

Entropy-Based Aggregation of Misspecified Asset Pricing Models ^{*}

Nikolay Gospodinov[†] and Esfandiar Maasoumi[‡]

February 16, 2016

Abstract

This paper proposes an entropy-based approach for aggregating information from misspecified asset pricing models. The mixing weights assigned to different models are determined either by minimizing the pricing errors of the aggregator or by minimizing the Hellinger distance of the densities of the aggregator and a suitably chosen pivot model. The proposed method relaxes the perfect substitutability of the candidate models, which is implicitly embedded in the linear pooling procedures, and ensures that the weights are selected by a proper distance measure that satisfies the triangular inequality. Our approach subsumes other pooling and model averaging approaches, including Bayesian and other methods which assume a "true model" exists (included or not). The empirical results illustrate the robustness and the pricing ability of the aggregation approach.

JEL Classification: C13, C52, G12.

Keywords: Entropy; Model aggregation; Asset pricing; Misspecified models; Hansen-Jagannathan distance.

^{*}We would like to thank Mark Fisher for useful discussions. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

[†]Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, Email: nikolay.gospodinov@atl.frb.org

[‡]Emory University, Department of Economics, Rich Memorial Building 324, 1602 Fishburne Drive, Atlanta, GA 30322-2240, Email: esfandiar.maasoumi@emory.edu

1 Introduction

All models are misspecified by design as they are constructed to approximate a complex, latent reality. The data generating processes are typically latent objects and models can be viewed as partial maps. This is especially true when these models are incompletely specified and are estimated by moment matching. Despite the obvious nature of the statement above, its accomodation remains inconsistent and even contradictory in many instances. In particular, the analysis of misspecified moment condition models is still in its infancy. It is often the case that there are several candidate models. One problem that arises in this setup is that the pseudo-true values that characterize these models are relative objects, depending on proposed models and even estimation criteria. Model selection and model averaging by comparing relative distances can be misconceived. This problem is somewhat mitigated in some situations, as in the context of comparing misspecified asset pricing models using the Hansen-Jagannathan (Hansen and Jagannathan, 1991, 1997) distance that uses the inverse of the second moment matrix of the test assets to weigh the pricing errors for all candidate models.

However, there is another problem with the standard model selection procedure and pooling of misspecified models. In general, the model selection procedure is designed to choose only one of these models and ignores the information in the remaining models. This will be valid and meaningful only if the ‘true’ DGP model is in the set of models considered and the model selection procedure is consistent. But, as argued above, this is a highly unrealistic situation. More realistically, all models are misspecified models.

Bernando and Smith (1994) characterize and offer a taxonomy of the different views regarding model comparison and selection. The first perspective, that includes Bayesian model averaging and frequentist model selection, is conditioning on one of the models being ‘true’. In this approach, the ambiguity about the true model is resolved asymptotically and in the limit, the mixture that summarizes the beliefs about the individual models assigns a weight of one to one of the models. Diebold (1991) provides an illuminating example of this in the context of Bayesian forecast combination. Another possibility is also to assume that a “true model” exists but it is too complicated or cumbersome to implement. In that respect, all of the candidate models are viewed as approximations of this fully-specified belief model and hence misspecified. The third view dispenses completely with the self-contradictory notion of a “true model” and treats the candidate models as genuinely misspecified either because they are believed to represent different aspects of the underlying data generating process or because the underlying structure is completely unknown. As pointed out in Maasoumi (1993), “if models are misspecified in an indeterminate manner, then we should not be aiming at the discovery of the ‘true data generating process’.” Good models reveal

information that is hopefully consistent with aspects of the underlying DGP.

In this paper, we take the view that the DGP/"true model" is likely not to be among the competing models although we do not need to differentiate between the different perspectives. This is a similar approach as the one adopted by Geweke and Amisano (2011, 2012) for prediction pooling of misspecified models. We develop a generalized entropy-based approach to mixing information from different models. The minimum Shannon entropy or Kullback-Leibler information criterion (KLIC) used by Geweke and Amisano (2011, 2012) and Hall and Mitchell (2007) is a special case of this framework. In this paper, our generalization is facilitated by the fact that we are not mixing densities so that the combination does not need to commute with any possible marginalization of the distributions involved (McConway, 1981; Genest, Weerahandi and Zidek, 1984). More importantly, unlike Geweke and Amisano (2011), we choose a divergence measure for selecting the mixture weights which is a proper measure of distance since it satisfies the triangular inequality. The generalized entropy also allows us to relax the perfect substitutability of the candidate models which is implicitly embedded in the linear pooling procedures. Finally, our closeness measure is appropriate for clustering models which might be particularly useful and informative if the set of candidate models is large. The model clustering will identify similar attributes across models and act effectively as a dimension reduction device by reducing the set of information-enhancing models.

Our contributions can be summarized as follows. On methodological side, we propose an information-theoretic approach to aggregating information in misspecified asset pricing models. The optimal aggregator takes a harmonic mean form with geometric and linear weighting schemes as special cases. The generalized entropy criterion that underlies our approach allows us to circumvent two serious drawbacks of the standard linear pooling. First, it ensures that the divergence measure between the densities of the pricing errors of candidate models is a proper distance measure that is positive, symmetric and satisfies the triangular inequality (Maasoumi, 1993). Second, the use of the harmonic mean as an aggregator relaxes the infinite substitutability assumption between models which is implicit in linear aggregation. On the practical side, our mixing procedure employs information from all models by assigning weights depending on the model's contribution to the overall reduction of the pricing errors. The weighted stochastic discount factor preserves the integrity of each structural model and pools the relevant information from each model in an efficient way. This stands in sharp contrast with the existing methods in the literature that either select factors from a set of candidate factors or choose a single ('least misspecified') model from a set of candidate models. Both of these cases result in loss of information from dropping factors or models. Our empirical analysis reports substantial improvements (in terms of pricing error reduction) from aggregation.

