
$(0.95, 0.5)$	Max.Abs.Err.	**0.19**	3.11	0.75	0.71	3.13	**0.09**	3.12	0.75	0.55	3.17
	Avg.Mom.Err.	**0.02**	2.62	0.08	0.07	2.61	**0.02**	2.62	0.08	0.05	2.61
	Avg.Abs.Err.	**0.04**	1.26	0.19	0.18	1.26	**0.02**	1.26	0.19	0.18	1.26
$(0.95, 0.95)$	Max.Abs.Err.	**0.18**	3.15	0.75	0.70	3.18	**0.09**	3.14	0.75	0.69	3.18
	Avg.Mom.Err.	**0.02**	2.62	0.08	0.07	2.62	**0.02**	2.62	0.08	0.07	2.62
						T=360					
	Avg.Abs.Err.	**0.03**	1.81	0.38	0.63	1.48	**0.02**	1.81	0.39	0.51	1.47
$(0, 0.5)$	Max.Abs.Err.	**0.12**	5.49	2.93	3.26	4.16	**0.07**	5.46	2.93	3.07	4.19
	Avg.Mom.Err.	**0.01**	4.86	0.38	0.63	3.58	**0.01**	4.86	0.39	0.50	3.58
	Avg.Abs.Err.	**0.04**	1.81	0.38	0.52	1.48	**0.02**	1.82	0.39	0.52	1.47
$(0, 0.95)$	Max.Abs.Err.	**0.13**	5.50	2.92	3.05	4.18	**0.08**	5.45	2.93	3.06	4.21
	Avg.Mom.Err.	**0.01**	4.86	0.38	0.51	3.58	**0.01**	4.86	0.39	0.50	3.58
	Avg.Abs.Err.	**0.06**	1.81	0.38	0.63	1.48	**0.03**	1.81	0.39	0.51	1.47
$(0.95, 0.5)$	Max.Abs.Err.	**0.23**	5.50	2.92	3.26	4.15	**0.13**	5.46	2.93	3.06	4.19
	Avg.Mom.Err.	**0.02**	4.87	0.39	0.64	3.60	**0.02**	4.87	0.40	0.51	3.59
	Avg.Abs.Err.	**0.06**	1.81	0.38	0.52	1.48	**0.03**	1.81	0.39	0.52	1.47
$(0.95, 0.95)$	Max.Abs.Err.	**0.21**	5.51	2.92	3.04	4.17	**0.13**	5.46	2.93	3.05	4.20
	Avg.Mom.Err.	**0.02**	4.88	0.39	0.51	3.59	**0.02**	4.87	0.40	0.51	3.59

Notes: For DGP 1 in Section 3, the simulation study is designed to compare the in-sample finite sample performance for different parameters: length of periods T (120 and 360), autocorrelation levels ρ_1 and ρ_2, and number of instrumental variables q (13300 and 33100). We compare RegGMM with the constant parameter GMM (CGMM), local constant GMM (LGMM), global series GMM (GGMM), and affine GMM (AGMM). We report three different norms of fitting errors: average estimation error, maximum estimation error, and average moment error, respectively.

Table 2: Simulation OOS Performance: Time Variation as a Function of Time

(ρ_1, ρ_2)	Criteria	RegGMM	CGMM	LGMM	GGMM	AGMM	RegGMM	CGMM	LGMM	GGMM	AGMM
				q=13300					q=33100		
						T=120					
	Avg.Abs.Err.	**0.71**	3.26	1.87	1.31	2.37	**0.69**	3.25	1.88	1.28	2.35
$(0, 0.5)$	Max.Abs.Err.	**2.28**	6.32	4.81	3.30	5.41	**2.32**	6.35	4.91	3.25	5.35
	Avg.Mom.Err.	**0.68**	13.19	4.62	2.40	6.93	**0.65**	13.13	4.68	2.31	6.82
	Avg.Abs.Err.	**0.71**	3.24	1.88	1.34	2.35	**0.69**	3.23	1.88	1.36	2.34
$(0, 0.95)$	Max.Abs.Err.	**2.30**	6.33	4.77	3.24	5.39	**2.34**	6.34	4.82	3.30	5.39
	Avg.Mom.Err.	**0.67**	13.06	4.63	2.43	6.89	**0.65**	13.04	4.64	2.65	6.80
	Avg.Abs.Err.	**0.69**	3.26	1.87	1.32	2.37	**0.70**	3.25	1.88	1.32	2.35
$(0.95, 0.5)$	Max.Abs.Err.	**2.23**	6.33	4.81	3.28	5.41	**2.30**	6.35	4.91	3.25	5.35
	Avg.Mom.Err.	**0.65**	13.21	4.62	2.41	6.92	**0.67**	13.14	4.68	2.44	6.82
	Avg.Abs.Err.	**0.69**	3.24	1.88	1.35	2.35	**0.70**	3.24	1.88	1.34	2.34
$(0.95, 0.95)$	Max.Abs.Err.	**2.26**	6.34	4.77	3.78	5.39	**2.32**	6.34	4.82	3.29	5.39
	Avg.Mom.Err.	**0.65**	13.08	4.63	2.54	6.88	**0.67**	13.06	4.64	2.58	6.80
						T=360					
	Avg.Abs.Err.	**0.35**	3.68	2.26	2.44	2.33	**0.35**	3.63	2.28	2.17	2.30
$(0, 0.5)$	Max.Abs.Err.	**0.63**	6.90	4.98	5.84	4.53	**0.61**	6.81	4.95	5.06	4.59
	Avg.Mom.Err.	**0.15**	18.10	6.50	7.85	7.09	**0.15**	17.66	6.61	6.45	6.95
	Avg.Abs.Err.	**0.35**	3.68	2.27	2.33	2.31	**0.35**	3.63	2.29	2.17	2.29
$(0, 0.95)$	Max.Abs.Err.	**0.64**	6.89	5.01	4.92	4.63	**0.64**	6.81	4.98	5.32	4.63
	Avg.Mom.Err.	**0.15**	18.07	6.54	6.91	6.93	**0.15**	17.64	6.64	6.92	6.91
	Avg.Abs.Err.	**0.35**	3.69	2.26	2.29	2.33	**0.35**	3.64	2.28	2.18	2.30
$(0.95, 0.5)$	Max.Abs.Err.	**0.67**	6.90	4.98	5.33	4.51	**0.63**	6.81	4.95	5.07	4.58
	Avg.Mom.Err.	**0.16**	18.16	6.49	6.61	7.11	**0.15**	17.69	6.61	6.55	6.96
	Avg.Abs.Err.	**0.36**	3.69	2.27	2.19	2.31	**0.36**	3.63	2.29	2.28	2.29
$(0.95, 0.95)$	Max.Abs.Err.	**0.66**	6.89	5.01	5.41	4.62	**0.65**	6.82	4.98	5.34	4.63
	Avg.Mom.Err.	**0.16**	18.13	6.54	6.70	6.95	**0.15**	17.67	6.64	7.29	6.92

