Subjective Beliefs Estimators and Their Properties *

Anisha Ghosh† Taisuke Otsu‡

October 31, 2025

Abstract

Information-theoretic methods recover investors' subjective beliefs by minimizing the statistical discrepancy between beliefs and the DGP, subject to assets' Euler constraints. We show that the estimated beliefs converges in probability to its pseudo-true value. Comparing estimators in the Cressie-Read family, we show that the exponential tilting and empirical likelihood estimators produce qualitatively similar estimates of the risk aversion levels and beliefs. The quadratic divergence estimator leads to negative subjective probabilities, implausibly large risk aversion levels, and underestimation of left tail risk. Our results suggest large institutional investors, like pension funds, have countercyclical beliefs about the market return, while extrapolative investors have procyclical beliefs. Our results offer an alternative explanation of the momentum effect in stock returns, help reconcile procyclical beliefs reported in individual investor surveys versus countercyclical beliefs implied by rational expectations representative agent models, and establish the information-theoretic approach as a powerful methodology for the recovery of beliefs.

Keywords: Rational Expectations, Subjective Beliefs, Pricing Kernel, Conditioning Set, Cressie-Read Discrepancy Minimization.

JEL Classification Codes: C51, E3, E70, G12, G14, G40

*We benefitted from helpful comments from Prosper Dovonon, Silvia Goncalves, Lars Hansen, Christian Julliard, and Yuichi Kitamura. All errors and omissions are our own.

†Desautels Faculty of Management, McGill University; anisha.ghosh@mcgill.ca, *Corresponding author*.

‡Department of Economics, London School of Economics; t.otsu@lse.ac.uk.

I Introduction

The risk preferences of households over the set of possible risky outcomes and their beliefs about the likelihood of occurrence of these outcomes are, perhaps, among the most ubiquitous concepts in economics. Although neither is directly observed, financial markets offer a natural laboratory to gain insights about the risk preferences and beliefs of financial market participants. Of particular interest in this context is the possibility that investors' beliefs may be subjective, i.e. their beliefs may differ from the true data generating process (DGP), hereafter referred to as the objective beliefs – a conjecture relied upon in the rapidly growing field of behavioral economics. Indeed, there is growing evidence that people, including professional forecasters, repeatedly make the same mistakes in their forecasts. A recent New York Times article[1] pointed out that in 12 of the last 15 years, professional forecasters have been too optimistic about future GDP growth and that these mistakes have had real costs. Using survey forecasts data, distorted subjective beliefs have been documented for macroeconomic variables (see, e.g., Bordalo, Gennaioli, Ma, and Shleifer (2020), Bianchi, Ludvigson, and Ma (2022)), stocks (see, e.g., Greenwood and Shleifer (2014) and Cocco, Gomes, and Lopes (2019)), bonds (see, e.g., Piazzesi, Salomao, and Schneider (2015)), real estate (see, e.g., Case and Shiller (2003)), and different asset classes including stocks, bonds, currencies, and commodities (see, e.g., Nagel and Xu (2022)).

Ghosh and Roussellet (2019) propose a novel information-theoretic approach, namely the *smoothed Cressie-Read discrepancy minimization* approach, that uses observed asset prices for the recovery of investors' beliefs. This approach does not rely on any parametric assumptions about the DGP or the form of investors' beliefs distortions (if any) relative to the DGP. Instead, it approximates the subjective beliefs distribution with a multinomial on the observed data sample. The multinomial probabilities are estimated such that the recovered beliefs are minimally distorted with respect to the DGP, within the class of mod-

[1] https://www.nytimes.com/2021/02/08/briefing/astrazeneca-vaccine-trump-impeachment-bruce-springsteen-jeep.html

els parametrized by a known stochastic discount factor (SDF) summarizing investors' risk preferences. Different Cressie-Read family members use different convex functions to characterize the divergence between the subjective and objective beliefs. Thus, beliefs refer to the investors' *distorted conditional distribution* of macroeconomic and financial variables of interest. The asset pricing Euler equations serve as constraints in the above minimization problem, thereby disciplining the estimated beliefs to be consistent with observed asset prices. Survey forecasts may be incorporated into the estimation as additional moment restrictions, thereby also requiring the beliefs to match certain aspects of survey forecasts. The recovered beliefs, therefore, offer a characterization of the *entire distribution* of variables of interest *as perceived by investors*.

Note that survey forecasts, an alternative to inferring investors' subjective beliefs, are much more restrictive, typically offering only the conditional means of the variables. Our methodology is also intended to overcome many of the other shortcomings of sole reliance on survey data, associated with their typically short time series, small sizes of the cross sections of respondents, and paucity of information on the identities of the respondents. Moreover, the method enables the recovery of not only investors' subjective beliefs, but also of the SDF parameters (e.g., the coefficient of risk aversion). Ghosh and Roussellet (2019) provide simulation evidence suggesting that the estimator works well in recovering beliefs in empirically realistic sample sizes. Whereas Ghosh and Roussellet (2019) focuses on a representative agent setting whereby the beliefs recovered may be interpreted as those of the representative investor in the aggregate equity market, Ghosh, Korteweg, and Xu (2019) extend the methodology to a heterogeneous investors setting that enables the recovery of the (potentially heterogeneous) beliefs of different investor types.

The first contribution of this paper is to establish the theoretical underpinnings of the smoothed Cressie-Read discrepancy minimization methodology. The divergence between the subjective and objective beliefs implies a form of model misspecification, whereby the investors' Euler equations hold under their subjective beliefs but not under the DGP. This

alters the theoretical properties of our estimator, much like a parametric maximum likelihood estimator has different properties in a misspecified setting relative to a correctly specified one. However, unlike in a misspecified parametric maximum likelihood setting, where the object of interest is typically a finite dimensional parameter vector, our primary object of interest is an infinite dimensional distribution function characterizing investors' beliefs. We derive the asymptotic properties of the Cressie-Read family of estimators of investors' beliefs. We show that, under relatively mild regularity conditions, the estimated subjective beliefs distribution converges in probability to its *pseudo-true* value, the latter being the distribution obtained by projecting the DGP on to the space of distributions satisfying the Euler equation constraints. In other words, the estimated beliefs correspond to the distribution that is minimally distorted with respect to the DGP, within the class of models parametrized by the known SDF. While our approach is nonparametric, not relying on parametric assumptions either about the DGP or the subjective beliefs distortions, the above property of our estimator parallels that of parametric maximum likelihood estimators for misspecified models (see, e.g., White (1982), Vuong (1989)).

An important advantage of this approach is that it does not require parametric assumptions on the DGP or the forms of beliefs distortions. At the same time, the technical assumptions underlying our framework encompass several of the commonly assumed specifications in the literature. Prominent examples include DGPs featuring persistent state variables such as the conditional mean of the real economic growth rate (see, e.g., Bansal and Yaron (2004)), habit levels (see, e.g., Campbell and Cochrane (1999)), or time-varying probabilities of rare catastrophic events (see, e.g., Wachter (2013)) and alternative subjective beliefs formation processes (see, e.g., Barberis, Greenwood, Jin, and Shleifer (2015) and Bordalo, Gennaioli, Porta, and Shleifer (2019)). Note that parametric specifications of the DGP and/or forms of beliefs distortions are often based on analytical tractability and, therefore, are likely to be misspecified. Moreover, there exists an infinitely large set of possible parametric specifications to choose from. Our non-parametric approach avoids the need for such parametric

assumptions. Although the recovered beliefs distribution converges to its pseudo true value (instead of the true distribution), in the absence of knowledge of the true beliefs, the pseudo true value offers a natural starting point in the analysis of subjective beliefs. The recovered beliefs may aid in providing guidance for the modeling of more plausible beliefs dynamics in structural models.

The second contribution of this paper is to compare empirically the beliefs estimated with different members of the Cressie-Read family of divergence measures. We consider the three most popular members of the family: (1) the smoothed empirical likelihood (SEL) estimator of Kitamura, Tripathi, and Ahn (2004), which constitutes the conditional counterpart of the empirical likelihood (EL) estimator of Owen (2001), (2) the smoothed exponentially tilted (SET) estimator, which is the conditional counterpart of the exponentially tilted (ET) estimator of Kitamura and Stutzer (1997), and (3) the smoothed Euclidean or quadratic divergence (SQD) estimator, which is related to the continuously updated GMM (QD) estimator of Hansen (1982). While the existing literature has focused on using the EL, ET, and QD methods for the recovery of unconditional distributions, smoothed counterparts of these estimators – namely, the SEL, SET, and SQD, respectively – are essential for the recovery of conditional distributions. The latter is crucial in our present setting, where the objective is to estimate investors' beliefs about the future, given their current information set.

We show that members of the Cressie-Read family that naturally impose non-negativity of the subjective probabilities, because of the presence of logarithms in their objective functions – namely, the SEL and SET estimators – are preferred to the SQD estimator, that does not. The latter produces negative estimates of the subjective probabilities (for about 5% of the possible states), particularly in the tails of the distributions of the variables of interest. This finding is reminiscent of the Hansen-Jagannathan variance minimizing SDF (see Hansen and Jagannathan (1997)), being linear in the asset returns used in its construction, having negative values at times corresponding to extreme return realizations. This issue with the quadratic divergence renders its use less desirable in our application since tail risk is widely

accepted as an important source of priced risk in financial markets and, therefore, inferring investors' subjective beliefs about tail events is of utmost importance.[2]

Also, the SQD approach produces large deviations between subjective and objective beliefs for economically plausible values of the risk aversion parameter. Since the SQD estimator chooses a point estimate of the SDF parameter that minimizes the discrepancy between subjective and objective beliefs, it leads to economically implausible values (in excess of 100) of the risk aversion coefficient. This issue does not arise with the SET or SEL approaches, which produce more plausible parameter estimates that are also similar to each other. The same pattern is observed for the beliefs estimates – while the beliefs recovered using the SEL and SET approaches are more similar to each other, there are more substantive differences between these beliefs and those recovered using the SQD estimator. Specifically, the SQD substantially underestimates tail risk compared to the SEL and SET estimators and exhibits excess sensitivity to small variations in the conditioning set.

The SQD objective function is quadratic in the ratio of the subjective and objective probabilities, i.e. the Radon-Nikodym derivative of the subjective distribution with respect to the objective one, and, therefore, estimates the subjective probabilities to minimize the variance of the divergence between the objective and subjective beliefs distributions. In other words, it tries to deviate as little as possible from a Gaussian distribution with mean and variance corresponding to those implied by the objective probabilities, while also satisfying the conditional Euler restrictions for the test assets. The SEL and SET objective functions, on the other hand, involve the logarithm of the Radon-Nikodym derivative of the objective and subjective measures and, therefore, depend on *all* its moments (not just the variance). In our empirical applications, the SQD estimator minimizes the variance of the divergence between the subjective and objective distributions by estimating implausibly large estimates of the risk aversion coefficient. What prevents the SEL and SET estimators from doing the

[2]In an unconditional setting, Chen, Hansen, and Hansen (2021) propose a slight modification to the QD approach whereby a non-negativity condition on the likelihood ratio is explicitly imposed to ensure non-negative subjective probabilities.

same is that their objective functions are weighted sums of all the moments of the Radon-Nikodym derivative and, therefore, need to satisfy the more stringent criteria that *all* the moments of the subjective distribution are 'close' to the corresponding objective ones.

As for the comparison between the SET and SEL, the latter has the attractive additional interpretation of being a nonparametric maximum likelihood estimator. However, its objective function is monotonically decreasing in the magnitude of discrepancy between the subjective and objective beliefs, which is theoretically undesirable in a misspecified setting such as the present one, especially when the DGP has unbounded support. Specifically, Chen, Hansen, and Hansen (2024) show that the pseudo-true beliefs may fail to exist in the case of the unconditional EL and that, therefore, the finite-sample estimates can converge to no beliefs distortion even when rational expectations is violated (see also Schennach (2007), Chen, Hansen, and Hansen (2020) and Chauduri, Renault, and Wahlstrom (2023) for robustness issues with the unconditional EL estimator under misspecification).

In our empirical analysis, the SET and SEL estimators produce similar estimates of both the SDF parameters as well as beliefs. The SET tends to produce a bit higher estimates of the risk aversion coefficient relative to those obtained with the SEL as well as higher overall levels of the expected market returns relative to the SEL. However, both methods identify pronounced heterogeneity between the beliefs of different investor types. Market investors as well as specialized investors that tilt towards value stocks exhibit strong countercyclicality in their expected aggregate stock market returns. Pension fund investors resemble the market and value investors in that they too exhibit countercyclical expected stock market returns, although they have higher risk aversion levels compared to the market and value investors. On the other hand, momentum investors perceive the expected market returns to be procyclical. Our results thereby offer an alternative explanation of the well documented momentum effect in stock returns on the one hand and helps reconcile the extant evidence of procyclical beliefs in individual investor surveys on the other. The pairwise correlations between the time series of the expected market returns obtained using the SET and SEL estimators are

over 80% in 9 out of 10 cases, compared to correlations as low as 2.2% observed between these two estimators and the SQD estimator. Nonetheless, given the theoretical shortcomings of the EL estimator highlighted in the literature (which we conjecture will likely extend to the SEL estimator), the SET offers an attractive candidate for the estimation of investors' subjective beliefs.

Our work builds upon Ghosh, Julliard, and Taylor (2017) that used relative entropy metrics to recover the missing component of the SDFs implied by a large class of representative agent consumption based models (see also Almeida and Garcia (2012), Almeida and Garcia (2016), Borovicka, Hansen, and Scheinkman (2016)). That this missing component may be interpreted as the beliefs distortions of investors relative to the DGP was pointed out in Hansen (2014). While Ghosh, Julliard, and Taylor (2017) focused on the recovery of unconditional distributions, Ghosh and Roussellet (2019) and Ghosh, Korteweg, and Xu (2019) extend the framework to the estimation of conditional distributions – a necessary extension in the context of the recovery of investors' beliefs – which requires a wholly different methodology. Egan, Mackay, and Yang (2022) (EMY) propose an approach to recover heterogeneous beliefs that yields the distribution of expected stock market returns across retail investors in S&P 500 tracking ETFs. This is similar to survey data that also provides the distribution of expected market returns for the cross-section of survey respondents. Our approach, in contrast, delivers the entire distribution of future returns as perceived by the heterogeneous investor types, not just the mean return. Our approach also does not require the mean-variance preference assumption needed in EMY. We can accommodate more flexible intertemporal preference specifications commonly assumed in the asset pricing literature. On the other hand, EMY use valuable flow data into S&P 500 tracking ETFs, whereas we rely on asset prices instead. A similarity between our paper and EMY is that neither uses the actual full portfolio holdings of different investor types.

Our work is also related to Chen, Hansen, and Hansen (2020) who argue that, in a misspecified setting, the minimally divergent conditional distribution with respect to the DGP

may not be the only distribution of interest. In an unconditional setting, they propose identifiable lower and upper bounds on moments of observable variables (e.g., the unconditional expected stock market return) subject to a relative entropy constraint on the maximum permissible divergence between investors' subjective beliefs and the DGP. Note that the maximal permissible divergence can be set higher than the minimum divergence between the two measures needed to satisfy the assets' Euler restrictions. Our point estimate of the beliefs distribution, which offers a natural starting point in the analysis of subjective beliefs, along with the moment bounds in Chen, Hansen, and Hansen (2020), offer a rich array of tools aimed at offering a data driven way of modeling investors' beliefs distortions relative to the DGP. We propose an extension of the Chen, Hansen, and Hansen (2020) bounds to a conditional framework, whereby lower and upper bounds on *conditional* moments of variables of interest may be obtained under investors' subjective beliefs.

Our work also contributes to a burgeoning literature trying to recover risk premia components with as few assumptions as possible. Ross (2015) argued that one can recover simultaneously investors' risk preferences and the DGP using a set of Arrow-Debreu securities. Borovicka, Hansen, and Scheinkman (2016) pointed out that data on asset prices alone were not sufficient, except in very special circumstances, to identify both the SDF and transition probabilities without imposing additional restrictions.[3] They further explained that this additional information could either be time-series evidence on the DGP or it could be information on the SDF. Our paper takes the latter approach by specifying an SDF summarizing investors' risk appetite. Although, the literature has progressed towards relaxing restrictive assumptions, to arrive at an (almost) model-free recovery (e.g., Schneider and Trojani (2019), Pazarbasi, Schneider, and Vilkov (2020)), this is akin to identifying the pricing kernel under rational expectations.[4] Unlike this literature, our approach attempts to

[3]see also Carr and Yu (2012), Jensen, Lando, and Pedersen (2019), Walden (2017) and Qin and Linetsky (2016, 2017) for related approaches.

[4]This literature is tightly linked to the identification of pricing kernel bounds as provided by financial instruments paying off statistical moments of index return. Notable contributions include Martin (2013, 2017), Kozhan, Neuberger, and Schneider (2013), Schneider (2015, 2018), Schneider and Trojani (2018) and Orlowski, Sali, and Trojani (2018).

recover investors' subjective beliefs that may differ from the true DGP.

Finally, a comment is in order regarding the interpretation of our approach. As highlighted above, the recovery of distorted beliefs inherently implies a misspecified setting, i.e. a scenario where the Euler equations hold under the distorted beliefs but not under the DGP. However, beliefs distortions need not be the only source of model misspecification. Another potential source of misspecification stems from the SDF that summarizes investors' risk preferences. Indeed, a wide variety of SDFs have been proposed in the literature to help explain different aspects of the macroeconomy and financial markets. Models relying on rational expectations make functional-form assumptions about the SDF and the true DGP, with the latter coinciding with investors' subjective beliefs. Behavioral models make additional assumptions on the dimensions of beliefs distortions relative to the DGP. All of these assumptions are prone to misspecification, which often manifests in the weaknesses in the empirical performance of these models. Our approach relies upon fewer assumptions than the extant literature in that, although we assume a functional form for the SDF, we do not need to take a stance on the functional-forms of either the DGP or the nature of beliefs distortions. In the absence of knowledge of the true beliefs, we offer a more data-driven approach to estimating investors' beliefs and subsequently assess the major dimensions of beliefs distortions relative to the DGP. Nonetheless, our results should be interpreted with the caveat that if our SDF is misspecified, then the estimated beliefs will reflect an amalgamation of the investors' beliefs and the components of the true (unknown) SDF missing from our assumed specification.

The remainder of the paper is organized as follows. Section II briefly describes the methodology for the simultaneous estimation of investors' beliefs and risk preferences. Section III presents the asymptotic properties of these estimators. The empirical results are presented in Section IV. Section V proposes an extension of the Chen, Hansen, and Hansen (2024) bounds to a conditional setting to enable a sensitivity analysis around our point estimates of subjective beliefs. Section VI concludes with suggestions for future research. The

Online Appendix contains the detailed derivations of our results and additional robustness checks.

II Identifying Subjective Beliefs

Our identification strategy works as follows. The absence of arbitrage opportunities implies the existence of a strictly positive SDF, denoted by M_{t+1}, such that the equilibrium gross returns $R_{t+1} \in \mathbb{R}^k$ of any set of k traded assets satisfy the conditional Euler equations:

$$\mathbb{E}^{\mathcal{P}}[M_{t+1}R_{t+1}|\mathcal{F}_t] = 1_k, \tag{1}$$

where 1_k is the k-dimensional vector of ones, $\mathcal{F}_t$ denotes the investors' information set at time t, and $\mathcal{P}$ denotes their subjective beliefs.

