

Estimating Conditional Asset Pricing Models: Efficiency and Robustness*

Kevin B. Proulx[†]

November 11, 2018

[Latest Version](#)

Abstract

This paper revisits the efficient estimation of conditional beta pricing models with traded risk factors. By applying the theory of redundant moments of Breusch, Qian, Schmidt, and Wyhowski (1999), we prove that conditional homoskedasticity of returns given the risk factors is sufficient for equilibrium pricing conditions to be redundant in the sense that they do not improve the semi-parametric efficiency bound for beta. With jointly elliptical returns and risk factors, we extend this to a necessary and sufficient condition for redundancy. Relatedly, we also prove under joint ellipticity that the optimal tuning parameter for the generalized Principal Components Analysis loadings estimator of Lettau and Pelger (2018) is the multivariate excess kurtosis coefficient of the joint distribution of the returns and risk factors. This result explains their finding that the optimal tuning parameter is zero when factors are strong and regression errors are normally distributed. A caveat for assuming a parametric model of conditional betas is the non-trivial risk of model misspecification. Motivated by the empirical work of Nagel and Singleton (2011), we proceed to evaluate constant and state-dependent parametric risk price models with an objective function that balances the level of unconditional pricing errors and the volatility of conditional pricing errors. As a benchmark for comparison, we also consider nonparametric state-dependent risk price specifications. An application to unconditional and conditional Fama and French (1993) three factor models suggests that state-dependence in risk prices delivers substantial reductions in the volatility of conditional pricing errors at a small expense to the level of unconditional errors.

*I gratefully acknowledge Eric Renault for his valuable guidance. I thank Susanne Schennach and Gauti Eggertsson for their helpful comments and suggestions. I also thank Michael Bedard, Simon Freyaldenhoven, and Florian Gunsilius for useful discussions. All errors are my own.

[†]kevin.proulx@brown.edu. Department of Economics, Brown University, Providence, RI 02912.

1 Introduction

Motivated by the recent work of Lettau and Pelger (2018), this paper revisits the efficient estimation of conditional beta pricing models under the assumption that systematic risk factors are traded excess returns. We specifically address the following question: in models with observable traded risk factors, when are equilibrium asset pricing conditions informative for estimating betas?

Beta pricing models formalize the notion that differences in expected returns in a cross-section of assets can be explained by the differential exposure of each asset to systematic risk factors. This is established by the equilibrium condition that the conditional expected excess return of an asset is equal to the inner product between a vector of risk prices and a vector of risk exposures. When the risk factors are also traded assets, the equilibrium condition defines the risk price of each factor as its conditional mean. This provides a link between betas and the first moments of returns and risk factors that can be exploited in estimation.

In a conditional regression formulation of beta pricing models with traded risk factors, the equilibrium condition allows us to eliminate conditional alphas and to set the focus on efficient estimation of time-invariant betas.¹ When the model is correctly specified, there are two sources of efficiency improvements associated with eliminating conditional alphas. The first improvement is simply due to the fact that regression through the origin is more efficient than unrestricted ordinary least squares when the true intercept is zero. The second improvement comes from using the equilibrium condition as an overidentifying restriction. The latter will be our primary focus of interest in this paper.

Using the theory of redundant moments established by Breusch, Qian, Schmidt, and Wyhowski (1999), our first contribution is to prove that conditional homoskedasticity of returns given the risk factors is a sufficient condition for the equilibrium pricing condition to be redundant in the sense that it does not improve the semi-parametric efficiency bound for beta. Our second contribution is to prove that conditional homoskedasticity is also a necessary condition for redundancy when the returns and risk factors are jointly elliptical. Restricting our analysis to the elliptical class of distributions is motivated by statistical and economic considerations.

From a statistical viewpoint, we have three reasons for considering elliptical distributions: the multivariate elliptical family is a generalization of the multivariate normal which allows for both thinner and fatter tails; the change in the thickness of the tails is captured conveniently by a single kurtosis parameter; and the only distribution in the elliptical family for which heteroskedasticity is never present is the normal. From an economic viewpoint, Berk (1997) proves that restricting returns to be elliptically distributed ensures that mean-variance analysis, and hence beta pricing, is consistent with expected utility maximization irrespective of the specific preferences of investors. Elliptic distributions in finance have notably been considered by Owen and Rabinovitch (1983), Chamberlain (1983), Hodgson, Linton, and Vorkink (2002), Peñaranda and Sentana (2012, 2015). The results presented herein are closely related to Peñaranda and Sentana (2012).

We develop intuition for the aforementioned results by characterizing the efficient GMM

¹We can also accommodate (affine) parametric specifications of dynamic betas. We present the constant beta case for ease of exposition.

estimator of beta as an optimally weighted average of two just-identified moment estimators: one derived from the statistical definition of beta as a regression coefficient; the other derived from the equilibrium pricing condition. Under joint ellipticity, we prove that the optimal weight on the model implied moment estimator is increasing in both the Sharpe ratio of the traded risk factors and the degree of conditional heteroskedasticity of returns given the risk factors. The latter is captured by the multivariate excess kurtosis coefficient of the joint elliptical distribution of the returns and risk factors.

Lettau and Pelger (2018) generalize Principal Component Analysis (PCA) for latent asset-pricing factor models to include a penalty that accounts for pricing errors in expected returns.² They prove that their loadings estimator is more efficient than conventional PCA when risk prices are non-zero, and that the optimal penalty on the pricing errors is zero when regression errors are i.i.d. normal. Through a GMM formulation of their estimator, they show that a zero penalty is equivalent to applying PCA to an uncentered covariance matrix. Under joint ellipticity of returns and risk factors, we prove in this paper that the optimal penalty for the generalized PCA loadings estimator of Lettau and Pelger (2018) is the multivariate excess kurtosis coefficient of the joint elliptical distribution. This explains why Lettau and Pelger (2018) find that the optimal penalty is zero when regressions are i.i.d. normal.

MacKinlay and Richardson (1991) were the first to use the CAPM equilibrium condition as an overidentifying restriction for estimating betas. Using the framework of the generalized method of moments (GMM) of Hansen (1982), they derive a correction for Wald tests that accounts for contemporaneous conditional heteroskedasticity in returns. Renault (1997) also studies the efficient estimation of beta in a model with traded risk factors. He shows in the case of Gaussian returns that the efficient estimator of beta is the ordinary least squares estimator in the regression through the origin of the returns on the risk factors. The non-trivial lesson implied by his result is that ordinary least squares is no longer generally efficient if returns are non-Gaussian.³

A caveat for modeling dynamic betas parametrically is the non-trivial risk of model misspecification, since it is often the case that economic theory does not provide guidance for how they should be modeled. This is troubling since we know from Ghysels (1998) that misspecifying the dynamics of betas can cause serious pricing errors that are often larger than those committed by constant beta models. As recently argued by Brandt and Chapman (2018), the same critique holds for the modeling of dynamic risk prices. This paper proceeds to re-evaluate constant risk price and state-dependent risk price models using the framework of conditionally affine stochastic discount factor (SDF) models. To hold fixed Ghysels (1998) critique, we remain nonparametric with respect to conditional betas and focus on the modeling and estimation of state-dependent risk prices.

With the mindset that parametric risk price models are misspecified, we advocate for defining an objective function that delivers a pseudo-true SDF that does the best job in a predefined sense. Nagel and Singleton (2011) recently argue that estimating conditional asset pricing models using unconditional moment restrictions delivers small average pricing

²Lettau and Pelger (2018) is closely related to Fan and Zhong (2018) who consider the more general problem of optimal subspace estimation with overidentifying restrictions.

³Relatedly, Jagannathan and Wang (1998) generalize the asymptotic distribution theory for the two-stage cross-sectional regression method of Shanken (1992) to account for conditional heteroskedasticity.

errors at the expense of “enormous time-variation in conditional pricing errors.” Motivated by this critique, we specify an objective function that trades off the level of average pricing errors with the volatility of conditional pricing errors.

Our focus on evaluating asset pricing models based on their ability to meet certain pre-specified objectives is in agreement with the philosophy of Ludvigson (2013) who prescribes that future econometric work should place “greater emphasis on methodologies that facilitate the comparison of multiple competing models, all of which are potentially misspecified” and “reduced emphasis on individual hypothesis tests of whether a single model is specified without error.” It is also in agreement with Balduzzi and Robotti (2010) who write “researchers are often interested in studying misspecified models and sometimes keep the issue of estimating risk premia separate from the issue of testing whether the underlying asset pricing model is literally true or not.”

Antoine, Proulx, and Renault (2018) also focus on the estimation and interpretation of a pseudo-true SDF defined as the argument that minimizes the conditional Hansen and Jagannathan (1997) distance as put forth by Gagliardini and Ronchetti (2018). They prove that the conditional HJ-distance delivers a loss function that trades off the level of unconditional pricing errors with the volatility of conditional pricing errors, and has the advantage that it delivers the exact pricing of traded risk factors. A difficulty with using the conditional HJ-distance, however, is that the weight matrix is a conditional second moment matrix of asset returns that must be nonparametrically estimated or modeled. With a misspecified model, the cost of a nonparametric weight matrix is severe because its bias and nonparametric rate of convergence pollutes estimation of the parameters of interest (e.g. see Hall and Inoue (2003) and Antoine et al. (2018)).

We avoid the complications of having a weight matrix that depends on the conditioning information by using the identity weight matrix. This is recommended by Ludvigson (2018) for applying estimation procedures that work directly with misspecified conditional moment restrictions (see also Chen, Favilukis, and Ludvigson (2014)). Ludvigson (2013) also advocates for the use of the identity weight matrix for evaluating asset pricing models in order to maintain the spread in average returns between the original test assets that the cross-sectional asset pricing model is meant to explain.

We employ our objective function to estimate various Fama and French (1993) three factors models with and without state-dependent risk prices. We consider two modeling approaches for the state-dependent case: first, we model the risk prices as affine parametric functions of a scalar conditioning variable; second, we consider the risk prices as nonparametric functions of a scalar conditioning variable. Since our objective function is comprised of conditional moments, we estimate each parametric (nonparametric) model with the local GMM estimator of Gospodinov and Otsu (2012) (Lewbel (2007)). Each local GMM estimator works directly with the conditional moment restrictions by formulating a GMM-like quadratic form of kernel estimates of conditional moments.

Our empirical work confirms that asset pricing models with state-dependent risk prices estimated with unconditional moments deliver small unconditional pricing errors at the expense of extremely large volatility in the conditional pricing errors. Nevertheless, when working directly with conditional moments, we find that state-dependence in risk prices can deliver both small average pricing errors and stable conditional pricing errors. As a benchmark for comparison, we also find that nonparametric risk prices deliver the least volatile

pricing errors, albeit at a small additional cost to the level of unconditional errors.

Roussanov (2014) also builds on Nagel and Singleton (2011) by testing the conditional implications of asset pricing models with nonparametrically estimated risk prices. His “non-parametric kernel regression” procedure is actually a local linear alternative to the local constant GMM of Lewbel (2007) used here. Other notable papers that estimate conditional asset pricing models nonparametrically are: Ferreira, Gil-Bazo, and Orbe (2011) who estimate nonparametric conditional beta pricing models; Fang, Ren, and Yuan (2011) who estimate a fully nonparametric SDF using the local GMM of Lewbel (2007); Li, Su, and Xu (2015) who develop a two-step semi-parametric estimator for conditional factor models that allows for a misspecified parametric first step; and Cai, Ren, and Sun (2015) who develop and apply a local linear estimator for conditionally affine SDF models.

Now that we have summarized our results and briefly reviewed the literature, we proceed to outline the rest of the paper. Section 2 revisits the efficient estimation of conditional beta pricing models under the assumptions of constant betas and traded risk factors. We start by reviewing the necessary and sufficient condition for redundancy of a conditional moment restriction as established by Breusch et al. (1999). We proceed to prove that if returns are contemporaneously conditionally homoskedastic given the risk factors, then equilibrium pricing conditions do not help to improve the semi-parametric efficiency bound for beta. We conclude this section with a characterization of the optimal tuning parameter for the generalized PCA loadings estimator of Lettau and Pelger (2018) when returns and risk factors are jointly elliptical. Section 3 reviews a general framework for conditionally affine SDF models and introduces our objective function for estimation of misspecified models. Since our objective is comprised of conditional moments that must be nonparametrically estimated, we also review two estimators that work directly with conditional moment restrictions: the local GMM of Gospodinov and Otsu (2012) and the nonparametric local GMM of Lewbel (2007). In Section 4, we estimate a variety of Fama-French three factor models with and without state-dependent risk prices. We find that state-dependent risk price models deliver both small unconditional pricing errors and stable conditional pricing errors when evaluated using our objective function. Section 5 concludes with a discussion and ideas for future research.

2 Overidentified Betas with Traded Risk Factors

This section addresses the question: when is the equilibrium condition of a conditional beta pricing model informative for the efficient estimation of betas? We start in the next subsection by considering two sets of conditional moment restrictions that define betas: statistical moments that define them as slope coefficients in regressions through the origin of excess returns on risk factors; and model implied moments derived from the equilibrium pricing conditions. We proceed to review a necessary and sufficient condition established by Breusch et al. (1999) for the redundancy of a conditional moment restriction. We say that a conditional moment restriction is redundant if it does not improve the semi-parametric efficiency bound for our parameters of interest. We use the redundancy condition to show that conditional homoskedasticity of returns given the risk factors is sufficient for redundancy of the model implied moments.

2.1 A General Redundancy Condition

We assume that the generating process for an N -vector of excess returns is:

$$\begin{aligned} R_{t+1}^e &= E[R_{t+1}^e | I(t)] + \beta(F_{t+1} - E[F_{t+1} | I(t)]) + u_{t+1} \\ &= \alpha_t + \beta F_{t+1} + u_{t+1} \end{aligned} \quad (2.1)$$

where F_{t+1} is a K -vector of orthogonal traded risk factors with $E[F_{t+1} | I(t)] \neq 0$, α_t is an implicitly defined N -vector of conditional alphas, $\beta = Cov(R_{t+1}^e, F_{t+1} | I(t)) Var(F_{t+1} | I(t))^{-1}$ is an $N \times K$ time-invariant matrix of assets' risk exposures, and the residuals u_{t+1} form a martingale difference sequence with respect to the investors' information set, $I(t)$, generated by the natural filtration of lagged returns and factors.⁴ Without any further restrictions on the return generating process, the conditional alphas and betas are identified from the following set of conditional moment restrictions:

$$E[\Psi_{t+1}^{\tilde{S}}(\alpha_t, \beta) | I(t)] = \begin{pmatrix} E[R_{t+1}^e - \alpha_t - \beta F_{t+1} | I(t)] \\ E[vec \{ (R_{t+1}^e - \alpha_t - \beta F_{t+1}) F_{t+1}' \} | I(t)] \end{pmatrix} = 0_{N+NK}. \quad (2.2)$$

The equilibrium implication of a conditional beta pricing model is that conditional expected excess returns are linear in beta and a K -vector of time-varying risk prices, λ_t :

$$E[R_{t+1}^e - \beta \lambda_t | I(t)] = 0_N. \quad (2.3)$$

Since we assume that the risk factors are traded and orthogonal, the equilibrium condition (2.3) also imposes the restriction that the vector of time-varying risk prices is equal to the conditional mean of the vector of risk factors:

$$E[F_{t+1} - \lambda_t | I(t)] = 0_K. \quad (2.4)$$

Equation (2.4) allows us to reformulate the equilibrium pricing condition (2.3) as:

$$E[\Psi_{t+1}^M(\beta) | I(t)] = E[R_{t+1}^e - \beta F_{t+1} | I(t)] = 0_N. \quad (2.5)$$

Since (2.5) is true if and only if $\alpha_t = 0$, we are able to eliminate the conditional alphas in order to set the focus on the efficient estimation of the beta coefficients. We arrive at the following set of $NK + N$ conditional moment restrictions for estimation of the NK unknown beta coefficients:

$$E[\Psi_{t+1}(\beta) | I(t)] = \begin{pmatrix} E[vec \{ (R_{t+1}^e - \beta F_{t+1}) F_{t+1}' \} | I(t)] \\ E[R_{t+1}^e - \beta F_{t+1} | I(t)] \end{pmatrix} = 0_{NK+N}. \quad (2.6)$$

where the moment function is $\Psi_{t+1}(\beta) = (\Psi_{t+1}^S(\beta)', \Psi_{t+1}^M(\beta)')'$.