Notes: For DGP 1 in Section 3, the simulation study is designed to compare the out-of-sample finite sample performance for different parameters: length of periods T (120 and 360), autocorrelation levels ρ_1 and ρ_2, and number of instrumental variables q (13300 and 33100). We compare RegGMM with the constant parameter GMM (CGMM), local constant GMM (LGMM), global series GMM (GGMM), and affine GMM (AGMM). We report three different norms of fitting errors based on 60 time periods in excess of T: average estimation error, maximum estimation error, and average moment error, respectively.

Table 3: Simulation in-sample Performance: Time Variation as a Function of Variables

(ρ_1, ρ_2)	Criteria	RegGMM	CGMM	LGMM	GGMM	AGMM	RegGMM	CGMM	LGMM	GGMM	AGMM
				q=13300					q=33100		
						T=120					
	Avg.Abs.Err.	**0.03**	0.71	0.62	0.71	0.71	**0.02**	0.71	0.62	0.64	0.71
$(0, 0.5)$	Max.Abs.Err.	**0.10**	2.45	2.16	2.45	2.42	**0.05**	2.45	2.16	2.15	2.41
	Avg.Mom.Err.	**0.01**	0.74	0.61	0.74	0.74	**0.01**	0.74	0.61	0.65	0.74
	Avg.Abs.Err.	**0.03**	0.97	0.66	0.68	0.96	**0.02**	0.97	0.67	0.65	0.96
$(0, 0.95)$	Max.Abs.Err.	**0.10**	2.57	2.55	2.50	2.64	**0.04**	2.59	2.52	2.46	2.61
	Avg.Mom.Err.	**0.01**	1.40	0.71	0.75	1.39	**0.01**	1.40	0.72	0.71	1.38
	Avg.Abs.Err.	**0.05**	0.71	0.62	0.71	0.71	**0.03**	0.71	0.62	0.64	0.71
$(0.95, 0.5)$	Max.Abs.Err.	**0.21**	2.45	2.16	2.45	2.42	**0.10**	2.45	2.15	2.15	2.42
	Avg.Mom.Err.	**0.02**	0.75	0.62	0.75	0.75	**0.02**	0.75	0.62	0.66	0.75
	Avg.Abs.Err.	**0.05**	0.97	0.66	0.68	0.96	**0.03**	0.97	0.67	0.65	0.96
$(0.95, 0.95)$	Max.Abs.Err.	**0.20**	2.58	2.55	2.51	2.65	**0.10**	2.59	2.52	2.46	2.62
	Avg.Mom.Err.	**0.02**	1.41	0.72	0.75	1.40	**0.02**	1.41	0.72	0.72	1.39
						T=360					
	Avg.Abs.Err.	**0.03**	0.82	0.72	0.82	0.79	**0.02**	0.82	0.72	0.79	0.79
$(0, 0.5)$	Max.Abs.Err.	**0.12**	2.67	2.66	2.67	2.76	**0.07**	2.66	2.68	2.75	2.75
	Avg.Mom.Err.	**0.01**	0.91	0.75	0.91	0.86	**0.01**	0.91	0.76	0.86	0.86
	Avg.Abs.Err.	**0.04**	0.90	0.73	0.90	0.86	**0.02**	0.90	0.73	0.84	0.86
$(0, 0.95)$	Max.Abs.Err.	**0.14**	2.65	2.34	2.65	2.54	**0.09**	2.65	2.30	2.59	2.55
	Avg.Mom.Err.	**0.01**	1.13	0.78	1.13	1.07	**0.01**	1.13	0.79	1.01	1.06
	Avg.Abs.Err.	**0.06**	0.82	0.72	0.82	0.79	**0.04**	0.82	0.72	0.79	0.79
$(0.95, 0.5)$	Max.Abs.Err.	**0.24**	2.67	2.66	2.67	2.76	**0.14**	2.67	2.68	2.75	2.75
	Avg.Mom.Err.	**0.02**	0.92	0.76	0.92	0.86	**0.02**	0.92	0.77	0.86	0.86
	Avg.Abs.Err.	**0.06**	0.90	0.73	0.90	0.86	**0.04**	0.90	0.73	0.84	0.86
$(0.95, 0.95)$	Max.Abs.Err.	**0.21**	2.66	2.34	2.66	2.55	**0.14**	2.65	2.30	2.60	2.55
	Avg.Mom.Err.	**0.02**	1.14	0.79	1.14	1.08	**0.02**	1.14	0.80	1.02	1.07

Notes: For DGP 2 in Section 3, the simulation study is designed to compare the in-sample finite sample performance for different parameters: length of periods T (120 and 360), autocorrelation levels ρ_1 and ρ_2, and number of instrumental variables q (13300 and 33100). We compare RegGMM with the constant parameter GMM (CGMM), local constant GMM (LGMM), global series GMM (GGMM), and affine GMM (AGMM). We report three different norms of fitting errors: average estimation error, maximum estimation error, and average moment error, respectively.