Suppose that $\mathcal{P}_0$ denotes the data generating measure (hereafter referred to interchangeably as DGP or objective beliefs). If $\mathcal{P} \neq \mathcal{P}_0$, i.e. investors' subjective beliefs are distorted relative to the objective beliefs, then the conditional moment restrictions in Equation (1) are misspecified in the sense that:

$$\mathbb{E}^{\mathcal{P}_0}[M_{t+1}R_{t+1}|\mathcal{F}_t] \neq 1_k, \tag{2}$$

i.e., the Euler equations hold under the investors' subjective beliefs but not under the physical DGP.

The central question here is how the investors' subjective beliefs, $\mathcal{P}$, may be recovered from observable data. In order to disentangle investors' risk preferences from their beliefs, we assume that the SDF, summarizing their risk appetite, is a parametric function of a set of observable variables, denoted by Y_{t+1}, and a vector of unknown parameters $\theta \in \Theta$:

$$M_{t+1} = M(Y_{t+1}; \theta_0), \tag{3}$$

where Θ denotes the admissible parameter space and θ_0 the true value of θ. The goal, therefore, becomes the simultaneous identification of the subjective beliefs, $\mathcal{P}$, and the finite-dimensional parameter, θ, summarizing investors' risk preferences.

For econometric tractability, we assume that the information set of the investors at time t can be summarized by a finite set of risk factors $X_t \in \mathbb{R}^d$. Let:

$$w_{ij} = \frac{\mathcal{K}_{ij}}{\sum_{j=1}^{T} \mathcal{K}_{ij}}, \qquad \mathcal{K}_{ij} = \mathcal{K}\left(\frac{X_i - X_j}{b_T}\right), \tag{4}$$

for a kernel function $\mathcal{K}$ and bandwidth b_T. The kernel weights $\{w_{i,j}\}_{i,j=1}^{T}$ provide a natural choice for the objective beliefs, $\mathcal{P}_0$, in that they do not require a parametric specification of the DGP and is, therefore, robust to potential misspecification of the same.[5] In Section III, we show that the recovered subjective beliefs converges to a probability limit that does not depend on the particular choice of the kernel function, $\mathcal{K}(.)$, in Equation (4).

For each $\theta \in \Theta$, the local or smoothed Cressie-Read criterion is defined as (see, Smith (2007))

$$\ell(\theta) = \min_{\{p_{ij}\}_{i,j=1}^{T}} \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{T} \sum_{j=1}^{T} w_{ij} \left[\left(\frac{p_{ij}}{w_{ij}} \right)^{\gamma+1} - 1 \right], \tag{5}$$

subject to

$$\sum_{j=1}^{T} p_{ij}[M(Y_j;\theta)R_j - 1] = 0_k, \quad \sum_{j=1}^{T} p_{ij} = 1, \quad \text{for each } i = 1, \ldots, T,$$

where 0_k is the k-dimensional vector of zeros. For ease of notation, we define $Z_j \equiv (Y_j, R_j)$ and $g(Z_j;\theta) \equiv M(Y_j;\theta)R_j - 1$. The objective function in Equation (5) represents the divergence between the subjective and objective probability distributions. It is always non-negative, attaining its minimum value of zero if and only if $p_{ij} = w_{ij}, \forall i, j = 1, 2, ..., T$, i.e. the objective and subjective distributions coincide. The approach, thus, estimates the

[5] Nonparametric kernel densities deliver estimates of probability densities that do not rely on specific parametric assumptions about the true underlying distribution of the data and have been shown to have well-behaved asymptotic properties for stationary observations (see, e.g., Chen, Linton, and Robinson (2001)).

subjective distribution that is minimally distorted relative to the objective one while also satisfying the Euler equations. In addition to the Euler restrictions, moment conditions based on survey forecasts of macroeconomic and financial variables can be included in the set of constraints (see, e.g., Ghosh and Roussellet (2019)).

The intuition behind the estimator may be described as follows. Suppose that investors are rational, i.e. their subjective beliefs coincide with the DGP: $\mathcal{P} = \mathcal{P}_0$. In this case, the Euler equations in Equation (1) are satisfied under the objective measure $\mathcal{P}_0$. Therefore, the estimator in Equation (5), that searches for a subjective measure that is minimally divergent from the objective one, simply sets the former equal to the latter thereby delivering the minimized objective function value of zero. This is, however, not possible in the scenario where investors' subjective beliefs do not coincide with the DGP. In this latter scenario, the Euler restrictions do not hold under the DGP and, therefore, the subjective beliefs, $\mathcal{P}$, need to be distorted relative to the objective measure, $\mathcal{P}_0$, in order to satisfy the Euler restrictions. In the implementation, this corresponds to the empirical objective weights, $\{w_{i,j}\}_{i,j=1}^{T}$, deviating from the corresponding subjective weights, $\{p_{i,j}\}_{i,j=1}^{T}$, thereby delivering a strictly positive minimized objective function value in Equation (5). The minimized Cressie-Read discrepancy estimator identifies the subjective weights that are the closest possible to the objective measure while also satisfying the Euler constraints and offer a natural starting point in the analysis of subjective beliefs.

The estimated subjective beliefs are given by the closed-form solution to Equation (5), for each value of the SDF parameter θ:

$$\hat{p}_{ij}(\theta) = w_{ij}[\gamma\{\hat{\mu}_i(\theta) + \hat{\lambda}_i(\theta)' g(Z_j;\theta)\}]^{1/\gamma},$$

for $i, j = 1, \ldots, T$, where $\hat{\mu}_i(\theta)$ and $\hat{\lambda}_i(\theta)$ are Lagrange multipliers that solve

$$\begin{aligned} \sum_{j=1}^{T} w_{ij}[\gamma\{\hat{\mu}_i(\theta) + \hat{\lambda}_i(\theta)' g(Z_j;\theta)\}]^{1/\gamma} &= 1, \\ \sum_{j=1}^{T} w_{ij}[\gamma\{\hat{\mu}_i(\theta) + \hat{\lambda}_i(\theta)' g(Z_j;\theta)\}]^{1/\gamma} g(Z_j;\theta) &= 0. \end{aligned}$$

Given the conditional probability distribution $\{\hat{p}_{ij}(\theta)\}_{i,j=1}^{T}$, the estimator of the preference parameter θ is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{T} \tau_i \sum_{j=1}^{T} w_{ij}[\gamma\{\hat{\mu}_i(\theta) + \hat{\lambda}_i(\theta)' g(Z_j;\theta)\}]^{\frac{\gamma+1}{\gamma}}, \tag{6}$$

where $\tau_i = \mathbb{I}\left\{\frac{1}{Tb_T^d} \sum_{j=1}^{T} \mathcal{K}_{ij} \geq b_T^{\delta}\right\}$, for some $\delta > 0$, is a trimming term to avoid the technical problem for low densities of X.

Thus, the methodology allows for the simultaneous estimation of risk preferences and subjective beliefs: $\left(\hat{\theta}, \{\hat{p}_{i,j}(\hat{\theta})\}_{i,j=1}^{T}\right)$. Provided the assumed SDF specification, $M(Y_j;\theta)$, summarizes reasonably well investors' risk preferences (i.e., it is correctly specified), the recovered subjective beliefs, $\{\hat{p}_{i,j}(\hat{\theta})\}_{i,j=1}^{T}$, represent the minimal distortions with respect to the DGP that are needed to match observed asset prices, i.e. satisfy the conditional Euler restrictions. It is in this sense that the recovered beliefs represent an estimate of the *pseudo-true beliefs* (a notion made more econometrically precise in Section III). Also, in this sense the true (unknown) subjective beliefs may be even farther from the objective DGP than what our estimate suggests, i.e. the recovered pseudo-true beliefs likely underestimate the extent of distortions between the true subjective beliefs and the DGP.

Different values of γ in Equation (5) lead to different choices of the convex function used to measure the divergence between the subjective and objective measures, i.e. different members of the Cressie-Read family, and, therefore, to different estimates of the subjective beliefs and the SDF parameter θ. For $\gamma = -1$, an application of L'Hopital's rule to Equation (5) delivers the smoothed empirical likelihood (SEL) estimator of Kitamura, Tripathi, and Ahn

(2004), with the objective function being the Kullback-Leibler divergence (KLIC) between the objective and subjective distributions:

$$\ell(\theta) = \min_{\{p_{ij}\}_{i,j=1}^{T}} \sum_{i=1}^{T}\sum_{j=1}^{T} w_{ij} \log\left(\frac{w_{ij}}{p_{ij}}\right). \tag{7}$$

Setting $\gamma = 0$, on the other hand, leads to the smoothed exponentially tilted (SET) estimator – an extension to the conditional setting of the exponentially tilted estimator proposed by Kitamura and Stutzer (1997). The objective function in this case is given by:

$$\ell(\theta) = \min_{\{p_{ij}\}_{i,j=1}^{T}} \sum_{i=1}^{T}\sum_{j=1}^{T} p_{ij} \log\left(\frac{p_{ij}}{w_{ij}}\right), \tag{8}$$

i.e., the SET swaps the roles of the objective and subjective distributions in the KLIC relative to the SEL. Setting $\gamma = 1$ gives the smoothed Euclidean divergence estimator (SQD), an extension to the conditional setting of the continuously updated GMM estimator (see, e.g., Hansen (1982)), with the (in this case, quadratic) objective function given by:[6]

$$\ell(\theta) = \min_{\{p_{ij}\}_{i,j=1}^{T}} \frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{T} w_{ij}\left[\left(\frac{p_{ij}}{w_{ij}}\right)^{2} - 1\right]. \tag{9}$$

Why do beliefs minimally divergent – in the chosen Cressie-Read discrepancy measure sense – from the DGP offer a reasonable estimate of investors' subjective beliefs? Recent advances in behavioral economics have advocated and found empirical support for subjective beliefs that represent *small departures* from rational expectations (see, e.g., Klaus, Marcet, and Nicolini (2016)). Our method offers a data-driven approach to identifying 'small' distortions in the beliefs needed to match asset prices, while making minimal assumptions about the true DGP or the nature of the subjective beliefs distortions relative to it. In the absence

[6]Other related methods include the exponentially tilted empirical likelihood estimator of Schennach (2007), that combines the empirical likelihood's discrepancy function with the exponential tilting's implied probabilities, and the exponentially tilted Hellinger distance estimator of Antoine and Dovonon (2021), that combines the exponential tilting and Hellinger distance estimators. The extensions of these estimators to their conditional counterparts and their empirical performance in the estimation of subjective beliefs is left for future research.

of knowledge of the true beliefs, our estimate, therefore, offers a natural starting point in the analysis of subjective beliefs. Furthermore, despite the minimum divergence criterion that stacks the odds against estimating large beliefs distortions, we show below that our empirically recovered beliefs are often substantially different from the DGP, e.g., procyclical beliefs for some investor types versus countercyclical beliefs for other types, in line with evidence from various survey forecasts (see, e.g., Greenwood and Shleifer (2014) versus Dahlquist and Ibert (2022)).

In the following sections, we present theoretical properties and empirical illustrations of our estimator of subjective beliefs for different values of γ. However, in a recent insightful paper, Chen, Hansen, and Hansen (2024) (Theorem 3.2) established a serious degeneracy problem for the Cressie-Read discrepancies with $\gamma < 0$ for the case of unconditional moment restrictions, which includes EL and Hellinger distance. In particular, if the (unconditional) moment function is misspecified and unbounded, the infimum for the population Cressie-Read discrepancy minimization may not be attained by any measure. We also refer to Schennach (2007) (Theorem 1) for another drawback of the EL estimator under misspecified unconditional moment restrictions (i.e., lack of $\sqrt{n}$-consistency of the EL estimator). Although a formal investigation is beyond the scope of this paper, we conjecture that an analogous degeneracy problem will emerge for the conditional moment restrictions considered in this paper. Therefore, although we hereafter present theoretical results and empirical illustrations for any value of γ, including the problematic case of $\gamma < 0$, researchers should be cautious to employ discrepancy measures with $\gamma < 0$.

III Properties of Recovered Beliefs

Kitamura, Tripathi, and Ahn (2004) derive the limiting properties of the SEL estimator under two critical assumptions: (1) correct model specification, i.e. the subjective and objective beliefs coincide, and (2) *i.i.d.* DGP. Both these assumptions are clearly violated in our framework, where the goal is to recover investors' subjective beliefs (implying model

misspecification) about the future given their current information set (implying intertemporal dependence). Also, Kitamura, Tripathi, and Ahn (2004) were primarily concerned with finite-dimensional parameter estimation – the estimation of the SDF parameter θ in our setting. We, on the other hand, are primarily interested in the estimation of the subjective beliefs distribution, alongside the SDF parameters. These critical differences necessitate establishing the asymptotic properties of the smoothed Cressie-Read discrepancy minimizing estimators in Equation (5), in the presence of weakly dependent time series data and model misspecification. In this section, we establish these properties for all the estimators in the Cressie-Read family.

Consider estimation of the subjective conditional probability for the event $\{Z \in A\}$ given $X = c$, where A is any measurable set and c is the value of the conditioning variables. For example, the conditional subjective distribution can be estimated by setting the event as $\{Z \leq z\}$. Our estimator for the subjective conditional probability is defined as

$$\hat{\mathbb{P}}(Z \in A | X = c) = \sum_{j=1}^{T} \hat{p}_j(\hat{\theta}) \mathbb{I}\{Z_j \in A\},$$

where

$$\begin{aligned} \hat{p}_j(\theta) &= w_j [\gamma\{\hat{\mu}(\theta) + \hat{\lambda}(\theta)' g(Z_j;\theta)\}]^{1/\gamma}, \\ w_j &= \frac{\mathcal{K}_j}{\sum_{j=1}^{T} \mathcal{K}_j}, \qquad \mathcal{K}_j = \mathcal{K}\left(\frac{c - X_j}{b_T}\right), \end{aligned}$$

and the Lagrange multipliers $\hat{\mu}(\theta)$ and $\hat{\lambda}(\theta)$ solve

$$\begin{aligned} \sum_{j=1}^{T} w_j [\gamma\{\hat{\mu}(\theta) + \hat{\lambda}(\theta)' g(Z_j;\theta)\}]^{1/\gamma} &= 1, \qquad (10) \\ \sum_{j=1}^{T} w_j [\gamma\{\hat{\mu}(\theta) + \hat{\lambda}(\theta)' g(Z_j;\theta)\}]^{1/\gamma} g(Z_j;\theta) &= 0. \end{aligned}$$

Note that the objects w_j, $\hat{p}_j(\theta)$, $\hat{\mu}(\theta)$, and $\hat{\lambda}(\theta)$ depend on the value c of the conditioning

variables (i.e., the current state i), but we suppress such dependence to simplify the notation.

First of all, the consistency of the point estimator $\hat{\theta}$ for the pseudo true value θ_* (defined in Equation (2) in Appendix A) is obtained as follows. A detailed proof is provided in Appendices A.1 – A.2.

Lemma 1 *Under the assumptions enumerated in Appendix A.1, we have* $\hat{\theta} \xrightarrow{p} \theta_*$.

We note that this consistency result is established under the relatively high-level conditions in Appendix A.1. As we discussed in the last section, the degeneracy problem for the case of $\gamma < 0$ in Chen, Hansen, and Hansen (2024) (Theorem 3.2) indicates that θ_* is typically not well-defined for the case of $\gamma < 0$ with unbounded moment functions (i..e., Assumption (ii) in Appendix A.1 is violated). Thus, we recommend to employ the Creassie-Read family with $\gamma \geq 0$ (e.g., SET), especially for the case of unbounded moment functions.

The asymptotic property of our estimator $\hat{\mathbb{P}}(Z \in A|X = c)$ for the conditional subjective probability is then obtained as follows. Specifically, the recovered subjective beliefs converge to its pseudo true value. A detailed proof is provided in Appendices A.3 – A.4.

Proposition 1 *Under the assumptions enumerated in Appendix A.4, we have*

$$\hat{\mathbb{P}}(Z \in A|X = c) \xrightarrow{p} \mathbb{E}^{\mathcal{P}_0}[[\gamma\{\mu_*(\theta_*) + \lambda_*(\theta_*)g(Z;\theta_*)\}]^{\frac{1}{\gamma}}\mathbb{I}\{Z \in A\}|X = c],$$

for each measurable set A *and* $c \in \mathbb{R}^d$*, where* θ_* *is the probability limit of* $\hat{\theta}$*, and*

$$\{\mu_*(\theta), \lambda_*(\theta)\} = \arg \max_{(\mu,\lambda)\in\Xi} \mathbb{E}^{\mathcal{P}_0}[[\gamma\{\mu + \lambda' g(Z;\theta)\}]^{1/\gamma}|X = c],$$

for a compact set Ξ.

Intuitively, in addition to the strict stationarity and mixing assumptions on the DGP, we require sufficient smoothness of the moment function $g(\cdot)$ and the density function of the conditioning variables X, and the existence of sufficiently higher moments for the observables (X, Z) and the moment function $g(Z;\theta)$ over $\theta \in \Theta$.

By setting $\{Z \in A\}$ as $\{Z \leq z\}$, we obtain an estimator for the conditional distribution and its probability limit as

$$\hat{F}(z|c) = \hat{\mathbb{P}}(Z \leq z|X = c) \xrightarrow{p} F_*(z|c) = \mathbb{E}^{\mathcal{P}_0}[[\gamma\{\mu_*(\theta_*) + \lambda_*'(\theta_*)g(Z;\theta_*)\}]^{\frac{1}{\gamma}}\mathbb{I}\{Z \leq z\}|X = c].$$

Note that, the limiting conditional distribution function $F_*(\cdot|\cdot)$ does not depend on the particular choice of the kernel function, $\mathcal{K}(.)$, in Equation (4). Under suitable regularity conditions, such as the ones in Komunjer and Ragusa (2016), the conditional distribution function $F_*(\cdot|\cdot)$ can be interpreted as the information projection by the Cressie-Read discrepancy (with γ) from the data generating conditional distribution function for Z given $X = c$ to the set of conditional distribution functions satisfying the conditional moment restrictions in Equation (1).

If the conditional moment restrictions are correctly specified (i.e., $\mathcal{P} = \mathcal{P}_0$), then θ_* is considered as the true parameter value, and our estimator $\hat{F}(\cdot|\cdot)$ will converge to the objective data generating distribution function. If, however, the model is misspecified, then θ_* becomes the pseudo-true parameter value, and $\hat{F}(\cdot|\cdot)$ is considered as an estimator of the investors' subjective beliefs distribution function minimizing the Cressie-Read discrepancy.

