Efficient Estimation of Pricing Kernels and Market-Implied Densities

Jeroen Dalderop*

January 7, 2021

Abstract

This paper studies the nonparametric identification and estimation of projected pricing kernels implicit in European option prices and underlying asset returns using conditional moment restrictions. The proposed series estimator avoids computing ratios of estimated risk-neutral and physical densities, and produces stable estimates even in regions with sparse observations. We consider efficient estimation by minimizing a Euclidean empirical likelihood or continuous-updating GMM criterion, which takes into account the relative informativeness of option prices with different strike prices beyond the included state variables. In a second step, this criterion can be used to extract implied predictive densities that match the informative part of a cross-section of option prices. The implied densities are related to density combinations and depend on the relative amount of time-variation in the physical return distribution and the pricing kernel.

Keywords: Option Prices, Risk Aversion, Density Forecasting, Empirical Likelihood
JEL Codes: C14, G13

1 Introduction

The pricing kernel or stochastic discount factor approach to asset pricing states that prices equal their expected payoff after stochastically discounting over states (Cochrane, 2009). By weighing state-dependent payoffs with their marginal rate of substitution, the pricing kernel translates the riskiness of assets to their expected return. The asset pricing implications of consumption-based equilibrium models can therefore be summarized by the specification of relevant state variables and how they enter the stochastic discount factor.

The conditional expectation of any valid stochastic discount factor given some traded asset return should be consistent with the observed risk premium for that asset. This defines the 'projected' pricing kernel as a possibly nonlinear function of the future asset return. While the

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. Email address: jdalderop@nd.edu.

linear projection of the pricing kernel on the space of asset returns is well studied, the nonlinear projected pricing kernel provides additional information about the shape of risk aversion (Aït-Sahalia and Lo, 2000).[1]

The projected pricing kernel is nonparametrically identified from European option prices, when the distribution of the underlying asset return is known. In particular, the cross-section of option prices at varying strike prices identifies the risk-neutral density of its underlying asset over a given horizon (Breeden and Litzenberger, 1978). For the purpose of pricing options and other derivatives it suffices to model the risk-neutral distribution. However, its decomposition into the objective density and the pricing kernel is useful for forecasting and risk management, or for understanding risk preferences in equilibrium models.

Density forecasts have so far mainly been obtained by specifying a parametric pricing kernel model to 'risk-adjust' the option-implied density. This introduces two types of misspecification; firstly, functional form misspecification regarding the shape of the stochastic discount factor, and secondly, dynamic misspecification from restricting the conditioning information to a set of observed variables. Pricing kernels are typically estimated by the ratio of the current option-implied density and a historical density estimator (Härdle et al., 2014; Beare and Schmidt, 2014). The critical assumption here is that the historical distribution of returns corresponds to the forward-looking distribution of the market. Any inconsistency due to omitted state variables will be erroneously ascribed to the pricing kernel. Empirically, pricing kernels have been reported to not be monotonically decreasing in wealth, in spite of standard preferences assumptions. The procedure of restricting the information set in computing the physical density led Linn et al. (2017) to suggest that this 'pricing kernel puzzle' might be a mere byproduct of econometric technique rather than a behavioral or economic phenomenon.

To avoid restrictive assumptions on the information set, this paper studies the nonparametric estimation of the pricing kernel from conditional moment restrictions on the net discounted payoff of option trades given the included state variables. These restrictions hold regardless of whether the option prices contain additional information on their discounted payoff beyond the observed conditioners. In fact, such additional dependence allows for efficiency gains by giving more weight to more informative option prices.