Table 4: Simulation OOS Performance: Time Variation as a Function of Variables

(ρ_1, ρ_2)	Criteria	RegGMM	CGMM	LGMM	GGMM	AGMM	RegGMM	CGMM	LGMM	GGMM	AGMM
				q=13300					**q=33100**		
						T=120					
(0, 0.5)	Avg.Abs.Err.	0.82	0.94	**0.73**	0.89	0.85	0.82	0.94	**0.73**	0.85	0.83
	Max.Abs.Err.	2.64	**2.38**	2.75	**2.38**	2.45	2.62	**2.38**	2.70	2.82	2.44
	Avg.Mom.Err.	1.09	1.07	**0.84**	1.00	0.93	1.10	1.07	**0.83**	1.12	0.91
(0, 0.95)	Avg.Abs.Err.	**0.51**	0.96	0.64	0.65	0.75	**0.51**	0.96	0.64	0.56	0.78
	Max.Abs.Err.	1.91	1.79	**1.63**	2.33	2.11	1.91	1.82	**1.68**	2.17	2.15
	Avg.Mom.Err.	**0.43**	1.09	0.56	0.64	0.78	**0.43**	1.10	0.56	0.51	0.84
(0.95, 0.5)	Avg.Abs.Err.	0.82	0.94	**0.73**	0.89	0.85	0.82	0.94	**0.73**	0.86	0.83
	Max.Abs.Err.	2.67	**2.38**	2.75	**2.38**	2.45	2.63	**2.38**	2.70	2.83	2.44
	Avg.Mom.Err.	1.10	1.08	**0.84**	1.00	0.93	1.10	1.07	**0.83**	1.14	0.91
(0.95, 0.95)	Avg.Abs.Err.	**0.51**	0.96	0.64	0.67	0.74	**0.51**	0.96	0.64	0.56	0.78
	Max.Abs.Err.	1.91	1.80	**1.63**	2.31	2.11	1.91	1.82	**1.68**	2.17	2.15
	Avg.Mom.Err.	**0.43**	1.09	0.56	0.69	0.77	**0.43**	1.10	0.56	0.52	0.84
						T=360					
(0, 0.5)	Avg.Abs.Err.	**0.63**	0.94	0.72	0.94	0.80	**0.62**	0.94	0.72	0.80	0.77
	Max.Abs.Err.	2.31	2.20	2.16	2.20	**2.12**	2.35	2.20	2.17	2.20	**2.06**
	Avg.Mom.Err.	**0.64**	1.05	0.79	1.04	0.87	**0.64**	1.04	0.79	0.89	0.83
(0, 0.95)	Avg.Abs.Err.	**0.70**	1.03	0.84	0.99	0.90	**0.70**	1.03	0.85	0.93	0.86
	Max.Abs.Err.	1.96	2.08	2.21	2.08	**1.93**	1.96	2.07	2.20	2.41	**1.94**
	Avg.Mom.Err.	**0.74**	1.29	1.01	1.24	1.03	**0.74**	1.29	1.02	1.18	0.98
(0.95, 0.5)	Avg.Abs.Err.	**0.63**	0.94	0.72	0.94	0.80	**0.63**	0.94	0.72	0.80	0.77
	Max.Abs.Err.	2.28	2.21	2.16	2.21	**2.13**	2.34	2.20	2.17	2.20	**2.06**
	Avg.Mom.Err.	**0.65**	1.05	0.79	1.04	0.87	**0.64**	1.04	0.79	0.89	0.83
(0.95, 0.95)	Avg.Abs.Err.	**0.70**	1.03	0.84	0.99	0.89	**0.70**	1.03	0.85	0.93	0.86
	Max.Abs.Err.	1.98	2.09	2.21	2.09	**1.92**	1.98	2.08	2.20	2.42	**1.93**
	Avg.Mom.Err.	**0.75**	1.30	1.01	1.25	1.02	**0.75**	1.29	1.02	1.18	0.98

Notes: For DGP 2 in Section 3, the simulation study is designed to compare the out-of-sample finite sample performance for different parameters: length of periods T (120 and 360), autocorrelation levels ρ_1 and ρ_2, and number of instrumental variables q (13300 and 33100). We compare RegGMM with constant parameter GMM (CGMM), local constant GMM (LGMM), global series GMM (GGMM), and affine GMM (AGMM). We report three different norms of fitting errors based on 60 time periods in excess of T: average estimation error, maximum estimation error, and average moment error, respectively.