To understand better the properties of the recovered conditional distribution, consider the quadratic divergence estimator (SQD), i.e. the Cressie-Read family member with $\gamma = 1$. In this case, the recovered conditional distribution has a closed-form solution. Specifically, the conditional probabilities are explicitly written as (see Smith (2007), Theorem 3.1):

$$\hat{p}_j(\theta) = w_j[1 + \hat{\lambda}(\theta)'\{g(Z_j;\theta) - \hat{g}(\theta)\}],$$

where the Lagrange multipliers have a closed-form solution: $\hat{\lambda}(\theta) = -[\hat{V}(\theta) - \hat{g}(\theta)\hat{g}(\theta)']^{-1}\hat{g}(\theta)$, $\hat{g}(\theta) = \sum_{j=1}^{T} w_j g(Z_j;\theta)$, and $\hat{V}(\theta) = \sum_{j=1}^{T} w_j g(Z_j;\theta)g(Z_j;\theta)'$. Therefore, in this case

$$F_*(z|c) = F_0(z|c)\{1 + g_*'[V_* - g_*g_*']^{-1}g_*\} - g_*'[V_* - g_*g_*']^{-1}g_{1*},$$

where

$$\begin{aligned} F_0(z|c) &= \mathbb{E}^{\mathcal{P}_0}[\mathbb{I}\{Z \leq z\}|X = c], \\ g_* &= \mathbb{E}^{\mathcal{P}_0}[g\left(Z_j;\, \theta_*\right)|X_t = c], \\ V_* &= \mathbb{E}^{\mathcal{P}_0}\left[g\left(Z_j;\, \theta_*\right) g\left(Z_j;\, \theta_*\right)'|X_t = c\right], \\ g_{1*} &= \mathbb{E}^{\mathcal{P}_0}[g\left(Z_j;\, \theta_*\right) \mathbb{I}\{Z \leq z\}|X_t = c]. \end{aligned}$$

If $\mathcal{P} = \mathcal{P}_0$, i.e. the conditional moment restrictions are correctly specified, then $\theta_* = \theta_0$, the true value of the parameter vector. Also, in this case, $g_* = \mathbb{E}^{\mathcal{P}_0}[g\left(Z_j;\, \theta_0\right)|X_t = c] = 0$. Therefore, we have $F_*(z|c) = F_0(z|c)$, i.e. the subjective beliefs coincide with the objective DGP and the estimation approach recovers this DGP. If, however, the model is misspecified, then $g_* \neq 0$ and the recovered beliefs are distorted relative to the objective DGP. Note that, because of the quadratic form of the divergence function, the distortions relate only to the mean and the covariance matrix of the vector of sample moments. For the SET and SEL estimators, on the other hand, while no closed-form solutions exist for the Lagrange multipliers, the beliefs distortions depend on *all* the moments, not just the mean and covariance matrix, of the vector of sample moments.

The recovered beliefs have several desirable properties. First, as shown in Proposition 1, they converge in probability to pseudo-true subjective beliefs. While the pseudo true may not equal the (unknown) true beliefs, in the absence of knowledge of the latter, the pseudo true distribution provides an estimate that relies on far fewer assumptions than typically needed in structural models. Second, for each choice of the Cressie-Read family member characterizing the divergence between the objective and subjective conditional distributions, a *unique* subjective distribution is recovered as the point estimate of beliefs. Third, the approach enables the simultaneous estimation of the subjective beliefs and the SDF parameters characterizing the risk appetite of investors, the only assumption needed being the functional form of the SDF. Finally, Ghosh and Roussellet (2019) provide extensive simu-

lation evidence to show that the proposed estimator performs well in recovering beliefs in empirically realistic sample sizes.

Finally, we note that our consistency results in Lemma 1 and Proposition 1 are derived under the assumptions of uniqueness of the pseudo true values θ_*, $\mu_*(\theta_*)$, and $\lambda_*(\theta_*)$ (see Assumption (ii) in Appendix A.1). However, recent papers report that for misspecified unconditional moment restriction models, some members of the Cressie-Read family (those with negative power coefficient γ in Equation (5)), like the EL, may violate such assumptions and lose their consistency. Chen, Hansen, and Hansen (2020) highlight potential degeneracy issues with the implied unconditional distributions for Cressie-Read members such as EL with negative power coefficients, but not for members with positive power coefficients like the ET, when the data has unbounded support (e.g., returns, excess returns). Chauduri, Renault, and Wahlstrom (2023) argue the lack of consistency of the estimated unconditional probability distribution in the presence of potentially unbounded disasters, for Cressie-Read members with negative power coefficients – a shortcoming not shared by members with positive power coefficients. In Section IV, we compare empirically the performance of Cressie-Read members with both positive and negative power coefficients in a broad range of settings, including in data samples with severe disasters.

IV Empirical Results

We illustrate our methodology to estimate investors' risk preferences and subjective beliefs in quarterly US data over 1972:$Q$1–2018:$Q$4.[7] We present results for the three most widely used members of the Cressie-Read family: $\gamma = 0$ in Equation (5) that delivers the SET estimator, $\gamma = -1$ that gives the SEL estimator, and $\gamma = 1$ that gives the SQD estimator. Section IV.1 describes the data and implementation details. Section IV.2 presents the empirical results.

[7] Very similar results are obtained over the entire post war sample period 1947:$Q$1–2018:$Q$4. These results are available from the authors upon request.

IV.1 Implementation Details

In order to disentangle risk preferences from beliefs, our methodology requires the parametric specification of an SDF summarizing the risk appetites of investors. Investor types differing from one another in terms of their risk preferences and/or beliefs have different optimal portfolio holdings. Based on this insight, we assume that the SDF of the representative type-l investor is given by

$$M_{t+1}^{(l)} = e^{-\theta^{(l)} \log\left(R_{W,t+1}^{(l)}\right)}, \tag{11}$$

where $R_{W,t+1}^{(l)}$ is the investor's gross return on total wealth and $\theta^{(l)}$ her risk aversion coefficient. Thus, the type-l investor's Euler equations are given by:

$$\mathbb{E}^{\mathcal{P}^{(l)}}\left[M_{t+1}^{(l)} R_{t+1} | \mathcal{F}_t\right] = 1_k. \tag{12}$$

The objective of the estimation is to simultaneously recover the risk preferences and beliefs, $\left(\theta^{(l)}, \mathcal{P}^{(l)}\right)$, for different investor types l.

Our beliefs estimation approach assumes that the SDF, summarizing investors' risk preferences, is correctly specified. If the SDF were misspecified, then the estimated conditional distribution would reflect contributions from both, the beliefs distortions relative to the objective DGP and the misspecification of the SDF. However, even with this caveat, our approach constitutes an advance with respect to the extant structural asset pricing literature. This is because rational models make functional-form assumptions about *both* the SDF and objective DGP (the latter coinciding with investors' beliefs). Behavioral models make additional assumptions about the form(s) of divergence of subjective beliefs from the objective DGP. Our approach, on the other hand, assumes a functional-form for only the SDF, while the objective DGP and forms of beliefs distortions are left fully unspecified.[8]

[8]Our SDF specification has been commonly used in the asset pricing literature. Such a specification, wherein the SDF depends only on terminal wealth, typically holds in single-period settings, such as the CAPM. It also obtains in certain intertemporal settings, such as in those with Epstein-Zin recursive preferences with a unitary risk aversion level (such that the consumption growth component of the SDF drops out). In general intertemporal frameworks, however, the SDF typically (also) depends on consumption growth.

Considering various optimal portfolios, $R_W^{(l)}$, in Equation (11), representative of different investor types, enables the recovery of their, potentially heterogeneous, preferences and beliefs. Note that, our method does not, apriori, restrict either risk preferences or beliefs to be heterogeneous across different investor types. Instead, the approach offers a data driven way of recovering the risk aversion parameter and the beliefs of each investor type such that their optimal portfolio choice is consistent with their preferences and beliefs, i.e. their Euler equations are satisfied. Whereas Ghosh, Korteweg, and Xu (2019) recover the beliefs for different investor types using the SEL approach alone, the contribution of this paper is to offer a comparison between the beliefs recovered using the SET, SEL, and SQD estimators.

We follow two approaches to characterizing investor types l, i.e., selecting $R_W^{(l)}$ in Equation (11). First, we consider broad investor types based on specific investment 'styles.' Specifically, we consider a representative investor who invests all her wealth in the aggregate market portfolio: $R_W^{(1)} = R^M$. This investor type (hereafter referred to as the 'Market' investor) is of broad interest in financial economics since a voluminous literature attempts to explain the behavior of the market portfolio in a representative agent setting, wherein such an agent buys and holds the aggregate market portfolio. Whereas structural models typically assume a parametric beliefs process – that either coincides with the DGP as in rational models or is distorted relative to the DGP in parametric ways as in behavioral models – our nonparametric methodology offers a data-driven way of identifying such an investor's subjective beliefs. The market investor type also serves as a reference point relative to which the preferences and beliefs of other investor types who tilt their holdings towards specialized portfolios or

The lack of availability of disaggregated consumption data led us to specify the SDF of each investor type as depending on her wealth portfolio return alone.

In terms of the robustness of the results to alternative specifications of the SDF, Ghosh, Julliard, and Taylor (2017) focused on the estimation of the (unconditional) 'missing' components of the SDFs for a broad set of SDF specifications stemming from representative agent consumption-based asset pricing models. Their results highlighted that the estimated 'missing' components were remarkably similar across very different SDF specifications, spanning power utility preferences with a constant coefficient of relative risk aversion, Epstein-Zin recursive preferences, external habit preferences, and preferences with complementarities in consumption of different goods. This analysis was possible in Ghosh, Julliard, and Taylor (2017) because their representative agent setting required only aggregate consumption data. Our setting, that invokes heterogeneity in SDFs and beliefs across different investor types, requires disaggregated consumption data, the absence of which prevents us from conducting a similar robutsness analysis.

investment styles can be assessed.

Among specialized investors, we consider those that allocate heavily to value stocks versus those that invest heavily in momentum stocks. Among value investors, we consider the investor type that invests in big market capitalization and value stocks, i.e. their optimal portfolio is $R_W^{(2)} = R^{BV}$ (hereafter referred to as the 'BV' investor), where BV consists of a value-weighted portfolio of stocks in the top quintiles in terms of both size and book-to-market-equity ratio. We also consider a type that invests in small market capitalization and value stocks, i.e. $R_W^{(3)} = R^{SV}$ (hereafter referred to as the 'SV' investor), where SV consists of stocks in the bottom quintile in terms of size and top quintile in terms of book-to-market-equity ratio. Among momentum investors, we consider a type whose optimal portfolio consists of the zero-cost winner-minus-loser (WML) portfolio (long in stocks that are in the top 30th percentile in terms of past performance and short in stocks that are in the bottom 30th percentile in terms of past performance) alongside the risk free asset: $R_W^{(4)} = R^{WML} + R^F$ (hereafter referred to as the 'MOM' investor). We also consider a momentum investor type who focuses on the long-short momentum strategy constructed from small market capitalization stocks alone: $R_W^{(5)} = R^{WML,S} + R^F$ (hereafter referred to as the 'S-MOM' investor).

We motivate our choices of the specialized investor types from the extant empirical evidence that different investors have distinct investment styles. For instance, value investing has been advocated for over a century, with the celebrated investor, Benjamin Graham, often dubbed as the 'father of value investing.' For the case of momentum investing, we motivate this choice from the observation of return extrapolation extensively documented in the recent empirical literature (see, e.g., Bacchetta, Mertens, and van Wincoop (2009), Amromin and Sharpe (2013), Greenwood and Shleifer (2014), Kuchler and Zafar (2019)). Da, Huang, and Jin (2021) present evidence of extrapolation from stocks' recent past returns and show that such extrapolation is more pronounced among nonprofessionals. Also, other than the equity premium, the value and momentum constitute arguably the two most important sources of

risk premia identified to date and have been documented not only in the United States but in most developed markets as well (see, e.g., Fama and French (2012), that also documents that these premia are higher among small market capitalization stocks). While there are competing rational explanations for the value premium, the momentum effect has proven very difficult to reconcile with rational models. Our methodology is aimed at assessing whether heterogeneous beliefs and risk preferences of different investor types can offer an alternative explanation of the value versus momentum effects. While the assumption of full exposure to the specialized portfolios is rather extreme, we show below that relaxing these extreme allocations, including allowing for allocation across a broad mix of risky equities as well as riskless short term Treasuries, does not change the results qualitatively.

Recall that our methodology requires two other inputs – the set of base assets and the investors' conditioning set. The base assets consist of the excess returns on the market, the SV and BV portfolios, and the W and L portfolios.[9] This ensures that the recovered beliefs and risk preferences of each investor type are consistent with the observed prices of *all* these portfolios, since they, by construction, satisfy the conditional Euler equation restrictions for these portfolios.

For the conditioning set, note that our estimation method, like most non-parametric methods, suffers from the curse of dimensionality, thereby necessitating the choice of a relatively modest number of conditioning variables. We consider two baseline choices of the conditioning set: (a) the averages of the last four quarters' returns on the market, SV, BV, W, and L portfolios, and (b) exponentially-weighted moving averages of the returns on these five portfolios.[10] We also present robustness results when the conditioning set includes variables, such as the market-wide price-dividend ratio and the volatility of the market return, that have been shown to have forecasting lower for stock returns as well as

[9]For the S-MOM investor type, the base assets consist of the excess returns on the market, the SV and BV portfolios, and the W_S and L_S portfolios.

[10]These two choices of the conditioning set are motivated by Greenwood and Shleifer (2014), on the one hand, who report evidence suggesting that investors rely primarily on information from the very recent past when forming their beliefs and Malmendier and Nagel (2011) on the other hand whose findings suggest that even experiences from several decades in the past influence future risk taking behavior.

variables, such as the GDP growth rate and the yield spread between long- and short-term government bonds, that are indicative of the overall state of the economy.

Monthly data on the market portfolio and the various style portfolios are obtained from Kenneth French's data library. The monthly risk free rate is the 1-month TBill rate, also obtained from the same source. Monthly returns within each quarter are compounded to obtain the quarterly returns. Excess returns on these portfolios are obtained by subtracting the risk free rate.

Note that the above investor types were motivated using well-documented heterogeneous investment styles across different investor types. The resulting choices of $R_W^{(l)}$ in Equation (11), however, do not constitute actual portfolio holdings of investors. Ideally, we would want to use actual portfolio holdings of different investor types in their SDFs. However, such data are not available for most investors, particularly over long time periods required for our nonparametric framework. Public pension plans, a type of very large institutional investor, offer a notable exception in that annual data on a large cross section of such funds are available from the turn of the century. Therefore, our second approach to characterizing investor types consists of setting $R_W^{(l)}$ in Equation (11) equal to the actual returns earned by a given public pension plan. The public plans data (PPD) is produced by the Center for Retirement Research at Boston college in partnership with the MissionSquare Research Institute, the National Association of State Retirement Administrators, and the Government Finance Officers Association.[11] Annual data on a large cross section of 247 public pension plans are obtained from this source from 2001 onwards. We extend this data series to start from 1992 by imputing the annual returns over 1992–2000. We impute returns for the years before 2001 from the reported 1-year and 10-year returns. For instance, the imputed return for the year 2000 is obtained as the reported 10-year return from 2000–2009 divided by the compounded 1-year returns from the years 2001 through 2009. Out of the universe of 247 plans, we focus on the 147 with June year ends.

[11] https://publicplansdata.org/

Our nonparametric methodology for the identification of the subjective beliefs distribution, while being robust to misspecification, relies on the availability of long time series of returns on both the total wealth portfolio characterizing the SDF and the base assets. The 1992–2018 annual sample over which pension fund returns data are available does not offer a sufficiently long time series for the accurate estimation of conditional distributions using nonparametric methods. We, therefore, construct factor-mimicking portfolios, or linear clones, for pension fund returns at the quarterly frequency using the following two-step approach.

In the first step, we perform linear regressions, at the annual frequency, of each pension fund's returns on different sets of widely used risk factors. Specifically, we consider combinations of the risk factors for hedge fund returns in Hasanhodzic and Lo (2007) and risk factors for equity markets in Fama and French (2015). Given the relatively short time series of pension fund returns, we search for the smallest set of risk factors that explain well the cross-section of pension fund returns. This leads to the choice of 7 risk factors, namely, (1) USD: the US Dollar Index return; (2) BOND: the return on the Corporate AA Intermediate Bond Index; (3) CREDIT: the spread between the BAA Corporate Bond Index and the Treasury Index; (4) SP500: the S&P 500 total return; (5) SMB: a diversified portfolio long in small market capitalization stocks and short in large market capitalization stocks; (6) RMW: a diversified portfolio long in stocks with robust profitability and short in stocks with weak profitability; and (7) CMA: a diversified portfolio long in stocks that invest conservatively and short in stocks that invest aggressively. This linear multi factor model produces an excellent fit with a median adjusted-R^2 = 94.1%, along with an interquartile range of $[92.1\%, 96.2\%]$, across the cross section of pension funds.

In the second step, we use the above 7 risk factors to construct linear clones for each pension fund returns. We do this by using a given fund's returns to estimate a set of portfolio weights corresponding to the risk factors used in the linear regression. In other words, we perform a linear regression (without intercept) of each fund's returns on the 7 risk factors, and constrain the beta coefficients to sum to unity. This yields portfolio

weights at the annual frequency. We then construct linear clones at the quarterly frequency using quarterly realizations of the risk factors and assuming that the beta estimates at the quarterly frequency equal the corresponding estimates obtained at the annual frequency.[12] We use these linear clones at the quarterly frequency as proxies for the corresponding pension fund returns. Thus, our data corresponds to linear clones of pension funds at the quarterly frequency over 1989:$Q2$–2018:$Q4$.[13]

Finally, our method also requires an estimate of the objective beliefs relative to which subjective beliefs distortions are measured. The objective beliefs are summarized by the kernel density estimator in Equation (4). This necessitates a choice for the kernel function and the bandwidth parameter. We choose an Epanechnikov kernel with bandwidth set to 3.[14]

IV.2 Recovered Preferences and Beliefs

Section IV.2.1 presents the results for the style investors while Section IV.2.3 does the same for the pension fund investors.

IV.2.1 Style Investors

The results are presented in Table 1. Each cell reports two numbers – that obtained with conditioning set (a) on the left and with conditioning set (b) on the right. Panel A considers the Market investor. Rows 1 and 2 show that the SEL and SET estimators deliver similar results for both the risk aversion parameter and the beliefs. Specifically, the estimated risk aversion parameters with the two approaches are 2.3 versus 4.2, respectively, for both choices of the conditioning set. The SET delivers higher estimates of the risk aversion parameter relative to those obtained with the SEL. But, with both these Cressie-Read family members,

[12]This assumption is reasonable in the absence of strong time-variation in the betas across quarters within a year, for which there is little empirical evidence.

[13]Our 1989:$Q2$ start date corresponds to the period from which the returns on all the 7 risk factors used to construct the linear clones are available.

[14]Very similar results are obtained with a Gaussian kernel and with the bandwidth parameter not larger than 4. These are omitted for brevity and are available from the authors upon request.

the estimated risk aversion parameter lies within its economically plausible range. We show later that the SET also implies higher overall levels of the expected market returns relative to the SEL. However, the recovered beliefs are strongly countercyclical with both approaches – the correlations between the time series of the expected market return and a recession dummy are 48.1% versus 47.1%, respectively, and 38.1% versus 39.7%, respectively, for the two conditioning sets; for the correlation between the conditional market return volatility and the recession dummy, the corresponding numbers are 48.1% versus 49.6%, respectively, and 41.4% versus 44.9%, respectively. The cyclicality of the expected return is more pronounced than that of the volatility, causing the expected Sharpe ratio to be countercyclical for both estimators: 34.8% versus 30.1%, respectively, and 26.4% versus 39.6%, respectively, for the two conditioning sets.