Since beta is now over-identified, we can begin to address the question of interest: when are the conditional moment restrictions imposed by the asset pricing model redundant in the sense that they do not improve the semi-parametric efficiency bound for beta? We start

⁴We can also accommodate (affine) parametric specifications of dynamic betas. We present the constant beta case for ease of exposition.

by establishing a general condition for redundancy of conditional moment restrictions. Let us define the conditional Jacobian:

$$\Gamma_t = \begin{bmatrix} \Gamma_{S,t} \\ \Gamma_{M,t} \end{bmatrix} = \begin{pmatrix} E \left[\frac{\partial \Psi_{t+1}^S(\beta_0)'}{\partial \beta} | I(t) \right] \\ E \left[\frac{\partial \Psi_{t+1}^M(\beta_0)'}{\partial \beta} | I(t) \right] \end{pmatrix} \quad (2.7)$$

and the conditional variance matrix:

$$\Sigma_t(\beta_0) = Var[\Psi_{t+1}(\beta_0) | I(t)] = \begin{bmatrix} \Sigma_{SS,t} & \Sigma_{SM,t} \\ \Sigma_{MS,t} & \Sigma_{MM,t} \end{bmatrix}. \quad (2.8)$$

Without loss of generality, we follow Breusch et al. (1999) by orthogonalizing the two sets of conditional moments restrictions to redefine them as:

$$E[\Phi_{t+1}(\beta) | I(t)] \begin{pmatrix} E[\Psi_{t+1}^S(\beta) | I(t)] \\ E[\Psi_{t+1}^M(\beta) - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Psi_{t+1}^S(\beta) | I(t)] \end{pmatrix} = 0_{NK+N} \quad (2.9)$$

where the moment function is $\Phi_{t+1}(\beta) = (\Phi_{t+1}^S(\beta)', \Phi_{t+1}^M(\beta)')$. This leads us to define another conditional Jacobian:

$$\tilde{\Gamma}_t(\beta_0) = \begin{bmatrix} \tilde{\Gamma}_{S,t} \\ \tilde{\Gamma}_{M,t} \end{bmatrix} = \begin{bmatrix} \Gamma_{S,t} \\ \Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t} \end{bmatrix} \quad (2.10)$$

and another conditional variance matrix:

$$\tilde{\Sigma}_t(\beta_0) = Var[\Phi_{t+1}(\beta_0) | I(t)] = \begin{bmatrix} \Sigma_{SS,t} & 0 \\ 0 & \Sigma_{MM,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Sigma_{SM,t} \end{bmatrix}. \quad (2.11)$$

Since the residuals u_{t+1} form a martingale difference sequence with respect to the filtration $I(t)$, the optimal estimating equations using the full set of conditional moment restrictions are:

$$E[Z_t(\beta_0) \Phi_{t+1}(\beta)] = 0 \quad (2.12)$$

where it is well known (e.g. Hansen (1985); Chamberlain (1987)) that the optimal choice of instruments, $Z_t(\beta_0)$, is:

$$Z_t(\beta_0) = \tilde{\Gamma}_t(\beta_0) \tilde{\Sigma}_t(\beta_0)^{-1}. \quad (2.13)$$

Similarly, the optimal estimating equations using only the statistical conditional moment restrictions $E[\Phi_{t+1}^S(\beta) | I(t)] = 0$ are:

$$E[Z_t^S(\beta_0) \Phi_{t+1}^S(\beta)] = 0 \quad (2.14)$$

where the optimal choice of instruments, $Z_t^S(\beta_0)$, is:

$$Z_t^S(\beta_0) = \tilde{\Gamma}_t^S(\beta_0) \Sigma_{SS,t}(\beta_0)^{-1}. \quad (2.15)$$

An estimator using a sample analog of (2.12) is defined as the solution, $\hat{\beta}_T$, to the system of

equations:

$$\frac{1}{T} \sum_{t=1}^{T-1} Z_t(\beta_0) \Phi_{t+1}(\hat{\beta}_T) + o_p(1) = 0. \quad (2.16)$$

A second estimator using the sample analog of (2.14) is defined as the solution, $\hat{\beta}_T^S$, to the system of equations:

$$\frac{1}{T} \sum_{t=1}^{T-1} Z_t^S(\beta_0) \Phi_{t+1}^S(\hat{\beta}_T^S) + o_p(1) = 0. \quad (2.17)$$

Note that in both cases the $o_p(1)$ term comes from using a first-step consistent estimator of the optimal instruments (e.g. Newey (1993)). We define the model implied moment restrictions as redundant given the statistical moment restrictions if $\hat{\beta}_T$ and $\hat{\beta}_T^S$ are asymptotically equivalent. The following proposition, which appears in Section 6 of Breusch et al. (1999), provides a necessary and sufficient condition for redundancy.

Proposition 1. *The almost sure equality*

$$\Gamma_{M,t}(\beta_0) = \Sigma_{MS,t}(\beta_0) \Sigma_{SS,t}(\beta_0)^{-1} \Gamma_{S,t}(\beta_0) \quad a.s. \quad (2.18)$$

is necessary and sufficient for the model implied conditional moment restrictions to be redundant given the statistical conditional moment restrictions.

Proof. The asymptotic variance of $\hat{\beta}_T^S$ is:

$$Avar(\sqrt{T}(\hat{\beta}_T^S - \beta_0)) = \Omega_S^{-1}$$

where:

$$\Omega_S = E[\Gamma_{S,t}(\beta_0) \Sigma_{SS,t}(\beta_0)^{-1} \Gamma_{S,t}(\beta_0)'].$$

The asymptotic variance of $\hat{\beta}_T$ is:

$$Avar(\sqrt{T}(\hat{\beta}_T - \beta_0)) = \Omega^{-1}$$

where:

$$\begin{aligned} \Omega &= E[\tilde{\Gamma}_t(\beta_0) \tilde{\Sigma}_t(\beta_0)^{-1} \tilde{\Gamma}_t(\beta_0)'] \\ &= E[\Gamma_{S,t}(\beta_0) \Sigma_{SS,t}(\beta_0)^{-1} \Gamma_{S,t}(\beta_0)'] + E[\tilde{\Gamma}_{M,t}(\beta_0) \tilde{\Sigma}_{MM,t}(\beta_0)^{-1} \tilde{\Gamma}_{M,t}(\beta_0)'] \\ &= \Omega_S + E \left[(\Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t}) \tilde{\Sigma}_{MM,t}^{-1} (\Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t})' \right] \end{aligned}$$

for $\tilde{\Sigma}_{MM,t} = \Sigma_{MM,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Sigma_{SM,t}$. It is easily seen that a necessary and sufficient condition for the asymptotic variances of $\hat{\beta}_T$ and $\hat{\beta}_T^S$ to coincide is that:

$$(\Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t}) \tilde{\Sigma}_{MM,t}^{-1} (\Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t})' = 0 \quad a.s.$$

Assuming that $\tilde{\Sigma}_{MM,t}$ is full-rank a.s., this is true if and only if:

$$\Gamma_{M,t} - \Sigma_{MS,t} \Sigma_{SS,t}^{-1} \Gamma_{S,t} = 0 \quad a.s.$$

Moreover, the system of equations in (2.16) and (2.17) coincide up to an $o_p(1)$ term if and only if $\Gamma_{M,t} - \Sigma_{MS,t}\Sigma_{SS,t}^{-1}\Gamma_{S,t} = 0$ a.s. This implies that the estimators are asymptotically equivalent; i.e. $\sqrt{T}(\hat{\beta}_T - \hat{\beta}_T^S) = o_p(1)$. \square

We now specialize the redundancy condition (2.18) to our problem of interest by characterizing the relevant conditional Jacobians and variance matrices using the full set of conditional moment restrictions (2.6).

Proposition 2. *The model implied conditional moment restrictions are redundant given the statistical conditional moment restrictions if and only if*

$$\left[E_t [F'_{t+1}] E_t [F_{t+1}F'_{t+1}]^{-1} \otimes I_N \right] = E_t [F'_{t+1} \otimes u_{t+1}u'_{t+1}] E_t [F_{t+1}F'_{t+1} \otimes u_{t+1}u'_{t+1}]^{-1} \quad (2.19)$$

almost surely.

Proof.

$$\begin{aligned} \Gamma_{M,t} &= E_t \left[-\frac{\partial((F'_{t+1} \otimes I_N)vec(\beta))}{\partial vec(\beta)'} \right] &&= E_t [-F'_{t+1} \otimes I_N] \\ \Gamma_{S,t} &= E_t \left[-\frac{\partial((F_{t+1}F'_{t+1} \otimes I_N)vec(\beta))}{\partial vec(\beta)'} \right] &&= E_t [-F_{t+1}F'_{t+1} \otimes I_N] \\ \Sigma_{MS,t} &= E_t [u_{t+1} ((F_{t+1} \otimes I_N)vec(u_{t+1}))'] &&= E_t [F'_{t+1} \otimes u_{t+1}u'_{t+1}] \\ \Sigma_{SS,t} &= E_t [vec(u_{t+1}F'_{t+1})vec(u_{t+1}F'_{t+1})'] &&= E_t [F_{t+1}F'_{t+1} \otimes u_{t+1}u'_{t+1}] \end{aligned}$$

The result follows from substitution into equation (2.18). \square

Equation (2.19) links the first and second moments of the risk factors with the third and fourth cross-moments of the risk factors and regression residuals. This form of the redundancy conditions leads us to search for restrictions on the conditional variance of the returns given the risk factors and/or the kurtosis of their joint distribution that might be sufficient for the redundancy of the model implied moments.

Proposition 3. *If excess returns are contemporaneously conditionally homoskedastic given the risk factors, then the model implied conditional moment restrictions are redundant given the statistical conditional moment restrictions.*

Proof. Define the extended information set $\tilde{I}(t) = I(t) \cup F_{t+1}$. Since:

$$Var[R_{t+1}^e | \tilde{I}(t)] = E[u_{t+1}u'_{t+1} | \tilde{I}(t)]$$

conditional homoskedasticity of the returns given the risk factors implies that:

$$E[u_{t+1}u'_{t+1} | \tilde{I}(t)] = E[u_{t+1}u'_{t+1} | I(t)].$$

Hence, we have:

$$\begin{aligned}
\Sigma_{MS,t}\Sigma_{SS,t}^{-1} &= E_t \left[F'_{t+1} \otimes E \left[u_{t+1}u'_{t+1} | \tilde{I}(t) \right] \right] E_t \left[F_{t+1}F'_{t+1} \otimes E \left[u_{t+1}u'_{t+1} | \tilde{I}(t) \right] \right]^{-1} \\
&= \left[E_t \left[F'_{t+1} \right] \otimes E_t \left[u_{t+1}u'_{t+1} \right] \right] \left[E_t \left[F_{t+1}F'_{t+1} \right]^{-1} \otimes E_t \left[u_{t+1}u'_{t+1} \right]^{-1} \right] \\
&= E_t \left[F'_{t+1} \right] E_t \left[F_{t+1}F'_{t+1} \right]^{-1} \otimes I_N \\
&= \Gamma_{M,t}\Gamma_{S,t}^{-1} \text{ a.s.}
\end{aligned}$$

The conclusion follows from Proposition 2. □

Since there is no conditional heteroskedasticity within a jointly normal vector, we immediately have the following corollary:

Corollary 1. *If the returns and risk factors are jointly normal conditional on the investors' information set, then the model implied condition moment restrictions are redundant given the statistical conditional moment restrictions.*

Proof. It is sufficient to prove that the conditional variance of the returns given the risk factors does not depend on the level of the risk factors. It is a standard textbook result that for a jointly normal vector:

$$Var[R_{t+1} | \tilde{I}(t)] = Var_t[R_{t+1}] - Cov_t[R_{t+1}, F_{t+1}]Var_t[F_{t+1}]^{-1}Cov_t[F_{t+1}, R_{t+1}]. \quad (2.20)$$

□

The intuition behind Proposition 3 and its corollary can be derived from a regression interpretation of the beta pricing model. Since the equilibrium pricing condition is equivalent to the restriction that the regression residuals are conditionally mean zero, it is intuitively informative for estimating beta (beyond imposing that the intercept is zero) only if the residuals provide additional information about the distribution of the risk factors. Conditional homoskedasticity of the returns given the risk factors delivers something close to independence between the residuals and the risk factors, in that the third and fourth cross-moments of the joint distribution of the residuals and the risk factors can be separated into products of first and second moments of the marginal distributions of the residuals and the risk factors. As discussed by Renault (1997), conditional heteroskedasticity ties together variance parameters with mean parameters, whereas conditional homoskedasticity allows for their independent estimation. Since the equilibrium pricing condition is comprised of conditional means, it is redundant precisely when mean parameters can be separated from variance parameters such as beta.

In principle there could be multivariate distributions of returns and risk factors that exhibit contemporaneous conditional heteroskedasticity and satisfy the redundancy condition (2.19). We proceed in the next section by re-characterizing the redundancy condition under an additional restriction that the returns and risk factors are jointly elliptical. We will prove in this case that conditional homoskedasticity is also necessary for redundancy of the equilibrium pricing condition.

2.2 Redundancy with Elliptically Distributed Returns

Restricting our analysis to the elliptical class of distributions is motivated by statistical and economic considerations. From a statistical viewpoint, we have three reasons for considering elliptical distributions: the multivariate elliptical family is a generalization of the multivariate normal which allows for both thinner and fatter tails; the change in the thickness of the tails is captured conveniently by a single kurtosis parameter; and the only distribution in the elliptical family for which heteroskedasticity is never present is the normal (Theorem 7, Kelker (1970)). Examples of elliptical distributions include the multivariate normal, Laplace, Student, Cauchy, and symmetric Stable distributions (Fang, Kotz, and Ng (1989)). We will restrict our attention however to distributions with finite variance.