The estimation approximates the pricing kernel by a series expansion, and combines the conditional moment restrictions for option prices, the underlying stock return, and the risk-free rate, into a linear sieve minimum-distance criterion (Ai and Chen, 2003). Its minimizer yields a consistent estimator when the number of time periods and number of basis functions increase appropriately. The series estimator can avoid some of the instability of density ratios

[1] In particular, the linear projection of the pricing kernel on the space of asset returns is minimum variance among all valid pricing kernels (Cochrane, 2009). The nonlinear projection does not need to be spanned by the asset returns. It can be formally defined as an orthogonal projection in an appropriate Hilbert space.

estimates based on dividing kernel-smoothed risk-neutral and physical densities. Moreover, shape constraints such as non-negativity and monotonicity can be directly imposed. However, as there are few liquid option prices in the tails of the distribution, there remains a need to regularize the tails of the fitted pricing kernels. After observing that option prices identify the first two partial moments of the risk-neutral distribution outside their liquidly traded range of strike prices, we propose to 'paste' parametric tails to the otherwise under- or weakly identified regions in order to stabilize the estimates.

Market-based predictive densities can then be extracted via the implied densities from a variant of the conditional empirical likelihood problem studied in Gagliardini et al. (2011). While their approach combines current option prices and historical stock returns, the method in this paper uses option prices and stock returns over the whole sampling period. More fundamentally, we do not assume all relevant state variables are observed and reflected in options prices. Instead, we introduce an option pricing error which can be decomposed into the unobserved variation in the pricing kernel and in the objective density, controlling for a set of observed conditioning variables. The relative importance of both sources of variation is measured by a tuning parameter for the partial smoothing of current and past option prices. The tuning parameter can be consistently determined by maximum predictive likelihood if the conditional physical density is a linear combination of the risk-neutral and unconditional density. It is also shown that choosing the pricing kernel by maximizing the predictive likelihood of adjusted risk-neutral densities directly, as in Linn et al. (2017), identifies a different function than the projected pricing kernel in the presence of missing conditioning variables. Alternatively, we can estimate the tuning parameter by regressing the discounted payoffs on option prices. We consider local regressions and find considerable heterogeneity in informativeness across moneyness levels, as call and ATM options are more informative about their payoff than put and OTM options.

The remainder of this paper is organized as follows. Section 2 introduces the projected pricing kernels, including their identification from conditional moment restrictions and their consistent estimation by series approximation. Section 3 studies the conditional empirical likelihood estimator that yields implied conditional densities that can be used for forecasting and computing expected returns. Section 4 concludes.

2 Nonparametric pricing kernel estimation

Standard no-arbitrage theory implies the existence of a stochastic discount factor M_t, such that the discounted price vector $M_t P_t$ is a martingale. This implies that the price of any traded asset can be represented as the conditional expectation

$$P_t = E_t \left(\frac{M_{t+\tau}}{M_t} P_{t+\tau} \right). \tag{1}$$

In consumption-based asset pricing models, the pricing equation (1) results from the first-order conditions for consumption and investment. In a representative agent with time-separable utility and discount rate β, the pricing kernel equals the marginal rate of substitution

$$\frac{M_{t+\tau}}{M_t} = \beta^\tau \frac{U'(C_{t+1})}{U'(C_t)}. \tag{2}$$

The traditional estimation approach lies on specifying some parametric utility function, such as power utility, and minimizing a GMM-criterion Hansen (1982). Numerous extensions of such preferences have been studied, such as habit formation and recursive preferences, to better match observed risk premia; for an overview see Ludvigson (2011). Linear factor models, which directly specify the pricing kernel as a linear function of some priced risk factors, are commonly used for matching cross-sectional returns. However, concerns about misspecifying the functional form of the pricing kernel have led Bansal and Viswanathan (1993) and Chapman (1997), among others, to investigate more general nonlinear functions of the priced risk factors.

The nonlinear payoff structure of option prices is particularly useful for recovering the shape of the pricing kernel as a function of the underlying asset return. Moreover, the dynamics of option prices are informative about how this shape varies over time. In this paper, we therefore project the pricing kernel on both the endogenous return $R_{t,t+\tau}$ as well as some exogenous state variables Z_t that describe the dynamics:

$$m(R_{t,t+\tau}, Z_t) := E(M_{t,t+\tau} \mid R_{t,t+\tau}, Z_t). \tag{3}$$

This projected pricing kernel has the same pricing properties as its original for payoffs that depend on R_{t+1}. In the absence of agreement on how an equilibrium-based stochastic discount factor $M_{t,t+\tau}$ should be specified, its conditional expectation given next period's return and current information variables can be determined from option price data, and provides a conditional moment that any proposed stochastic discount factor should satisfy.