However, Panel A, Row 3 shows that the SQD estimator gives markedly different results compared to the SEL and SET estimators. First, the estimated risk aversion coefficients are implausible at 100.0 (hitting the upper bound of the admissible parameter space) for both conditioning sets. Second, the recovered beliefs exhibit much less cyclicality – for the first conditioning set, the correlation between the time series of the expected market return and the recession dummy is 26.3% (close to one-half the magnitudes obtained with the SEL and SET estimators) and that between the conditional Sharpe ratio and the recession dummy is close to 0 at 5.0%. The beliefs also display more sensitivity to the precise choice of the conditioning set, unlike those obtained with the SEL and SET estimators – the correlation between the Sharpe ratio and the recession dummy is 4.5% with the first conditioning set whereas it is an order of magnitude bigger at 18.3% for the second conditioning set.

The results in Panels B and C for the BV and SV investors, respectively, are very similar to those obtained for the market investors in Panel A. Specifically, similar (and economically plausible) estimates of the risk aversion coefficient are obtained with the SEL and SET approaches, whereas the SQD implies unrealistically large values of this parameter. The expected market returns and the market's Sharpe ratio are strongly countercyclical with

Table 1: Beliefs About the Stock Market and Risk Preferences For Different Investor Types

	$\theta^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(\sigma^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$
		Panel A: $R_W = R_M$		
SEL	2.3/2.3	48.1/38.1	48.1/41.4	34.8/26.4
SET	4.2/4.2	47.1/39.7	49.6/44.9	30.1/39.6
SQD	100.0/100.0	26.3/36.1	55.3/48.4	5.0/18.3
		Panel B: $R_W = R_{BV}$		
SEL	2.0/2.0	47.5/38.7	47.6/40.3	35.2/21.2
SET	4.3/4.2	38.8/39.8	49.1/43.1	14.1/25.2
SQD	100.0/99.5	24.7/36.4	50.8/45.2	−1.8/19.2
		Panel C: $R_W = R_{SV}$		
SEL	2.1/2.0	46.6/38.7	48.8/44.0	36.4/26.5
SET	3.1/3.1	41.6/39.9	48.6/44.8	13.9/37.2
SQD	99.6/99.6	25.2/25.1	53.2/48.0	−1.6/0.73
		Panel D: $R_W = R_{WML} + R_F$		
SEL	5.3/5.2	−42.6/ − 34.1	43.2/41.6	−46.9/ − 34.0
SET	11.7/11.5	−43.6/ − 32.2	52.2/45.1	−47.6/ − 33.3
SQD	100.0/100.0	−33.1/ − 16.5	54.4/46.0	−51.5/ − 26.5
		Panel E: $R_W = R_{WML_S} + R_F$		
SEL	6.6/5.1	−19.6/ − 18.3	30.3/33.8	−25.1/ − 13.2
SET	11.9/11.8	−23.3/ − 13.4	47.2/40.5	−29.8/ − 14.3
SQD	100.0/100.0	−8.9/ − 12.1	51.5/44.5	−27.8/ − 22.1

The table reports the point estimates of the SDF parameter $\theta^{(l)}$ (Column 2), and correlations between a NBER recession dummy, I_{bc}, and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. Panels A–E present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value stocks, small market capitalization and value stocks, the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset, and the the zero-cost winner-minus-loser portfolio constructed from small cap stocks alone (WML_S) alongside the risk free asset. Rows 1–3 in each panel present the results for the SEL, SET, and SQD approaches, respectively. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The two choices of the conditioning set include (a) the average returns of the last four quarters of the market, BV, SV, W, and L portfolios (left number in each cell), and (b) exponentially-weighted moving averages of the returns these five portfolios (right number in each cell). The sample is quarterly covering the period 1972:Q1-2018:Q4.

both the SEL and SET approaches. The beliefs recovered with the SQD, on the other hand, exhibit much less countercyclicality in these moments of the market return as well as higher sensitivity to the choice of the conditioning set (e.g., correlation between the Sharpe ratio and the recession dummy of -1.8% versus 19.2% for the two choices of instruments).

Panel D, that presents the results for the MOM investors, once again shows a similar comparison between the three estimators – the SEL and SET produce results close to each other for both conditioning sets whereas the SQD delivers large and implausible estimates of the risk aversion parameter and less cyclicality in the expected market return compared to the former two. Note, however, that, regardless of the choice of the Cressie-Read family member, the recovered beliefs for this investor type exhibit pro cyclicality of the expected market return, counter cyclicality of the conditional volatility of the market return, and, therefore, pro cyclicality of the conditional Sharpe ratio of the market. In fact, the beliefs of momentum investors about the Winner (W) and Loser (L) portfolios separately are also procyclical – e.g., correlations of -56.9% and -58.0%, respectively, for the W and L portfolios with the recession dummy, for the first choice of conditioning set. Very similar results are obtained in Panel E for the S-MOM investors.

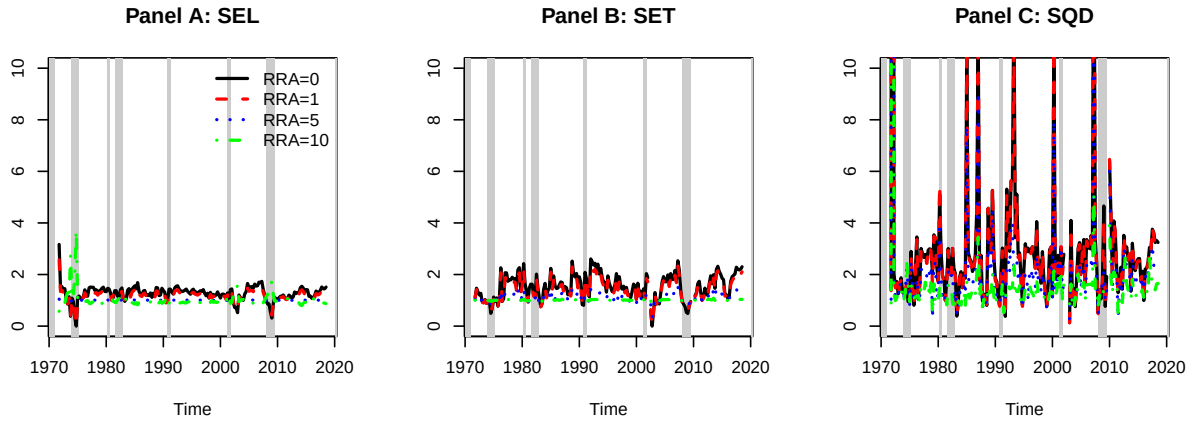
Why does the SEL and SET lead to reasonable estimates of the risk aversion coefficient whereas the SQD produces implausibly large estimates of the parameter? Note that, for each Cressie-Read divergence measure, our methodology delivers a point estimate of the subjective beliefs for each value of the risk aversion parameter, θ. To shed light on the above question, Figure 1 plots, for the momentum (MOM) investor type, the time series of the Cressie-Read divergence between the subjective beliefs – recovered using the SEL, SET, and SQD approaches – and the objective DGP, for a range of values of θ. We consider several values of θ in the economically plausible range, namely, 0, 1, 5, and 10. For each divergence measure, we report the ratio of the Cressie-Read divergence at each of these risk aversion values to the divergence at the point estimate of the parameter. This facilitates comparison between the three divergence measures. The figure makes clear the cause of the

implausibly large point estimates of the risk aversion parameter obtained with SQD, namely that values of the parameter in the economically plausible $0-10$ range lead to several-fold larger divergences between the subjective and objective beliefs. Specifically, the ratio of the divergence at a particular value of θ to that at its point estimate of 100 declines with increase in θ: average ratios of 3.7, 3.1, and 1.8 are obtained at θ =1, 5, and 10, respectively. For the SEL, on the other hand, the divergence at $\theta = 5$ relative to that at its point estimate of $\theta = 5.2$ has an average value of 1.0. As the magnitude of the difference between θ and its point estimate increases, so does the divergence – the average divergence at θ =1 and θ =10 relative to the divergence at the point estimate of $\theta = 5.2$ are 1.2 and 0.99, respectively. The results with the SET are similar to those obtained with the SEL.[15]

Figure 2 plots the time series of the expected market returns recovered using the SEL (black-solid line), SET (red-dashed line), and SQD (blue-dotted line) approaches for three of the investor types considered. Panels A–C present the results for the market, BV, and MOM investors, respectively. Immediately obvious from the figure is the counter cyclicality of the beliefs about the expected market returns for the market and BV investors (Panels A–B): the expected market returns are high during recessionary episodes (shaded rectangles) and low during the expansionary phase of the business cycle. The beliefs of the MOM investors (Panel C), on the other hand, are pro cyclical: their expected market returns being high during expansions and low during recessions. Our results suggest that this conclusion is robust to the particular choice of the Cressie-Read family of statistical divergence measure used to recover these investors' beliefs. These findings help reconcile the seemingly contradictory evidence, documented in Greenwood and Shleifer (2014), of procyclical expected market returns found in individual investor survey data versus the countercyclical expected returns

[15]Table 5 in Appendix B reports the estimates of subjective beliefs obtained with the SEL, SET, and SQD approaches, when the risk aversion parameter is set to unity. The results shows that, with unitary risk aversion, the SQD-implied beliefs are close to those obtained with the SEL and SET approaches (particularly the SET), in terms of the correlations between the recovered beliefs and the business cycle. This is in contrast to the beliefs recovered at the point estimates of the risk aversion parameter in Table 1, where the SEL and SET produce similar beliefs estimates but these estimates are markedly different from those obtained using the SQD approach. These results point to the undesirability of using the SQD relative to the SEL or SET.

Figure 1: Cressie-Read Divergence Between Subjective and Objective Beliefs at Alternative Risk Aversion Levels

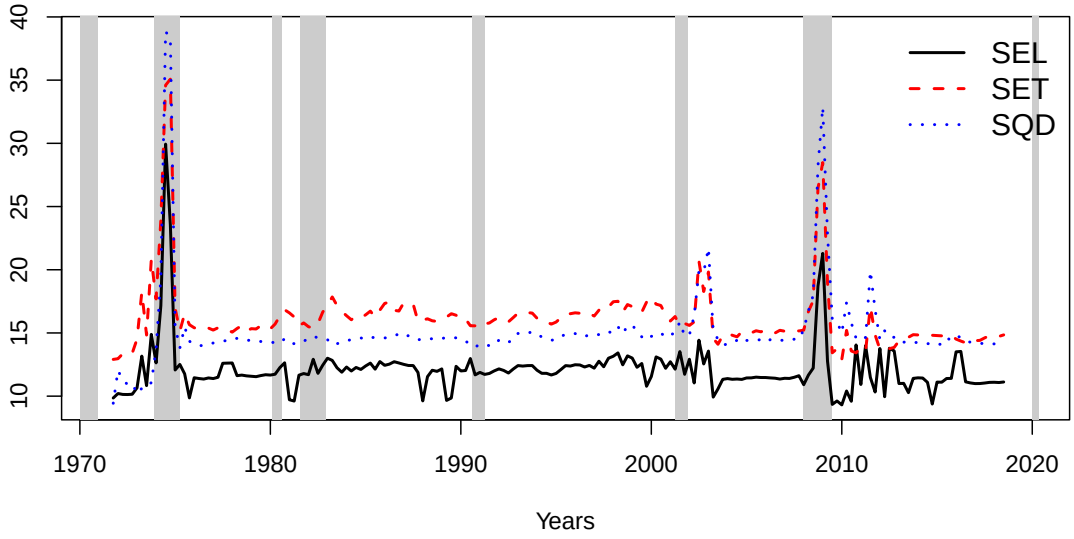


Note: The figure plots, for the momentum (MOM) investor type, the time series of the ratio of the Cressie-Read divergence between the subjective beliefs – recovered using the SEL (Panel A), SET (Panel B), and SQD (Panel C) approaches – and the objective DGP, for a range of values of the risk aversion parameter between 0-10 and the point estimate of the parameter. The vertical shaded areas denote NBER recession episodes. The pricing kernel is exponentially affine in the MOM investors' chosen portfolio, namely the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The conditioning set includes exponentially-weighted moving averages of the returns of these five portfolios. The sample is quarterly covering the period 1972:Q1-2018:Q4.

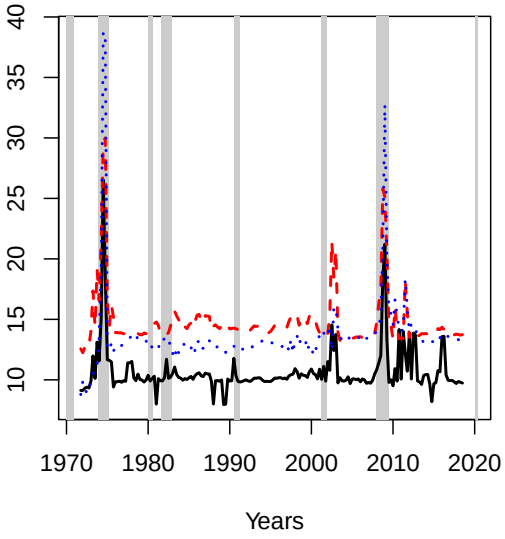
implied by rational expectations models. The persistent differences in beliefs across different investor types, an implicit assumption in our approach, is also consistent with the recent survey-based evidence in Giglio, Maggiori, Stroebel, and Utkus (2021).

The close similarities between the beliefs recovered using the SEL and SET estimators for all investor types considered and the more substantive differences between these beliefs and those recovered using the SQD estimator are also evident from Table 2. This table reports the pairwise correlations between the time series of the expected market returns obtained using the three estimators for the different investor types. The correlations between the expected returns recovered using the SEL and SET approaches varies from 80.2%–86.3%, when the first conditioning set is used, corresponding to an average of 85.0%. The corresponding average correlation between the SEL and SQD estimators is only 48.4% and that between the SET

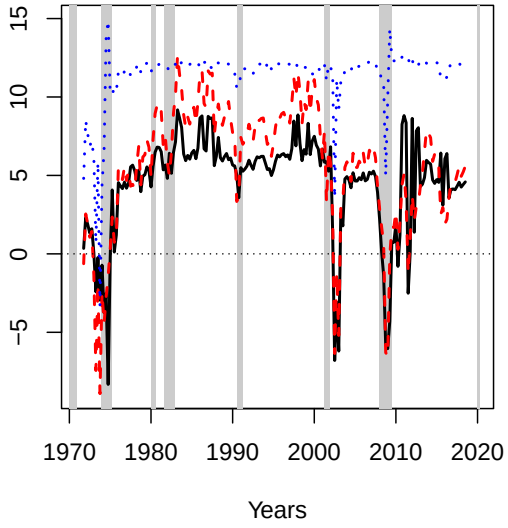
Figure 2: Beliefs About Expected Market Returns



Panel A. Market Investors



Panel B. BV Investors



Panel C. MOM Investors

Note: The figure plots the time series of the expected market return for three different investor types, obtained with the SEL (black-solid line), SET (red-dashed line), and SQD (blue-dotted line) approaches. Panels A–C present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value (BV) stocks, and the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset. Each panel present the results for the SEL, SET, and SQD approaches, respectively. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The conditioning set includes exponentially-weighted moving averages of the returns of these five portfolios. The sample is quarterly covering the period 1972:Q1-2018:Q4.

Table 2: Correlations Between the Beliefs of Different Investor Types

	Market	BV	SV	MOM	MOM-S
		$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$			
SEL & SET	84.0/87.9	83.4/83.0	86.3/91.3	91.2/88.1	80.2/52.1
SEL & SQD	65.1/81.4	61.4/81.8	70.7/74.6	41.3/49.5	3.7/26.4
SET & SQD	81.4/87.0	78.8/82.4	76.5/73.8	36.6/62.3	2.2/76.7

This table reports the pairwise correlations between the time series of the expected market returns obtained using the SEL, SET, and SQD estimators. The investor types consist of those who allocate to the market portfolio (Column 2), large market capitalization and value stocks (Column 3), small market capitalization and value stocks (Column 4), the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset (Column 5), and the the zero-cost winner-minus-loser portfolio constructed from small cap stocks alone (WML_S) alongside the risk free asset (Column 6). For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The conditioning set includes either the averages of the last four quarters' returns on the market, BV, SV, W, and L portfolios (left number in each cell) or exponentially-weighted moving averages of the returns of these portfolios (right number in each cell). The sample is quarterly covering the period 1972:Q1-2018:Q4.

and SQD only 55.1%. The average correlations between the SQD-implied expected market returns and those obtained with the SEL and SET improves with the second conditioning set, but still remains smaller than the average correlation between the SEL and SET.

Note that our approach recovers the *entire conditional distribution* of the market return, as perceived by the different investor types. In order to compare further the beliefs recovered using the three estimation approaches, Figure 3 plots the one quarter ahead conditional densities of the (log) market return recovered using the SEL (black-solid line), SET (red-dashed line), SQD (blue-dotted line) approaches for two extreme states: (i) 2008:Q3, that corresponds to a bad state characterized by low economic growth and a severe stock market downturn (Panels A–C), and (ii) 1998:Q3, that corresponds to a good state characterized by high economic growth and unprecedented stock market boom (Panels D–F).

The figure shows that the SEL and SET densities are very close to, and often indistinguishable from, each other for both states and for all investor types. The SQD density,

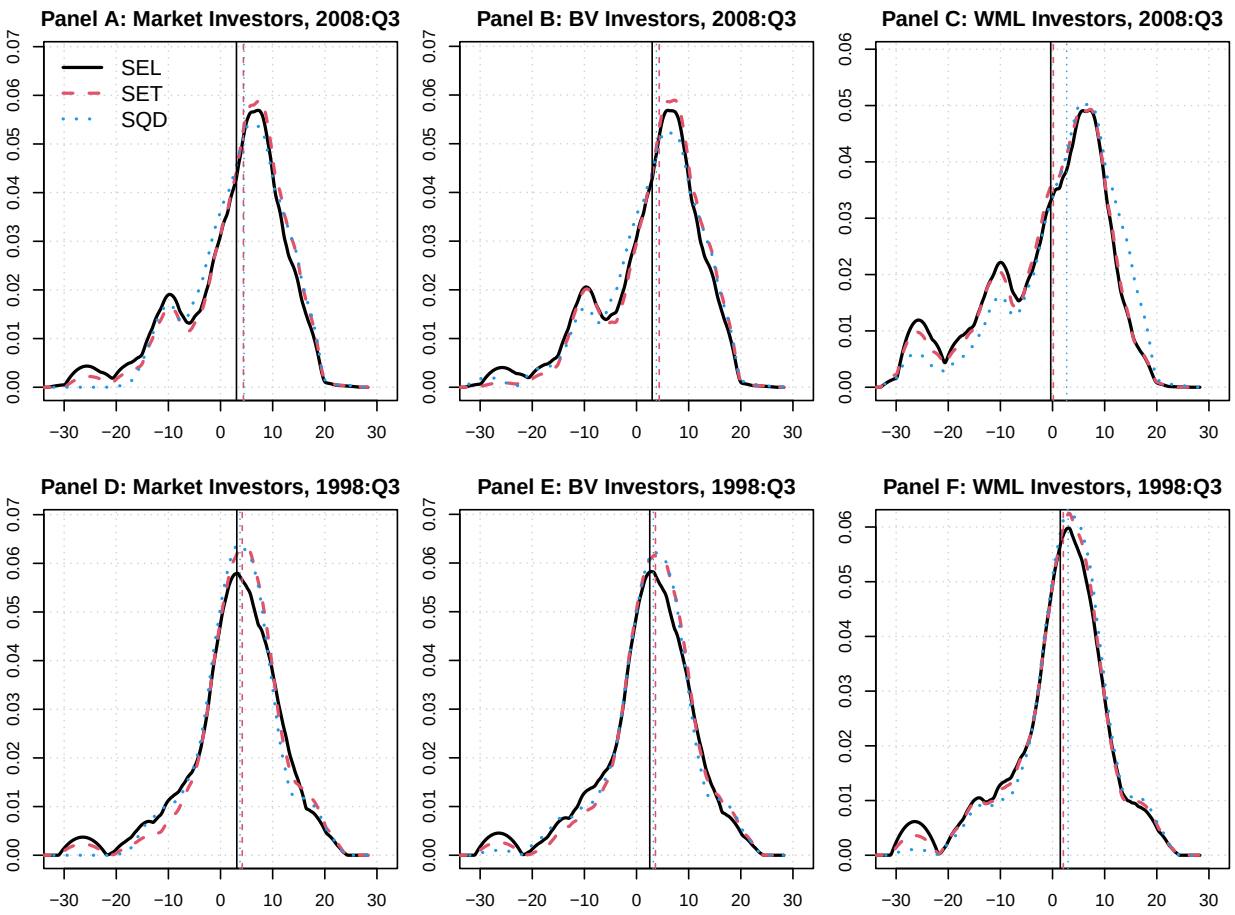
on the other hand, systematically underweights the left tail relative to the SEL and SET densities. Consider, for instance, the MOM investors (Panels C and F for the bad and good states, respectively). In the bad state, the SEL density assigns a probability mass of 10.3% to events in the left 2% tail of the realized market returns (this corresponds to a quarterly log market return below -17.8%). The SET density assigns a very similar probability weight of 9.0% to the bottom 2% left tail. The SQD approach, on the other hand, assigns a weight of only 5.8% to such tail risk – only about half of those obtained with the SEL and SET estimators. Similar conclusions are obtained for the good state – the probabilities assigned to the bottom 2% realized market returns are 3.7%, 2.2%, and 0.64%, respectively, for the SEL, SET, and SQD estimators.