Ultimately, the reason so many find that almost all kinds of pooling and mixing methods "per-

form well" can be readily gleaned from the classical results in classical linear regression. Constraints (such as omitted components), even false constraints, are variance (uncertainty) reducing, with a cost on correct centering (bias). But the latter has an uncertain characterization when the true DGP/model is not known.

The rest of the paper proceeds as follows. Section 2 discusses the main setup for evaluating asset pricing models and introduces the approach to model aggregation. Section 3 describes the candidate consumption-based asset pricing models and presents the empirical results. Section 4 concludes.

2 Models and Aggregation

2.1 SDF and Hansen-Jagannathan Distance

Let R denote the returns on N test assets and $m \in \mathcal{M}$ be an admissible stochastic discount factor (SDF) that prices the test assets correctly,

$$E[Rm] = q, \tag{1}$$

where q denotes an $N \times 1$ vector of payoffs (a vector of ones if R are gross returns). Furthermore, let $y(\gamma)$ be a candidate stochastic discount factor that depends on a k -vector of unknown parameters $\gamma \in \Gamma$, where Γ is the parameter space of γ . If $y(\gamma)$ prices the N test assets correctly, then the vector of pricing errors, $e(\gamma)$, of the test assets is exactly zero:

$$e(\gamma) = E[Ry(\gamma)] - q = 0_N. \tag{2}$$

However, the pricing errors are nonzero when the asset-pricing model is misspecified. The squared Hansen-Jagannathan (Hansen and Jagannathan, 1991, 1997) distance

$$\delta^2 = \min_{\gamma \in \Gamma} \min_{m \in \mathcal{M}} E[(y(\gamma) - m)^2] \tag{3}$$

provides a misspecification measure of $y(\gamma)$ and can be used for estimating the unknown parameters γ . It is sometimes more convenient to solve the following dual problem:

$$\delta^2 = \min_{\gamma \in \Gamma} \max_{\lambda \in \mathbb{R}^N} E[y(\gamma)^2 - (y(\gamma) - \lambda'R)^2] - 2\lambda'q, \tag{4}$$

where λ is an $N \times 1$ vector of Lagrange multipliers. Note that $\lambda'R$ provides the smallest correction, in mean squared sense, to $y(\gamma)$ in order to make it an admissible SDF. Note that for a given SDF $y(\gamma)$ and γ , the vector of Lagrange multipliers and the squared Hansen-Jagannathan distance can be expressed as

$$\lambda = U^{-1}e(\gamma), \tag{5}$$

and

$$\delta^2(\gamma) = e(\gamma)'U^{-1}e(\gamma), \quad (6)$$

where $U = E[RR']$.

Importantly, Hansen and Jagannathan (1991) provide a maximum pricing error interpretation of the distance $\delta(\gamma)$. Consider a portfolio a with unit second moment, i.e., $a'Ua = 1$. By the Cauchy-Schwartz inequality, the squared pricing error of this portfolio is

$$(a'e(\gamma))^2 = (a'U^{\frac{1}{2}}U^{-\frac{1}{2}}e(\gamma))^2 \leq (a'Ua)(e(\gamma)'U^{-1}e(\gamma)) = \delta^2(\gamma). \quad (7)$$

Specifically, the portfolio $a = U^{-1}e(\gamma)/\delta(\gamma)$ has a pricing error $\delta(\gamma)$. Then,

$$\max_{a: a'Ua=1} |a'e(\gamma)| = \delta(\gamma), \quad (8)$$

and $\delta(\gamma)$ can be interpreted as the maximum pricing error that one can obtain from using $y(\gamma)$ to price the test assets.

The Hansen-Jagannathan distance has an information-theoretic interpretation too. Let P and Q be two probability measures. The generalized entropy (minimum contrast) or Cressie-Read (Cressie and Read, 1984) divergence from Q to P is given by

$$I_\pi(P, Q) = \int \phi_\pi(dQ/dP) dQ, \quad (9)$$

where

$$\phi_\pi(x) = \frac{1}{\pi(\pi+1)} (x^{\pi+1} - 1) \quad (10)$$

is the Cressie-Read power divergence family of functions. Almeida and Garcia (2012) show that for a fixed vector of parameters γ , the primal and dual problems in the SDF framework can be written as

$$\delta_\pi(\gamma) = \min_{m \in \mathcal{M}} E \left[\frac{(1+m-y(\gamma))^{\pi+1}}{\pi(\pi+1)} \right]$$

and

$$\delta_\pi(\gamma) = \max_{\lambda \in \mathbb{R}^N} \lambda'q - E \left[\frac{(\pi\lambda'R)^{\pi+1}}{\pi+1} + (y(\gamma)-1)\lambda'R + \frac{1}{\pi(\pi+1)} \right],$$

respectively. The dual problem for the Hansen-Jagannathan distance is obtained for $\pi = 1$ (see Almeida and Garcia, 2012).

2.2 Aggregation

Suppose there are M proposed misspecified models, $\hat{y}_i = y_i(\hat{\gamma}_i)$, $i = 1, \dots, M$, for the unknowable true model m . We allow for both linear and nonlinear SDF specifications as well as nested and non-nested SDFs. For the sake of argument, we assume that the model parameters for each model

are estimated by minimizing the Hansen-Jagannathan distance. Our approach in this paper is to treat each model as an incomplete ‘indicator’ of the latent DGP. Then, a model averaging rule would aggregate information from all of these models and construct a pseudo-true model \tilde{y} .