From an economic viewpoint, Berk (1997) proves that restricting returns to be elliptically distributed ensures that mean-variance analysis, and hence beta pricing, is consistent with expected utility maximization irrespective of the specific preferences of investors. Elliptic distributions in finance have notably been considered by Owen and Rabinovitch (1983), Chamberlain (1983), Hodgson et al. (2002), Peñaranda and Sentana (2012, 2015).

Suppose the $M \times 1$ vector $X_{t+1} = (R'_{t+1}, F'_{t+1})'$ follows an elliptical distribution conditional on the information set, $I(t)$. Following Peñaranda and Sentana (2012, 2015), we define an elliptical distribution by means of an affine transformation

$$X_{t+1} = \nu_t + \Sigma_t^{\frac{1}{2}} e_{t+1} u_{t+1}$$

where $\text{rank}(\Sigma_t) = M$, u_{t+1} is uniformly distributed on the unit sphere in \mathbb{R}^M with $\text{Cov}_t(u_{t+1}) = I_M/M$, and the generating variate e_{t+1} is a non-negative scalar-valued random variable independent of u_{t+1} with normalization $E_t[e_{t+1}^2] = M$. These properties imply that $E_t[e_{t+1}u_{t+1}] = 0$, $\text{Var}_t(e_{t+1}u_{t+1}) = I_M$, $E_t[X_{t+1}] = \nu_t$, and $\text{Var}_t(X_{t+1}) = \Sigma_t$. Lastly, if $E_t[e_{t+1}^4] < \infty$, then by the Cauchy-Schwartz inequality and $E_t[e_{t+1}^2] = M$ the coefficient of multivariate excess kurtosis, $\kappa = E_t[e_{t+1}^4]/[M(M+2)] - 1$, will be bounded below by $\kappa \geq -2/(M+2)$.

We denote the elliptical distribution conditional on the information set at time t as $\mathcal{E}_M(\nu_t, \Sigma_t, F_e)$, where F_e is the CDF of the generating variate, and ν_t and Σ_t are partitioned as:

$$\nu_t = (E_t[R_{t+1}]', E_t[F_{t+1}]')'$$

and

$$\Sigma_t = \begin{bmatrix} \Sigma_{RR,t} & \Sigma_{RF,t} \\ \Sigma_{FR,t} & \Sigma_{FF,t} \end{bmatrix}.$$

It is well known that, when it exists, the conditional variance matrix $\text{Var}(R_{t+1}|\tilde{I}(t))$ is:

$$\text{Var} [R_{t+1}|\tilde{I}(t)] = \frac{1}{N} \left(E[e_{t+1}^2|\tilde{I}(t)] - d_{\Sigma_{FF,t}}^2(F_{t+1}, E_t(F_{t+1})) \right) \Sigma_{R|F,t}$$

where

$$\Sigma_{R|F,t} = \Sigma_{RR,t} - \Sigma_{RF,t} \Sigma_{FF,t}^{-1} \Sigma_{FR,t}$$

and

$$d_{\Sigma_t}(X_{t+1}, \nu_t) = \sqrt{(X_{t+1} - \nu_t)' \Sigma_t^{-1} (X_{t+1} - \nu_t)}.$$

Since heteroskedasticity is present only through a scalar function, we subsequently denote the conditional variance of interest as:

$$\text{Var} \left[R_{t+1} | \tilde{I}(t) \right] = \rho(F_{t+1}) \Sigma_{R|F,t}$$

where $\rho(\cdot)$ is a scalar function of the factors. We now turn to a lemma that will be useful for proving the main proposition of this section. These results can also be found in Lemma D.3. of Peñaranda and Sentana (2012), but we present the proofs for self-containedness.

Lemma 1. *With the normalization $E_t[e_{t+1}^2] = M$ and by the symmetry of the elliptical distribution, $E_t[\rho(F_{t+1})] = 1$ and $E_t[\rho(F_{t+1})F_{t+1}] = E_t[F_{t+1}]$.*

Proof.

$$\begin{aligned} \text{Var}_t(R_{t+1}) &= E_t[\text{Var}(R_{t+1} | \tilde{I}(t)) + \text{Var}_t(E[R_{t+1} | \tilde{I}(t)])] \\ &= E_t[\rho(F_{t+1}) \Sigma_{R|F,t} + \Sigma_{RF,t} \Sigma_{FF,t}^{-1} \Sigma_{FR,t}] \\ &= \Sigma_{R|F,t} (E_t[\rho(F_{t+1})] - 1) + \text{Var}_t(R_{t+1}) \end{aligned}$$

where the second line follows from $E[R_{t+1} | \tilde{I}(t)] = \beta F_{t+1}$ and the definition of β . It follows that $E_t[\rho(F_{t+1})] = 1$ provided that $\Sigma_{R|F,t}$ is full rank.

By the symmetry of the elliptical distribution,

$$\begin{aligned} E_t[(F_{t+1} - E_t[F_{t+1}])' \otimes (R_{t+1} - E_t[R_{t+1}])(R_{t+1} - E_t[R_{t+1}])] &= 0 \\ \iff E_t[(F_{t+1} - E_t[F_{t+1}])' \otimes \text{Var}(R_{t+1} | \tilde{I}(t))] &= 0 \\ \iff E_t[(F_{t+1} - E_t[F_{t+1}])' \otimes \rho(F_{t+1}) \Sigma_{R|F,t}] &= 0 \\ \iff E_t[F_{t+1}' \rho(F_{t+1})] &= E_t[F_{t+1}'] \end{aligned}$$

where the last line follows by $E_t[\rho(F_{t+1})] = 1$. □

We are now ready to prove the main result of this section.

Proposition 4. *Within the class of elliptical distributions, the model implied conditional moment restriction is redundant given the statistical conditional moment restriction if and only if the joint distribution of the returns and factors exhibits no multivariate excess kurtosis.*

Proof. By Lemma 1,

$$\begin{aligned} \Sigma_{MS,t} &= E_t[F_{t+1}' \otimes u_{t+1} u_{t+1}'] = E_t \left[F_{t+1}' \otimes \text{Var} \left[u_{t+1} | \tilde{I}(t) \right] \right] \\ &= E_t \left[F_{t+1}' \rho(F_{t+1}) \right] \otimes \Sigma_{uu,t} \\ &= E_t \left[F_{t+1}' \right] \otimes \Sigma_{uu,t} \end{aligned}$$

From Appendix D. of Peñaranda and Sentana (2012), the fourth moment of the elliptical

distribution is:

$$\begin{aligned}
E_t [(F_{t+1} - E_t[F_{t+1}])(F_{t+1} - E_t[F_{t+1}])' \otimes u_{t+1}u_{t+1}'] &= (\kappa + 1)(\Sigma_{FF,t} \otimes \Sigma_{uu,t}) \\
&+ (\kappa + 1)(\Sigma_{Fu,t} \otimes \Sigma_{uF,t})K_{N,K} \\
&+ (\kappa + 1)\text{vec}(\Sigma_{Fu,t})\text{vec}(\Sigma_{Fu,t})' \\
&= (\kappa + 1)(\Sigma_{FF,t} \otimes \Sigma_{uu,t})
\end{aligned}$$

where we assume that the multivariate excess kurtosis coefficient κ of the conditional elliptical distribution is time-invariant, and $K_{N,K}$ is the $NK \times NK$ commutation matrix that, for any $N \times K$ matrix A , transforms $\text{vec}(A)$ into $\text{vec}(A')$.

It also follows from Lemma 1 that the fourth moment of the elliptical distribution is equal to:

$$E_t[F_{t+1}F_{t+1}' \otimes u_{t+1}u_{t+1}'] - E_t[F_{t+1}]E_t[F_{t+1}'] \otimes \Sigma_{uu,t}$$

Combining the previous two equations:

$$\begin{aligned}
\Sigma_{SS,t} &= E_t[F_{t+1}F_{t+1}' \otimes u_{t+1}u_{t+1}'] \\
&= (\kappa + 1)(\Sigma_{FF,t} \otimes \Sigma_{uu,t}) + E_t[F_{t+1}]E_t[F_{t+1}'] \otimes \Sigma_{uu,t} \\
&= \kappa(\Sigma_{FF,t} \otimes \Sigma_{uu,t}) + E_t[F_{t+1}F_{t+1}'] \otimes \Sigma_{uu,t}
\end{aligned}$$

Now if we impose the redundancy condition (2.18) that $\Gamma_{M,t}\Gamma_{S,t}^{-1} = \Sigma_{MS,t}\Sigma_{SS,t}^{-1}$ a.s., we arrive at:

$$\begin{aligned}
E_t[F_{t+1}']E_t[F_{t+1}F_{t+1}']^{-1} \otimes I_N &= E_t [F_{t+1} \otimes u_{t+1}u_{t+1}'] E_t [F_{t+1}F_{t+1}' \otimes u_{t+1}u_{t+1}']^{-1} \\
&= (E_t[F_{t+1}'] \otimes \Sigma_{uu,t}) [\kappa(\Sigma_{FF,t} \otimes \Sigma_{uu,t}) + E_t[F_{t+1}F_{t+1}'] \otimes \Sigma_{uu,t}]^{-1}.
\end{aligned}$$

Since we have assumed at the outset that $E_t[F_{t+1}] \neq 0$, this equation can only hold if the multivariate excess kurtosis coefficient, κ , is zero. \square

2.3 Optimally Weighted Moment Estimators

In an unconditional, one factor, elliptically symmetric setting, we characterize how the efficient GMM estimator of beta using the full set of moments (2.6) optimally weights two estimators: a method of moments estimator using the just-identified equilibrium pricing condition; and a method of moments estimator using the just-identified statistical moment condition.⁵ This allows us to develop further intuition for why the equilibrium pricing condition is useful for improving our beta estimates when returns are conditionally heteroskedastic given the factors. Let us define the GMM estimator as a weighted average estimator:

$$\hat{\beta}_{GMM} = \omega \hat{\beta}_S + (1 - \omega) \hat{\beta}_M$$

⁵It is a general result that efficient GMM optimally weights estimators based on subsets of just-identified moments (e.g. Chen, Jacho-Chávez, and Linton (2016)).

where the method of moments estimators are:

$$\hat{\beta}_S = \frac{\sum_{t=1}^T R_t^e F_t}{\sum_{t=1}^T F_t^2}, \quad \text{and} \quad \hat{\beta}_M = \frac{\sum_{t=1}^T R_t^e}{\sum_{t=1}^T F_t}.$$

and ω is an infeasible weight that we will choose to minimize the asymptotic variance of $\hat{\beta}_{GMM}$. It is a straightforward exercise to show that:

$$\sqrt{T}(\hat{\beta}_{GMM} - \beta) = \omega \left(\frac{1}{T} \sum_{t=1}^T F_t^2 \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t F_t + (1 - \omega) \left(\frac{1}{T} \sum_{t=1}^T F_t \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t$$

and that under a joint Central Limit Theorem (CLT):

$$\left[\begin{array}{c} T^{-1/2} \sum_{t=1}^T u_t \\ T^{-1/2} \sum_{t=1}^T u_t F_t \end{array} \right] \xrightarrow{d} \mathcal{N} \left[0, \begin{pmatrix} E[u_t^2] & E[u_t^2 F_t] \\ E[u_t^2 F_t] & E[u_t^2 F_t^2] \end{pmatrix} \right]$$

the asymptotic variance of $\sqrt{T}(\hat{\beta}_{GMM} - \beta)$ is (by Slutsky's lemma):

$$Avar(\hat{\beta}_{GMM}) = \omega^2 \frac{E[u_t^2 F_t^2]}{E[F_t^2]^2} + (1 - \omega)^2 \frac{E[u_t^2]}{E[F_t]^2} + 2\omega(1 - \omega) \frac{E[u_t^2 F_t]}{E[F_t^2]E[F_t]}.$$

We now derive the optimal weight that minimizes $Avar(\hat{\beta}_{GMM})$.

Proposition 5. *If the risk factors and regression errors are elliptically symmetric then the weight ω^* that minimizes $Avar(\hat{\beta}_{GMM})$ is:*

$$\begin{aligned} \omega^* &= \frac{1}{1 + CV(F_t)^2 \left(\frac{E[u_t^2 F_t^2]}{E[u_t^2]E[F_t^2]} - 1 \right)} \\ &= \frac{E[F_t^2]}{E[F_t^2] + \kappa E[F_t]^2} \end{aligned}$$

where $CV(F_t) = \frac{E[F_t]}{\sqrt{Var(F_t)}}$ is the coefficient of variation (Sharpe ratio) of the risk factor and κ is the multivariate excess kurtosis coefficient of the joint distribution of the risk factors and regressions errors.

Proof. From the first order condition and elliptical symmetry:

$$\omega^* \left[\frac{E[u_t^2 F_t^2]}{E[F_t^2]^2} - \frac{E[u_t^2]}{E[F_t]^2} + \frac{E[u_t^2]}{E[F_t]^2} - \frac{E[u_t^2]}{E[F_t]^2} \right] = \left[\frac{E[u_t^2]}{E[F_t]^2} - \frac{E[u_t^2]}{E[F_t]^2} \right]$$

Straightforward algebra leads to:

$$\frac{1}{\omega^*} = 1 + \frac{E[F_t]^2}{Var[F_t]} \left[\frac{E[u_t^2 F_t^2]}{E[u_t^2]E[F_t^2]} - 1 \right].$$

Using a result from the proof of Proposition 4 that $E[u_t^2 F_t^2] = \kappa Var(F_t)E[u_t]^2 + E[F_t^2]E[u_t^2]$,

this simplifies further to:

$$\frac{1}{\omega^*} = 1 + \kappa \frac{E[F_t]^2}{E[F_t^2]}.$$

□

We learn from Proposition 5 that the equilibrium pricing condition becomes more informative as both the the Sharpe ratio of the risk factor and the degree of dependence between u_t^2 and F_t^2 increase. As in Proposition 3, conditional homoskedasticity of the regression residuals delivers the zero linear dependence between u_t^2 and F_t^2 that is sufficient for redundancy of the equilibrium pricing condition.