IV.2.2 Robustness

We first show that our conclusions about the risk preferences and beliefs of the different investor types are robust to specifications of the conditioning set. In Appendix B, Table 6, the conditioning set includes the aggregate log price-dividend ratio and the realized variance of the market return as instruments, in addition to the market return and the returns on the momentum winner and loser portfolios. In other words, the conditioning set replaces the returns on the BV and SV portfolios from the baseline choice with the log price-dividend ratio and the realized variance of the market return.[16] The results in Table 6 are similar to those obtained with our baseline choice of instruments in Table 1 – namely that investors allocating heavily to momentum-type strategies have procyclical beliefs about the expected market return and its Sharpe ratio, whereas those allocating to other equity strategies such as large market cap and/or value stocks have countercyclical beliefs. In Table 7, the conditioning set includes the GDP growth rate and the term spread as instruments, in addition to the market return and the returns on the momentum winner and loser portfolios. The results in Table

[16]The quarterly price-dividend ratio is calculated as the ratio of the price at the end of the quarter and the sum of the dividends over the past year. The realized variance is obtained as the sample variance of the monthly market returns over the past two years.

Figure 3: Beliefs About the Market Returns Distribution



Note: The figure plots the one quarter ahead conditional densities of the (log) market return recovered using the SEL (black solid line), SET (red-dashed line), and SQUAD (blue-dotted line) approaches for two states: (a) 2008:Q3 in Panels A–C, and (b) 1998:Q3 in Panels D–F. Panels A (D)–C (F) present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value (BV) stocks, and the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset. The vertical lines in each panel are the expected market returns recovered with the three estimators. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The conditioning set includes exponentially-weighted moving averages of the returns of these five portfolios. The sample is quarterly covering the period 1972:Q1-2018:Q4.

7 are, once again similar to those obtained with our baseline choice of instruments. Overall, the results suggest that our main findings are largely robust to alternative specifications of the conditioning set.

Next we assess whether our main conclusions about subjective beliefs estimates are robust

across non-overlapping sub periods. Since we focus on investor types based on 'anomaly' portfolios, this raises a natural question given the observation that anomaly performance has generally degraded over the past 20–30 years. Table 8 in Appendix B reports the estimated risk aversion parameter and beliefs of the momentum (MOM) investors, over the three non-overlapping 25-year sub periods since 1947:Q1. The table shows that the procyclical beliefs of the momentum investors, i.e. the negative correlations between the recession dummy and the expected aggregate stock market returns (and the Sharpe ratio of the market), obtain in all sub periods. The magnitudes of the negative correlations remain almost identical across all sub periods for the SET. For the SEL, the magnitudes of the negative correlations decline in the third sub period relative to the first two (only for the baseline conditioning set (a) but not for set (b)), but still remains substantive. We also note that the momentum effect remains statistically significant in all three sub periods, although its magnitude is diminished in the last sub period.[17]

Finally, the analysis in Section IV.2.1 focused on investor types with single corner portfolios as their total wealth portfolios. We next show that our main empirical finding on beliefs heterogeneity obtains even when the investors' wealth portfolios are expanded beyond the single corner portfolios to include a much broader mix of risky equities. We consider investors who allocate simultaneously to the five equity strategies in Fama and French (2015) as well as to the momentum WML strategy. Specifically, we consider a 50-50 base mix of the market portfolio and short term Treasuries, along with various levels of exposure to the SMB (long-short factor mimicking portfolio proxying for risk related to size), HML (long-short factor mimicking portfolio proxying for risk related to the book-to-market-equity ratio), RMW (long-short factor mimicking portfolio proxying for risk related to operating profitability), CMA (long-short factor mimicking portfolio proxying for risk related to investment), and

[17] Over the first sub period (1947:Q1-1971:Q4), the WML premia is 9.7% annualized and is highly economically and statistically significant. Over the second sub period (1972:Q1-1996:Q4), the WML premia remains similar in magnitude, at 10.6% annualized, and highly significant. In the most recent sub period (1997:Q1-2018:Q4), the WML premia reduces by about half to 4.4% annualized and loses statistical significance, but still remains economically sizeable.

WML strategies. We estimate the risk aversion levels and beliefs about the market returns for such investors using the three Cressie-Read family members. The test assets consist of the excess market returns and the returns on the SMB, HML, RMW, CMA, and WML portfolios. The conditioning set consists of the averages of the last four quarters' returns on these six portfolios.

The results are reported in Table 3. Panel A presents the results obtained with the SEL. Row 1 shows that a 100% exposure to each of the SMB, HML, RMW, CMA, and WML strategies leads to an estimated risk aversion coefficient of 3.5 and procyclical beliefs about the market return – the correlation between the expected market returns and the recession dummy is -18.0% and that between the market's Sharpe ratio and the recession dummy is -27.4%. Row 2 shows that this procyclicality is driven by the momentum (WML) portfolio, i.e. removal of the momentum exposure of the portfolio leads to countercyclical beliefs with correlation coefficients of 27.3% and 23.9%, respectively, for the expected returns and Sharpe ratio with the recession indicator. Rows 3–6 show that the above results, namely procyclical beliefs in the presence of the momentum exposure and countercyclical beliefs in its absence, still obtain when the magnitudes of exposure to all the long-short strategies are reduced. Not surprisingly, the extent of procyclicality is reduced with decrease in the momentum exposure of the portfolio.

Panel B presents the results obtained with the SET. The results are qualitatively similar to those obtained with the SEL in Panel A. In fact, the SET produces even stronger procyclical beliefs for investors with momentum exposure, in the sense that the correlations between the recession indicator and both the subjective expected market return and the market's Sharpe ratio are larger in magnitude and more negative, compared to those obtained with the SEL. The SET estimates larger risk aversion coefficients relative to the SEL for investors with or without momentum exposure, reminiscent of the results obtained with the investors with single corner portfolios in Table 1.

Finally, Panel C shows that the SQD produces implausibly large estimates of the risk

Table 3: Risk Preferences and Stock Market Beliefs for Investors with Optimal Portfolio: $R_W = x_1 R_{WML} + x_2 R_{SMB} + x_3 R_{HML} + x_4 R_{RMW} + x_5 R_{CMA} + 0.5(R_M - R_F)$

	$\gamma^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)$	$\rho\left(Vol^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)$
		Panel A: SEL		
$x_1 = x_2 = ... = x_5 = 1.0$	3.5	−18.0	43.2	−27.4
$x_1 = 0, x_2 = ... = x_5 = 1.0$	2.8	27.3	36.9	23.9
$x_1 = x_2 = ... = x_5 = 0.5$	6.3	−8.9	43.6	−24.1
$x_1 = 0, x_2 = ... = x_5 = 0.5$	4.1	30.4	36.7	26.8
$x_1 = x_2 = ... = x_5 = 0.3$	6.9	6.4	43.5	−14.0
$x_1 = 0, x_2 = ... = x_5 = 0.3$	4.8	32.1	35.9	28.9
		Panel B: SET		
$x_1 = x_2 = ... = x_5 = 1.0$	4.5	−22.7	43.2	−33.0
$x_1 = 0, x_2 = ... = x_5 = 1.0$	5.1	17.0	36.9	5.11
$x_1 = x_2 = ... = x_5 = 0.5$	7.8	−16.5	43.6	−35.0
$x_1 = 0, x_2 = ... = x_5 = 0.5$	7.5	25.7	36.7	9.43
$x_1 = x_2 = ... = x_5 = 0.3$	10.5	−2.29	46.8	−34.5
$x_1 = 0, x_2 = ... = x_5 = 0.3$	8.8	30.2	44.8	15.9
		Panel C: SQD		
$x_1 = x_2 = ... = x_5 = 1.0$	99.9	−25.8	43.2	−27.0
$x_1 = 0, x_2 = ... = x_5 = 1.0$	100	−1.91	40.3	−10.2
$x_1 = x_2 = ... = x_5 = 0.5$	100	−19.7	41.7	−21.4
$x_1 = 0, x_2 = ... = x_5 = 0.5$	100	19.7	39.8	−2.89
$x_1 = x_2 = ... = x_5 = 0.3$	100	15.0	29.2	7.74
$x_1 = 0, x_2 = ... = x_5 = 0.3$	100	28.4	32.5	8.93

The table reports the point estimates of the SDF parameter $\gamma^{(l)}$ (Column 2), and correlations (%) between a NBER recession dummy, I_{rec}, and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. Panel A-C present results for the SEL, SET, and SQD approaches, respectively. The SDF is assumed to be exponentially affine in the return on total wealth that consists of a 50-50 base mix of the market portfolio and short-term Treasuries, along with varying degrees of exposure to long-short factor-mimicking portfolios proxying for risk factors related to size (SMB), boob-to-market-equity ratio (HML), operating profitability (RMW), investment (CMA), and momentum (WML). The test assets consist of the excess market returns and the returns on the SMB, HML, RMW, CMA, and WML portfolios. The conditioning set consists of the averages of the last four quarters' returns on these six portfolios. The sample is quarterly covering the period 1972:Q1-2018:Q4.

aversion coefficient in all cases. Also, it estimates beliefs that often correlate poorly with those obtained with the SEL and SET approaches. These features of the SQD estimator are, once again, similar to those obtained earlier in Table 1.

IV.2.3 Pension Fund Investors

Next we turn to the estimation of beliefs and risk aversion levels for pension fund investors. We use the SEL, SET, and SQD approaches to recover the subjective beliefs and risk aversion levels for each pension fund. In line with the specifications of the SDFs of the different investor types in Section IV.2.1, the SDF of a pension fund is assumed to be exponentially affine in the return on its linear clone. Given the shorter sample (1989:$Q2$–2018:$Q4$) over which quarterly returns on the linear pension fund clones are available, our base assets consist of the returns on the market portfolio and the return on the linear pension fund clone only.[18] The conditioning set consists of either the averages of the last 4 quarters' returns on these two base assets or exponentially-weighted moving averages of the returns on the two assets.

The results are presented in Table 4, Panel A. Column 2 reports the average estimated risk aversion level, across the cross section of pension funds, along with its 95% confidence interval (reported in square brackets below).[19] With the SEL approach, the average risk aversion is 5.9 and 3.9, respectively, for the two choices of the conditioning set. The corresponding estimates obtained with the SET are higher at 9.1 and 6.4, respectively. With both estimators, the confidence intervals are very tight. As in Table 1, the SET tends to produce larger estimates of the risk aversion parameter relative to the SEL. But both approaches lead to estimates in the economically plausible range.

[18] If the set of base assets is expanded to include the market, BV, SV, W, L, and the pension fund clone and the conditioning set is similarly expanded to include past returns on these six assets, then, for a subset of the pension funds and in a subset of possible states (e.g., the 2008-09 financial crisis), the estimation approach fails to satisfy the Euler restrictions. Excluding the funds where such numerical instability is encountered, the results are qualitatively similar to those reported in Table 4. These unreported results are available from the authors upon request.

[19] The 95% confidence intervals are obtained as: $\bar{x} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}$, where $\bar{x}$ denotes the mean of the estimates of x across the funds, σ_x denotes the standard deviation of the estimates of x across the funds, and n the number of funds.

Table 4: Risk Preferences and Stock Market Beliefs of Pension Funds

	$\gamma^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)\%$	$\rho\left(Vol^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)\%$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{rec}\right)\%$
		Panel A: $R_W = R_{PF}$		
SEL	5.92/3.93	54.8/50.2	63.4/57.5	50.8/46.3
	[5.38,6.46]/[3.47,4.39]	[51.0,58.7]/[46.4,54.1]	[62.3,64.4]/[56.5,58.4]	[46.8,54.7]/[42.1,50.4]
SET	9.07/6.35	59.6/56.9	64.7/58.8	61.7/57.8
	[7.77,10.38]/[5.65,7.06]	[56.1,63.1]/[54.4,59.5]	[64.2,65.3]/[58.3,59.2]	[60.9,62.6]/[57.5,58.2]
SQD	86.95/90.11	26.3/21.8	64.2/58.9	$-22.5/-27.3$
	[81.72,92.18]/[85.50,94.72]	[21.9,30.7]/[17.0,26.6]	[63.7,64.7]/[58.3,59.4]	[−28.8,−16.2]/[−33.0,−21.7]
		Panel B: $R_W = R_M$		
SEL	2.98/2.29	56.0/52.5	62.2/57.6	49.8/46.1
	[2.85,3.12]/[2.09,2.50]	[53.0,59.0]/[49.7,55.3]	[60.8,63.7]/[56.7,58.5]	[45.9,53.8]/[42.3,50.0]
SET	4.92/3.99	60.9/58.0	58.5/59.3	62.2/58.4
	[4.49,5.35]/[3.56,4.41]	[57.8,64.0]/[55.9,60.1]	[63.2,65.6]/[58.3,59.2]	[61.1,63.3]/[58.0,58.8]
SQD	92.85/94.57	22.1/24.9	63.9/58.5	$-21.5/-23.1$
	[89.17,96.52]/[91.22,97.91]	[16.9,27.3]/[19.9,30.0]	[63.4,64.4]/[57.9,59.1]	[−27.2,−15.7]/[−28.1,−18.0]

The table reports the point estimates of the SDF parameter $\gamma^{(l)}$ (Column 2), and correlations between a NBER recession dummy, I_{rec}, and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. The SDF is exponentially affine in the return on a pension fund's linear clone (Panel A) or the market portfolio (Panel B). The base assets are the returns on the market portfolio and the return on the pension fund's linear clone. The conditioning set consists of the averages of the last 4 quarters' returns on these two base assets (left number in each cell) or exponentially-weighted moving averages of the returns on the two assets (right number in each cell). The sample is quarterly over $1989:Q2$–$2018:Q4$.

To facilitate comparison, Table 4, Panel B presents results for the market investor over the same sample period and for the same choices of base assets and conditioning set as in Panel A. In other words, the only difference between Panels A and B lies in the SDF – the SDF depends on the return of a linear pension fund clone in Panel A whereas, in Panel B, the SDF depends on the return on the market portfolio. The table shows that the market investor is much less risk averse than the average pension fund investor. Specifically, with the first (second) conditioning set, the SEL and SET produces average risk aversion coefficient estimates of 3.0 (2.3) and 4.9 (4.0), respectively, for the market investor versus 5.9 (3.9) and 9.1 (6.4), respectively, for the average PF investor. Pension Funds are expected to be quite risk averse because of obligations to meet their pension liabilities and the estimated risk aversion coefficients are in line with this premise. The tight confidence intervals are also in line with expectation – the PFs are similar in terms of their risk appetite.

Columns 3–5 show that the pension funds' beliefs about the expected market returns, obtained with the SEL and SET approaches, are strongly countercyclical. The correlation between the time series of the expected market returns and a recession dummy has averages of 54.8% and 59.6%, respectively, for the first conditioning set and 50.2% and 56.9%, respectively, for the second conditioning set. The associated 95% confidence intervals are very tight. The conditional volatility of the market return and the market's conditional Sharpe ratio are also strongly countercyclical, and the SEL and SET approaches generate very similar correlations of these conditional moments with the recession dummy. Panel B shows that very similar cyclical properties of beliefs obtain with the SEL and SET approaches for the market investor.

As with the other (style) investor types in Table 1, the SQD approach produces markedly different results compared to the SEL and SET. For the pension fund investors, Panel A shows that the average estimated risk aversion coefficient is implausibly large at 86.7 and 90.1, respectively, for the two conditioning sets. The median risk aversion coefficient across the funds hits the upper bound of 100 for both choices of the conditioning set. Similar results

obtain for the market investor in Panel B. The cyclical properties of the expected market returns obtained by the SQD estimator are also starkly different from those obtained with the SEL and SET approaches. Specifically, the SQD implies procyclical Sharpe ratio of the market return for both conditioning sets and both investor types, in stark contrast with the strongly countercyclical Sharpe ratio obtained with the SEL and SET approaches. The expected market return implied by the SQD estimator is countercyclical, like those obtained with the SEL and SET approaches, but the magnitudes of the correlations with the NBER recession dummy are less than half of those produced with the SEL and SET estimators.

Overall, this section compares empirically the beliefs and risk preferences recovered using the three most popular members of the smoothed Cressie-Read family. The SQD estimator suffers from two major shortcomings. First, it leads to economically implausible estimates of the risk aversion parameter that often hits the upper bound of the admissible parameter space instead of interior solutions. Second, it leads to negative subjective probabilities for up to 5% of states, particularly those in the tails of the distributions of the variables of interest. Furthermore, it underestimates left tail risk relative to the SEL and SET estimators and delivers estimates of the risk aversion parameter and subjective beliefs that are often substantially different from those obtained with the SEL and SET estimators.