Here, we follow Maasoumi (1986) in characterizing the solution for \tilde{y} . Let $y_t = (\hat{y}_{1,t}, \dots, \hat{y}_{M,t})'$ be the i -th row of the $T \times M$ matrix Y and $\tilde{y} = h(\hat{y}_1, \dots, \hat{y}_M)$, where h is an aggregator or index function. Note that it might be more convenient to work with the estimated pricing errors $e_i(\hat{\gamma}_i)$, $i = 1, \dots, M$, instead of \hat{y}_i 's. We are interested in finding the aggregator \tilde{y}_t with a distribution that is as close as possible to the multivariate distribution of \hat{y}_i 's. Information-theoretic or entropy-based approaches to general measures of divergence between distributions are readily provided by information or entropy theory. Maasoumi (1986) shows that generalizing the pairwise criteria of divergence to a general multivariate context results in the following measure of divergence:

$$D_\rho(\tilde{y}, Y; w) = \sum_{i=1}^M w_i \left\{ \sum_{t=1}^T \tilde{y}_t \left[\left(\frac{\tilde{y}_t}{y_{i,t}} \right)^\rho - 1 \right] / \rho(\rho + 1) \right\},$$

The aggregator that minimizes $D_\rho(\tilde{y}, Y; w)$ subject to $\sum_{i=1}^M w_i = 1$ is given by

$$\tilde{y}_t^* \propto \left[\sum_{i=1}^M w_i y_{i,t}^{-\rho} \right]^{-1/\rho}.$$

Note that a linear pooling of models is obtained as a special case when $\rho = -1$.

In order to implement the above aggregation scheme, we need to estimate the unknown parameters $w = (w_1, \dots, w_M)'$ and ρ . We propose two methods for estimating these parameters.

The first method is, for given $(\hat{y}_{1,t}, \dots, \hat{y}_{M,t})'$ obtained in a preliminary step by minimizing the Hansen-Jagannathan distance for each model and candidate values for w and ρ , construct the pricing errors of the aggregator

$$\tilde{e}_T(w, \rho) = \frac{1}{T} \sum_{t=1}^T R_t \left[\sum_{i=1}^M w_i \hat{y}_{i,t}^{-\rho} \right]^{-1/\rho} - q.$$

Then, the unknown parameters $\theta = (w', \rho)'$ are obtained as

$$\hat{\theta} = \arg \min \tilde{e}_T(\theta)' \left(\frac{1}{T} \sum_{t=1}^T R_t R_t' \right)^{-1} \tilde{e}_T(\theta)$$

subject to the restrictions $w_i \geq 0$ for $i = 1, \dots, M$ and $\sum_{i=1}^M w_i = 1$. Note also that these parameters can be estimated by any member of the Cressie-Read divergence family. For example, Kitamura, Otsu and Evdokimov (2013) show the robustness of the Hellinger-distance estimator (discussed below) for misspecified moment condition models. We use the Hansen-Jagannathan distance estimator due to its computational simplicity and maximum pricing error interpretation.

The other possibility is to estimate θ by minimizing the distance between two distributions. Let P be a probability measure associated with some pivot with density p and q denote the density of the aggregator $\tilde{y}_t(\theta) = \left[\sum_{i=1}^M w_i y_{i,t}^{-\rho} \right]^{-1/\rho}$. Using the generalized entropy (Cressie-Read) divergence from Q to P defined in (9)-(10) and imposing $\pi = -1/2$, we obtain the scaled Hellinger distance $\mathcal{H} \propto I_{-1/2}(P, Q)$ given by (Granger, Maasoumi and Racine, 2004)

$$\mathcal{H} = \frac{1}{2} \int \left(p^{1/2}(x) - q^{1/2}(x) \right)^2 dx. \quad (11)$$

Unlike the other measures in the Cressie-Read divergence family, the Hellinger distance is a proper measure of distance since it satisfies the triangular inequality. Minimizing \mathcal{H} with respect to the parameters θ , subject to the relevant restrictions, provides an estimate of θ . In the practical implementation, we estimate p and q by a kernel density estimator and the integral in (11) is evaluated numerically. The choice of a pivot is discussed in the next section.

3 Empirical Analysis

3.1 Data and Asset-Pricing Models

We analyze seven popular nonlinear asset-pricing models. The SDF for the first six models is log-linear in the factors and takes the form $y_t(\gamma) = \exp(\gamma' \tilde{f}_t)$. For the last model, the SDF contains conditional expectations that are approximated by a second-order Taylor series expansion under the assumption that consumption growth is normally distributed with constant mean and variance.

1. CAPM of Brown and Gibbons (1985):

$$y_t^{CAPM}(\alpha, \beta) = \beta(1 - k)^{-\alpha} R_{m,t}^{-\alpha}$$

or

$$\ln(y_t^{CAPM}(\gamma)) = \gamma_0 + \gamma_1 \ln(R_{m,t}), \quad (12)$$

where R_m is the gross market return, β is the discount rate, $\alpha > 0$ is the coefficient of relative risk aversion, k is the proportion of wealth consumed in every period, $\gamma_0 = -\alpha \ln(\beta(1 - k))$ and $\gamma_1 = -\alpha$.

2. Consumption CAPM (CCAPM):

$$y_t^{CCAPM}(\alpha, \beta) = \beta \left(\frac{C_t}{C_{t-1}} \right)^{-\alpha}$$

or

$$\ln(y_t^{CCAPM}(\gamma)) = \gamma_0 + \gamma_1 c_t, \quad (13)$$

where C denotes real per capita consumption of non-durable goods (seasonally adjusted), $c_t = \ln(C_t) - \ln(C_{t-1})$ is the growth rate in nondurable consumption, $\gamma_0 = \ln(\beta)$ and $\gamma_1 = -\alpha$.

3. Ultimate consumption (UC) model of Parker and Julliard (2005):

$$y_t^{UC}(\alpha, \beta) = \beta \left(\frac{C_{t+s}}{C_{t-1}} \right)^{-\alpha}$$

or

$$\ln(y_t^{UC}(\gamma)) = \gamma_0 + \gamma_1 c_t^s, \quad (14)$$

where $c_t^s = \ln(C_{t+s}) - \ln(C_{t-1})$ and $s > 0$.