Table 5: Simulation Robustness on Tuning Parameters

		q=13300			q=33100		
Periods	(ρ_1, ρ_2)	Tuning	CV.Err.	p-value	Tuning	CV.Err.	p-value
				DGP1			
T=120	(0, 0.5)	$\lambda^* = e^{-12}$	1.425E-5	1	$\lambda^* = e^{-12.5}$	1.609E-5	0.99
		$\lambda = 0$	2.697E1	0	$\lambda = 0$	1.611E-5	0
		$\lambda = 1$	3.921E-3	0	$\lambda = 1$	4.100E-3	0
	(0, 0.95)	$\lambda^* = e^{-12}$	1.444E-5	1	$\lambda^* = e^{-12.5}$	1.617E-5	1
		$\lambda = 0$	3.911E2	0	$\lambda = 0$	1.619E-5	0
		$\lambda = 1$	4.074E-3	0	$\lambda = 1$	4.188E-3	0
	(0.95, 0.5)	$\lambda^* = e^{-12.5}$	3.279E-5	1	$\lambda^* = e^{-12}$	3.417E-5	0.95
		$\lambda = 0$	2.702E1	0	$\lambda = 0$	3.423E-5	0
		$\lambda = 1$	3.919E-3	0	$\lambda = 1$	4.106E-3	0
	(0.95, 0.95)	$\lambda^* = e^{-12.5}$	3.211E-5	0.98	$\lambda^* = e^{-12}$	3.350E-5	0.74
		$\lambda = 0$	3.919E2	0	$\lambda = 0$	3.356E-5	0
		$\lambda = 1$	4.073E-3	0	$\lambda = 1$	4.195E-3	0
T=360	(0, 0.5)	$\lambda^* = e^{-16}$	4.640E-6	1	$\lambda^* = e^{-16}$	5.359E-6	1
		$\lambda = 0$	7.596E1	0	$\lambda = 0$	4.151E5	0
		$\lambda = 1$	2.469E-3	0	$\lambda = 1$	2.511E-3	0
	(0, 0.95)	$\lambda^* = e^{-16}$	4.597E-6	1	$\lambda^* = e^{-16}$	5.285E-6	1
		$\lambda = 0$	4.079E3	0	$\lambda = 0$	2.183E1	0
		$\lambda = 1$	2.533E-3	0	$\lambda = 1$	2.568E-3	0
	(0.95, 0.5)	$\lambda^* = e^{-15.5}$	1.008E-5	1	$\lambda^* = e^{-15.5}$	1.105E-5	1
		$\lambda = 0$	7.597E1	0	$\lambda = 0$	4.152E5	0
		$\lambda = 1$	2.476E-3	0	$\lambda = 1$	2.517E-3	0
	(0.95, 0.95)	$\lambda^* = e^{-15.5}$	9.943E-6	1	$\lambda^* = e^{-15.5}$	1.096E-5	1
		$\lambda = 0$	4.079E3	0	$\lambda = 0$	2.184E1	0
		$\lambda = 1$	2.543E-3	0	$\lambda = 1$	2.575E-3	0
				DGP2			
T=120	(0, 0.5)	$\lambda^* = e^{-17.5}$	1.433E-5	0.68	$\lambda^* = e^{-19}$	1.611E-5	0.56
		$\lambda = 0$	3.525E1	0	$\lambda = 0$	1.611E-5	0.36
		$\lambda = 1$	1.239E-3	0	$\lambda = 1$	1.235E-3	0
	(0, 0.95)	$\lambda^* = e^{-17}$	1.452E-5	1	$\lambda^* = e^{-18.5}$	1.619E-5	0.63
		$\lambda = 0$	5.738E2	0	$\lambda = 0$	1.619E-5	0.30
		$\lambda = 1$	2.395E-3	0	$\lambda = 1$	2.349E-3	0
	(0.95, 0.5)	$\lambda^* = e^{-17}$	3.294E-5	0.90	$\lambda^* = e^{-18.5}$	3.423E-5	0.52
		$\lambda = 0$	3.530E1	0	$\lambda = 0$	3.423E-5	0.37
		$\lambda = 1$	1.259E-3	0	$\lambda = 1$	1.255E-3	0
	(0.95, 0.95)	$\lambda^* = e^{-16.5}$	3.225E-5	1	$\lambda^* = e^{-18}$	3.356E-5	0.61
		$\lambda = 0$	5.747E2	0	$\lambda = 0$	3.356E-5	0.29
		$\lambda = 1$	2.416E-3	0	$\lambda = 1$	2.365E-3	0
T=360	(0, 0.5)	$\lambda^* = e^{-20}$	4.649E-6	1	$\lambda^* = e^{-19.5}$	5.366E-6	1
		$\lambda = 0$	7.780E1	0	$\lambda = 0$	4.256E5	0
		$\lambda = 1$	5.568E-4	0	$\lambda = 1$	5.252E-4	0
	(0, 0.95)	$\lambda^* = e^{-19.5}$	4.606E-6	1	$\lambda^* = e^{-19}$	5.293E-6	0.65
		$\lambda = 0$	3.868E3	0	$\lambda = 0$	2.072E1	0
		$\lambda = 1$	6.535E-4	0	$\lambda = 1$	6.394E-4	0
	(0.95, 0.5)	$\lambda^* = e^{-19}$	1.012E-5	1	$\lambda^* = e^{-19}$	1.108E-5	1
		$\lambda = 0$	7.781E1	0	$\lambda = 0$	4.258E5	0
		$\lambda = 1$	5.633E-4	0	$\lambda = 1$	5.294E-4	0
	(0.95, 0.95)	$\lambda^* = e^{-18.5}$	9.974E-6	1	$\lambda^* = e^{-18.5}$	1.098E-5	1
		$\lambda = 0$	3.869E3	0	$\lambda = 0$	2.072E1	0
		$\lambda = 1$	6.558E-4	0	$\lambda = 1$	6.438E-4	0

Notes: For a given tuning parameter value, this table reports the cross-validation error and the p-value under the null hypothesis that the tuning parameter is optimal, for different parameters: length of periods T (120 and 360), autocorrelation levels ρ_1 and ρ_2, and number of instrumental variables q (13300 and 33100). The calculation of the p-value is given in Appendix C.1. λ^* is based on the 5-fold repeated CV criterion.

Table 6: Asset Pricing Performance: Hansen-Jagannathan R^2

	Bi-Sort Portfolios	Uni-Sort Portfolios	Bi-Sort Portfolios	Uni-Sort Portfolios
	10 Factors		Five Factors	
Constant GMM	76.5%	72.4%	59.8%	63.0%
		330 Instrumental Variables		
RegGMM	89.4%	88.7%	74.5%	77.9%
Local GMM	73.0%	83.5%	61.0%	71.2%
Global GMM	52.3%	50.6%	23.9%	25.7%
Affine GMM	82.1%	78.9%	59.4%	66.6%
		110 Instrumental Variables		
RegGMM	91.0%	89.4%	78.4%	79.5%
Local GMM	75.0%	83.8%	64.1%	72.1%
Global GMM	55.5%	58.9%	34.1%	39.0%
Affine GMM	84.6%	80.2%	60.2%	68.5%

Notes: This table reports the asset pricing performance of RegGMM for the monthly training sample from 1972 to 2011. As described in section 4.1, we provide results for different test assets (bivariate and univariate sorted portfolios), different factor models (10 and 5 factors), and different numbers of instrumental variables (330 and 110). Other comparison methods are introduced in section 3. The formula for HJ R^2 is in equation (12).