The SEL and SET estimators do not suffer from the above shortcomings. In our empirical applications, the SEL and SET approaches estimate similar beliefs and risk aversion levels. While these empirical results are obtained in a conditional setting, whereby the primary object of interest is the subjective conditional distribution of stock returns, earlier literature reported even greater degrees of similarity between the results obtained in unconditional settings with the empirical likelihood (EL) and exponential tilting (ET) estimators. Specifically, Julliard and Ghosh (2012) show that estimates of the unconditional distributions of stock returns as well as risk aversion levels obtained with the EL and ET approaches are almost indistinguishable from each other in long samples dating back to 1890 that include disaster

episodes such as World Wars I and II, the World War II aftermath, and the Great Depression. Note that the latter two disasters are two out of the 65 major rare economic disasters of the twentieth century identified in Barro (2006). In addition, Ghosh, Julliard, and Taylor (2017) show that the EL and ET approaches deliver near identical 'missing' components of the SDF in a broad class of representative agent consumption-based asset pricing models. These results obtain despite the data consisting of variables, such as gross returns and excess returns, with theoretically unbounded support. To summarize, the theoretical issues with the EL-type estimators highlighted in recent literature do not seem to arise empirically, at least for the range of applications summarized above. Identification of empirical scenarios where the theoretical shortcomings of the EL estimator would get realized is outside the scope of the present paper and constitutes an interesting area for future research. Nevertheless, in light of the theoretical shortcomings of the EL/SEL approach highlighted in the literature and the shortcomings of the QD/SQD approach we document in the present manuscript, the ET/SET offers a promising candidate for the estimation of subjective beliefs.

V Rationale Behind the Subjective Beliefs Estimator and Extensions

Our methodology enables the identification of a point estimate of investors' subjective beliefs. This is accomplished using the following steps. First, we use the historical sample of the conditioning variables to approximate the objective DGP, $F_0(z|c)$, with a nonparametric kernel density estimator. Second, to disentangle risk preferences from beliefs, we assume a functional form for the SDF that summarizes investors' risk preferences. Third, given the objective beliefs and the SDF, the subjective beliefs are estimated as the distribution, $F_*(z|c)$, that is minimally divergent from the objective one, $F_0(z|c)$, subject to the constraint that the estimated subjective beliefs satisfy the Euler equations for the test assets.

Since different functions (corresponding to different members of the Cressie-Read family)

can be used to measure the divergence between two distributions, the estimated subjective beliefs depend on the particular choice of this divergence function. For a given divergence function, however, the methodology delivers a unique point estimate of the conditional distribution of the relevant macroeconomic and financial variables under the subjective measure. This point estimate offers a natural starting point in the analysis of subjective beliefs. With the subjective conditional distribution in hand, the implied (subjective) moments of different variables of interest, e.g., the expected market return, can be determined for each possible value of the state (conditioning) vector.

While the minimally divergent conditional distribution, $F_*(z|c)$, offers a natural starting point in the analysis of subjective beliefs, it would constitute a fruitful direction of research to develop some framework for sensitivity analysis instead of fixating on $F_*(z|c)$ alone. In a recent insightful paper, Chen, Hansen, and Hansen (2024) developed a novel methodology to explore beliefs distributions other than the one implied by minimal divergence. In particular, they search over a neighborhood of the divergence minimizing (population) pseudo-true distribution and derive bounds for subjective expectations of variables interest that are valid for any beliefs within this neighborhood. Chen, Hansen, and Hansen (2024) focused mainly on the unconditional joint distribution of (Z, X). Although a full investigation is beyond the scope of this paper, one may extend their methodology to our context, which focuses on the conditional distribution $Z|X = c$.

Suppose that the object of interest is the conditional probability $\mathbb{P}(Z \in A|X = c)$ for some measurable set A, and fix the SDF parameter value θ to simplify the presentation. A conditional version of Problem 4.1 in Chen, Hansen, and Hansen (2024) with the ET (relative entropy) divergence function may be formulated as:

$$\mathbb{K}(\mathbb{I}\{Z \in A\}) \equiv \min_{\varrho \geq 0} \mathbb{E}[\varrho \mathbb{I}\{Z \in A\}|X = c], \tag{13}$$

subject to

$$\mathbb{E}[\varrho \log \varrho | X = c] \leq \kappa, \quad \mathbb{E}[\varrho\{M(Y;\theta)R - 1\}|X = c] = 0, \quad \mathbb{E}[\varrho|X = c] = 1,$$

where $\kappa \geq \underline{\kappa}$ is a tuning constant, and $\underline{\kappa}$ is the minimized value of the (conditional) relative entropy defined as

$$\underline{\kappa} \equiv \inf_{\varrho > 0} \mathbb{E}[\varrho \log \varrho | X = c] \quad \text{subject to } \mathbb{E}[\varrho\{M(Y;\theta)R - 1\}|X = c] = 0, \quad \mathbb{E}[\varrho|X = c] = 1.$$

Note that this paper focuses on the unique conditional measure that estimates the minimum value $\underline{\kappa}$. On the other hand, the minimization problem in Equation (13) characterizes possible values of the conditional probability of $Z|X = c$ under the constraints of the Euler equations and neighborhood around the minimum divergence measure. Based on this formulation, the possible values of the conditional probability $\mathbb{P}(Z \in A|X = c)$ can be obtained as (see, Remark 4.3 of Chen, Hansen, and Hansen (2024)):

$$[\mathbb{K}(\mathbb{I}\{Z \in A\}), -\mathbb{K}(-\mathbb{I}\{Z \in A\})]. \tag{14}$$

The target parameter of our estimator $\hat{\mathbb{P}}(Z \in A|X = c)$ is contained in this interval, which corresponds to the minimum divergence conditonal measure. Instead of aiming for point estimation, the above interval characterizes the values of $\mathbb{P}(Z \in A|X = c)$ which are compatible with conditional measures in a neighborhood of the minimum divergence one.

We conjecture that the interval (14) can be estimated by adapting the dual representation of Theorem 4.7 in Chen, Hansen, and Hansen (2024) with replacements of the unconditional expectation ($\mathbb{E}[\cdot]$) with the conditional one ($\mathbb{E}[\cdot|X = c]$). Then the conditional version of the Chen, Hansen, and Hansen (2024) interval estimator can be developed by employing nonparametric kernel smoother as in this paper. We leave this extension for future research.

VI Conclusion

Information-theoretic methods have recently been proposed in the literature to estimate investors' subjective beliefs. These methods estimate subjective beliefs to minimize the statistical divergence between subjective beliefs and the objective DGP, subject to the constraint that the subjective beliefs satisfy assets' Euler equations. Each specific choice of divergence function in the Cressie-Read family delivers a point estimate of investors' subjective beliefs. The method also delivers a point estimate of the SDF parameters, summarizing the risk preferences of investors. Distortions in subjective beliefs constitutes a form of model misspecification, whereby the properties of these moment-based estimators are altered relative to the correctly-specified model scenario. We derive the properties of these estimators in such a misspecified setting and show that the subjective beliefs estimate converges in probability to its pseudo true value, the latter defined as the distribution obtained by projecting the DGP on to the space of distributions satisfying the Euler equations.

Applying the methodology to the three most popular members of the smoothed Cressie-Read family – namely, the SEL, SET, and SQD – we find that the SEL and SET produce empirically similar estimates of subjective beliefs and risk aversion levels. The SQD, on the other hand, leads to economically implausible estimates of the risk aversion parameter (often in excess of 100). The SQD also produces negative estimates of subjective probabilities, particularly in the tails of the distribution, and underestimates tail risk relative to the SEL and SET. Given these issues with the SQD and given the theoretical shortcomings of the SEL highlighted in recent literature, we recommend the use of the SET.

The point estimate of the subjective conditional distribution of macro and financial variables, along with the proposed conditional interval estimator, offer a rich array of tools aimed at offering a data driven way of modeling investors' beliefs distortions relative to the DGP. Our approach offers a useful supplement to survey forecasts that, on the one hand, may be argued as providing a more direct estimate of beliefs, but, on the other hand, are often associated with shortcomings stemming from their short time series, paucity of information

on the respondents, and limited evidence of the association between reported beliefs and investment actions. Our approach also offers a way to integrate survey forecasts by including them as moment restrictions in the estimation, in addition to the assets' Euler equations. This is left for future research.

References

Almeida, C., and R. Garcia (2012): "Assessing misspecified asset pricing models with empirical likelihood estimators," *Journal of Econometrics*, 170(2), 519–537.

Almeida, C., and R. Garcia (2016): "Economic Implications of Nonlinear Pricing Kernels," *Management Science*, 63(10), 3147–3529.

Amromin, G., and S. A. Sharpe (2013): "From the Horse's Mouth: Economic Conditions and Investor Expectations of Risk and Return," *Management Science*, 60, 845–866.

Antoine, B., and P. Dovonon (2021): "Robust estimation with exponentially tilted Hellinger distance," *Journal of Econometrics,*, 224, 330–344.

Bacchetta, P., E. Mertens, and E. van Wincoop (2009): "Predictability in financial markets: What do survey expectations tell us?," *Journal of International Money and Finance*, 28, 406–426.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015): "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115, 1–24.

Barro, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3), 823–866.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022): "Belief Distortions and Macroeconomic Fluctuations," *American Economic Review*, 112, 2269?2315.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–82.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019): "Diagnostic Expectations and Stock Returns," *Journal of Finance*, 74(6), 2839?2874.

Borovicka, J., L. P. Hansen, and J. A. Scheinkman (2016): "Misspecified Recovery," *Journal of Finance*, 71(6), 2493–2544.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–51.

Carr, P., and J. Yu (2012): "Risk, Return, and Ross Recovery," *Journal of Derivatives*, 20(1), 38–59.

Case, K., and R. Shiller (2003): "Is There a Bubble in the Housing Market?," *Brookings Papers on Economic Activity*, 34, 299–362.

Chauduri, S., E. Renault, and O. Wahlstrom (2023): "Identification of Beliefs in the Presence of Disaster Risk and Misspecification," *Advances in Econometrics*, 45, 261–290.

Chen, X., L. P. Hansen, and P. G. Hansen (2020): "Robust Identification of Investor Beliefs," *Proceedings of the National Academy of Science*, 117, 33130–33140.

——— (2021): "Robust Inference for Moment Condition Models without Rational Expectations," University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2021-122.

——— (2024): "Robust Inference for Moment Condition Models without Rational Expectations," *Journal of Econometrics*, 243, 1–32.

Chen, X., H. Hong, and M. Shum (2007): "Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models," *Journal of Econometrics*, 141, 109–140.

Chen, X., O. Linton, and P. M. Robinson (2001): "THE ESTIMATION OF CONDITIONAL DENSITIES," London School of Economics Working Paper.

Cocco, J. F., F. Gomes, and P. Lopes (2019): "Evidence on Expectations of Household Finances," Discussion paper, London Business School.

Da, Z., X. Huang, and L. Jin (2021): "Extrapolative beliefs in the cross-section: What can we learn from the crowds?," *Journal of Financial Economics*, 140, 175–196.

Dahlquist, M., and M. Ibert (2022): "Equity Return Expectations and Portfolios: Evidence from Large Asset Managers," Working Paper.

Egan, M., A. Mackay, and H. Yang (2022): "Recovering Investor Expectations from Demand for Index Funds," *Review of Economic Studies*, 89, 2559?2599.

Fama, E. F., and K. R. French (2012): "Size, value, and momentum in international stock returns," *The Journal of Financial Economics*, 105, 457–472.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

Ghosh, A., C. Julliard, and A. Taylor (2017): "What is the Consumption-CAPM missing? An information-Theoretic Framework for the Analysis of Asset Pricing Models," *Review of Financial Studies*, 30(2), 442–504.

Ghosh, A., A. Korteweg, and Q. Xu (2019): "Recovering Heterogeneous Beliefs and Preferences from Asset Prices," http://dx.doi.org/10.2139/ssrn.3650199.

Ghosh, A., and G. Roussellet (2019): "Identifying Beliefs from Asset Prices," http://dx.doi.org/10.2139/ssrn.3400005.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021): "Five Facts About Beliefs and Portfolios," *American Economic Review*, 111(5), 1481–1522.

Greenwood, R., and A. Shleifer (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27(3), 714–746.

Hansen, B. E. (2008): "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, 726–748.

Hansen, L. P. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

——— (2014): "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122(5), 945 – 987.

Hansen, L. P., and R. Jagannathan (1997): "Asssessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52, 557–590.

Hasanhodzic, J., and A. W. Lo (2007): "CAN HEDGE-FUND RETURNS BE REPLICATED?: THE LINEAR CASE," *JOURNAL OF INVESTMENT MANAGEMENT*, 5, 5?45.

Hong, H., B. Preston, and M. Shum (2003): "Generalized empirical likelihood-based model selection criteria for moment condition models," *Econometric Theory*, 19, 923–943.

Jensen, C., D. Lando, and L. Pedersen (2019): "Generalized Recovery," *Journal of Financial Economics*, 133.

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative To Generalized Method Of Moments Estimation," *Econometrica*, 65(4), 861–874.

Kitamura, Y., G. Tripathi, and H. Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72(6), 1667–1714.

Klaus, A., A. Marcet, and J. P. Nicolini (2016): "Stock Market Volatility and Learning," *Journal of Finance*, 71, 33–82.

Komunjer, I., and G. Ragusa (2016): "Existence and Characterization of Conditional Density Projections," *Econometric Theory*, 32, 947–987.

Kozhan, R., A. Neuberger, and P. Schneider (2013): "The Skew Risk Premium in the Equity Index Market," *Review of Financial Studies*, 26, 2174–2203.

Kristensen, D. (2009): "Uniform Convergence Rates Of Kernel Estimators With Heterogeneous Dependent Data," *Econometric Theory*, 25, 1433–1445.

Kuchler, T., and B. Zafar (2019): "Personal Experiences and Expectations about Aggregate Outcomes," *Journal of Finance*, 74(5), 2491–2542.

Malmendier, U., and S. Nagel (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?," *The Quarterly Journal of Economics*, 126, 373–416.

Martin, I. (2013): "Consumption-Based Asset Pricing with Higher Cumulants," *Review of Economic Studies*, 80(2), 745–773.

——— (2017): "What is the Expected Return on the Market?," *Quaterly Journal of Economics*, 132(1), 367–433.

Nagel, S., and Z. Xu (2022): "Dynamics of Subjective Risk Premia," Working paper, University of Chicago.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F.Engle, and D. McFadden, vol. 4. Elsevier Press.

Orlowski, P., A. Sali, and F. Trojani (2018): "Arbitrage Free Dispersion," Discussion paper.

Owen, A. B. (2001): *Empirical Likelihood*. Chapman and Hall.

Pazarbasi, A., P. Schneider, and G. Vilkov (2020): "Sentimental Recovery," Discussion paper.

Piazzesi, M., J. Salomao, and M. Schneider (2015): "Trend and Cycle in Bond Premia," mimeo.

Qin, L., and V. Linetsky (2016): "Positive Eigenfunctions of Markovian Pricing Operators: Hansen-Scheinkman Factorization, Ross Recovery and Long-Term Pricing," *Operations Research*, 64(1), 99–117.

Qin, L., and V. Linetsky (2017): "Long Term Risk: A Martingale Approach," *Econometrica*, 85(1).

Ross, S. A. (2015): "The Recovery Theorem," *The Journal of Finance*, 70(2), 615–648.

Schennach, S. (2007): "Point estimation with exponentially tilted empirical likelihood," *Annals of Statistics*, 35, 634–672.

Schneider, P. (2015): "Generalized risk premia," *Journal of Financial Economics*, 116(3), 487–504.

——— (2018): "An Anatomy of the Market Return," *Journal of Financial Economics*, Forthcoming.

Schneider, P., and F. Trojani (2018): "Divergence and the Price of Uncertainty," *The Journal of Financial Econometrics*, 1(56).

——— (2019): "(Almost) Model-Free Recovery," *Journal of Finance*, 74.

Smith, R. (2007): "Efficient information theoretic inference for conditional moment restrictions," *Journal of Econometrics*, 138, 430–460.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, pp. 307–333.

Wachter, J. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68(3), 987–1035.

Walden, J. (2017): "Recovery with Unbounded Diffusion Processes," *Review of Finance*, 21(4), 1403–1444.

White, H. (1982): "Maximum Likelihood of Misspecified Models," *Econometrica*, 50, 1–25.

A Appendix A: Derivations

In this appendix, we provide assumptions and the proof for Proposition 1 for the case of $\gamma \neq 0, -1$. The cases of $\gamma = 0$ and -1 are treated in analogous ways.

We use the following notation. Letting $h(Z;\theta) = [1, g(Z;\theta)']'$ and $\eta = (\mu, \lambda')'$, the parameter estimator $\hat{\theta}$ in Equation (6) can be written as

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \max_{\eta_i\in\Xi} \sum_{i=1}^{T} \tau_i \sum_{j=1}^{T} w_{ij} \{\gamma \eta_i' h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}, \tag{15}$$

where the set Ξ is assumed to be compact. Letting

$$\eta_*(X;\theta) = \arg\max_{\eta\in\Xi} E[[\gamma\{\eta' h(Z;\theta)\}]^{1/\gamma}|X],$$

the population counterpart of $\hat{\theta}$ is defined as

$$\theta_* = \arg\min_{\theta\in\Theta} E[E[[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{1/\gamma}|X]]. \tag{16}$$

Also, based on Equation (7), define $\hat{\eta}(\theta) = (\hat{\mu}(\theta), \hat{\lambda}(\theta)')'$ and $\eta_*(\theta) = \eta_*(c;\theta)$. Let $H(Z;\theta) = \partial h(Z;\theta)/\partial\theta'$. Finally, "w.p.a.1" means "with probability approaching one".

A.1 Assumptions for Lemma 1

(i) $\{Z_t, X_t\}$ is strictly stationary and strong mixing with mixing coefficients α_m satisfying $\alpha_m \leq Am^{-a}$ for some $0 < A, a < \infty$. $\mathbb{E}[|X|^c] < \infty$ for some $c > 1$. The density $f(x)$ of X, and its first and second derivatives are uniformly bounded over the support $\mathcal{X}$ of X.

(ii) Θ is compact, and θ_* defined in (16) is unique. $g(Z,\theta)$ is almost surely continuous at each $\theta \in \Theta$, $\mathbb{E}[\sup_{\theta\in\Theta} |[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{\frac{\gamma+1}{\gamma}}|] < \infty$, and $\frac{\partial^2}{\partial x \partial x'}\{\mathbb{E}[[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{\frac{\gamma+1}{\gamma}}|X = x]f(x)\}$ is uniformly bounded over $\Theta \times \mathcal{X}$. Note that this assumption does not require

X to be compact.

(iii) The kernel function $\mathcal{K}(\cdot)$ is compactly supported, differentiable on the support, and $\int \mathcal{K}(u)du = 1$. The bandwidth b_T satisfies $b_T \to 0$ and $T^{1-\delta} b_T^{2(d+\delta)} \to \infty$ for some $\delta > 0$ as $T \to \infty$.

Assumption (i) is on the data structure and allows weakly dependent data. Assumption (ii) contains conditions for the pseudo true value θ_*, moment function $g(Z, \theta)$ and population version of the Lagrange multiplier $\eta_*(X; \theta)$. Uniqueness of θ_* is commonly imposed in misspecification analysis (see, e.g., Chen, Hong, and Shum (2007) and Hong, Preston, and Shum (2003)). Other conditions are used to apply the uniform law of large numbers in Kristensen (2009). Assumption (iii) contains standard conditions for the kernel function $\mathcal{K}(\cdot)$ and bandwidth b_T.