4. Non-expected utility (EZ) model of Epstein and Zin (1989, 1991) and Weil (1989):

$$y_t^{EZ}(\alpha, \beta, \sigma) = \beta^{\frac{1-\alpha}{1-\sigma}} \left(\frac{C_t}{C_{t-1}} \right)^{-\sigma \left(\frac{1-\alpha}{1-\sigma} \right)} R_{m,t}^{\frac{\sigma-\alpha}{1-\sigma}}, \quad (15)$$

where $1/\sigma \geq 0$ is the elasticity of intertemporal substitution. Note that the restriction $\alpha = \sigma$ reduces the model to the standard expected utility model (nonlinear CCAPM). The logarithm of the SDF is given by

$$\ln(y_t^{EZ}(\gamma)) = \gamma_0 + \gamma_1 c_t + \gamma_2 \ln(R_{m,t}),$$

where $\gamma_0 = 1 - \ln(\beta)$, $\gamma_1 = -\frac{(1-\alpha)(\sigma(1-\phi)+\phi)}{1-\sigma}$, and $\gamma_2 = \frac{\sigma-\alpha}{1-\sigma}$.

5. Durable consumption CAPM (D-CCAPM) of Yogo (2006):

$$y_t^{D-CAPM}(\alpha, \beta, \sigma, \phi) = \beta^{\frac{1-\alpha}{1-\sigma}} \left(\frac{C_t}{C_{t-1}} \right)^{-\sigma \left(\frac{1-\alpha}{1-\sigma} \right)} \left(\frac{C_{d,t}/C_t}{C_{d,t-1}/C_{t-1}} \right)^{\phi(1-\alpha)} R_{m,t}^{\frac{\sigma-\alpha}{1-\sigma}}, \quad (16)$$

where C_d is consumption of durable goods and $\phi \in [0, 1]$ is the budget share of durable consumption. When $\phi = 0$, we have the classical non-expected (Epstein-Zin) utility model. By imposing the additional restriction $\alpha = \sigma$, we obtain the standard expected utility model (nonlinear CCAPM). After taking logarithms, we have

$$\ln(y_t^{D-CAPM}(\gamma)) = \gamma_0 + \gamma_1 c_t + \gamma_2 c_{d,t} + \gamma_3 \ln(R_{m,t}), \quad (17)$$

where $\gamma_0 = 1 - \ln(\beta)$, $\gamma_1 = -\frac{(1-\alpha)(\sigma(1-\phi)+\phi)}{1-\sigma}$, $\gamma_2 = \phi(1-\alpha)$, and $\gamma_3 = \frac{\sigma-\alpha}{1-\sigma}$.

6. External habit (EH) model of Abel (1990):

$$y_t^{EH}(\alpha, \beta, \tau) = \beta \left(\frac{C_t}{C_{t-1}} \right)^{-\alpha} \left(\frac{C_{t-1}}{C_{t-2}} \right)^{\tau(\alpha-1)}, \quad (18)$$

where $\tau \geq 0$ is time-separability parameter, or

$$\ln(y_t^{EH}(\gamma)) = \gamma_0 + \gamma_1 c_t + \gamma_2 c_{t-1}, \quad (19)$$

where $\gamma_0 = \ln(\beta)$, $\gamma_1 = -\alpha$ and $\gamma_2 = \tau(\alpha - 1)$. When $\tau = 0$, EH model reduces to the nonlinear CCAPM.

7. Internal habit-persistence (IH) model of Constantinides (1990) and Ferson and Constantinides (1991):

$$y^{IH}(\alpha, \beta, \omega, \mu_c, \sigma_\epsilon^2) = \frac{\beta e^{-\alpha c_{t-1}} \{(e^{c_t} + \omega)^{-\alpha} + \beta \omega e^{-\alpha c_t} E_t[(e^{c_{t+1}} + \omega)^{-\alpha}]\}}{(e^{c_{t-1}} + \omega)^{-\alpha} + \beta \omega e^{-\alpha c_{t-1}} E_{t-1}[(e^{c_t} + \omega)^{-\alpha}]}, \quad (20)$$

where

$$E_t[(e^{c_{t+1}} + \omega)^{-\alpha}] \approx (e^{\mu_c} + \omega)^{-\alpha} [1 + 0.5 \sigma_\epsilon^2 \alpha e^{\mu_c} (e^{\mu_c} + \omega)^{-2} (\alpha e^{\mu_c} - \omega)].$$

using the assumption in Balduzzi and Kallal (1997) that $c_{t+1} = \mu_c + \epsilon_{t+1}$ with $\epsilon_{t+1} \sim N(0, \sigma_\epsilon^2)$.

When $\omega = 0$, IH model reduces to the nonlinear CCAPM.

In summary, the traditional CCAPM is nested within the EZ, EH, and IH models (when $\alpha = \sigma$, $\tau = 0$, and $\omega = 0$, respectively). The EZ, EH, and IH models are overlapping with the overlapping part being the traditional CCAPM. The UC model is strictly non-nested with all the other models.

As a benchmark model ('pivot') for computing the Hellinger distance between the densities of the scaled pricing errors of two models, we use the three-factor (FF3) model of Fama and French (1993)

$$y_t^{FF3}(\gamma) = \gamma_0 + \gamma_1 r_{m,t} + \gamma_2 smb_t + \gamma_3 hml_t, \quad (21)$$

where r_m denotes the excess return on the market portfolio, smb is the return difference between portfolios of stocks with small and large market capitalizations, and hml is the return difference between portfolios of stocks with high and low book-to-market ratios ("value" and "growth" stocks, respectively). The constant SDF model is the least favorable specification for pricing the test assets but it provides a robust pivot. The FF3 model is one of the most successful empirical models and the information contained in the smb and hml factors is somewhat orthogonal to the information in the consumption-based CAPM models considered above.