The projected pricing kernel links the 'risk-neutral' probability measure $\mathbb{Q}$ and the physical distribution $\mathbb{P}$ via

$$\frac{d\mathbb{Q}(R_{t+1} \mid Z_t)}{d\mathbb{P}(R_{t+1} \mid Z_t)} = R^f_{t,t+\tau} m(R_{t,t+\tau}, R^f_{t,t+\tau}, Z_t), \tag{4}$$

where $R^f_{t,t+\tau}$ is the risk free rate. The density of the risk-neutral distribution can be recovered from option prices with varying strike price Aït-Sahalia and Lo (1998). Equation (4) therefore suggests that the projected pricing kernel can be estimated as the ratio of estimates of the conditional risk-neutral and physical densities. This approach has been taken by Rosenberg and Engle (2000); Giacomini and Härdle (2008); Grith et al. (2012); and Beare and Schmidt (2014), among others. A difficulty is that this involves division by a density estimator, whose small values in the tails may lead to erratic behaviour of the estimated ratio and corresponding inference procedure. In particular, it has typically lead to non-monotonic pricing kernels, which are difficult to rationalize with standard preference assumptions. In the following we avoid separate estimation of the densities, by instead formulating conditional moment restrictions that identify the projected pricing kernel.

2.1 Conditional moment restrictions

In a market with a stock S_t with return $R_{t+1} = \frac{S_{t+1}}{S_t}$ and a one-period bond B_t with risk-free rate R_t^f the general pricing model (1) implies:

$$1 = E_t\left(M_{t,t+1}R_{t+1}\right) \tag{5}$$

$$1 = E_t\left(M_{t,t+1}R_t^f\right), \tag{6}$$

where $M_{t,t+\tau} \equiv \frac{M_{t+\tau}}{M_t}$. A useful assumption is that the projected pricing kernel is separable in the risk free rate R_t^f, which can be included in Z_t. In this case we write

$$m(R_{t,t+\tau}, R_{t,t+\tau}^f, Z_t) = e^{-r_{t,t+\tau}^f}\mu(R_{t,t+\tau}, Z_t), \tag{7}$$

in terms of the log risk free rate $r_{t,t+\tau}^f = \log R_{t,t+\tau}^f$. In particular, the risk free rate pricing equation (6) implies

$$E(\mu(R_{t,t+\tau}, Z_t) \mid Z_t) = 1. \tag{8}$$

Similarly, the no-arbitrage price of a European option with strike price K and time to maturity τ is given by the conditional expectation

$$C_t = E_t\left(M_{t,t+\tau}(P_{t+\tau} - K)^+\right)$$

$$\Leftrightarrow \tilde{C}_t = E_t\left(M_{t,t+\tau}(R_{t+\tau} - \kappa_t)^+\right), \tag{9}$$

where $\tilde{C}_t \equiv \frac{C_t}{P_t}$ is the normalized option price, and $\kappa_t \equiv \frac{K}{P_t}$ is referred to as the 'moneyness' of the option. In most models, the price level P_t does not affect the distribution of the return $R_{t+\tau}$, so that the option price is homogeneous in K and P_t and the normalized option prices are stationary.

When estimating the conditional expectation in (9) by specifying a set of conditioning variables Z_t, we generally introduce an error term:

$$\tilde{C}_t = E\left(e^{-r^f_{t+1}}\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ \mid Z_t\right) + \varepsilon_t, \qquad E(\varepsilon_t|Z_t) = 0 \tag{10}$$

The error term ε_t would disappear if the true model is Markovian in the specified state variables. In general, it is a serially correlated residual that captures the dynamics of the omitted state variables. Moreover, the error term may account for observational error in recorded option prices, typically described as market microstructural noise.

Several methods for cross-sectional option pricing require their estimators to match observed option prices exactly under the Markov assumption (e.g. Gagliardini et al., 2011). However, when pooling option prices from different time periods it is natural to allow for mispricing due to conditioning on a restricted number of observed state variables.