A.2 Proof of Lemma 1

Pick any γ. Let $\hat{\eta}_i(\theta) = \arg\max_{\eta_i \in \Xi} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \eta_i' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}}$. The definition of $\hat{\eta}_i(\theta)$ implies

$$\begin{aligned} \hat{Q}(\theta) &:= \frac{1}{T} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \hat{\eta}_i(\theta)' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}} \\ &\geq \frac{1}{T} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \eta_*(X_i; \theta)' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}}, \end{aligned}$$

for each $\theta \in \Theta$. Under our assumptions, the uniform law of large numbers implies that the right hand side of the above display converges to $Q_*(\theta) = \mathbb{E}[\{\gamma \eta_*(X; \theta)' h(Z; \theta)\}^{\frac{\gamma+1}{\gamma}}]$ uniformly over $\theta \in \Theta$. Thus, we have the one-sided uniform convergence: $\inf_{\theta \in \Theta} \{\hat{Q}(\theta) - Q_*(\theta)\} \geq 0$ w.p.a.1. Furthermore, since our assumptions guarantee $\max_{1 \leq i \leq T} \tau_i |\hat{\eta}_i(\theta_*) - \eta_*(X_i; \theta_*)| \xrightarrow{p} 0$, we obtain $\hat{Q}(\theta_*) \xrightarrow{p} Q_*(\theta_*)$. Combining these results, the same argument in the proof of Theorem 2.1 of Newey and McFadden (1994) yields the conclusion.

A.3 Assumptions for Proposition 1

(i) $\{Z_t, X_t\}$ is strictly stationary and strong mixing with mixing coefficients α_m satisfying $\alpha_m \leq Am^{-a}$ for some $0 < A, a < \infty$. The density $f(x)$ of X_t is positive and continuous at $x = c$, and there exists $p' < \infty$ such that for all $p \geq p'$ the joint density $f_p(x_0, x_p)$ of (X_0, X_p) is positive and continuous at $(x_0, x_p) = (c, c)$.

(ii) (15) holds true with a compact set Ξ, θ_* defined in (16) is unique, and $\hat{\theta} \xrightarrow{p} \theta_*$.

(iii) $g(Z, \theta)$ is almost surely differentiable with respect to θ in a neighborhood $\mathcal{N}_{\theta_*}$ around θ_*. For each $k = 1, \ldots, \dim g$ and $\theta \in \mathcal{N}_{\theta_*}$, there exists $s > 2$ such that $E[|g_k(Z;\theta)^{\frac{\gamma+1}{\gamma}}|^s] < \infty$ and $E[|\partial g_k(Z,\theta)^{\frac{\gamma+1}{\gamma}}/\partial\theta'|^s] < \infty$, and $E[|g_k(Z;\theta)^{\frac{\gamma+1}{\gamma}}|X = x]$ and $E[|\partial g_k(Z,\theta)^{\frac{\gamma+1}{\gamma}}/\partial\theta'||X = x]$ are continuous at $x = c$. Furthermore, $a > \frac{1+(s-1)\{1+(d+\dim\theta)/q+d+\dim\theta\}}{s-2}$ for some $q > 0$.

(iv) Let $B_0(Z) = \{\gamma\eta_*(\theta_*)'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}$, and there exist measurable functions $B_1(Z)$ and $B_2(Z)$ such that

$$\begin{aligned} \max_{k=1,\ldots,\dim h} \sup_{\theta\in\mathcal{N}_{\theta_*}} |h_k(Z;\theta)^{\frac{1}{\gamma}} H(Z;\theta)| &\leq B_1(Z), \\ \max_{k=1,\ldots,\dim h} \sup_{\theta\in\mathcal{N}_{\theta_*}} |h_k(Z;\theta)^{\frac{1-\gamma}{\gamma}} h(Z;\theta)| &\leq B_2(Z). \end{aligned}$$

For $l = 0, 1, 2$, $B_l(Z)$ satisfies $E[|B_l(Z)|^{s_l}] < \infty$ for some $s_l > 2$, $E[|B_l(Z)|^{s_l}|X = x]$ is continuous at $x = c$, and there exists $p' < \infty$ such that for all $p \geq p'$, $E[|B_l(Z_0)B_l(Z_0)||X_0 = x_0, X_p = x_p]$ is continuous at $(x_0, x_p) = (c, c)$. Also $a > \frac{2s_l-2}{s_l-2}$ for $l = 0, 1, 2$.

(v) The kernel function $\mathcal{K}(\cdot)$ is bounded and integrable. The bandwidth b_T satisfies $b_T \to 0$ and $Tb_T^d \to \infty$ as $T \to \infty$.

Assumption (i) is on the data structure and allows weakly dependent data. Assumption (ii) contains technical conditions for Ξ and θ_*, and a high-level assumption on the consistency of $\hat{\theta}$ to the pseudo true value θ_*. The compactness of Ξ and uniqueness of θ_* are commonly

imposed in misspecification analysis (see, e.g., Chen, Hong, and Shum (2007) and Hong, Preston, and Shum (2003)). Although the formal presentation is somewhat tedious, a similar argument can be adapted for the Cressie-Read family. Assumption (iii) is on smoothness and boundedness of the moment function $g(Z, \theta)$. This assumption can be verified for each application. We impose this assumption to apply the uniform convergence theorem in Kristensen (2009). Assumption (iv) lists additional regularity conditions to control remainder terms, which are also mild and can be verified for each application. Assumption (v) contains standard conditions for the kernel function $\mathcal{K}(\cdot)$ and bandwidth b_T.

A.4 Proof of Proposition 1

First, we show that

$$\hat{\eta}(\hat{\theta}) \xrightarrow{p} \eta_*(\theta_*). \tag{17}$$

Let $\mathcal{N}^{\delta}_{\eta_*(\theta_*)}$ be an open neighborhood around $\eta_*(\theta_*)$ with radius $\delta > 0$. Since $\hat{\eta}(\hat{\theta})$ maximizes $\sum_{j=1}^n w_j \{\gamma \eta' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}}$ with respect to $\eta \in \Xi$, it is sufficient for (17) to show that for each $\delta > 0$, there exists $c_\delta > 0$ such that

$$\sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \geq \sup_{\eta \in \Xi \setminus \mathcal{N}^{\delta}_{\eta_*(\theta_*)}} \sum_{j=1}^n w_j \{\gamma \eta' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} + c_\delta, \tag{18}$$

w.p.a.1.

By an expansion around $\hat{\theta} = \theta_*$, it holds

$$\begin{aligned}
& \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \\
= & \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \theta_*)\}^{\frac{\gamma+1}{\gamma}} + (\gamma+1)\eta_*(\theta_*)' \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \tilde{\theta})\}^{\frac{1}{\gamma}} H(Z_j; \tilde{\theta})(\hat{\theta} - \theta_*) \\
= & \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \theta_*)\}^{\frac{\gamma+1}{\gamma}} + o_p(1) \\
= & E[\{\gamma \eta_*(\theta_*)' h(Z; \theta_*)\}^{\frac{\gamma+1}{\gamma}} | X = c] + o_p(1),
\end{aligned} \tag{19}$$

where $\tilde{\theta}$ is a point on the line joining $\hat{\theta}$ and θ_*, the second equality follows from $\hat{\theta} \xrightarrow{p} \theta_*$ (Assumption (ii)) and $\sum_{j=1}^n w_j B_1(Z_j) = O_p(1)$ (by applying Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v)), and the last equality also follows from Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v).

Pick any $\delta > 0$, and define

$$M_\delta = E[\{\gamma\eta_*(\theta_*)'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c] - \sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c]. \tag{20}$$

The definition of $\eta_*(\theta_*)$ guarantees $M_\delta > 0$. Now, observe that

$$\begin{aligned}
&\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} \sum_{j=1}^n w_j\{\gamma\eta'h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \\
\leq\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} \left\{\sum_{k=1}^n w_k\{\gamma\eta'h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} - E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}}\right\} \\
&+ \sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}} \\
=\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}} + o_p(1) \\
=\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c] + o_p(1),
\end{aligned} \tag{21}$$

where the first equality follows from the uniform convergence (obtained by a specialization of Theorem 1 of Kristensen (2009) for the pointwise consistency at $X = c$ under Assumptions (i), (iii), and (v))

$$\sup_{\theta\in\mathcal{N}_{\theta_*}} \sup_{\eta\in\Xi} \left|\sum_{k=1}^n w_k\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}} - E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\right| \xrightarrow{p} 0,$$

and the second equality follows from $\hat{\theta} \xrightarrow{p} \theta_*$ (Assumption (ii)) and Assumption (iv).

Combining (18)-(21), we obtain for each $\epsilon > 0$,

$$\begin{aligned}\sum_{j=1}^{n} w_j \{\gamma \eta_*(\theta_*)' h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} &\geq E[\{\gamma \eta_*(\theta_*)' h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}} | X = c] - \frac{\epsilon}{2} \\ &\geq \sup_{\eta \in \Xi \setminus \mathcal{N}^{\delta}_{\eta_*(\theta_*)}} \sum_{j=1}^{n} w_j \{\gamma \eta' h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} + (M_\delta - \epsilon),\end{aligned}$$

w.p.a.1. By choosing ϵ small enough to satisfy $M_\delta - \epsilon > 0$, we obtain (18), which implies (17).

Next, we obtain the probability limit of $\hat{P}(Z \in A | X = c)$. An expansion around $(\hat{\eta}(\hat{\theta}), \hat{\theta}) = (\eta_*(\theta_*), \theta_*)$ yields

$$\begin{aligned}\hat{P}(Z \in A | X = c) &= \sum_{j=1}^{T} \hat{p}_j(\hat{\theta}) \mathbb{I}\{Z_j \in A\} \\ &= \sum_{j=1}^{T} w_j [\gamma \{\hat{\eta}(\hat{\theta})' h(Z_j;\hat{\theta})\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} \\ &= \sum_{j=1}^{n} w_j [\gamma \{\eta_*(\theta_*)' h(Z_j;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} \\ &\quad + \frac{1}{\gamma} \sum_{j=1}^{n} w_j [\gamma \{\bar{\eta}' h(Z_j;\bar{\theta})\}]^{\frac{1-\gamma}{\gamma}} h(Z_j;\bar{\theta}) \mathbb{I}\{Z_j \in A\} \{\hat{\eta}(\hat{\theta}) - \eta_*(\theta_*)\} \\ &\quad + \frac{1}{\gamma} \bar{\eta}' \sum_{j=1}^{T} w_j [\gamma \{\bar{\eta}' h(Z_j;\bar{\theta})\}]^{\frac{1-\gamma}{\gamma}} H(Z_j;\bar{\theta}) \mathbb{I}\{Z_j \in A\} (\hat{\theta} - \theta_*) \\ &= \sum_{j=1}^{n} w_j [\gamma \{\eta_*(\theta_*)' h(Z_j;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} + o_p(1) \\ &= E[[\gamma \{\eta_*(\theta_*)' h(Z;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z \in A\} | X = c] + o_p(1),\end{aligned}$$

where $(\bar{\eta}, \bar{\theta})$ is a point on the line joining $(\hat{\eta}(\hat{\theta}), \hat{\theta})$ and $(\eta_*(\theta_*), \theta_*)$, the fourth equality follows from $\hat{\theta} - \theta_* \xrightarrow{p} 0$ (Assumption (ii)), $\hat{\eta}(\hat{\theta}) - \eta_*(\theta_*) \xrightarrow{p} 0$ (eq. (17)), $\sum_{j=1}^{n} w_j B_1(Z_j) = O_p(1)$, and $\sum_{j=1}^{n} w_j B_2(Z_j) = O_p(1)$ (obtained by applying Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v)), and the last equality follows from the consistency of the kernel estimator (obtained by a specialization of Theorem 1 of Kristensen (2009) under

Assumptions (i), (iii), and (v)). Therefore, we obtain the conclusion.

B Appendix B: Additional Empirical Results

Table 5: Beliefs About the Stock Market For Different Investor Types with RRA=1

	$\theta^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(\sigma^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$
		Panel A: $R_W = R_M$		
SEL	1.0/1.0	30.8/21.1	38.0/36.6	22.2/9.9
SET	1.0/1.0	49.3/26.5	43.7/40.6	36.0/37.2
SQD	1.0/1.0	48.6/36.7	48.1/40.6	38.6/34.1
		Panel B: $R_W = R_{BV}$		
SEL	1.0/1.0	33.5/27.9	37.6/36.3	25.2/11.9
SET	1.0/1.0	50.1/35.9	43.4/40.4	28.5/25.4
SQD	1.0/1.0	48.4/37.8	47.9/41.0	36.7/28.3
		Panel C: $R_W = R_{SV}$		
SEL	1.0/1.0	31.8/27.2	39.3/38.1	22.5/12.6
SET	1.0/1.0	47.4/31.3	44.2/41.3	31.0/35.7
SQD	1.0/1.0	49.8/39.8	47.8/41.0	40.4/35.9
		Panel D: $R_W = R_{WML} + R_F$		
SEL	1.0/1.0	$-23.8/-16.6$	42.2/39.9	$-23.1/-13.4$
SET	1.0/1.0	$-39.0/-20.7$	50.1/41.2	$-50.4/-39.2$
SQD	1.0/1.0	$-27.0/-14.4$	52.3/41.7	$-38.3/-31.5$
		Panel E: $R_W = R_{WML_S} + R_F$		
SEL	1.0/1.0	$-8.7/-5.0$	34.2/34.8	$-7.6/-2.4$
SET	1.0/1.0	$-27.4/-9.5$	43.4/41.0	$-38.0/-29.4$
SQD	1.0/1.0	$-28.2/-5.2$	50.1/41.9	$-38.3/-19.6$

The table reports the correlations between the NBER recession dummy and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. The SDF parameter $\theta^{(l)}$ is set equal to unity (Column 2). Panels A–E present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value stocks, small market capitalization and value stocks, the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset, and the winner-minus-loser portfolio constructed from small cap stocks alone (WML_S) alongside the risk free asset. Rows 1–3 in each panel present the results for the SEL, SET, and SQD approaches, respectively. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The two choices of the conditioning set include (a) the average returns of the last four quarters of the market, BV, SV, W, and L portfolios (left number), and (b) exponentially-weighted moving averages of the returns these five portfolios (right number). The sample is quarterly covering the period 1972:Q1-2018:Q4.

Table 6: Beliefs and Risk Preferences with Alternative Instruments: $\left(\log\left(\frac{P}{D}\right), \text{Var}\left(R_M\right)\right)$

	$\theta^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(\sigma^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$
		Panel A: $R_W = R_M$		
SEL	2.6/2.4	39.9/24.2	49.6/43.8	25.8/12.9
SET	3.4/3.5	52.2/38.6	58.2/48.7	37.0/37.4
SQD	100.0/100.0	53.8/44.6/	57.5/48.2/	39.5/33.7
		Panel B: $R_W = R_{BV}$		
SEL	2.1/1.9	38.0/39.4	49.8/41.1	27.6/27.2
SET	3.2/3.4	57.1/46.8	57.6/46.6	48.0/33.9
SQD	99.5/100.0	38.6/42.0	59.1/47.3	−9.4/9.8
		Panel C: $R_W = R_{SV}$		
SEL	2.2/2.1	40.3/32.1	50.6/45.9	27.1/20.1
SET	2.6/2.8	51.0/40.6	58.3/48.6	33.9/38.3
SQD	99.8/98.4	37.5/34.2	57.5/48.9	−19.9/ − 0.1
		Panel D: $R_W = R_{WML} + R_F$		
SEL	5.5/5.4	−50.6/ − 35.7	55.7/35.2	−54.3/ − 38.8
SET	9.9/8.9	−43.1/ − 29.6	54.1/43.9	−46.8/ − 33.2
SQD	99.8/99.6	−6.2/ − 3.0	58.4/48.2	−37.4/ − 21.1
		Panel E: $R_W = R_{WML_S} + R_F$		
SEL	5.1/5.2	−31.5/ − 19.9	44.9/34.5	−35.8/ − 21.7
SET	11.7/13.0	1.3/ − 10.2	45.4/39.3	−2.0/ − 11.4
SQD	100.0/99.1	16.1/ − 1.8	56.2/48.1	−15.1/ − 19.4

The table reports the SDF parameter $\theta^{(l)}$ (Column 2) and the correlations between the NBER recession dummy and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. Panels A–E present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value stocks, small market capitalization and value stocks, the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset, and the winner-minus-loser portfolio constructed from small cap stocks alone (WML_S) alongside the risk free asset. Rows 1–3 in each panel present the results for the SEL, SET, and SQD approaches, respectively. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The two choices of the conditioning set include (a) the average returns of the last four quarters of the market, W, and L portfolios, the aggregate log price-dividend ratio, and the realized variance of the market return (left number), and (b) exponentially-weighted moving averages of the five variables in (a). The sample is quarterly covering the period 1972:Q1-2018:Q4.

Table 7: Beliefs and Risk Preferences with Alternative Instruments: (ΔGDP, Term Spr.)

	$\theta^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(\sigma^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$
		Panel A: $R_W = R_M$		
SEL	2.6/2.3	46.2/39.0	60.0/53.7	30.2/14.9
SET	3.8/3.6	60.8/52.1	66.4/59.3	27.5/37.3
SQD	100.0/100.0	42.3/48.3	68.4/59.8	−4.0/7.3
		Panel B: $R_W = R_{BV}$		
SEL	2.0/1.9	31.1/43.2	50.3/52.5	17.9/14.9
SET	4.5/3.4	50.3/58.7	59.9/56.2	2.1/13.5
SQD	99.9/100.0	9.7/30.1	67.0/52.5	−20.7/ − 12.6
		Panel C: $R_W = R_{SV}$		
SEL	2.2/2.1	31.5/40.7	55.6/55.8	15.5/14.1
SET	2.2/2.0	45.9/48.3	57.7/56.7	10.1/41.9
SQD	99.8/100.0	19.0/47.4	68.0/56.6	−27.5/ − 21.6
		Panel D: $R_W = R_{WML} + R_F$		
SEL	6.0/5.4	−60.5/ − 27.0	59.3/42.1	−65.0/ − 39.4
SET	10.3/9.1	−61.4/ − 22.3	66.3/54.7	−63.0/ − 40.4
SQD	99.9/99.8	−38.9/6.0	70.2/57.7	−48.2/ − 31.1
		Panel E: $R_W = R_{WML_S} + R_F$		
SEL	6.1/5.2	−40.6/ − 2.9	45.8/45.9	−41.7/ − 15.3
SET	11.6/13.2	−26.8/18.0	60.4/50.0	−27.5/ − 0.5
SQD	100.0/98.8	−17.1/19.5	69.5/57.6	−37.7/ − 24.1

The table reports the SDF parameter $\theta^{(l)}$ (Column 2) and the correlations between the NBER recession dummy and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for different investor types. Panels A–E present results, respectively, for investors who allocate to the market portfolio, large market capitalization and value stocks, small market capitalization and value stocks, the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset, and the winner-minus-loser portfolio constructed from small cap stocks alone (WML_S) alongside the risk free asset. Rows 1–3 in each panel present the results for the SEL, SET, and SQD approaches, respectively. For each investor type, the pricing kernel is exponentially affine in their chosen portfolio. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The two choices of the conditioning set include (a) the average returns of the last four quarters of the market, W, and L portfolios, the GDP growth rate, and the term spread (left number), and (b) exponentially-weighted moving averages of these five conditioning variables (right number). The sample is quarterly covering the period 1972:Q1-2018:Q4.