The test asset returns are the monthly gross returns on the value-weighted 25 Fama-French size and book-to-market ranked portfolios, and the 17 industry portfolios from Kenneth French's website. The sample period is February 1959 to December 2015. The consumption data that is used to construct the growth rates c_t , c_t^s and $c_{d,t}$, is real per capita, seasonally adjusted consumption of non-durable and durable goods from the Bureau of Economic Analysis. The excess return $r_{m,t}$ on the value-weighted stock market index (NYSE-AMEX-NASDAQ) is obtained from Kenneth French's website. The gross market return is constructed by adding the one-month T-bill rate to the excess return. The data for the smb and hml factors also come from Kenneth French's website. For the UC model of Parker and Julliard (2005), we use $s = 23$.

Since the use of the UC model results in loss of observations (last 23 months of the sample), we also present results when the aggregation is performed over the set of models that excludes the UC model.

The unknown parameters are estimated by minimizing the Hansen-Jagannathan distance in (4) which is equivalent to

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} e_T(\gamma)' \left(\frac{1}{T} \sum_{t=1}^T R_t R_t' \right)^{-1} e_T(\gamma),$$

where $e_T(\gamma)$ denotes the sample pricing errors of the model. Plugging in the estimated parameters, the sample Hansen-Jagannathan distance is given by

$$\hat{\delta} = \sqrt{e_T(\hat{\gamma})' \left(\frac{1}{T} \sum_{t=1}^T R_t R_t' \right)^{-1} e_T(\hat{\gamma})}.$$

3.2 Results

For starting values, we use the inverse of the Hansen-Jagannathan distance for each model and linear pooling, i.e., $\hat{w}_i = (1/\hat{\delta}_i) / \sum_{i=1}^M (1/\hat{\delta}_i)$ for $i = 1, \dots, M$ and $\rho = -1$. We refer to ‘method1’ in the tables as the method for estimating $\theta = (w', \rho)'$ by minimizing the Hansen-Jagannathan distance for the aggregator SDF. We refer to ‘method2’ as the method that minimizes the Hellinger distance between the densities of the aggregator and the pivot (3-factor Fama-French model).

The tables report the estimates of the Hansen-Jagannathan distances of the 7 consumption-based asset pricing models, the benchmark (FF3) model and the two aggregators. They also report the estimated weights that the two aggregators assign to each individual model as well as the estimate of ρ . It should be noted that the specification test (Hansen-Jagannathan distance test) rejects overwhelming the null of correct specification for all models. Thus, the aggregation is performed for misspecified models.

In order to assess the robustness of our aggregation procedure across different portfolios of test assets, we consider the following portfolios: (1) 25 Fama-French and 17 industry portfolios, (2) 25 Fama-French portfolios, (3) 17 industry portfolios and (4) 25 Fama-French, 17 industry portfolios and the gross return on the risk-free asset. As documented in the literature, the 3-factor Fama-French model performs best for pricing the 25 Fama-French portfolios. This should present a challenge for our aggregation since none of the consumption-based models provide proxies of the *smb* and *hml* factors in the FF3 model. Also, unlike the portfolio returns, the gross return on the risk-free asset is highly persistent and most models have difficulties pricing this asset.

The results are presented in Tables 1, 2 and 3. Table 1 reports results when the aggregation is performed over all consumption-based asset pricing models. Since the UC model results in loss of observations at the end of the sample, Table 2 presents results when this model is not included in the aggregation. Finally, we include the benchmark (FF3) model in the Hansen-Jagannathan distance type of aggregation and report the results in Table 3.

The results in Table 1 illustrate clearly the advantages of our aggregation method. The aggregation reduces the pricing errors of the individual models in all four cases. It also fares very favorably to the empirically most successful Fama-French model. Figures 1 and 2 plot the SDFs for each model and the weighted (aggregated) SDF that uses information from all models for the 25 Fama-French and 17 industry portfolios. The aggregator SDF based on minimizing the Hansen-Jagannathan distance strikes a nice balance between the volatility of the different models. Although it assigns the largest weight to the D-CCAPM, it reduces its volatility and its pricing errors. The second aggregation method shrinks the SDF towards the SDF of the FF3 model although it cannot match fully the performance of this pivot.¹ It is also interesting to see that the aggregator based on the Hansen-Jagannathan distance closely resembles the dynamics and performance (in terms of pricing errors) of the benchmark model despite of the different information sets.

The aggregation methods are also quite robust to different sets of test assets as they adapt and recalibrate the weights across the different models. While the D-CCAPM receives the largest weight for the 25 FF portfolio returns, the EH model weight dominates for the 17 industry portfolio returns for the first aggregation method. Dropping the UC model does not deteriorate and even improves the performance of the aggregator.

Including the FF3 model in the aggregation (Table 3) delivers the smallest pricing errors across all test asset returns. While the aggregator is dominated by the FF3 model, mixing information from the consumption-based models seems to be beneficial, especially for the 17 industry portfolio returns. Overall, the robust performance of the proposed method suggests that combining information from different, possibly misspecified models, may offer substantial advantages.