We conclude this section by characterizing the optimal penalty for the generalized PCA loadings estimator of Lettau and Pelger (2018) for a strong factor model with elliptically distributed returns and risk factors.⁶

Proposition 6. *If the returns and risk factors are jointly elliptically distributed, then the optimal tuning parameter for the generalized PCA loadings estimator of Lettau and Pelger (2018), γ , is equal to the multivariate excess kurtosis coefficient of the elliptical joint distribution, κ .*

Proof. From Theorem 1 of Lettau and Pelger (2018), the asymptotic variance of their generalized PCA loadings estimator in a one factor model is:

$$\frac{E[F_{t+1}^2 u_{t+1}^2]}{[E[F_{t+1}^2] + \gamma E[F_{t+1}]^2]^2} + \frac{\gamma^2 E[F_{t+1}]^2 E[u_{t+1}^2]}{[E[F_{t+1}^2] + \gamma E[F_{t+1}]^2]^2} + \frac{2\gamma E[F_{t+1}] E[F_{t+1} u_{t+1}^2]}{[E[F_{t+1}^2] + \gamma E[F_{t+1}]^2]^2}$$

where γ is a penalty on the equilibrium pricing condition. From Proposition 5, the asymptotic variance of our efficient GMM estimator is:

$$\omega^{*2} \frac{E[F_{t+1}^2 u_{t+1}^2]}{E[F_{t+1}^2]^2} + (1 - \omega^*)^2 \frac{E[u_{t+1}^2]}{E[F_{t+1}]^2} + 2\omega^*(1 - \omega^*) \frac{E[F_{t+1} u_{t+1}^2]}{E[F_{t+1}^2] E[F_{t+1}]}.$$

where $\omega^* = \frac{E[F_{t+1}^2]}{E[F_{t+1}^2] + \kappa E[F_{t+1}]^2}$. It is immediately seen by a direct comparison of these two formulas that the optimal tuning parameter is $\gamma = \kappa$. □

Lettau and Pelger (2018) show in their Lemma 1 that if $E[F_{t+1}] \neq 0$, then it is inefficient to use the centered covariance matrix for estimating loadings. Since the choice of $\gamma = -1$ coincides with using the centered covariance matrix and the multivariate excess kurtosis coefficient is bounded below as $\kappa \geq -2/(M+2)$, we actually find that efficient GMM will never coincide with unrestricted ordinary least squares. Furthermore, Lettau and Pelger (2018) show in their corollary 1 that the optimal tuning parameter is $\gamma = 0$ when the regression errors are normal. We have learned in this paper that the optimal tuning parameter is zero

⁶With large N, T asymptotics the asymptotic distribution of the generalized PCA loadings estimator is derived as if the factors are observable. This allows us to apply the theory developed here for the estimation of beta pricing models with observable factors to the unobservable factor setting considered in Lettau and Pelger (2018).

when the regressions errors are normal because the normal distribution exhibits no multivariate excess kurtosis; a necessary and sufficient for the equilibrium pricing condition to be redundant.

2.4 Numerical Evidence

We conclude our analysis on the efficient estimation of beta pricing models with a brief simulation based study. We simulate 2,500 draws of sample size $T = 7,500$ of one return from the return generating process:

$$R_{t+1}^e = \beta F_{t+1} + u_{t+1}$$

where (F_{t+1}, u_{t+1}) is jointly t-distributed with mean $(E[F_{t+1}], E[u_{t+1}]) = (0.5, 0.0)$, variances $Var(F_{t+1}) = Var(u_{t+1}) = 1$, and covariance $Cov(F_{t+1}, u_{t+1}) = 0$, and the true beta coefficient is 1.0. We vary the degree of freedom parameter, ν , of the joint t-distribution from $\nu = 4.2$ to $\nu = \infty$ in order to generate a range of the multivariate excess kurtosis coefficient, $\kappa = 2/(\nu - 4)$, from 10 to 0. The full set of moment conditions is:

$$E \begin{bmatrix} \Psi_{t+1}^M(\beta) \\ \Psi_{t+1}^S(\beta) \end{bmatrix} = E \begin{bmatrix} R_{t+1}^e - \beta F_{t+1} \\ (R_{t+1}^e - \beta F_{t+1})F_{t+1} \end{bmatrix} = 0.$$

We calculate four different beta estimates: the efficient GMM estimator using the full set of moment restrictions and the identity matrix as an initial weight matrix; a method of moments estimator based on $E[\Psi_{t+1}^M(\beta)] = 0$; a method of moments estimator based on $E[\Psi_{t+1}^S(\beta)] = 0$ (regression through the origin); and an ordinary least squares estimator with unrestricted intercept.

In Table 1 of Appendix A, we present the efficiency of each method of moments estimator relative to the efficient GMM estimator, as well as the skewness and kurtosis of each estimators' sampling distribution. The first takeaway is that the uncentered statistical method of moments estimator always delivers an efficiency gain relative to regression with an intercept. The more interesting takeaway, however, is that the efficient GMM estimator is generally more efficient than using the uncentered statistical moment on its own. Also, as expected from our analysis of redundancy, we clearly see that the efficiency improvements disappear as the degree of conditional heteroskedasticity weakens.⁷

3 Conditionally Affine SDF Models and the Local GMM

A caveat with the aforementioned analysis is the non-trivial risk of model misspecification due to the parametric modeling of conditional betas. Although conditional asset pricing models should, in principle, perform better in terms of minimizing pricing errors than their

⁷From equation (14) of MacKinlay and Richardson (1991):

$$Var(R_{t+1}|F_{t+1}) = \left[\frac{\nu - 2}{\nu - 4} \right] \left[1 + \frac{(F_{t+1} - E[F_{t+1}])^2}{(\nu - 2)Var(F_{t+1})} \right] Var(u_{t+1})$$

unconditional counterparts, it is well known from Ghysels (1998) and Brandt and Chapman (2018) that this is often not the case. Economic theory simply does not provide any guidance for how the dynamics of conditional betas and conditional risk prices should be modeled.

In this section, we will briefly summarize a framework for specifying and evaluating conditionally affine stochastic discount factor models with and without state-dependent risk prices. For the state-dependent case, we will model the risk prices in two ways: first, as affine parametric functions of a scalar conditioning variables; second, as an unknown nonparametric function of a scalar conditioning variable. In order to estimate our models, we will also review the local GMM approaches of Lewbel (2007) and Gospodinov and Otsu (2012). Each approach works directly with the conditional moment restrictions by formulating a GMM-like quadratic form of kernel estimates of conditional moments.

3.1 Asset Pricing Framework

Consider a vector of gross returns in \mathbb{R}^n at date $t + 1$, $R_{t+1} = (R_{1,t+1}, \dots, R_{n,t+1})'$, that we assume belong to the linear space $L^2[I(t + 1)]$ of real random variables with finite second moment and are measurable with respect to an information set $I(t + 1)$. We know from Hansen and Richard (1987) that under the assumption of no-arbitrage (and some regularity conditions) there exists a Stochastic Discount Factor (SDF) process, $M_{t,t+1} \in L^2[I(t + 1)]$, that at any date $t = 1, 2, \dots$, satisfies the conditional no-arbitrage restrictions:

$$E[M_{t,t+1}R_{t+1}|I(t)] = 1_n. \quad (3.1)$$

If we assume the existence of a risk-free asset, then it follows from (3.1) that the risk-free rate, r_t^f , is:

$$r_t^f = -E[M_{t,t+1}|I(t)]^{-1}. \quad (3.2)$$

Applying the definition of a conditional covariance, we see that the economic content of (3.1) is that the expected excess return of any asset i , $R_{i,t+1}^e = R_{i,t+1} - r_t^f$, is governed by it's conditional covariance with the stochastic discount factor process:

$$E[R_{i,t+1}^e|I(t)] = -r_t^f Cov[M_{t,t+1}, R_{i,t+1}|I(t)]. \quad (3.3)$$

Although (3.3) implies that all expected excess returns can be explained by one systematic risk factor, it is typically the case that researchers in empirical finance specify linear models of excess returns with multiple risk factors. This can be justified theoretically if we assume that the stochastic discount factor process is spanned by a small set of risk factors. Following Nagel and Singleton (2011), we define a conditionally affine parametric SDF model as:

$$m_{t,t+1}(F_{t+1}; \theta) = \phi_t^0(\theta) + \phi_t^F(\theta)'F_{t+1} \quad (3.4)$$

where F_{t+1} is a K -vector of risk factors, and

$$(\phi_t^0, \phi_t^{F'}) = (\phi_t^0, \phi_t^1, \dots, \phi_t^K)$$

is a $(K + 1)$ -vector of state-dependent risk price parameters that we assume belong to the

econometrician’s information set.⁸ In addition, we assume that state-dependence is captured solely through a conditioning variable z_t :

$$m_{t,t+1}(Y_{t+1}; \theta) = \phi^0(z_t; \theta) + \phi^F(z_t; \theta)' F_{t+1} \quad (3.5)$$

where:

$$(\phi^0(z_t; \theta), \phi^{F'}(z_t; \theta)) = (\phi^0(z_t; \theta), \phi^1(z_t; \theta), \dots, \phi^K(z_t; \theta))$$

is a $(K + 1)$ –vector of parametric risk price functions and $Y_{t+1} = (F_{t+1}, z_t)$. This modeling approach for state-dependent risk prices was popularized by Jagannathan and Wang (2002), Lettau and Ludvigson (2001), and Santos and Veronesi (2005). They all consider linearized consumption-based SDF models wherein the risk price functions are affine in a scalar conditioning variable z_t .⁹

3.2 Pseudo-true SDFs

The main assertion of Nagel and Singleton (2011) is that “considerable latitude remains for enhanced model discrimination by more efficiently exploiting the economic content of the dynamic pricing relation” given by (3.1). They illustrate this point by estimating and testing the linearized consumption-based SDF models of Jagannathan and Wang (2002), Lettau and Ludvigson (2001), Santos and Veronesi (2005) with unconditional and conditional moment restrictions. The primary result of their empirical work is that the “small average pricing errors that are obtained when estimation is based on unconditional moment restrictions hide enormous time-variation in conditional pricing errors.”

We rationalize this result by confirming empirically that the choice of objective function is important when evaluating potentially misspecified asset pricing models. From Hall and Inoue (2003) we know that the definition and interpretation of the pseudo-true parameters of a moment-based model depend directly on the choice of weight matrix used for its evaluation. Also, as argued by Antoine et al. (2018), we know that a properly defined objective function will deliver a pseudo-true SDF that does the best job in a predefined sense.

The unconditional GMM objective with identity weight matrix delivers a pseudo-true SDF that does the best job minimizing unconditional pricing errors since, by definition, it is equal to the sum of squared pricing errors:

$$\begin{aligned} Q^{gmm}(\theta) &= E[m_{t,t+1}(Y_{t+1}; \theta) R_{t+1}^e]' E[m_{t,t+1}(Y_{t+1}; \theta) R_{t+1}^e] \\ &= \sum_{i=1}^N E[e_i(I(t); \theta)]^2 \end{aligned} \quad (3.6)$$

where we define an N –vector of conditional pricing errors, $e(I(t); \theta)$, as:

$$e(I(t); \theta) = (E[m_{t,t+1}(Y_{t+1}; \theta) R_{1,t+1}^e | I(t)], \dots, E[m_{t,t+1}(Y_{t+1}; \theta) R_{N,t+1}^e | I(t)])'$$

⁸We use the terminology risk price parameter here to avoid confusion with risk prices from the equivalent beta representation of expected excess returns. We will use the terminology interchangeably, hereafter.

⁹Their models differ based on the choice of z_t : the corporate bond spread (def) for Jagannathan and Wang (2002), the consumption-wealth ratio (cay) for Lettau and Ludvigson (2001), and the labor-income ratio (yc) for Santos and Veronesi (2005).

If we are instead interested in obtaining both small average pricing errors and stable conditional pricing errors, then a suitable objective function should work directly with the dynamic pricing restrictions (3.1). The following GMM-like objective function delivers exactly that:

$$\begin{aligned}
Q^{local}(\theta) &= E \left[E[m_{t,t+1}(Y_{t+1}; \theta) R_{t+1}^e | I(t)]' E[m_{t,t+1}(Y_{t+1}; \theta) R_{t+1}^e | I(t)] \right] \\
&= E \left[\sum_{i=1}^N E[m_{t,t+1}(Y_{t+1}; \theta) R_{i,t+1}^e | I(t)]^2 \right] \\
&= \sum_{i=1}^N [Var[e_i(I(t); \theta)] + E[e_i(I(t); \theta)]^2].
\end{aligned} \tag{3.7}$$

The objective function, $Q^{local}(\theta)$, places the emphasis directly on both unconditional pricing errors, $E[e_i(I(t); \theta)]$, and the volatility of conditional pricing errors, $Var[e_i(I(t); \theta)]$, since by definition it minimizes the mean-squared conditional pricing errors. As a nonparametric benchmark, we also consider a date-by-date GMM objective function composed of the sum of squared conditional pricing errors:

$$\begin{aligned}
Q^{np}(I(t), \phi_t) &= E[m_{t,t+1}(Y_{t+1}; \phi_t) R_{t+1}^e | I(t)]' E[m_{t,t+1}(Y_{t+1}; \phi_t) R_{t+1}^e | I(t)] \\
&= \sum_{i=1}^N e_i(I(t); \phi_t)^2
\end{aligned} \tag{3.8}$$

whose minimization delivers a nonparametric risk price parameter ϕ_t . This objective was originally employed by Roussanov (2014) for the estimation and testing of conditionally affine asset pricing models.

An estimation procedure based on the objective functions $Q^{local}(\theta)$ and $Q^{np}(\phi_t)$ must deal with the fact that they are comprised of conditional moments that need to be nonparametrically estimated. In the next section we review the local GMM estimators of Gospodinov and Otsu (2012) and Lewbel (2007) for the estimation of parametric and functional parameters defined by conditional moment restrictions. Both of these estimators work directly with local conditional moment restrictions. An alternative estimation procedure could be based off of the global sieve minimum distance of Ai and Chen (2003, 2007). This approach to estimating possibly misspecified SDFs models with conditional moment restrictions is taken by Chen et al. (2014).

3.3 Local GMM

We consider the estimation of an unknown state-invariant parameter, θ_0 , defined by uniform conditional moment restrictions:

$$E[g(Y_{t+1}; \theta_0) | Z_t] = 0_q \quad \forall z \in \mathcal{Z}, P - a.s. \tag{3.9}$$

where Z_t is a $d \times 1$ dimensional state variable process, $g(\cdot)$ is a known $q \times 1$ vector of moment functions where $q \geq p$, and the unknown parameter vector of interest θ_0 is in the set

$\Theta \subset \mathbb{R}^p$. Estimation of θ_0 by the local GMM of Gospodinov and Otsu (2012) proceeds by minimizing the following objective composed of an average of T GMM-like quadratic forms of the estimated local moments:

$$Q_T^{local}(\theta) = \frac{1}{T} \sum_{t=1}^T Q_T^{np}(z_t, \theta) \quad (3.10)$$

where W_T is a (possibly) state-dependent positive definite matrix of size $q \times q$

$$Q_T^{np}(z_t, \theta) = g_T(z_t, \theta)' W_T g_T(z_t, \theta) \quad (3.11)$$

and the estimated local moments are:

$$g_T(z_t, \theta) = \hat{E}[g(Y_{t+1}; \theta) | Z_t = z_t]. \quad (3.12)$$

Following Lewbel (2007), Gagliardini, Gourieroux, and Renault (2011), and Gospodinov and Otsu (2012), we restrict attention to the local constant estimator:

$$\begin{aligned} \hat{E}[g(Y_{t+1}; \theta) | Z_t = z_t] &= \frac{\sum_{s=1}^T g(y_{s+1}; \theta) K\left(\frac{z_s - z_t}{h_T}\right)}{\sum_{j=1}^T K\left(\frac{z_j - z_t}{h_T}\right)} \\ &\equiv \sum_{s=1}^T \omega_{s,t} g(y_{s+1}; \theta). \end{aligned} \quad (3.13)$$

With an appropriate choice of weight matrix, local GMM belongs to the class of local generalized empirical likelihood estimators developed by Kitamura, Tripathi, and Ahn (2004), Smith (2007), and Antoine, Bonnal, and Renault (2007). Gospodinov and Otsu (2012) studies the first- and higher- asymptotic properties of local GMM for stationary and geometrically ergodic Markov processes.