Table 7: Risk-Adjusted Investment Performance

	In-Sample (1972-2011)		Out-of-Sample (2012-2021)	
	Panel A: Annualized Sharpe Ratio			
	Bi-Sort Port	Uni-Sort Port	Bi-Sort Port	Uni-Sort Port
MKTRF	0.33	0.33	1.16	1.16
CGMM_F5	0.96	1.06	0.73	0.83
RegGMM_F5_110	1.50	1.43	1.04	0.94
RegGMM_F5_330	1.50	1.38	0.96	0.76
MVE_F5	1.18	1.18	0.77	0.77
EW_F5	0.98	0.98	0.56	0.56
RegGMM_RW_F5_110	-	-	0.97	0.94
RegGMM_RW_F5_330	-	-	0.93	0.75
CGMM_F10	1.45	1.59	0.75	0.72
RegGMM_F10_110	1.84	1.85	1.23	1.00
RegGMM_F10_330	1.96	1.83	1.26	0.98
MVE_F10	1.84	1.84	1.23	1.23
EW_F10	1.46	1.46	0.96	0.96
RegGMM_RW_F10_110	-	-	1.19	0.97
RegGMM_RW_F10_330	-	-	1.27	0.96
	Panel B: One-Year Maximum Drawdown (%)			
	Bi-Sort Port	Uni-Sort Port	Bi-Sort Port	Uni-Sort Port
MKTRF	46.34	46.34	20.51	20.51
CGMM_F5	8.43	9.78	5.41	3.90
RegGMM_F5_110	5.21	6.23	5.39	7.49
RegGMM_F5_330	5.64	7.25	4.37	8.98
MVE_F5	7.71	7.71	6.61	6.61
EW_F5	12.49	12.49	12.70	12.70
RegGMM_RW_F5_110	-	-	7.18	6.96
RegGMM_RW_F5_330	-	-	4.53	8.70
CGMM_F10	8.99	9.15	10.17	8.25
RegGMM_F10_110	9.22	9.68	8.52	16.95
RegGMM_F10_330	7.68	8.98	7.66	17.48
MVE_F10	7.74	7.74	6.72	6.72
EW_F10	10.72	10.72	8.99	8.99
RegGMM_RW_F10_110	-	-	8.66	17.32
RegGMM_RW_F10_330	-	-	7.79	18.08

Notes: This table reports the risk-adjusted investment performance (annualized Sharpe ratio and one-year maximum drawdown) for factor investing using the SDF model. We estimate the SDF model from 1972 to 2011 and perform an out-of-sample analysis from 2012 to 2021. In addition to RegGMM SDF strategies, we include the constant parameter GMM (CGMM), the market factor, the mean-variance efficient portfolio (MVE), and the equal-weighted portfolio (EW), for comparison. We also report results for the out-of-sample rolling window (RW) updated RegGMM. As described in section 4.1, we report results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets, 5- and 10-factor SDF models, as well as 110 and 330 instrumental variables.

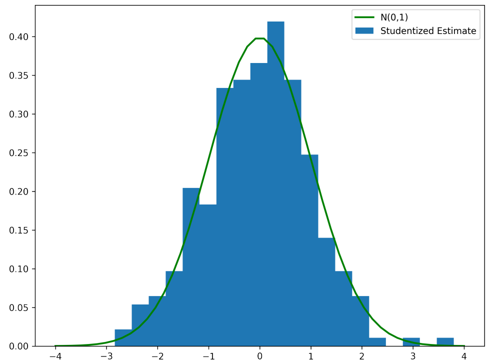
Table 8: Model-Adjusted Investment Performance

	In-Sample (1972-2011)		Out-of-Sample (2012-2021)	
	Panel A: Jensen's Alpha (%)			
	Bi-Sort Port	Uni-Sort Port	Bi-Sort Port	Uni-Sort Port
CGMM_F5	0.33***	0.32***	0.21	0.14
RegGMM_F5_110	0.38***	0.42***	0.26*	0.14
RegGMM_F5_330	0.39***	0.42***	0.14	0.04
MVE_F5	0.33***	0.33***	0.05	0.05
EW_F5	0.29***	0.29***	−0.10	−0.10
RegGMM_RW_F5_110	-	-	0.38**	0.14
RegGMM_RW_F5_330	-	-	0.21*	0.03
CGMM_F10	0.38***	0.36***	0.25*	0.13
RegGMM_F10_110	0.54***	0.57***	0.61***	0.32*
RegGMM_F10_330	0.53***	0.57***	0.57***	0.30
MVE_F10	0.43***	0.43***	0.26***	0.26***
EW_F10	0.44***	0.44***	0.09	0.09
RegGMM_RW_F10_110	-	-	0.69**	0.32*
RegGMM_RW_F10_330	-	-	0.60***	0.28
	Panel B: FF3 Alpha (%)			
	Bi-Sort Port	Uni-Sort Port	Bi-Sort Port	Uni-Sort Port
CGMM_F5	0.32***	0.29***	0.16	0.13
RegGMM_F5_110	0.36***	0.39***	0.14	0.09
RegGMM_F5_330	0.36***	0.39***	0.11	0.06
MVE_F5	0.23***	0.23***	0.10	0.10
EW_F5	0.13***	0.13***	0.05	0.05
RegGMM_RW_F5_110	-	-	0.23*	0.09
RegGMM_RW_F5_330	-	-	0.15	0.06
CGMM_F10	0.35***	0.31***	0.33***	0.21**
RegGMM_F10_110	0.54***	0.56***	0.61***	0.44**
RegGMM_F10_330	0.52***	0.57***	0.59***	0.43**
MVE_F10	0.38***	0.38***	0.32***	0.32***
EW_F10	0.33***	0.33***	0.17***	0.17***
RegGMM_RW_F10_110	-	-	0.69***	0.43**
RegGMM_RW_F10_330	-	-	0.63***	0.41**

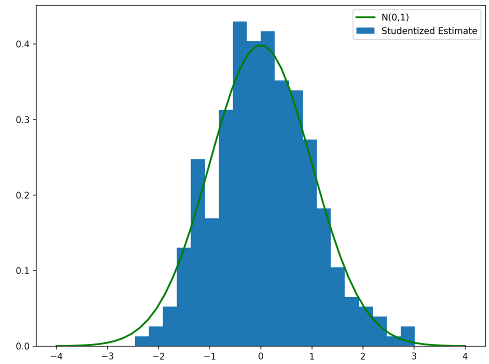
Notes: This table reports the risk-adjusted investment performance (alphas from CAPM and the Fama-French 3-factor model) for factor investing using the SDF model. We estimate the SDF model from 1972 to 2011 and perform an out-of-sample analysis from 2012 to 2021. In addition to RegGMM SDF strategies, we include the constant parameter GMM (CGMM), the mean-variance efficient portfolio (MVE), and the equal-weighted portfolio (EW), for comparison. We also report results for the out-of-sample rolling window (RW) updated RegGMM. As described in section 4.1, we report results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets, 5- and 10-factor SDF models, as well as 110 and 330 instrumental variables.