4 Conclusions

Economic models are misspecified by design as they try to approximate a complex and often an unknown true data generating process. Instead of selecting a single model for pricing assets, decision making or forecasting, aggregating information from all these models may adapt better to the underlying uncertainty and result in a more robust approximation. Information theory and generalized entropy provide the natural theoretical foundation for dealing with these types of uncertainty and partial specification. We capitalize on some insights from the information-theoretic approach and propose a mixture method for aggregating information from different misspecified asset pricing models. The optimal aggregator takes a harmonic mean form with geometric and linear weighting schemes as special cases. In addition, the generalized entropy criterion that underlies

¹In unreported results, we relax the positivity constraint on w which allows some poorly behaved models to receive a negative weight in the aggregation procedure. Interestingly, this provides further, and often substantial, reduction of the pricing errors which is accompanied by a much higher volatility of the pricing kernel.

our approach allows us to circumvent some serious drawbacks of the standard linear pooling. The application of the aggregator to combining consumption-based asset pricing models demonstrates the advantages of our approach. Density forecasting using a large set of diverse, partially specified models is another natural application of the proposed method.

References

- Abel, A. B., 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review* 80, 38–42.
- Almeida, C., and R. Garcia, 2012, Assessing misspecified asset pricing models with empirical likelihood estimators, *Journal of Econometrics* 170, 519–537.
- Balduzzi, P., and H. Kallal, 1997, Risk premia and variance bounds, *Journal of Finance* 52, 1913–1949.
- Bernardo, J., and A. Smith, 1994, *Bayesian Theory*, Wiley.
- Brown, D. P., and M. Gibbons, 1985, A simple econometric approach for utility-based asset pricing models, *Journal of Finance* 40, 359–381.
- Chen, X., and S. C. Ludvigson, 2009, Land of addicts? An empirical investigation of habit-based asset pricing models, *Journal of Applied Econometrics* 24, 1057–1093.
- Constantinides, G. M., 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.
- Cressie, N., and T. Read, 1984, Multinomial goodness of fit tests, *Journal of the Royal Statistical Society B* 46, 440–464.
- Diebold, F. X., 1991, A note on Bayesian forecast combination procedures, in *Economic Structural Change: Analysis and Forecasting* (P. Hackl and A. H. Westlund, eds.), 225–232.
- Epstein, L. G., and S. E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–968.
- Epstein, L. G., and S. E. Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical investigation, *Journal of Political Economy* 99, 555–576 .
- Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Ferson, W. E., and G. M. Constantinides, 1991, Habit persistence and durability in aggregate consumption, *Journal of Financial Economics* 29, 199–240.
- Genest, C., S. Weerahandi, and J. V. Zidek, Aggregating opinions through logarithmic pooling, *Theory and Decision* 17, 61–70.

- Geweke, J., and G. Amisano, 2011, Optimal prediction pools, *Journal of Econometrics* 164, 130–141.
- Geweke, J., and G. Amisano, 2012, Prediction with misspecified models, *American Economic Review: Papers & Proceedings* 102, 482–486.
- Granger, C. W., E. Maasoumi, and J. C. Racine, 2004, A dependence metric for possibly nonlinear processes, *Journal of Time Series Analysis* 25, 649–669.
- Hall, S. G., and J. Mitchell, 2007, Combining density forecasts, *International Journal of Forecasting* 23, 1–13.
- Hansen, L. P., and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.
- Hansen, L. P., and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.
- Kitamura, Y., T. Otsu, and K. Evdokimov, Robustness, infinitesimal neighborhoods, and moment restrictions, *Econometrica* 81, 1185–1201.
- Maasoumi, E., 1986, The measurement and decomposition of multi-dimensional inequality, *Econometrica* 54, 991–997.
- Maasoumi, E., 1993, A compendium to information theory in economics and econometrics, *Econometric Reviews* 12, 137–181.
- Maasoumi, E., and J. S. Racine, 2002, Entropy and predictability of stock market returns, *Journal of Econometrics* 107, 291–312.
- McConway, K. J., 1981, Marginalization and linear opinion pools, *Journal of the American Statistical Association* 76, 410–414.
- Parker, J. A., and C. Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.
- Weil, P., 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.
- Yogo, M., 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.

Table 1: Aggregation with the ultimate consumption (UC) model.

	CAPM	CCAPM	UC	EZ	D-CCAPM	EH	IH	FF3	method1	method2
25 Fama-French + 17 industry portfolios										
$\hat{\delta}$	0.4258	0.4324	0.4348	0.4147	0.4117	0.4287	0.4416	0.3956	0.4094	0.4186
$\hat{\theta}_1$	0.0015	0.0000	0.0000	0.0915	0.9066	0.0000	0.0003		12.1623	
$\hat{\theta}_2$	0.0001	0.0003	0.0001	0.3586	0.0014	0.6393	0.0002			-0.9985
25 Fama-French portfolios										
$\hat{\delta}$	0.3150	0.3332	0.3338	0.3092	0.2978	0.3305	0.3412	0.2702	0.2988	0.3049
$\hat{\theta}_1$	0.0364	0.0216	0.0213	0.0457	0.8274	0.0315	0.0162		0.1429	
$\hat{\theta}_2$	0.1724	0.3213	0.4160	0.0047	0.0314	0.0441	0.0101			-0.9967
17 industry portfolios										
$\hat{\delta}$	0.1393	0.1379	0.1387	0.1373	0.1366	0.1330	0.1491	0.1340	0.1328	0.1377
$\hat{\theta}_1$	0.0005	0.0006	0.0004	0.0013	0.1634	0.8336	0.0003		-0.4760	
$\hat{\theta}_2$	0.1116	0.0033	0.0384	0.0019	0.0006	0.0909	0.7532			-0.9914
25 Fama-French + 17 industry portfolios + risk-free asset										
$\hat{\delta}$	0.5152	0.5224	0.5212	0.5120	0.5109	0.5173	0.5330	0.4840	0.5092	0.5166
$\hat{\theta}_1$	0.0002	0.0001	0.1660	0.0005	0.6528	0.1804	0.0000		-0.9998	
$\hat{\theta}_2$	0.0012	0.2040	0.0005	0.0725	0.0027	0.7186	0.0005			-0.9948

Notes: This table reports the estimates for the Hansen-Jagannathan distance $\hat{\delta}$ and the aggregation parameters $\hat{\theta} = (\hat{w}', \hat{\rho})'$ for the method based on minimizing the Hansen-Jagannathan distance (method1) and on minimizing the Hellinger distance between the densities of the aggregator and the FF3 model (method2).