We also consider the estimation of a functional (nonparametric) parameter, $\gamma_t \equiv \gamma(Z_t)$, defined by uniform conditional moment restrictions:

$$E[g(Y_{t+1}; \gamma(Z_t)) | Z_t] = 0_q \quad \forall z \in \mathcal{Z}, P - a.s. \quad (3.14)$$

Estimation of γ_t by the local GMM of Lewbel (2007) proceeds by the date-by-date minimization of the quadratic form:

$$Q_T^{np}(z_t, \gamma(z_t)) = g_T(z_t, \gamma(z_t))' W_T g_T(z_t, \gamma(z_t)). \quad (3.15)$$

The local GMM of Lewbel (2007) can be seen as an extension of the nonparametric conditional method of moments introduced by Brandt (1999) that allows for overidentification. Lewbel (2007) develops the first-order asymptotic theory in the i.i.d. case. Nevertheless, it can easily be shown that the first-order asymptotic theory extends to strictly stationary and geometrically strongly mixing data since this estimator is nested in the kernel moment

estimator of Gagliardini et al. (2011).¹⁰

4 Conditional Fama-French Three Factor Models

In this section we employ the three objective functions (3.6)-(3.8) to estimate various conditional Fama-French three factor models. The models considered differ based on the choice of conditioning variable and the specification of state-dependent risk prices. Wang (2003) was the first to estimate a fully nonparametric stochastic discount factor formulation of the Fama-French three factor model. We differ from his work by allowing for overidentification.

The empirical work that follows is also motivated by Adrian, Crump, and Moench (2015). They also estimate conditional Fama-French models with parametric state-dependent risk prices, and find that time variation in risk prices is helpful for minimizing squared pricing errors. We use a subset of their test assets and conditioning variables to corroborate their findings. In contrast with Adrian et al. (2015), the primary focus of our empirical work is to illustrate that objective functions based on conditional moments deliver pseudo-true SDFs that optimally balance stability in conditional pricing errors with the level of unconditional pricing errors.

4.1 Data

We use the returns on the ten size-sorted portfolios for US equities from Kenneth French's data library as test assets. We compute excess returns over the one-month Treasury bill yield obtained from the Center for Research in Securities Prices (CRSP). The risk factors of the Fama and French (1993) three factor model are the excess return on the market portfolio (MKT), the small minus big portfolio (SMB), and the high minus low portfolio (HML). Since the risk factors are portfolios, we include them in the vector of test assets.

The ten-year Treasury yield (TSY10), the spread between the constant maturity ten-year Treasury yield and the three-month Treasury bill (TERM), and the Baa corporate bond yield relative to the constant maturity ten-year Treasury yield (DEF) are separately considered as state variables. Adrian et al. (2015) also model time variation in risk prices with TSY10 and TERM as factors. The corporate bond spread serves as a proxy for default risk, and it is the state variable used in the consumption-based CAPM model of Jagannathan and Wang (1996). The ten-year Treasury yield and the three-month Treasury bill rate are obtained from the Federal Reserve Statistical Release H.15. The corporate bond spread is obtained from the Federal Reserve Economic Data (FRED). The sample period is 1964:01-2017:12 minus the 2008:09-2009:04 financial crisis, for a total of $T = 639$ observations.

4.2 Models and Estimation

We consider a stochastic discount factor formulation of the Fama-French three factor model:

$$m_{t,t+1}(f_{t+1}; \phi_t^f(z_t)) = 1 - \phi_t^{f'}(z_t)f_{t+1} \quad (4.1)$$

¹⁰One might also consider semi-parametric specifications that combine state-invariant and functional parameters. We sketch an extension of semi-parametric local GMM in Appendix B for moment functions with linearly separable parameters.

where $f_{t+1} = (mkt_{t+1}, smb_{t+1}, hml_{t+1})'$ is the vector of risk factors,

$$\phi_t^f(z_t) = (\phi_t^{mkt}(z_t), \phi_t^{smb}(z_t), \phi_t^{hml}(z_t))'$$

is a vector of risk price parameters, and z_t is a scalar conditioning variable. We consider separately three conditioning variables (TSY10, TERM, and DEF) in order to avoid the well known curse of dimensionality that comes with nonparametric estimation of conditional moments. Lastly, we normalize the intercept to one since it is not identified from a vector of excess returns.

We base our estimation on the local conditional moment restrictions:

$$E[m_{t,t+1}(f_{t+1}; \phi_t(z))r_{t+1}^e | z_t = z] = 0_N \quad (4.2)$$

where r_{t+1}^e is the $N \times 1$ vector of test assets and 0_N is an $N \times 1$ vector of zeros. We estimate three stochastic discount factor models for each conditioning variable. First, we consider an unconditional asset pricing model wherein the risk prices are state-invariant parameters:

$$m_{t,t+1}^u(f_{t+1}; \phi) = 1 - \phi_f' f_{t+1} \quad (4.3)$$

where $\phi_f = (\phi_{mkt}, \phi_{smb}, \phi_{hml})'$. Second, we consider a conditionally affine scaled multi-factor model, wherein state-dependent risk prices are modeled as affine functions of the conditioning variable:

$$m_{t,t+1}^{tv}(f_{t+1}, z_t; \phi) = 1 - (\phi_f + \phi_{z,f} z_t)' f_{t+1} \quad (4.4)$$

We refer to the coefficients $\phi_f = (\phi_{mkt}, \phi_{smb}, \phi_{hml})$ as factor coefficients, and the coefficients $\phi_{z,f} = (\phi_{z,mkt}, \phi_{z,smb}, \phi_{z,hml})$ as scaled factor coefficients. Third, we consider a nonparametric conditionally affine stochastic discount factor:

$$m_{t,t+1}^{np}(f_{t+1}; \phi_t(z_t)) = 1 - \phi_t^f(z_t)' f_{t+1} \quad (4.5)$$

wherein the state-dependent risk prices $\phi_t^f(z_t) = (\phi_t^{mkt}(z_t), \phi_t^{smb}(z_t), \phi_t^{hml}(z_t))$ are modeled as unknown functional parameters. We estimate the parametric specifications (4.3) and (4.4) by unconditional GMM and the local GMM of Gospodinov and Otsu (2012). We estimate the nonparametric model (4.5) by the local GMM of Lewbel (2007).¹¹

We have to choose a kernel function and bandwidth parameters for the implementation of local GMM. We use the Epanechnikov kernel function. For the bandwidth we choose one per state variable model for simplicity, even though there are $N = 13$ conditional moments composed of conditional mean returns and nonparametric betas. We choose it by taking the median bandwidth from least-squares cross-validation for the nonparametric beta estimates. Fortunately, cross-validation delivered bandwidths for all of the nonparametric betas in a narrow range. Roussanov (2014) treats the bandwidth as an additional parameter of his local linear GMM objective function, whereas Cai et al. (2015) choose it to minimize the

¹¹In Appendix C, we also consider the estimation of the parametric specifications with the smooth minimum distance (SMD) estimator of Laverne and Patilea (2013). This estimator was recently studied for estimating misspecified asset pricing models by Antoine et al. (2018). The primary advantages of SMD relative to local GMM are also summarized.

weighted integrated asymptotic means squared error of their local linear estimator.

4.3 Estimation Results

Tables 2 and 3 in Appendix A report the point estimates, standard errors, and 95% confidence intervals for the risk price parameters in models (4.3) and (4.4), respectively.¹² The main takeaway here is that local GMM delivers more stable estimates of the factor coefficients in the presence of the scaled factors. This can be attributed to the use of conditioning information in the objective function which serves to aid in the identification of the model parameters. We also find some evidence for significant risk prices for the market risk factor and the HML factor. We do not find any evidence in favor of the SMB factor.

The second point to note is that all of the confidence intervals in Table 3 are wide and do not provide any evidence in favor of state-dependence in the risk price parameters. Since there are six unknown parameters and the ten size sorted portfolios are highly correlated, it may be the case that the parameters are not strongly identified by the thirteen moment conditions.

Table 4 reports summary statistics for the nonparametric risk price estimates. The time-series average of the estimates coincides closely with the constant risk price estimates. This is sensible because in the event that the true risk prices were constant, the time-series average of the nonparametric estimates would be a \sqrt{T} -consistent estimator. The time-series standard deviations of the estimates are all relatively small. This can be attributed to over-smoothing of the conditional moments. Nevertheless, the range of the nonparametric estimates is non-trivial. Also, the average of the point-wise standard errors for the nonparametric estimates are larger than the standard errors for the constant risk price estimates by GMM, but smaller than the standard errors for the constant risk estimates by local GMM.

Figure 1 plots the functional risk price estimates with 95% point-wise confidence intervals that ignore the bias component. Due to the poor boundary behavior of the Nadaraya-Watson estimator, it is clear that the functional parameters are not estimated well at the boundaries of the state space. Since the confidence intervals are point-wise, no statements can be made regarding the uniform pricing of the risk factors over the state domain. However, it can be seen that the high minus low risk price is significant, date-by-date, across all specifications.

Figure 2 compares the nonparametric risk prices with the affine state-dependent risk prices estimated by local GMM. There is preliminary evidence of nonlinearities in the risk prices. The market factor risk price is concave in TSY10, the small minus big risk price is concave in DEF, and the high minus low risk price is concave in TERM.

4.3.1 Average and Conditional Pricing Errors

The main findings of our empirical work can be found in Tables 5 - 8. Table 5 reports unconditional pricing errors of the constant risk price model (4.3) estimated by GMM and local GMM. As expected, the GMM objective delivers parameter estimates that produce the smallest unconditional pricing errors. This also holds in Table 6 for the state-dependent risk price models. The more interesting conclusion reached from Tables 5 and 6 is that, holding

¹²We do not report J-tests because the asymptotic distribution of the J-statistic is nonstandard for an inefficient choice of weight matrix. We leave this to later work.

Figure 1: Nonparametric Risk Price Estimates with 95% CIs

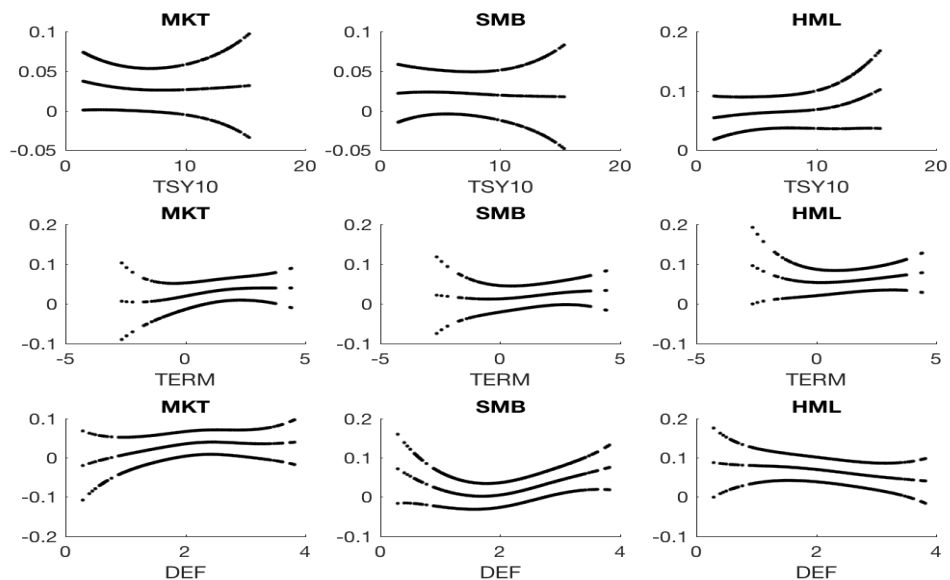


Figure 1 presents nonparametric risk price estimates for three models that differ based on the choice of state variables: TSY10, TERM, and DEF.

Figure 2: Nonparametric and Parametric Risk Price Estimates

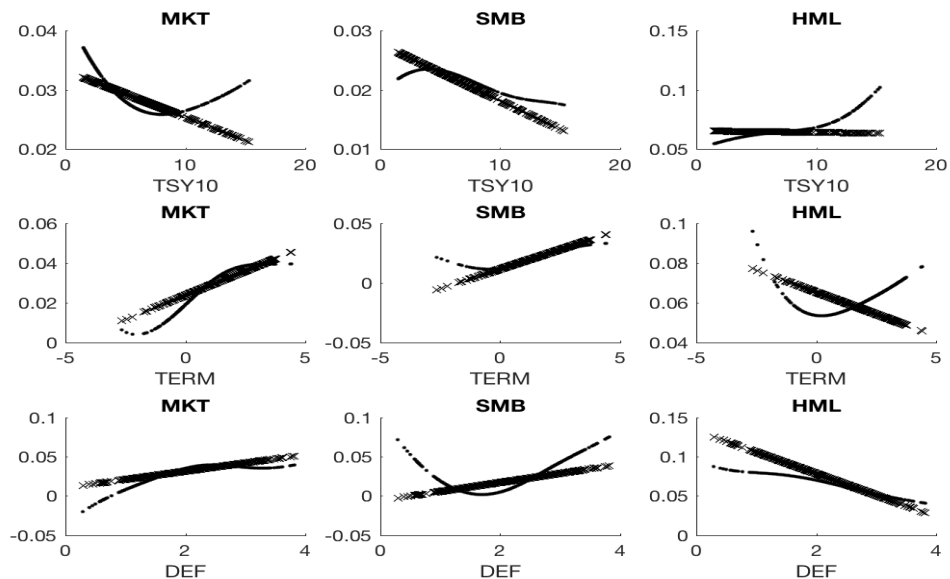


Figure 2 presents nonparametric and parametric risk price estimates for three models that differ based on the choice of state variable: TSY10, TERM, and DEF. The thick 'x' marked lines are the parametric risk price estimates.

the estimating procedure fixed, the introduction of state-dependence in the risk prices leads to a considerable reduction in the unconditional pricing errors. However, Table 8 confirms the critique of Nagel and Singleton (2011) that this reduction comes at the expense of volatile conditional pricing errors when the model is estimated with unconditional moments.

This is not a criticism of state-dependent risk price models though, but rather of the estimation procedure. It is clear from a comparison of Tables 7 and 8 that the state-dependent risk price models estimated by local GMM produce stable conditional pricing errors. Moreover, a comparison of Tables 6 and 8 illustrates that the nonparametric specification delivers a drastic reduction in the volatility of conditional pricing errors at minimal expense in terms of unconditional pricing errors.

A caveat for all of these findings is that we do not consider the sampling variation of the pricing errors. Future work should report confidence intervals for the summary statistics of the conditional pricing errors. Formal comparison of conditional asset pricing models has notably been considered by Gagliardini and Ronchetti (2018).