Figure 1: Simulation: Studentized Estimates for RegGMM

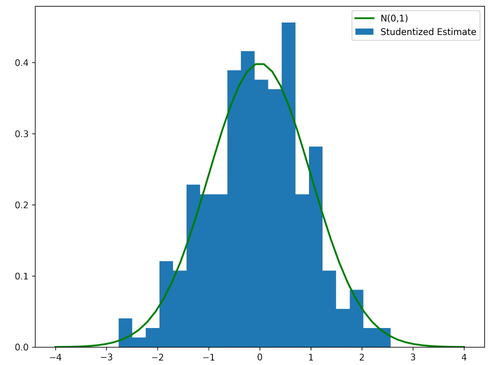
(a) DGP 1 (T=120)



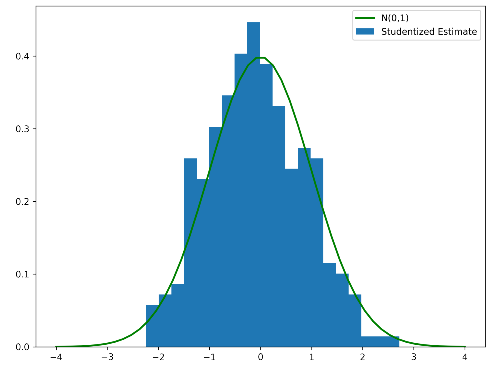
(b) DGP 2 (T=120)



(c) DGP 1 (T=360)



(d) DGP 2 (T=360)



Notes: This figure shows the studentized RegGMM time-varying estimates for $\frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{2,t}$ from bootstrap samples.