The pricing restrictions can be combined into the conditional moment vector:

$$\rho(Z_t, \mu) := E\left(g(R_{t+1}, Z_t, \mu) - P_t \mid Z_t\right) = 0, \tag{11}$$

where

$$g(R_{t+1}, Z_t, \mu) = e^{-r^f_{t+1}}\mu(R_{t+1}, Z_t)h(R_{t+1}, R^f_t, \bar{\kappa}_t), \tag{12}$$

with the payoff vector and price vector, respectively,

$$h(R_{t+1}, R^f_t, \bar{\kappa}_t) = \begin{pmatrix} (R_{t+1} - \bar{\kappa}_t)^+ \\ R_{t+1} \\ R^f_t \end{pmatrix}, \qquad P_t = \begin{pmatrix} \bar{C}_t \\ 1 \\ 1 \end{pmatrix}. \tag{13}$$

where $\bar{C}_t = (\tilde{C}_{t1}, ..., \tilde{C}_{tn})^T$ is the vector of normalized traded option prices at time t with normalized strike prices $\bar{\kappa}_t = (\kappa_{t1}, ..., \kappa_{tn})^T$. We assume for now that the data form a balanced panel of n option prices per time point. This can easily be generalized to an unbalanced panel in line with time-varying trading activity.

The conditional moment restrictions (11) form the basis for identification, estimation, and testing of the projected pricing kernel. The equations should hold for any stochastic discount factor model, as long as there is no arbitrage and the martingale pricing equation (1) holds. In particular, we do not assume that the selected state variables are a Markov process that fully describe the information set. The pricing condition should hold for any set of conditioning variables $Z_t \in \mathcal{F}_t$. However, the more of the time-variation in the option prices is captured by Z_t, the smaller the pricing errors, and the more variation in the projected pricing kernel can be detected. The identification and consistent estimation of $\mu(R_{t+1}, Z_t)$ relies on observing the option prices and state variables over time, merely requiring stationarity of (R_{t+1}, Z_t). Typical state variables would be volatility proxies, lagged market returns, term structure variables, or proxies related to the business cycle. Alternatively, Z_t could include a time argument, which could be dealt with by local-in-time smoothing under a local stationarity assumption.

2.2 Identification

The existence of a pricing kernel that satisfies (1) follows from no arbitrage, and implies that there exists a solution μ_0 that satisfies $\rho(Z_t, \mu_0) = 0$ almost surely. However, in incomplete market models, there may be not be a unique M_{t+1} that satisfies (1), which could lead to multiple projected pricing kernels $\mu(R_{t+1}, Z_t)$ that satisfy (11). This happens when the market return is driven by more state variables than there are traded securities, so that option prices cannot be perfectly hedged.[2] However, the more option prices with different strike prices are traded, the closer to completeness the market for the underlying asset return becomes. In particular, when a continuum of strike prices or moneyness ratios can be traded, any nonlinear function of the market return is replicable, so there must be a unique projected pricing kernel.

In reality, strike prices are traded at a discrete grid, and only within a subset of the range of the market return. When the available strike prices are close to each other, the identified set is likely to be small. However, in the tails of the return distribution there are few liquid option prices, which may lead to lack of identification of the tails of the projected pricing kernel. This could lead to volatile estimates, and could explain the various shapes in empirical studies (e.g. Rosenberg and Engle, 2000; Beare and Schmidt, 2014).

The stability of the estimators can be increased by switching to a semiparametric approach that imposes some functional form in regions which are poorly identified. Therefore our identification strategy is to treat the normalized strike prices κ_t as a continuous but truncated random variable, and to assume a parametric functional form for the tail of the projected pricing kernel. The randomness of κ_t implies that the projected pricing kernel is point identified in the actively traded range. The parametric functional form implies that the tail risk parameters are point identified by adding the pricing equations for stock and risk free returns. These parameters can still be a unspecified functions of the exogenous variables Z_t.