4.3.2 Stochastic Discount Factors

Since the SDF is the central object of interest for an asset pricing model, we conclude this section by presenting summary statistics of the pseudo-true SDFs. Table 9 presents the mean, variance, skewness, and kurtosis of the SDFs for the constant risk price models. We see that there is almost no discernible discrepancy between the SDFs of the constant risk price models due to the choice of estimation procedure. However, we do see in Table 10 that the SDFs of the state-dependent parametric models estimated by unconditional GMM clearly display too much volatility. The excessive volatility of the SDF translates into excessive volatility of the conditional pricing errors.

5 Conclusion

The primary contribution of this paper has been to explore the informational content of conditional no-arbitrage restrictions. This content was summarized by an asset pricing model's ability to improve the precision of beta estimates. However, in order to set focus on efficient estimation, we have maintained the strong assumption of time-invariant conditional betas. We have proceeded then to explore the trade offs involved in the specification of constant versus state-dependent parameter models. Nevertheless, in our exploration of this trade off, we have found ourselves at the other end of the modeling spectrum with fully nonparametric asset pricing models.

Future research should bridge the gap between these two extremes. A fruitful approach could make use of the large econometrics literature on generalized empirical likelihood (GEL) (e.g. Newey and Smith (2004)) and its localized versions developed by Kitamura, Tripathi, and Ahn (2004), Antoine, Bonnal, and Renault (2007), and Smith (2007). The primary lesson of this literature for asset pricing is that when conditional no-arbitrage restrictions overidentify the model's structural parameters, they may bring information about aspects of the return distribution not already captured by our model.

The framework of Gagliardini, Gourieroux, and Renault (2011) is particularly useful for asset pricing as it allows for both fixed and state-dependent parameters subject to both

uniform and local conditional no-arbitrage restrictions. It may allow us to make use of the overidentifying information in equilibrium pricing conditions to efficiently estimate nonparametric betas (risk prices) in the presence of parametrically specified risk prices (betas).

Another natural extension of this paper could follow by characterizing the optimal penalty for the generalized PCA of Lettau and Pelger (2018) when risk factors are weak (the primary situation studied by these authors). It is conjectured that the redundancy conditions of Breusch et al. (1999) could be extended to accommodate these situations.

Appendix A: Tables and Figures

A.1 Numerical Evidence

Table 1: Finite Sample Properties of Efficient GMM and the Method of Moments

		Relative Efficiency										
	ν	4.2	4.4	4.6	4.8	5	6	7	8	9	10	∞
Efficient GMM		-	-	-	-	-	-	-	-	-	-	-
Uncentered		2.04	1.83	1.75	1.69	1.35	1.08	1.05	1.03	1.01	1.01	1.00
Model		2.07	2.24	2.34	2.34	2.53	2.98	3.28	3.75	3.83	4.05	4.99
Centered		2.99	2.64	2.58	2.35	1.92	1.47	1.41	1.38	1.33	1.34	1.25

		Skewness										
	ν	4.2	4.4	4.6	4.8	5	6	7	8	9	10	∞
Efficient GMM		-0.06	0.00	0.00	-0.07	0.03	0.04	-0.08	-0.05	0.02	-0.07	-0.06
Uncentered		2.36	-1.49	1.10	1.50	-0.79	0.08	-0.08	-0.06	0.02	-0.04	-0.06
Model		0.01	-0.07	-0.06	0.03	0.00	0.06	-0.02	0.05	-0.04	-0.08	0.01
Centered		2.30	-1.16	1.09	1.11	-0.76	0.06	-0.05	-0.05	0.02	0.04	-0.06

		Kurtosis										
	ν	4.2	4.4	4.6	4.8	5	6	7	8	9	10	∞
Efficient GMM		2.94	3.15	3.13	3.11	2.90	2.90	2.95	3.05	3.11	2.94	2.94
Uncentered		55.96	53.95	33.64	61.58	11.57	2.90	2.99	3.02	3.08	2.96	2.94
Model		2.90	3.05	3.03	2.94	2.98	3.01	2.92	2.97	3.15	2.95	3.04
Centered		54.67	46.57	34.42	50.69	11.06	2.91	3.05	2.98	3.00	3.06	2.94

Table 1 presents the efficiency of the method of moments estimators relative to efficient GMM, as well as the skewness and kurtosis of each estimators' sampling distribution. We vary the degree of freedom parameter, ν , of the joint distribution of the returns and risk factors. The method of moments estimators are constructed from the uncentered statistical moment (uncentered), the asset pricing model moment (model), and the centered statistical moment (centered).

A.2 Fama-French Three Factor Models

Table 2: Estimation Results for Constant Risk Price Models

	GMM	LGMM		
		TSY10	TERM	DEF
$\hat{\phi}_{mkt}$	0.030 (0.010) [0.010,0.049]	0.029 (0.031) [-0.033,0.090]	0.040 (0.026) [-0.011,0.091]	0.046 (0.026) [-0.005,0.097]
$\hat{\phi}_{smb}$	0.021 (0.013) [-0.005,0.047]	0.024 (0.052) [-0.078,0.126]	0.014 (0.030) [-0.046,0.073]	0.000 (0.029) [-0.057,0.058]
$\hat{\phi}_{hml}$	0.066 (0.013) [0.040,0.092]	0.076 (0.143) [-0.204,0.355]	0.095 (0.081) [-0.064,0.255]	0.106 (0.075) [-0.042,0.254]

Table 2 presents point estimates, standard errors, and 95 percent confidence intervals for the constant risk price parameter model estimated by the generalized method of moments (GMM) and the local generalized method of moments (LGMM). The model is estimated by LGMM separately for each conditioning variable: TSY10, TERM, and DEF.

Table 3: Estimation Results for Parametric State-Dependent Risk Price Models

	TSY10			TERM			DEF
	GMM	LGMM	GMM	LGMM	GMM	LGMM	
$\hat{\phi}_{mkt}$	0.001 (0.213) [-0.417,0.419]	0.033 (0.094) [-0.151,0.217]	-0.165 (0.162) [-0.483,0.153]	0.024 (0.042) [-0.059,0.106]	0.030 (0.260) [-0.479,0.540]	0.010 (0.080) [-0.146,0.167]	
$\hat{\phi}_{smb}$	0.001 (0.260) [-0.508,0.509]	0.028 (0.039) [-0.048,0.103]	0.182 (0.153) [-0.118,0.482]	0.012 (0.035) [-0.057,0.080]	0.144 (0.236) [-0.317,0.606]	-0.006 (0.034) [-0.073,0.061]	
$\hat{\phi}_{hml}$	0.002 (0.197) [-0.384,0.388]	0.065 (0.422) [-0.761,0.892]	0.007 (0.101) [-0.190,0.204]	0.065 (0.090) [-0.112,0.243]	0.069 (0.413) [-0.740,0.878]	0.133 (0.279) [-0.413,0.679]	
$\hat{\phi}_{z,mkt}$	0.005 (0.032) [-0.059,0.068]	-0.001 (0.014) [-0.028,0.027]	0.123 (0.102) [-0.077,0.322]	0.005 (0.021) [-0.036,0.046]	0.000 (0.119) [-0.234,0.234]	0.011 (0.035) [-0.058,0.079]	
$\hat{\phi}_{z,smb}$	0.004 (0.039) [-0.074,0.081]	-0.001 (0.001) [-0.002,0.000]	-0.111 (0.108) [-0.323,0.101]	0.007 (0.008) [-0.009,0.022]	-0.062 (0.113) [-0.283,0.160]	0.012 (0.008) [-0.004,0.027]	
$\hat{\phi}_{z,hml}$	0.010 (0.032) [-0.052,0.073]	0.000 (0.066) [-0.129,0.129]	0.016 (0.064) [-0.108,0.141]	-0.004 (0.077) [-0.156,0.147]	0.000 (0.185) [-0.362,0.363]	-0.027 (0.117) [-0.256,0.201]	

Table 3 presents point estimates, standard errors, and 95 percent confidence intervals for the affine state-dependent risk price models estimated by the generalized method of moments (GMM) and the local generalized method of moments (LGMM). Each model differs based on choice of state variable: TSY10, TERM, and DEF.

Table 4: Summary Statistics of the Nonparametric Risk Price Estimates

	TSY10				TERM				DEF		
	$\hat{\phi}_{mkt}$	$\hat{\phi}_{smb}$	$\hat{\phi}_{hml}$	$\hat{\phi}_{mkt}$	$\hat{\phi}_{smb}$	$\hat{\phi}_{hml}$	$\hat{\phi}_{mkt}$	$\hat{\phi}_{smb}$	$\hat{\phi}_{hml}$	$\hat{\phi}_{mkt}$	$\hat{\phi}_{hml}$
mean	0.028	0.022	0.064	0.031	0.021	0.062	0.030	0.016	0.067	0.030	0.016
std.	0.002	0.001	0.005	0.005	0.004	0.003	0.006	0.006	0.004	0.006	0.006
min	0.026	0.018	0.056	0.011	0.015	0.059	0.012	0.011	0.054	0.012	0.011
max	0.035	0.023	0.094	0.039	0.029	0.071	0.039	0.042	0.078	0.039	0.042
avg. s.e.	0.015	0.019	0.020	0.014	0.019	0.019	0.015	0.020	0.020	0.015	0.020

Table 4 presents the time-series average, standard deviation, minimum, and maximum of the date-by-date parameter estimates. It also reports the average standard error across all date-by-date parameter estimates. Each model differs based on choice of state variable: TSY10, TERM, and DEF.

Table 5: Unconditional Pricing Errors for the Constant Risk Price Model

	GMM	LGMM		
		TSY10	TERM	DEF
size1	0.0139	0.0160	0.0044	0.1158
size2	-0.0549	-0.0431	-0.0704	0.0274
size3	0.0199	0.0345	-0.0109	0.0645
size4	-0.0162	0.0013	-0.0451	0.0228
size5	0.0134	0.0335	-0.0295	0.0176
size6	-0.0071	0.0160	-0.0555	-0.0249
size7	0.0107	0.0375	-0.0445	-0.0292
size8	-0.0241	0.0038	-0.0881	-0.0860
size9	-0.0517	-0.0214	-0.1211	-0.1400
size10	-0.0630	-0.0197	-0.1271	-0.1747
mkt-rf	0.1590	0.1956	0.1004	0.0809
smb	0.0206	0.0098	0.0943	0.2167
hml	0.0057	-0.0668	-0.1936	-0.2805
$\frac{1}{N} \sum_{i=1}^N e_i(\hat{\theta})^2$	0.0029	0.0038	0.0083	0.0162

Table 5 presents the unconditional pricing errors for each of the test assets and the average squared unconditional pricing error across the test assets for the constant risk price model estimated by the generalized method of moments (GMM) and the local generalized method of moments (LGMM).

Table 6: Unconditional Pricing Errors for the State-Dependent Risk Price Models

	TSY10					TERM					DEF	
	GMM	LGMM	NP	GMM	LGMM	NP	GMM	LGMM	NP	GMM	LGMM	NP
size1	0.0098	0.0326	0.0440	-0.0097	0.0020	-0.0043	-0.0084	0.0727	0.0938	-0.0084	0.0727	0.0938
size2	-0.0453	-0.0359	-0.0226	-0.0265	-0.0839	-0.0851	-0.0346	-0.0091	0.0063	-0.0346	-0.0091	0.0063
size3	0.0263	0.0411	0.0542	0.0343	-0.0133	-0.0106	0.0274	0.0568	0.0719	0.0274	0.0568	0.0719
size4	-0.0123	0.0051	0.0180	-0.0365	-0.0512	-0.0497	-0.0176	0.0226	0.0317	-0.0176	0.0226	0.0317
size5	0.0115	0.0360	0.0474	0.0020	-0.0217	-0.0203	0.0183	0.0363	0.0457	0.0183	0.0363	0.0457
size6	-0.0107	0.0161	0.0276	0.0020	-0.0388	-0.0365	-0.0145	0.0099	0.0208	-0.0145	0.0099	0.0208
size7	0.0083	0.0358	0.0483	0.0141	-0.0214	-0.0198	0.0062	0.0221	0.0292	0.0062	0.0221	0.0292
size8	-0.0297	0.0000	0.0110	-0.0287	-0.0569	-0.0588	-0.0231	-0.0260	-0.0171	-0.0231	-0.0260	-0.0171
size9	-0.0561	-0.0269	-0.0142	-0.0540	-0.0814	-0.0838	-0.0583	-0.0607	-0.0556	-0.0583	-0.0607	-0.0556
size10	-0.0564	-0.0402	-0.0276	-0.0539	-0.0929	-0.0998	-0.0572	-0.0900	-0.0921	-0.0572	-0.0900	-0.0921
mkt-rf	0.1541	0.1834	0.1946	0.1542	0.1275	0.1221	0.1588	0.1467	0.1485	0.1588	0.1467	0.1485
smb	0.0113	0.0164	0.0146	0.0275	0.0181	0.0220	0.0241	0.0819	0.0865	0.0241	0.0819	0.0865
hml	0.0038	0.0081	0.0056	0.0079	0.0483	0.0389	0.0076	-0.0449	-0.0043	0.0076	-0.0449	-0.0043
$\frac{1}{N} \sum_{i=1}^N e_i(\hat{\theta})^2$	0.0026	0.0033	0.0039	0.0027	0.0038	0.0038	0.0028	0.0041	0.0046	0.0028	0.0041	0.0046

Table 6 presents the unconditional pricing errors for each of the test assets and the average squared unconditional pricing error across the test assets for the state-dependent risk price models estimated by the generalized method of moments (GMM), the local generalized method of moments (LGMM), and the nonparametric local generalized method of moments (NP).

Table 7: Variance of Conditional Pricing Errors for the Constant Risk Price Model

	TSY10		TERM		DEF	
	GMM	LGMM	GMM	LGMM	GMM	LGMM
size1	0.0150	0.0121	0.0763	0.0579	0.1663	0.1025
size2	0.0058	0.0039	0.0614	0.0466	0.1530	0.0989
size3	0.0068	0.0049	0.0625	0.0506	0.1302	0.0844
size4	0.0044	0.0030	0.0673	0.0545	0.1250	0.0803
size5	0.0047	0.0034	0.0582	0.0471	0.0990	0.0627
size6	0.0079	0.0062	0.0705	0.0592	0.1227	0.0871
size7	0.0054	0.0039	0.0552	0.0453	0.1009	0.0697
size8	0.0085	0.0068	0.0406	0.0317	0.0891	0.0649
size9	0.0067	0.0052	0.0325	0.0255	0.0720	0.0573
size10	0.0036	0.0026	0.0241	0.0171	0.0373	0.0330
mkt-rf	0.0019	0.0013	0.0315	0.0231	0.0535	0.0426
smb	0.0004	0.0003	0.0094	0.0077	0.0291	0.0170
hml	0.0036	0.0032	0.0026	0.0022	0.0163	0.0268
$\frac{1}{N} \sum_{i=1}^N Var(\hat{e}(I(t); \hat{\theta}))$	0.0057	0.0044	0.0455	0.0360	0.0919	0.0636
$MSE(\hat{e}(I(t); \hat{\theta}))$	0.0086	0.0082	0.0484	0.0444	0.0947	0.0798

Table 7 presents the variance of the estimated conditional pricing errors for each of the test assets, the average variance across the test assets, and the MSE of the conditional pricing errors for the constant risk price model estimated by the generalized method of moments (GMM) and the local generalized method of moments (LGMM).