Figure 2: Year-by-Year Annualized Returns for Factors

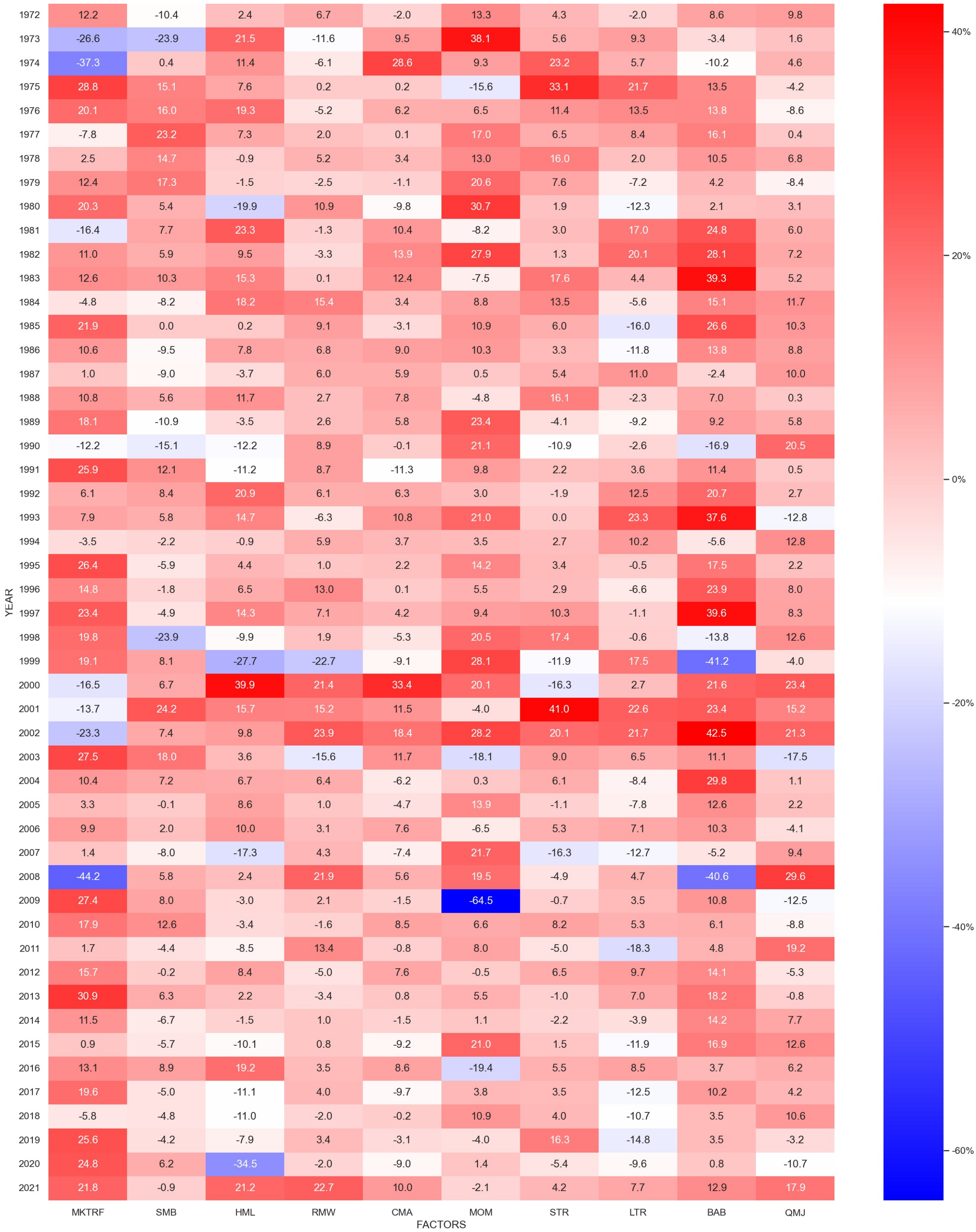
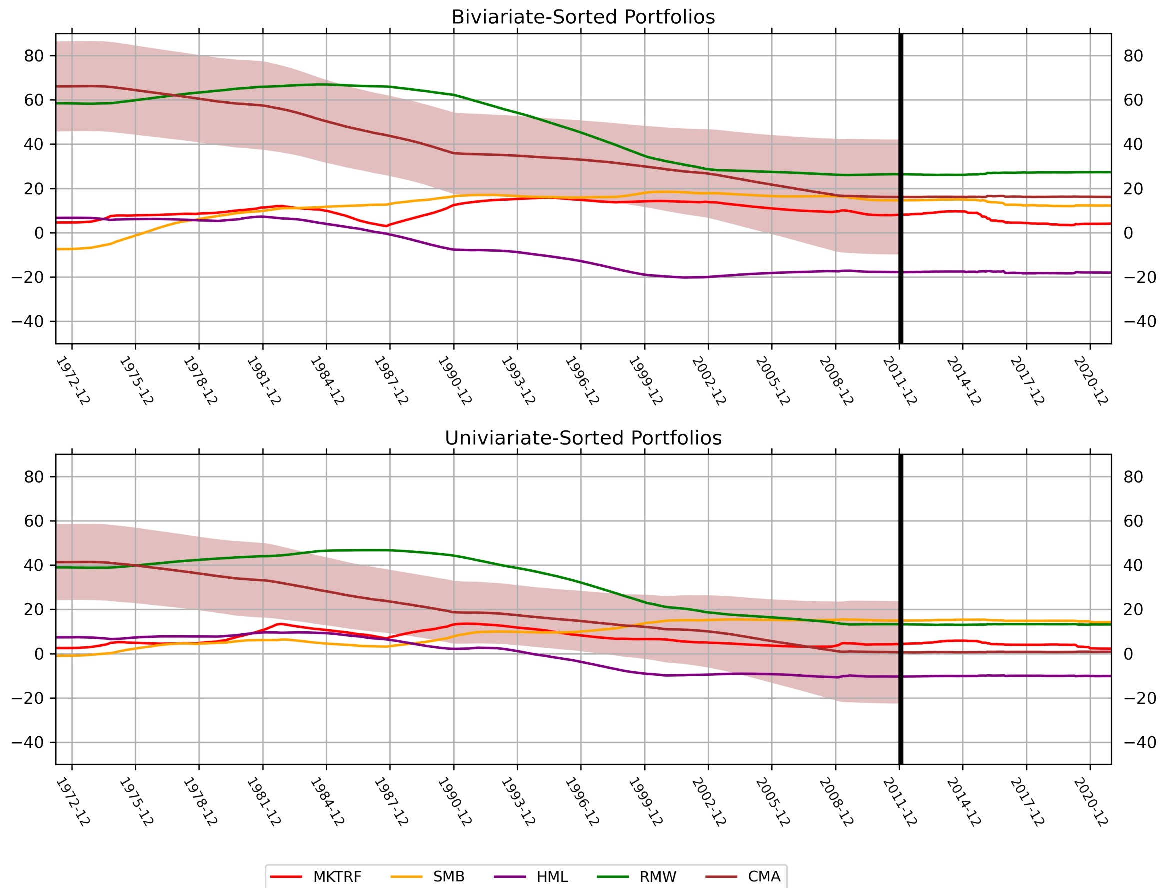
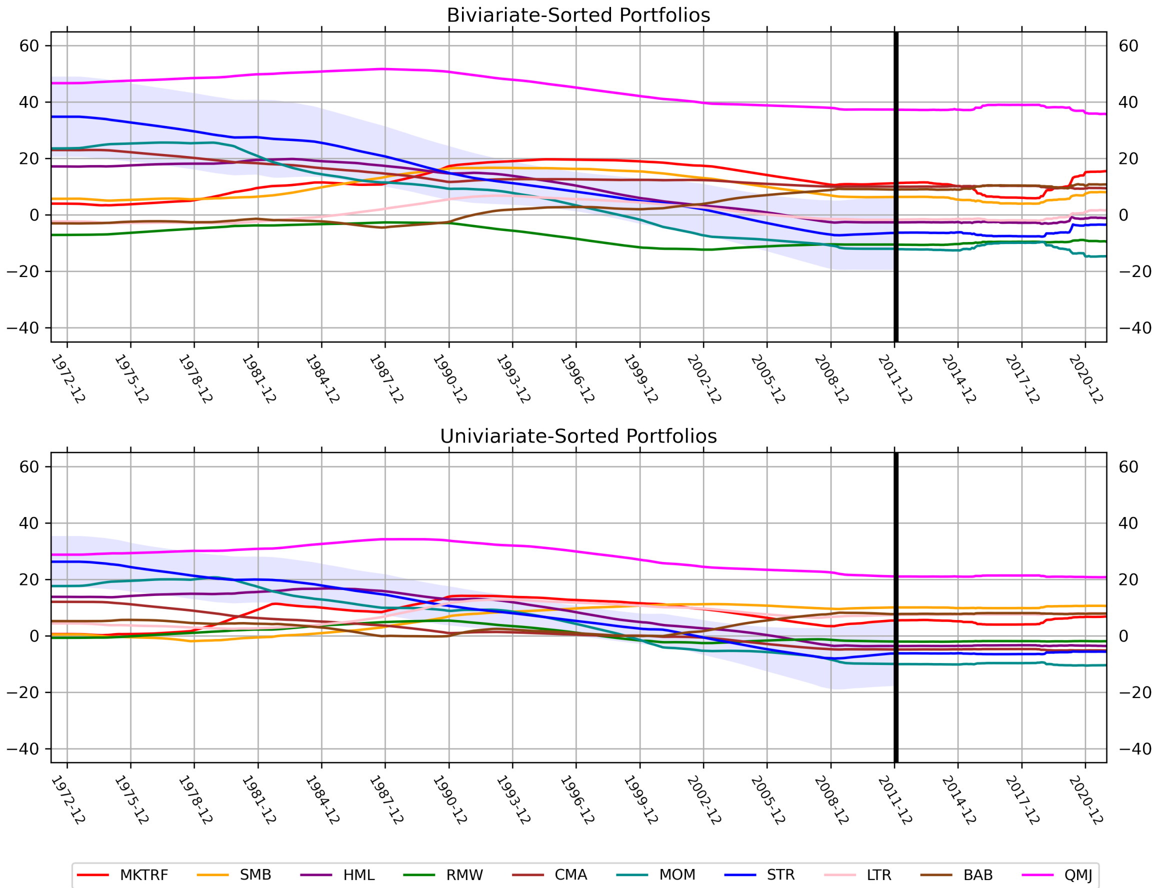