Table 2: Aggregation without the ultimate consumption (UC) model.

	CAPM	CCAPM	EZ	D-CCAPM	EH	IH	FF3	method1	method2
25 Fama-French + 17 industry portfolios									
$\hat{\delta}$	0.4187	0.4237	0.4079	0.4054	0.4214	0.4361	0.3939	0.4022	0.4266
$\hat{\theta}_1$	0.0003	0.0000	0.1009	0.8987	0.0000	0.0001		17.9530	
$\hat{\theta}_2$	0.0003	0.4211	0.0035	0.0718	0.0040	0.4992			-0.9639
25 Fama-French portfolios									
$\hat{\delta}$	0.3107	0.3283	0.3044	0.2930	0.3254	0.3399	0.2739	0.2919	0.3013
$\hat{\theta}_1$	0.0056	0.0000	0.0004	0.9939	0.0000	0.0001		9.6285	
$\hat{\theta}_2$	0.2634	0.0002	0.5105	0.2254	0.0002	0.0003			-0.9989
17 industry portfolios									
$\hat{\delta}$	0.1407	0.1398	0.1394	0.1393	0.1379	0.1554	0.1343	0.1378	0.1405
$\hat{\theta}_1$	0.0011	0.0013	0.0177	0.1901	0.7893	0.0004		-0.6099	
$\hat{\theta}_2$	0.5330	0.0989	0.0045	0.0031	0.1777	0.1828			-0.7963
25 Fama-French + 17 industry portfolios + risk-free asset									
$\hat{\delta}$	0.5009	0.5077	0.4974	0.4965	0.5048	0.5204	0.4744	0.4939	0.5027
$\hat{\theta}_1$	0.0012	0.0000	0.1174	0.8805	0.0008	0.0000		23.2510	
$\hat{\theta}_2$	0.0219	0.2811	0.1327	0.0027	0.5521	0.0095			-0.9963

Table 3: Aggregation without the ultimate consumption (UC) model but with FF3 model.

	CAPM	CCAPM	EZ	D-CCAPM	EH	IH	FF3	method1
25 Fama-French + 17 industry portfolios								
$\hat{\delta}$	0.4187	0.4237	0.4079	0.4054	0.4214	0.4361	0.3939	0.3862
$\hat{\theta}_1$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.9998	28.5482
25 Fama-French portfolios								
$\hat{\delta}$	0.3107	0.3283	0.3044	0.2930	0.3254	0.3399	0.2739	0.2696
$\hat{\theta}_1$	0.0007	0.0000	0.0000	0.0005	0.0000	0.0002	0.9986	16.2997
17 industry portfolios								
$\hat{\delta}$	0.1407	0.1398	0.1394	0.1393	0.1379	0.1554	0.1343	0.1327
$\hat{\theta}_1$	0.0006	0.0004	0.0116	0.0243	0.0368	0.0000	0.9263	10.6390
25 Fama-French + 17 industry portfolios + risk-free asset								
$\hat{\delta}$	0.5009	0.5077	0.4974	0.4965	0.5048	0.5204	0.4744	0.4695
$\hat{\theta}_1$	0.0000	0.0000	0.0033	0.0032	0.0001	0.0000	0.9933	19.3069

Notes: This table reports the estimates for the Hansen-Jagannathan distance $\hat{\delta}$ and the aggregation parameters $\hat{\theta} = (\hat{w}', \hat{\rho})'$ for the method based on minimizing the Hansen-Jagannathan distance (method1).