Table 8: Variance of Conditional Pricing Errors for the State-Dependent Risk Price Models

	TSY10			TERM			DEF		
	GMM	LGMM	NP LGMM	GMM	LGMM	NP LGMM	GMM	LGMM	NP LGMM
size1	0.1892	0.0014	0.00201	0.6773	0.0028	0.00028	1.0883	0.0188	0.00143
size2	0.1504	0.0007	0.00018	1.5135	0.0055	0.00145	1.0282	0.0059	0.00052
size3	0.1450	0.0012	0.00009	2.0061	0.0035	0.00032	0.8472	0.0048	0.00030
size4	0.1200	0.0018	0.00061	2.0374	0.0025	0.00008	0.7665	0.0058	0.00010
size5	0.1161	0.0011	0.00019	2.6216	0.0019	0.00015	0.6017	0.0041	0.00077
size6	0.1154	0.0014	0.00018	2.8237	0.0035	0.00206	0.5564	0.0050	0.00121
size7	0.0968	0.0008	0.00001	3.4516	0.0030	0.00028	0.4392	0.0037	0.00035
size8	0.1039	0.0015	0.00041	3.8949	0.0014	0.00032	0.3600	0.0048	0.00034
size9	0.0770	0.0014	0.00030	4.0676	0.0013	0.00014	0.2325	0.0044	0.00041
size10	0.0375	0.0006	0.00010	5.1231	0.0050	0.00073	0.0748	0.0069	0.00067
mkt-rf	0.0478	0.0012	0.00074	4.3563	0.0030	0.00011	0.1827	0.0040	0.00041
smb	0.0136	0.0002	0.00012	0.6749	0.0006	0.00013	0.3058	0.0122	0.00011
hml	0.0116	0.0023	0.00001	0.0490	0.0051	0.00002	0.0365	0.0029	0.00002
$\frac{1}{N} \sum_{i=1}^N Var(\hat{e}(I(t); \hat{\theta}))$	0.0942	0.0012	0.00038	2.5613	0.0030	0.00047	0.5015	0.0064	0.00051
$MSE(\hat{e}(I(t); \hat{\theta}))$	0.0968	0.0045	0.0042	2.5640	0.0068	0.0043	0.5043	0.0105	0.0051

Table 8 presents the variance of the estimated conditional pricing errors for each of the test assets, the average variance across the test assets, and the MSE of the conditional pricing errors for the state-dependent risk price models estimated by GMM, local GMM (LGMM), and nonparametric local GMM (NP).

Table 9: SDF Summary Statistics for the Constant Risk Price Model

	LGMM			
	GMM	TSY10	TERM	DEF
mean	0.954	0.950	0.939	0.934
std.	0.196	0.217	0.266	0.296
skewness	-0.458	-0.521	-0.463	-0.384
kurtosis	5.999	5.824	5.283	4.941

Table 9 presents the average, standard deviation, skewness, and kurtosis of the stochastic discount factor for the constant risk price model estimated by GMM and local GMM (LGMM).

Table 10: SDF Summary Statistics for the State-Dependent Risk Price Models

	TSY10				TERM				DEF					
	GMM	LGMM	NP	GMM	NP	GMM	LGMM	NP	GMM	LGMM	NP	GMM	LGMM	NP
mean	0.950	0.954	0.953	0.928	0.952	0.949	0.962	0.944	0.962	0.944	0.942	0.944	0.944	0.942
std.	0.217	0.193	0.193	0.677	0.193	0.197	0.236	0.220	0.236	0.220	0.208	0.220	0.220	0.208
skewness	-0.556	-0.468	-0.436	-0.131	-0.221	-0.307	0.237	-0.424	0.237	-0.424	-0.604	-0.424	-0.424	-0.604
kurtosis	9.973	5.746	5.560	9.157	6.152	6.280	5.266	5.299	5.266	5.299	7.288	5.299	5.299	7.288

Table 10 presents the average, standard deviation, skewness, and kurtosis of the stochastic discount factor for the state-dependent risk price models estimated by GMM, local GMM (LGMM), and nonparametric local GMM (NP).

Appendix B: Semi-Parametric Local GMM

This section focuses on the semi-parametric estimation of fixed risk price parameters defined by uniform conditional moment restrictions with moment functions that also contain unknown functional risk price parameters:

$$E[g(Y_{t+1}; \theta_0, \gamma_0(Z_t)) | Z_t] = 0_q \quad \forall z \in \mathcal{Z}, P - a.s. \quad (\text{B.1})$$

The unknown fixed parameter θ_0 is in the set $\Theta \subset \mathbb{R}^{p_1}$, and at a given date t , the unknown time-varying parameter $\gamma_{0,t} \equiv \gamma_0(z_t)$ is in the set $\Gamma(z_t) \subset \mathbb{R}^{p_2}$ for a given constant realization, z_t , of the $d \times 1$ dimensional state variable process Z_t . Lastly, g is a known $q \times 1$ function where $q \geq p = p_1 + p_2$. Ideally, the parameter vector of interest is $\theta_0^* = (\theta_0, \gamma_{0,1}, \dots, \gamma_{0,T})$, however, we will treat the unknown functional risk price parameters as nuisance parameters for the remainder of this section.¹³

We consider estimating the parameter vector $\theta_0^* = (\theta, \gamma_1, \gamma_2, \dots, \gamma_T)$ by minimizing the following objective function composed of an average of T GMM-like quadratic forms of the estimated local moments :

$$Q_T(\theta, \gamma_1, \gamma_2, \dots, \gamma_T) = \frac{1}{T} \sum_{t=1}^T \tilde{Q}_T(z_t, \theta, \gamma_t) \quad (\text{B.2})$$

where:

$$\tilde{Q}_T(z_t, \theta, \gamma_t) = g_T(z_t, \theta, \gamma_t)' W_T g_T(z_t, \theta, \gamma_t) \quad (\text{B.3})$$

W_T is a possibly state-dependent positive definite matrix of size $q \times q$, and

$$g_T(z_t, \theta, \gamma_t) = \hat{E}[g(Y_{t+1}; \theta, \gamma_t) | Z_t = z_t]. \quad (\text{B.4})$$

In particular, we restrict attention to the case where the local moments are estimated by the local constant estimator:

$$\begin{aligned} \hat{E}[g(Y_{t+1}; \theta, \gamma_t) | Z_t = z_t] &= \frac{\sum_{s=1}^T g(y_{s+1}; \theta, \gamma_t) K\left(\frac{z_s - z_t}{h_T}\right)}{\sum_{j=1}^T K\left(\frac{z_j - z_t}{h_T}\right)} \\ &\equiv \sum_{s=1}^T \omega_{s,t} g(y_{s+1}; \theta, \gamma_t) \end{aligned} \quad (\text{B.5})$$

Moreover, we assume in this section that the moment functions are affine in the parameters θ and γ_t so that the conditional Jacobian will be independent of the parameter values.

We provide a heuristic derivation of the asymptotic variance of the semi-parametric estimator of θ . Expand the first-order condition with respect to the fixed parameter θ

¹³Semi-nonparametric methods discussed in Chen (2007) would be more suitable for estimation and inference on both the fixed and functional risk price parameters.

around θ_0^* :

$$\frac{1}{T} \sum_{t=1}^T \left[\frac{\partial \tilde{Q}_T(\theta_{0,t}^*)}{\partial \theta} + \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \theta \partial \gamma_t'} (\hat{\gamma}_{t,T} - \gamma_{0,t}) + \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \theta \partial \theta'} (\hat{\theta}_T - \theta_0) \right] = 0 \quad (\text{B.6})$$

Scale by the parametric rate of convergence \sqrt{T} and evaluate each component of (B.6):

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \tilde{Q}_T(\theta_{0,t}^*)}{\partial \theta} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial g_T'}{\partial \theta} W_T g_T(\theta_{0,t}^*) \\ &\equiv \frac{1}{\sqrt{T}} \sum_{t=1}^T A_{1,T} g_T(\theta_{0,t}^*) \end{aligned} \quad (\text{B.7})$$

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \theta \partial \gamma_t'} (\hat{\gamma}_{t,T} - \gamma_{0,t}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial g_T'}{\partial \theta} W_T \frac{\partial g_T}{\partial \gamma_t'} (\hat{\gamma}_{t,T} - \gamma_{0,t}) \\ &\equiv \frac{1}{\sqrt{T}} \sum_{t=1}^T A_{2,T} (\hat{\gamma}_{t,T} - \gamma_{0,t}) \end{aligned} \quad (\text{B.8})$$

$$\begin{aligned} \frac{1}{\sqrt{T}} \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \theta \partial \theta'} (\hat{\theta}_T - \theta_0) &= \frac{1}{T} \sum_{t=1}^T \frac{\partial g_T'}{\partial \theta} W_T \frac{\partial g_T}{\partial \theta'} \sqrt{T} (\hat{\theta}_T - \theta_0) \\ &\equiv \frac{1}{T} \sum_{t=1}^T A_{3,T} \sqrt{T} (\hat{\theta}_T - \theta_0) \end{aligned} \quad (\text{B.9})$$

$$(\text{B.10})$$

Expand the first-order conditions with respect to the state-dependent parameter γ_t around $\theta_{0,t}^*$ for all $t = 1, 2, \dots, T$. Scale by the nonparametric rate of convergence $\sqrt{Th_T^d}$ and evaluate each component of (B.11) $\forall t = 1, 2, \dots, T$:

$$\frac{\partial \tilde{Q}_T(\theta_{0,t}^*)}{\partial \gamma_t} + \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \gamma_t \partial \theta'} (\hat{\theta}_T - \theta_0) + \frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \gamma_t \partial \gamma_t'} (\hat{\gamma}_{t,T} - \gamma_{0,t}) = 0 \quad \forall t = 1, 2, \dots, T \quad (\text{B.11})$$

$$\begin{aligned} \sqrt{Th_T^d} \frac{\partial \tilde{Q}_T(\theta_{0,t}^*)}{\partial \gamma_t} &= \frac{\partial g_T'}{\partial \gamma_t} W_T \sqrt{Th_T^d} g_T(\theta_{0,t}^*) \\ &\equiv B_{1,T} \sqrt{Th_T^d} g_T(\theta_{0,t}^*) \end{aligned} \quad (\text{B.12})$$

$$\begin{aligned}
\frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \gamma_t \partial \theta'} \sqrt{Th_T^d}(\hat{\theta}_T - \theta_0) &= \left[\frac{\partial g_T'}{\partial \gamma_t} W_T \frac{\partial g_T}{\partial \theta'} \right] \sqrt{Th_T^d}(\hat{\theta}_T - \theta_0) \\
&\equiv B_{2,T} \sqrt{Th_T^d}(\hat{\theta}_T - \theta_0) \\
&= O_p(h_T^{d/2})
\end{aligned} \tag{B.13}$$

$$\begin{aligned}
\frac{\partial^2 \tilde{Q}_T(\bar{\theta}_t^*)}{\partial \gamma_t \partial \gamma_t'} \sqrt{Th_T^d}(\hat{\gamma}_{t,T} - \gamma_{0,t}) &= \frac{\partial g_T'}{\partial \gamma_t} W_T \frac{\partial g_T}{\partial \gamma_t'} \sqrt{Th_T^d}(\hat{\gamma}_{t,T} - \gamma_{0,t}) \\
&\equiv B_{3,T} \sqrt{Th_T^d}(\hat{\gamma}_{t,T} - \gamma_{0,t})
\end{aligned} \tag{B.14}$$

Combine (B.12)-(B.14) to arrive at:

$$\sqrt{Th_T^d}(\hat{\gamma}_{t,T} - \gamma_{0,t}) = -B_{3,T}^{-1} B_{1,T} \sqrt{Th_T^d} g_T(\theta_{0,t}^*) + o_p(1) \quad \forall t = 1, 2, \dots, T \tag{B.15}$$

Lastly, combine (B.7)-(B.9) and (B.15) to arrive at:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = - \left[\frac{1}{T} \sum_{t=1}^T H_T \right]^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T C_T g_T(\theta_{0,t}^*) \right] + o_p(1) \tag{B.16}$$

where:

$$\begin{aligned}
H_T &= A_{3,T} - A_{2,T} B_{3,T}^{-1} B_{2,T} \\
C_T &= A_{1,T} - A_{2,T} B_{3,T}^{-1} B_{1,T}.
\end{aligned}$$

Under regularity conditions found in Gagliardini et al. (2011) or Gospodinov and Otsu (2012), one can show that:

$$\begin{aligned}
H_T &\xrightarrow{p} H_{0,t} = A_{3,t} - A_{2,t} B_{3,t}^{-1} A'_{2,t} \\
C_T &\xrightarrow{p} C_{0,t} = A_{1,t} - A_{2,t} B_{3,t}^{-1} B_{1,t}
\end{aligned}$$

where:

$$\begin{aligned}
A_{1,t} &\equiv \text{plim}_{T \rightarrow \infty} A_{1,T} = J_{\theta,t} W \\
B_{1,t} &\equiv \text{plim}_{T \rightarrow \infty} B_{1,T} = J'_{\gamma_t} W
\end{aligned}$$

$$A_{2,t} \equiv \text{plim}_{T \rightarrow \infty} A_{2,T} = J'_{\theta,t} W J_{\gamma_t}$$

$$B_{3,t} \equiv \text{plim}_{T \rightarrow \infty} B_{3,T} = J'_{\gamma_t} W J_{\theta,t}$$

$$A_{3,t} \equiv \text{plim}_{T \rightarrow \infty} A_{3,T} = J'_{\theta,t} W J_{\theta,t}$$

$$B_{2,t} \equiv \text{plim}_{T \rightarrow \infty} B_{2,T} = J'_{\gamma_t} W J_{\theta,t}$$

and:

$$W = \text{plim}_{T \rightarrow \infty} W_T$$

$$J_{\theta,t} \equiv E \left[\frac{\partial g(Y_{t+1}, \theta_{0,t}^*)}{\partial \theta'} \middle| Z_t = z_t \right]$$

$$J_{\gamma_t} \equiv E \left[\frac{\partial g(Y_{t+1}, \theta_{0,t}^*)}{\partial \gamma'_t} \middle| Z_t = z_t \right]$$

The asymptotic distribution of the semi-parametric local GMM estimator will depend behavior of:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T C_{0,t} g_T(\theta_{0,t}^*) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{s=1}^T C_{0,t} \omega_{s,t} g(y_{s+1}; \theta_{0,t}^*).$$