Figure 3: Time-Varying Risk Price for the 5-Factor SDF Model



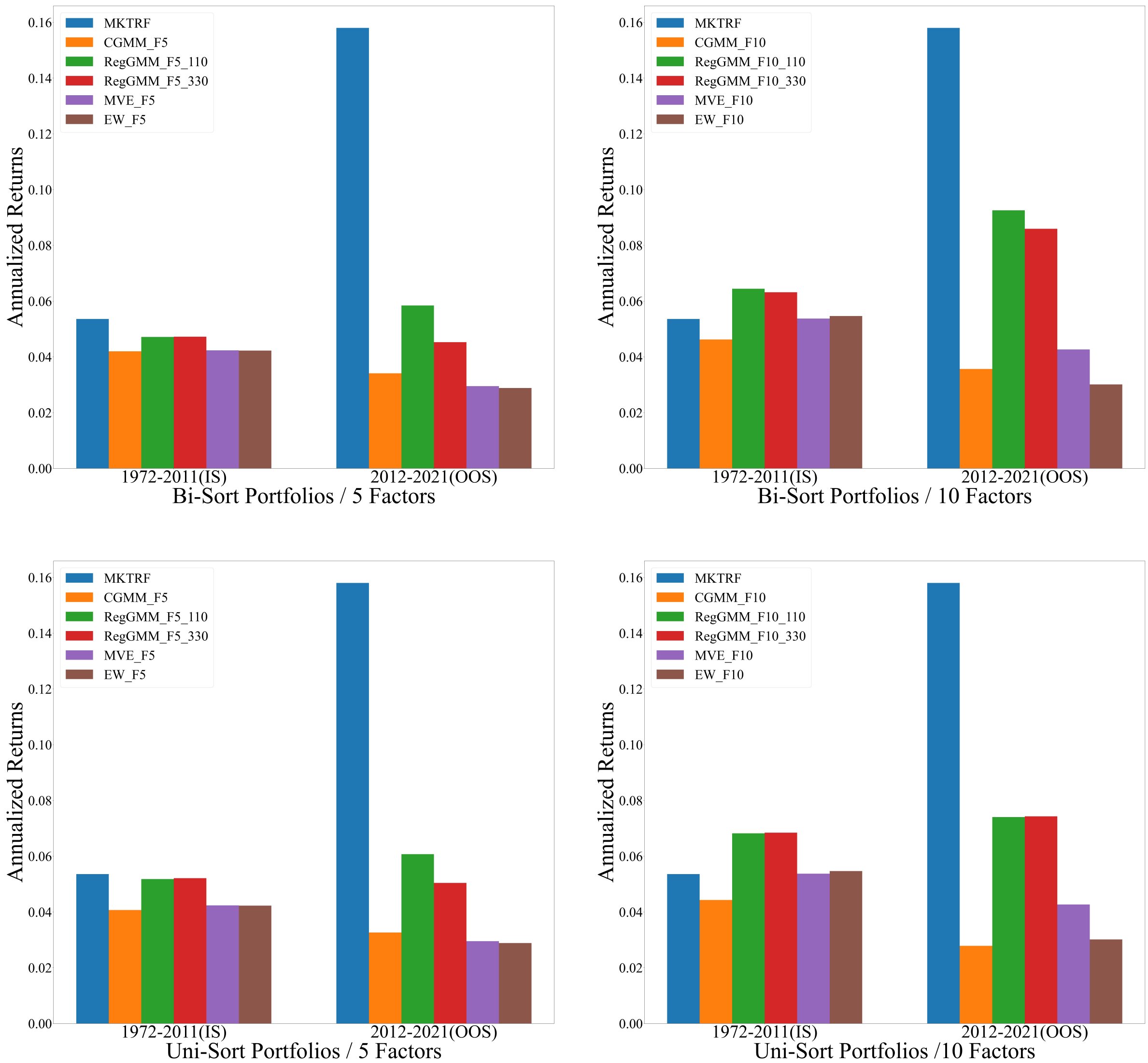
Notes: This figure shows the time-varying factor risk price (SDF loadings) for the Fama-French 5 factors. The training sample is the first 40 years, and the test sample is the last 10 years. The out-of-sample rolling-window update scheme is discussed in Section 4. The shaded area covers the 95% confidence interval for the investment factor, CMA. As described in section 4.1, we report results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets.

Figure 4: Time-Varying Risk Price for the 10-Factor SDF Model



Notes: This figure shows the time-varying factor risk price (SDF loadings) for Fama-French five factors, plus additional five factors introduced in section 4.1. The training sample is the first 40 years, and the test sample is the last 10 years. The out-of-sample rolling-window update scheme is discussed in Section 4. The shaded area covers the 95% confidence interval for the investment factor, STR. As described in section 4.1, we report results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets.

Figure 5: Absolute Investment Performance



Notes: This figure shows the annualized average returns for factor investing using the SDF model. We estimate the model from 1972 to 2011 and perform an out-of-sample analysis from 2012 to 2021. In addition to RegGMM SDF strategies, we include the constant GMM (CGMM), the market factor (MKTRF), the mean-variance efficient portfolio (MVE), and the equal-weighted portfolio (EW), for comparison. As described in section 4.1, we report results for 360 bivariate- and 610 univariate-sorted portfolios as different sets of test assets, 5- and 10-factor models, as well as 110 and 330 instrumental variables.