Table 8: Beliefs and Risk Preferences of Momentum Investors in Subperiods

	$\theta^{(l)}$	$\rho\left(E^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(\sigma^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$	$\rho\left(SR^{\mathbb{P}^{(l)}}(R_m), I_{bc}\right)$
		Panel A: 1947:Q1 – 1971:Q4		
SEL	14.1/16.8	$-51.1/-39.1$	51.8/48.4	$-51.5/-37.2$
SET	26.3/29.5	$-51.5/-29.2$	52.2/46.8	$-51.4/-31.5$
SQD	99.8/100	1.70/46.0	52.3/39.4	-28.9/33.4
		Panel B: 1972:Q1 – 1996:Q4		
SEL	10.4/10.5	$-51.0/-39.6$	46.7/39.4	$-52.1/-42.0$
SET	23.2/22.9	$-54.0/-42.4$	53.5/46.8	$-54.2/-43.1$
SQD	100.0/99.9	$-9.4/-24.1$	45.1/47.4	$-14.7/-29.0$
		Panel C: 1997:Q1 – 2018:Q4		
SEL	2.6/2.4	$-37.6/-40.6$	71.4/50.0	$-25.5/-36.9$
SET	11.2/4.1	$-50.6/-31.1$	69.4/52.5	$-48.2/-8.2$
SQD	99.9/99.9	$-46.5/-42.3$	65.5/38.2	$-51.0/-39.1$

The table reports the SDF parameter $\theta^{(l)}$ (Column 2) and the (%) correlations between the NBER recession dummy and the recovered beliefs about the market's expected return (Column 3), volatility (Column 4), and Sharpe ratio (Column 5), for the type-MOM investors. Panels A-C present results for subperiods 1947:Q1 – 1971:Q4, 1972:Q1 – 1996:Q4, and 1997:Q1 – 2018:Q4, respectively. Rows 1–3 in each panel present the results for the SEL, SET, and SQD approaches, respectively. The pricing kernel is exponentially affine in the investor's chosen portfolio, namely the zero-cost winner-minus-loser (WML) portfolio alongside the risk free asset. The test assets consist of the excess returns on the market, BV, SV, W, and L portfolios. The two choices of the conditioning set include (a) the average returns of the last four quarters of the market, BV, SV, W, and L portfolios, (left number), and (b) exponentially-weighted moving averages of the returns of these portfolios (right number).

Online Apendix for "Subjective Beliefs Estimators and Their Properties" *

Anisha Ghosh† Taisuke Otsu‡

October 31, 2025

Abstract

This Online Appendix provides additional methodological details, tables, figures supporting the main text.

*We benefited from helpful comments from Jaroslav Borovicka, George Constantinides, Janice Eberly, Deborah Lucas, Sydney Ludvigson, and seminar participants at the NBER's Summer Institute Asset Pricing Workshop and the Capital Markets and the Economy (EFEL) Workshop. All errors and omissions are our own.

†Desautels Faculty of Management, McGill University; anisha.ghosh@mcgill.ca, *Corresponding author*.

‡Department of Economics, London School of Economics; t.otsu@lse.ac.uk.

A Internet Appendix

In this appendix, we provide assumptions and the proof for Proposition 1 for the case of $\gamma \neq 0, -1$. The cases of $\gamma = 0$ and -1 are treated in analogous ways.

We use the following notation. Letting $h(Z;\theta) = [1, g(Z;\theta)']'$ and $\eta = (\mu, \lambda')'$, the parameter estimator $\hat{\theta}$ in Equation (6) can be written as

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \max_{\eta_i\in\Xi} \sum_{i=1}^{T} \tau_i \sum_{j=1}^{T} w_{ij} \{\gamma \eta_i' h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}, \tag{1}$$

where the set Ξ is assumed to be compact. Letting

$$\eta_*(X;\theta) = \arg\max_{\eta\in\Xi} E[[\gamma\{\eta' h(Z;\theta)\}]^{1/\gamma}|X],$$

the population counterpart of $\hat{\theta}$ is defined as

$$\theta_* = \arg\min_{\theta\in\Theta} E[E[[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{1/\gamma}|X]]. \tag{2}$$

Also, based on Equation (7), define $\hat{\eta}(\theta) = (\hat{\mu}(\theta), \hat{\lambda}(\theta)')'$ and $\eta_*(\theta) = \eta_*(c;\theta)$. Let $H(Z;\theta) = \partial h(Z;\theta)/\partial\theta'$. Finally, "w.p.a.1" means "with probability approaching one".

A.1 Assumptions for Lemma 1

(i) $\{Z_t, X_t\}$ is strictly stationary and strong mixing with mixing coefficients α_m satisfying $\alpha_m \leq Am^{-a}$ for some $0 < A, a < \infty$. $\mathbb{E}[|X|^c] < \infty$ for some $c > 1$. The density $f(x)$ of X, and its first and second derivatives are uniformly bounded over the support $\mathcal{X}$ of X.

(ii) Θ is compact, and θ_* defined in (2) is unique. $g(Z,\theta)$ is almost surely continuous at each $\theta \in \Theta$, $\mathbb{E}[\sup_{\theta\in\Theta} |[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{\frac{\gamma+1}{\gamma}}|] < \infty$, and $\frac{\partial^2}{\partial x \partial x'}\{\mathbb{E}[[\gamma\{\eta_*(X;\theta)' h(Z;\theta)\}]^{\frac{\gamma+1}{\gamma}}|X = x]f(x)\}$ is uniformly bounded over $\Theta \times \mathcal{X}$. Note that this assumption does not require

X to be compact.

(iii) The kernel function $\mathcal{K}(\cdot)$ is compactly supported, differentiable on the support, and $\int \mathcal{K}(u)du = 1$. The bandwidth b_T satisfies $b_T \to 0$ and $T^{1-\delta} b_T^{2(d+\delta)} \to \infty$ for some $\delta > 0$ as $T \to \infty$.

Assumption (i) is on the data structure and allows weakly dependent data. Assumption (ii) contains conditions for the pseudo true value θ_*, moment function $g(Z, \theta)$ and population version of the Lagrange multiplier $\eta_*(X; \theta)$. Uniqueness of θ_* is commonly imposed in misspecification analysis (see, e.g., Chen, Hong, and Shum (2007) and Hong, Preston, and Shum (2003)). Other conditions are used to apply the uniform law of large numbers in Kristensen (2009). Assumption (iii) contains standard conditions for the kernel function $\mathcal{K}(\cdot)$ and bandwidth b_T.

A.2 Proof of Lemma 1

Pick any γ. Let $\hat{\eta}_i(\theta) = \arg\max_{\eta_i \in \Xi} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \eta_i' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}}$. The definition of $\hat{\eta}_i(\theta)$ implies

$$\begin{aligned} \hat{Q}(\theta) &:= \frac{1}{T} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \hat{\eta}_i(\theta)' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}} \\ &\geq \frac{1}{T} \sum_{i=1}^T \tau_i \sum_{j=1}^T w_{ij} \{\gamma \eta_*(X_i; \theta)' h(Z_j; \theta)\}^{\frac{\gamma+1}{\gamma}}, \end{aligned}$$

for each $\theta \in \Theta$. Under our assumptions, the uniform law of large numbers implies that the right hand side of the above display converges to $Q_*(\theta) = \mathbb{E}[\{\gamma \eta_*(X; \theta)' h(Z; \theta)\}^{\frac{\gamma+1}{\gamma}}]$ uniformly over $\theta \in \Theta$. Thus, we have the one-sided uniform convergence: $\inf_{\theta \in \Theta} \{\hat{Q}(\theta) - Q_*(\theta)\} \geq 0$ w.p.a.1. Furthermore, since our assumptions guarantee $\max_{1 \leq i \leq T} \tau_i |\hat{\eta}_i(\theta_*) - \eta_*(X_i; \theta_*)| \xrightarrow{p} 0$, we obtain $\hat{Q}(\theta_*) \xrightarrow{p} Q_*(\theta_*)$. Combining these results, the same argument in the proof of Theorem 2.1 of Newey and McFadden (1994) yields the conclusion.

A.3 Assumptions for Proposition 1

(i) $\{Z_t, X_t\}$ is strictly stationary and strong mixing with mixing coefficients α_m satisfying $\alpha_m \leq Am^{-a}$ for some $0 < A, a < \infty$. The density $f(x)$ of X_t is positive and continuous at $x = c$, and there exists $p' < \infty$ such that for all $p \geq p'$ the joint density $f_p(x_0, x_p)$ of (X_0, X_p) is positive and continuous at $(x_0, x_p) = (c, c)$.

(ii) (1) holds true with a compact set Ξ, θ_* defined in (2) is unique, and $\hat{\theta} \xrightarrow{p} \theta_*$.

(iii) $g(Z, \theta)$ is almost surely differentiable with respect to θ in a neighborhood $\mathcal{N}_{\theta_*}$ around θ_*. For each $k = 1, \ldots, \dim g$ and $\theta \in \mathcal{N}_{\theta_*}$, there exists $s > 2$ such that $E[|g_k(Z;\theta)^{\frac{\gamma+1}{\gamma}}|^s] < \infty$ and $E[|\partial g_k(Z,\theta)^{\frac{\gamma+1}{\gamma}}/\partial\theta'|^s] < \infty$, and $E[|g_k(Z;\theta)^{\frac{\gamma+1}{\gamma}}|X = x]$ and $E[|\partial g_k(Z,\theta)^{\frac{\gamma+1}{\gamma}}/\partial\theta'||X = x]$ are continuous at $x = c$. Furthermore, $a > \frac{1+(s-1)\{1+(d+\dim\theta)/q+d+\dim\theta\}}{s-2}$ for some $q > 0$.

(iv) Let $B_0(Z) = \{\gamma\eta_*(\theta_*)'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}$, and there exist measurable functions $B_1(Z)$ and $B_2(Z)$ such that

$$\begin{aligned} \max_{k=1,\ldots,\dim h} \sup_{\theta\in\mathcal{N}_{\theta_*}} |h_k(Z;\theta)^{\frac{1}{\gamma}} H(Z;\theta)| &\leq B_1(Z), \\ \max_{k=1,\ldots,\dim h} \sup_{\theta\in\mathcal{N}_{\theta_*}} |h_k(Z;\theta)^{\frac{1-\gamma}{\gamma}} h(Z;\theta)| &\leq B_2(Z). \end{aligned}$$

For $l = 0, 1, 2$, $B_l(Z)$ satisfies $E[|B_l(Z)|^{s_l}] < \infty$ for some $s_l > 2$, $E[|B_l(Z)|^{s_l}|X = x]$ is continuous at $x = c$, and there exists $p' < \infty$ such that for all $p \geq p'$, $E[|B_l(Z_0)B_l(Z_0)||X_0 = x_0, X_p = x_p]$ is continuous at $(x_0, x_p) = (c, c)$. Also $a > \frac{2s_l-2}{s_l-2}$ for $l = 0, 1, 2$.

(v) The kernel function $\mathcal{K}(\cdot)$ is bounded and integrable. The bandwidth b_T satisfies $b_T \to 0$ and $Tb_T^d \to \infty$ as $T \to \infty$.

Assumption (i) is on the data structure and allows weakly dependent data. Assumption (ii) contains technical conditions for Ξ and θ_*, and a high-level assumption on the consistency of $\hat{\theta}$ to the pseudo true value θ_*. The compactness of Ξ and uniqueness of θ_* are commonly

imposed in misspecification analysis (see, e.g., Chen, Hong, and Shum (2007) and Hong, Preston, and Shum (2003)). Although the formal presentation is somewhat tedious, a similar argument can be adapted for the Cressie-Read divergence. Assumption (iii) is on smoothness and boundedness of the moment function $g(Z, \theta)$. This assumption can be verified for each application. We impose this assumption to apply the uniform convergence theorem in Kristensen (2009). Assumption (iv) lists additional regularity conditions to control remainder terms, which are also mild and can be verified for each application. Assumption (v) contains standard conditions for the kernel function $\mathcal{K}(\cdot)$ and bandwidth b_T.

A.4 Proof of Proposition 1

First, we show that

$$\hat{\eta}(\hat{\theta}) \xrightarrow{p} \eta_*(\theta_*). \tag{3}$$

Let $\mathcal{N}^{\delta}_{\eta_*(\theta_*)}$ be an open neighborhood around $\eta_*(\theta_*)$ with radius $\delta > 0$. Since $\hat{\eta}(\hat{\theta})$ maximizes $\sum_{j=1}^n w_j \{\gamma \eta' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}}$ with respect to $\eta \in \Xi$, it is sufficient for (3) to show that for each $\delta > 0$, there exists $c_\delta > 0$ such that

$$\sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \geq \sup_{\eta \in \Xi \setminus \mathcal{N}^{\delta}_{\eta_*(\theta_*)}} \sum_{j=1}^n w_j \{\gamma \eta' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} + c_\delta, \tag{4}$$

w.p.a.1.

By an expansion around $\hat{\theta} = \theta_*$, it holds

$$\begin{aligned}
& \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \\
= & \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \theta_*)\}^{\frac{\gamma+1}{\gamma}} + (\gamma+1)\eta_*(\theta_*)' \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \tilde{\theta})\}^{\frac{1}{\gamma}} H(Z_j; \tilde{\theta})(\hat{\theta} - \theta_*) \\
= & \sum_{j=1}^n w_j \{\gamma \eta_*(\theta_*)' h(Z_j; \theta_*)\}^{\frac{\gamma+1}{\gamma}} + o_p(1) \\
= & E[\{\gamma \eta_*(\theta_*)' h(Z; \theta_*)\}^{\frac{\gamma+1}{\gamma}} | X = c] + o_p(1),
\end{aligned} \tag{5}$$

where $\tilde{\theta}$ is a point on the line joining $\hat{\theta}$ and θ_*, the second equality follows from $\hat{\theta} \xrightarrow{p} \theta_*$ (Assumption (ii)) and $\sum_{j=1}^n w_j B_1(Z_j) = O_p(1)$ (by applying Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v)), and the last equality also follows from Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v).

Pick any $\delta > 0$, and define

$$M_\delta = E[\{\gamma\eta_*(\theta_*)'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c] - \sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c]. \tag{6}$$

The definition of $\eta_*(\theta_*)$ guarantees $M_\delta > 0$. Now, observe that

$$\begin{aligned}
&\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} \sum_{j=1}^n w_j\{\gamma\eta'h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} \\
\leq\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} \left\{\sum_{k=1}^n w_k\{\gamma\eta'h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} - E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}}\right\} \\
&+ \sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}} \\
=\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\Big|_{\theta=\hat{\theta}} + o_p(1) \\
=\ &\sup_{\eta\in\Xi\setminus\mathcal{N}^\delta_{\eta_*(\theta_*)}} E[\{\gamma\eta'h(Z_j;\theta_*)\}^{\frac{\gamma+1}{\gamma}}|X=c] + o_p(1),
\end{aligned} \tag{7}$$

where the first equality follows from the uniform convergence (obtained by a specialization of Theorem 1 of Kristensen (2009) for the pointwise consistency at $X = c$ under Assumptions (i), (iii), and (v))

$$\sup_{\theta\in\mathcal{N}_{\theta_*}}\sup_{\eta\in\Xi}\left|\sum_{k=1}^n w_k\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}} - E[\{\gamma\eta'h(Z_j;\theta)\}^{\frac{\gamma+1}{\gamma}}|X=c]\right| \xrightarrow{p} 0,$$

and the second equality follows from $\hat{\theta} \xrightarrow{p} \theta_*$ (Assumption (ii)) and Assumption (iv).

Combining (4)-(7), we obtain for each $\epsilon > 0$,

$$\begin{aligned}\sum_{j=1}^{n} w_j \{\gamma \eta_*(\theta_*)' h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} &\geq E[\{\gamma \eta_*(\theta_*)' h(Z;\theta_*)\}^{\frac{\gamma+1}{\gamma}} | X = c] - \frac{\epsilon}{2} \\ &\geq \sup_{\eta \in \Xi \setminus \mathcal{N}^{\delta}_{\eta_*(\theta_*)}} \sum_{j=1}^{n} w_j \{\gamma \eta' h(Z_j;\hat{\theta})\}^{\frac{\gamma+1}{\gamma}} + (M_\delta - \epsilon),\end{aligned}$$

w.p.a.1. By choosing ϵ small enough to satisfy $M_\delta - \epsilon > 0$, we obtain (4), which implies (3).

Next, we obtain the probability limit of $\hat{P}(Z \in A | X = c)$. An expansion around $(\hat{\eta}(\hat{\theta}), \hat{\theta}) = (\eta_*(\theta_*), \theta_*)$ yields

$$\begin{aligned}\hat{P}(Z \in A | X = c) &= \sum_{j=1}^{T} \hat{p}_j(\hat{\theta}) \mathbb{I}\{Z_j \in A\} \\ &= \sum_{j=1}^{T} w_j [\gamma \{\hat{\eta}(\hat{\theta})' h(Z_j;\hat{\theta})\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} \\ &= \sum_{j=1}^{n} w_j [\gamma \{\eta_*(\theta_*)' h(Z_j;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} \\ &\quad + \frac{1}{\gamma} \sum_{j=1}^{n} w_j [\gamma \{\bar{\eta}' h(Z_j;\bar{\theta})\}]^{\frac{1-\gamma}{\gamma}} h(Z_j;\bar{\theta}) \mathbb{I}\{Z_j \in A\} \{\hat{\eta}(\hat{\theta}) - \eta_*(\theta_*)\} \\ &\quad + \frac{1}{\gamma} \bar{\eta}' \sum_{j=1}^{T} w_j [\gamma \{\bar{\eta}' h(Z_j;\bar{\theta})\}]^{\frac{1-\gamma}{\gamma}} H(Z_j;\bar{\theta}) \mathbb{I}\{Z_j \in A\} (\hat{\theta} - \theta_*) \\ &= \sum_{j=1}^{n} w_j [\gamma \{\eta_*(\theta_*)' h(Z_j;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z_j \in A\} + o_p(1) \\ &= E[[\gamma \{\eta_*(\theta_*)' h(Z;\theta_*)\}]^{\frac{1}{\gamma}} \mathbb{I}\{Z \in A\} | X = c] + o_p(1),\end{aligned}$$

where $(\bar{\eta}, \bar{\theta})$ is a point on the line joining $(\hat{\eta}(\hat{\theta}), \hat{\theta})$ and $(\eta_*(\theta_*), \theta_*)$, the fourth equality follows from $\hat{\theta} - \theta_* \xrightarrow{p} 0$ (Assumption (ii)), $\hat{\eta}(\hat{\theta}) - \eta_*(\theta_*) \xrightarrow{p} 0$ (eq. (3)), $\sum_{j=1}^{n} w_j B_1(Z_j) = O_p(1)$, and $\sum_{j=1}^{n} w_j B_2(Z_j) = O_p(1)$ (obtained by applying Theorem 1 of Hansen (2008) under Assumptions (i), (iv), and (v)), and the last equality follows from the consistency of the kernel estimator (obtained by a specialization of Theorem 1 of Kristensen (2009) under Assumptions (i), (iii), and (v)). Therefore, we obtain the conclusion.

References

Chen, X., H. Hong, and M. Shum (2007): "Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models," *Journal of Econometrics*, 141, 109–140.

Hansen, B. E. (2008): "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, 726–748.

Hong, H., B. Preston, and M. Shum (2003): "Generalized empirical likelihood-based model selection criteria for moment condition models," *Econometric Theory*, 19, 923–943.

Kristensen, D. (2009): "Uniform Convergence Rates Of Kernel Estimators With Heterogeneous Dependent Data," *Econometric Theory*, 25, 1433–1445.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F.Engle, and D. McFadden, vol. 4. Elsevier Press.