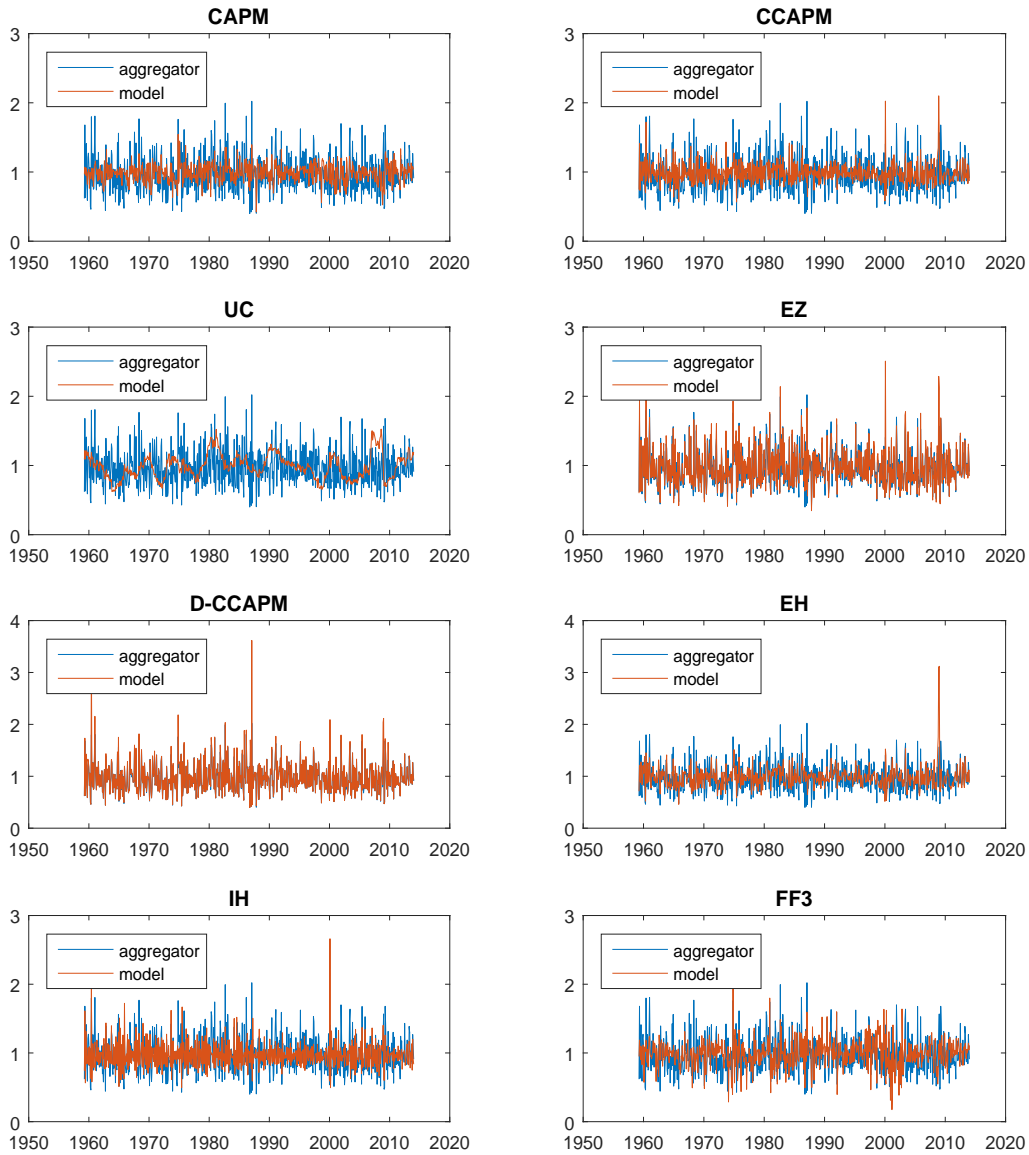


Figure 1: SDFs for different models and aggregator based on the first method.

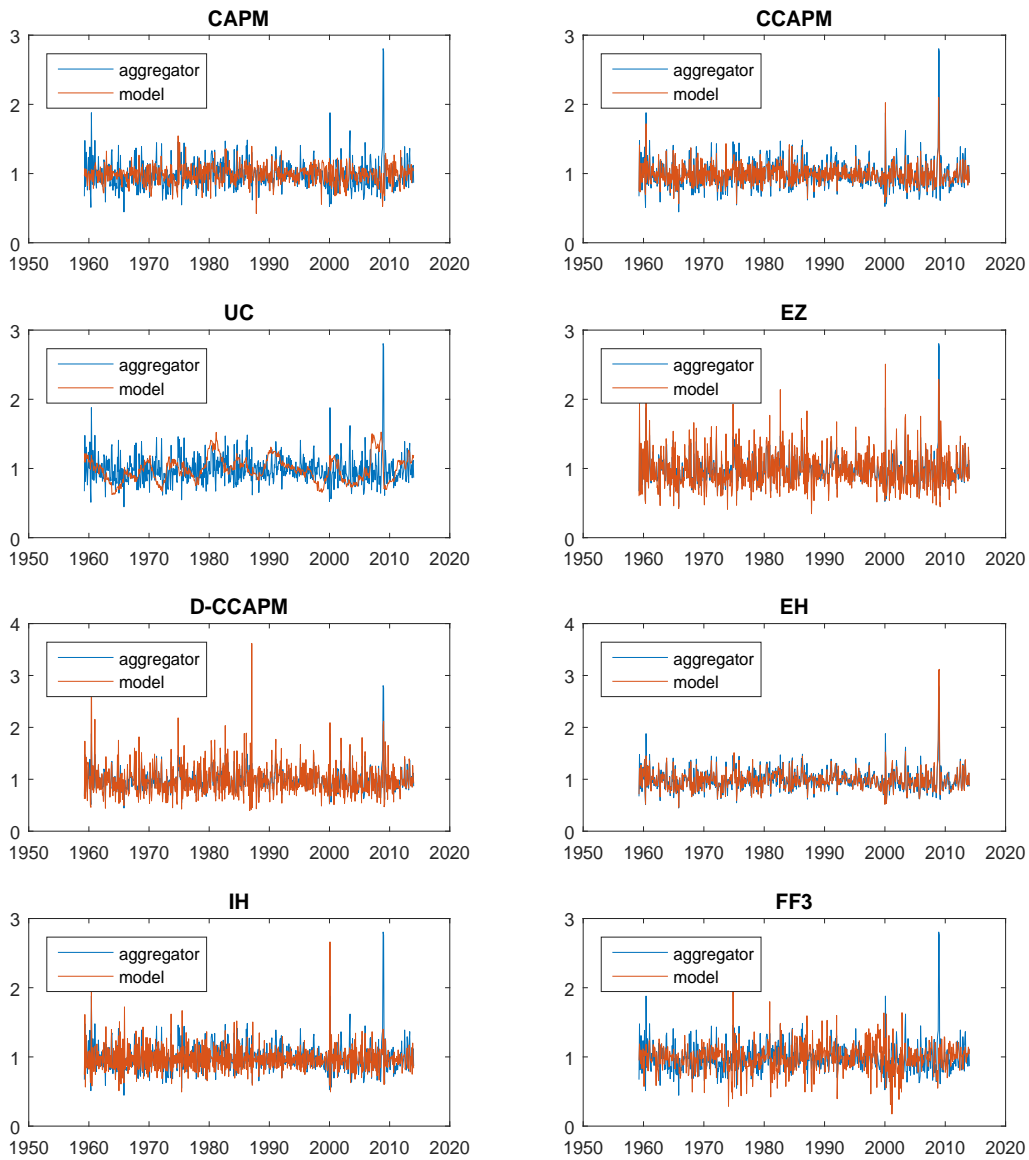


Figure 2: SDFs for different models and aggregator based on the second method.