Define the following:

$$\Sigma_{0,t} \equiv \text{Var}[g(Y_{t+1}; \theta_{0,t}^*) | Z_t = z_t]$$

$$\omega^2 \equiv \int K(u)^2 du < \infty$$

Following arguments in the proofs of Theorem 1 of Gospodinov and Otsu (2012) and Lemma B.2. of Kitamura et al. (2004), combined with a central limit theorem for dependent U-statistics (e.g. Fan and Li (1999), Dehling and Wendler (2010)), the asymptotic variance of the parametric component is:

$$\text{Avar}[\sqrt{T}(\hat{\theta}_T - \theta_0)] = \frac{\omega^2}{f_Z(z_t)} H_0^{-1} V_0 H_0^{-1}$$

where:

$$H_0 = E[H_{0,t}]$$

$$V_0 = E[C'_{0,t} \Sigma_{0,t} C_{0,t}]$$

If the weight matrix is chosen such that $\text{plim}_{T \rightarrow \infty} W_T = \Sigma_{0,t}^{-1}$, then H_0 and V_0 simplify to an equivalent expression:

$$H_0^* = V_0^* = E[\mathcal{I}_{\theta,\theta|t} - \mathcal{I}_{\theta,\gamma_t} \mathcal{I}_{\gamma_t,\gamma_t}^{-1} \mathcal{I}_{\gamma_t,\theta}]$$

where:

$$\begin{aligned}\mathcal{I}_{\theta,\theta|t} &= J'_{\theta,t} \Sigma_{0,t}^{-1} J_{\theta,t} \\ \mathcal{I}_{\theta,\gamma_t} &= J_{\theta,t} \Sigma_{0,t}^{-1} J_{\gamma_t} \\ \mathcal{I}_{\gamma_t,\gamma_t} &= J'_{\gamma_t} \Sigma_{0,t}^{-1} J_{\gamma_t}\end{aligned}$$

The asymptotic variance resembles the semi-parametric efficiency bound in Chamberlain (1992) derived for the estimation of finite dimensional parameters defined by conditional moment restrictions with a nonparametric nuisance component.

$$Avar^*[\sqrt{T}(\hat{\theta}_T - \theta_0)] = \frac{\omega^2}{f_Z(z_t)} E[\mathcal{I}_{\theta,\theta|t} - \mathcal{I}_{\theta,\gamma_t} \mathcal{I}_{\gamma_t,\gamma_t}^{-1} \mathcal{I}_{\gamma_t,\theta}]^{-1}. \quad (\text{B.17})$$

A concern for semi-parametric estimators is that \sqrt{T} -consistency of the parametric component is not always granted due to the slower than \sqrt{T} rate of convergence of the nonparametric component. A rigorous proof of \sqrt{T} -consistency and asymptotic normality should work off of Theorem 8.1 of Newey and McFadden (1994). However, for the class of affine models considered in this paper, it is possible to concentrate out the functional parameters to obtain the concentrated objective function:

$$Q_T^c(\theta) = \frac{1}{T} \sum_{t=1}^T \tilde{Q}_T^c(z_t, \theta)$$

where:

$$\tilde{Q}_T^c(z_t, \theta) = \tilde{g}_T(z_t, \theta)' \tilde{W}_T \tilde{g}_T(z_t, \theta)$$

and:

$$\begin{aligned}\tilde{W}_T &= [I_q - B_{1,t} W_T^{-1} B_{3,t}^{-1} B_{1,t}]' W_T [I_q - B_{1,t} W_T^{-1} B_{3,t}^{-1} B_{1,t}] \\ &= W_T - B'_{1,t} B_{3,t}^{-1} B_{1,t}\end{aligned}$$

is a state-dependent weight matrix that is independent of the parameters. Applying Theorem 1 of Gospodinov and Otsu (2012), the asymptotic variance of $\hat{\theta}$ is shown to be given by equation (B.17). The above argument is essentially an application of the Frisch-Waugh-Lovell Theorem. The same argument was made by Chamberlain (1992) for the evaluation of a fixed-effects estimator of a model with linearly separable structural and nuisance parameters.

Appendix C: SMD Estimation of Conditional Fama-French Models

In this section, we re-estimate the constant and time-varying risk price models studied in Section 4 with the Smooth Minimum Distance (SMD) estimator of Lavergne and Patilea (2013). Recently, Antoine et al. (2018) advocate for SMD as an alternative to local GMM when conditional moment restrictions are misspecified.

For an identity weight matrix, the SMD estimator is motivated by the following population objective function:

$$Q^{smd}(\theta) = E[(m_{t,t+1}(Y_{t+1}; \theta)R_{t+1}^e)' E[m_{t,t+1}(Y_{t+1}; \theta)R_{t+1}^e | z_t] f(z_t)].$$

where $f(z_t)$ is the density function of the conditioning variable. A sample counterpart of the objective function is then obtained by replacing the conditional expectation and the density function by a jackknife kernel estimator and a kernel density estimator, respectively, and by replacing the outer expectation with a sample average. Specifically, the SMD estimator is defined as:

$$\hat{\theta}_T(h) \equiv \arg \min_{\theta \in \Theta} \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \Psi_{t+1}(\theta)' \Psi_{s+1}(\theta) K\left(\frac{z_t - z_s}{h}\right)$$

where $\Psi_{t+1}(\theta) = m_{t,t+1}(Y_{t+1}; \theta)R_{t+1}^e$. Although it is based on kernel smoothing, the primary benefit of the SMD approach is that it avoids instabilities and the need for trimming displayed by local GMM estimators when the density of the conditioning variable is arbitrarily small. In addition, the SMD estimator alleviates the curse of dimensionality associated with kernel smoothing because it is shown by Lavergne and Patilea (2013) to be consistent for a fixed bandwidth. Also, as illustrated by Antoine and Lavergne (2014), the SMD estimator can be robustified to deal with nearly-weak identification situations. Finally, the SMD estimator is asymptotically normal with an asymptotic variance matrix that is characterized for the well-specified case in Lavergne and Patilea (2013) and for the misspecified case in Antoine et al. (2018).

The estimation results are presented in Table 11 for the same choice of bandwidth as the local GMM estimator considered above. The point estimates for the constant risk price model estimated by SMD are remarkably stable across specifications. Moreover, the confidence intervals for the state-dependent models are considerably narrower than those obtained from GMM and local GMM. However, at the point estimates delivered by SMD, each of the models performs worse in terms of MSE of conditional pricing errors than when they are evaluated at the point estimates delivered by GMM and local GMM.

Table 11: Smooth Minimum Distance Estimation Results for Constant and State-dependent Risk Price Models

	Smooth Minimum Distance					
	TSY10		TERM		DEF	
$\hat{\phi}_{mkt}$	0.034 (0.011) [0.013,0.055]	0.076 (0.033) [0.012,0.140]	0.042 (0.014) [0.014,0.070]	0.028 (0.018) [0.007,0.063]	0.048 (0.012) [0.025,0.071]	0.053 (0.037) [-0.028,0.108]
$\hat{\phi}_{smb}$	0.022 (0.016) [-0.010,0.054]	-0.012 (0.059) [-0.128,0.103]	0.014 (0.016) [-0.018,0.047]	-0.010 (0.022) [-0.054,0.033]	-0.014 (0.016) [-0.042,0.021]	-0.092 (0.067) [-0.215,0.042]
$\hat{\phi}_{hml}$	0.109 (0.029) [0.053,0.165]	0.193 (0.096) [0.004,0.381]	0.108 (0.039) [-0.032,0.184]	0.103 (0.032) [0.041,0.165]	0.102 (0.043) [0.026,0.176]	0.393 (0.174) [0.058,0.593]
$\hat{\phi}_{z,mkt}$		-0.008 (0.005) [-0.018,0.003]		0.000 (0.011) [-0.022,0.022]		-0.009 (0.018) [-0.035,0.028]
$\hat{\phi}_{z,smb}$		0.005 (0.009) [-0.012,0.022]		0.029 (0.013) [0.002,0.055]		0.055 (0.032) [-0.009,0.114]
$\hat{\phi}_{z,hml}$		-0.018 (0.015) [-0.048,0.012]		-0.048 (0.029) [-0.104,0.009]		-0.152 (0.073) [-0.232,-0.011]
$\frac{1}{N} \sum_{i=1}^N e_i(\hat{\theta})^2$	0.0138	0.0038	0.0121	0.0076	0.0337	0.0150
$\frac{1}{N} \sum_{i=1}^N var(\hat{e}_i(I(t); \hat{\theta}))$	0.0019	0.0202	0.0331	0.1123	0.0593	0.1250

Table 11 presents point estimates, standard errors, and 95 percent confidence intervals for the constant and affine state-dependent risk price models estimated by smooth minimum distance (SMD). Each model differs based on choice of state variable: TSY10, TERM, and DEF. The table also presents the average squared unconditional pricing errors and the average variance of the estimated conditional pricing errors.

References

- Adrian, T., R. K. Crump, and E. Moench (2015). Regression-based estimation of dynamic asset pricing models. *Journal of Financial Economics* 118(2), 211–244.
- Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71(6), 1795–1843.
- Ai, C. and X. Chen (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics* 141(1), 5–43.
- Antoine, B., H. Bonnal, and E. Renault (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics* 138(2), 461–487.
- Antoine, B. and P. Lavergne (2014). Conditional moment models under semi-strong identification. *Journal of Econometrics* 182(1), 59–69.
- Antoine, B., K. Proulx, and E. Renault (2018). Pseudo-true SDFs in conditional asset pricing models. *Journal of Financial Econometrics (online)*.
- Balduzzi, P. and C. Robotti (2010). Asset pricing models and economic risk premia: A decomposition. *Journal of Empirical Finance* 17(1), 54–80.
- Berk, J. B. (1997). Necessary conditions for the CAPM. *Journal of Economic Theory* 73(1), 245–257.
- Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *The Journal of Finance* 54(5), 1609–1645.
- Brandt, M. W. and D. A. Chapman (2018). Linear approximations and tests of conditional pricing models. *Review of Finance* 22(2), 455–489.
- Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999). Redundancy of moment conditions. *Journal of Econometrics* 91(1), 89–111.
- Cai, Z., Y. Ren, and L. Sun (2015). Pricing kernel estimation: A local estimating equation approach. *Econometric Theory* 31(3), 560–580.
- Chamberlain, G. (1983). A characterization of the distributions that imply mean—variance utility functions. *Journal of Economic Theory* 29(1), 185–201.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34(3), 305–334.
- Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica* 60(3), 567–596.

- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics* 6, 5549–5632.
- Chen, X., J. Favilukis, and S. C. Ludvigson (2014). An estimation of economic models with recursive preferences. *Quantitative Economics* 4(1), 39–83.
- Chen, X., D. T. Jacho-Chávez, and O. Linton (2016). Averaging of an increasing number of moment condition estimators. *Econometric Theory* 32(1), 30–70.
- Dehling, H. and M. Wendler (2010). Central limit theorem and the bootstrap for u-statistics of strongly mixing data. *Journal of Multivariate Analysis* 101(1), 126–137.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.
- Fan, J. and Y. Zhong (2018). Optimal subspace estimation using overidentifying vectors via generalized method of moments. *arXiv preprint arXiv:1805.02826*.
- Fan, Y. and Q. Li (1999). Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10(3), 245–271.
- Fang, K.-T., S. Kotz, and K.-W. Ng (1989). *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall.
- Fang, Y., Y. Ren, and Y. Yuan (2011). Nonparametric estimation and testing of stochastic discount factor. *Finance Research Letters* 8(4), 196–205.
- Ferreira, E., J. Gil-Bazo, and S. Orbe (2011). Conditional beta pricing models: A nonparametric approach. *Journal of Banking & Finance* 35(12), 3362–3382.
- 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 (2018). Comparing asset pricing models by the conditional Hansen-Jagannathan distance. *Working paper*.
- 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.
- Hall, A. R. and A. Inoue (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114(2), 361–394.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.

- Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30(1-2), 203–238.
- Hansen, L. P. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. *The 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.
- Hodgson, D. J., O. Linton, and K. Vorkink (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *Journal of Applied Econometrics* 17(6), 617–639.
- Jagannathan, R. and Z. Wang (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance* 51(1), 3–53.
- Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *The Journal of Finance* 53(4), 1285–1309.
- Jagannathan, R. and Z. Wang (2002). Empirical evaluation of asset-pricing models: A comparison of the sdf and beta methods. *The Journal of Finance* 57(5), 2337–2367.
- Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, 419–430.
- Kitamura, Y., G. Tripathi, and H. Ahn (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72(6), 1667–1714.
- Lavergne, P. and V. Patilea (2013). Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory. *Journal of Econometrics* 177(1), 47–59.
- Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109(6), 1238–1287.
- Lettau, M. and M. Pelger (2018). Estimating latent asset-pricing factors. *No. w24618. National Bureau of Economic Research*.
- Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters* 94(1), 124–128.
- Li, Y., L. Su, and Y. Xu (2015). A combined approach to the inference of conditional factor models. *Journal of Business & Economic Statistics* 33(2), 203–220.
- Ludvigson, S. C. (2013). Advances in consumption-based asset pricing: Empirical tests. In G. Constantinides, M. Harris, and R. Stulz (Eds.), *Handbook of the Economics of Finance*, Volume 2B, pp. 799–906. Amsterdam, North-Holland: Elsevier.

- Ludvigson, S. C. (2018). Comments on: “Pseudo-true SDFs in conditional asset pricing models” by B. Antoine, K. Proulx, and E. Renault. *Journal of Financial Econometrics (Online)*.
- MacKinlay, A. C. and M. P. Richardson (1991). Using generalized method of moments to test mean-variance efficiency. *The Journal of Finance* 46(2), 511–527.
- Nagel, S. and K. J. Singleton (2011). Estimation and evaluation of conditional asset pricing models. *The Journal of Finance* 66(3), 873–909.
- Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In G. Maddala, C. Rao, and H. Vinod (Eds.), *Handbook of Statistics*, Volume 11, Econometrics. Elsevier, Amsterdam.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4, 2111–2245.
- Newey, W. K. and R. J. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.
- Owen, J. and R. Rabinovitch (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance* 38(3), 745–752.
- Peñaranda, F. and E. Sentana (2012). Spanning tests in return and stochastic discount factor mean-variance frontiers: A unifying approach. *Journal of Econometrics* 170(2), 303–324.
- Peñaranda, F. and E. Sentana (2015). A unifying approach to the empirical evaluation of asset pricing models. *Review of Economics and Statistics* 97(2), 412–435.
- Renault, E. (1997). Économétrie de la finance: la méthode des moments généralisés. In Y. Simon (Ed.), *Encyclopédie des Marchés Financiers*, Volume 1, Chapter 19, pp. 330–407. Economica, Paris.
- Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section of stock returns. *Journal of Financial Economics* 111(2), 352–380.
- Santos, T. and P. Veronesi (2005). Labor income and predictable stock returns. *The Review of Financial Studies* 19(1), 1–44.
- Shanken, J. (1992). On the estimation of beta-pricing models. *The Review of Financial Studies* 5(1), 1–33.
- Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics* 138(2), 430–460.
- Wang, K. Q. (2003). Asset pricing with conditioning information: A new test. *The Journal of Finance* 58(1), 161–196.