Suppose that the moneyness κ_t of traded option prices is a continuous random variable, bounded in the range (κ^l, κ^u). Although in practice only a fixed set of strike prices is traded, their normalization by the stock price leads to a continuous distribution in a range centered around one. Since κ_t is observed at time t, we can write the conditional moments in (11) associated with option prices as

$$\rho^C(Z_t, \kappa_t, \mu) := E\left(\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ - \tilde{C}_t \mid Z_t, \kappa_t\right) = 0, \tag{14}$$

[2]Formally, completeness requires a rank condition on the exposure matrix of the asset returns to the driving state variables.

Provided R_{t+1} is conditionally independent of κ_t after controlling for Z_t, the normalized strike prices only affects the moment via the known payoff function.[3] Hence

$$E\left(\mu(R_{t+1}, Z_t)(R_{t+1} - \kappa_t)^+ | Z_t = z, \kappa_t = \kappa\right) = \int_{\kappa}^{\infty} \mu(y, z)(y - \kappa) f(y|z) dy, \tag{15}$$

in terms of the conditional density $f_{R_{t+1}|Z_t}(y|z)$. Following Breeden and Litzenberger (1978), differentiating twice with respect to κ yields

$$\frac{\partial^2}{\partial^2 \kappa} \int_{\kappa}^{\infty} \mu(y, z)(y - \kappa) f(y|z) dy = \mu(\kappa, z) f(\kappa|z), \qquad \kappa \in [\kappa^l, \kappa^u],\ z \in \mathcal{Z}, \tag{16}$$

so that (14) yields

$$\mu_0(\kappa, z) = \frac{\partial^2 E\left(\tilde{C}_t \mid Z_t = z, \kappa_t = \kappa\right) / \partial^2 \kappa}{f(\kappa|z)}, \qquad \kappa \in [\kappa^l, \kappa^u],\ z \in \mathcal{Z}. \tag{17}$$

This shows that $\mu(R_{t+1}, Z_t)$ is point identified for $R_{t+1} \in [\kappa^l, \kappa^u]$. The projected pricing kernel in the lower and upper tails of the market return can be identified from the stock, the risk free return, and put and call option prices at the boundary of the optionable range. That is, put options with $\kappa_t = \kappa^l$ and payout $(\kappa^l - R_{t+1})^+$ and call options with $\kappa_t = \kappa^u$. To see this, denote

$$G(z) = \int_{\kappa^l}^{\kappa^u} \mu(y, z) y f(y|z) dy,$$
$$H(z) = \int_{\kappa^l}^{\kappa^u} \mu(y, z) f(y|z) dy.$$

Both of these are identified by the option prices in the actively traded range. After substituting the identified region, the conditional moment restrictions for the stock (5) and risk free rate (6) can be written as

$$1 = \int_0^{\kappa^l} \mu(y, z) y f(y|z) dy + G(z) + \int_{\kappa^u}^{\infty} \mu(y, z) y f(y|z) dy,$$
$$1 = \int_0^{\kappa^l} \mu(y, z) f(y|z) dy + H(z) + \int_{\kappa^u}^{\infty} \mu(y, z) f(y|z) dy.$$

Furthermore, for a put option with moneyness κ^l and a call option with moneyness κ^u we have

$$E(P_t \mid Z_t, \kappa^l) = \int_0^{\kappa^l} \mu(y, z)(\kappa^l - y) f(y|z) dy,$$
$$E(C_t \mid Z_t, \kappa^u) = \int_{\kappa^u}^{\infty} \mu(y, z)(y - \kappa^u) f(y|z) dy.$$

[3] In an efficient and liquid market the strike prices that are traded are unlikely to have any predictive power for the market return, as asset prices would adjust immediately.

Substituting the latter into the stock condition gives

$$1 + E(P_t \mid Z_t, \kappa^l) - E(C_t \mid Z_t, \kappa^u u) - G(z) = \kappa^l \int_0^{\kappa^l} \mu(y,z) f(y|z) dy + \kappa^u \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy,$$

which together with the risk free rate condition yields

$$\begin{pmatrix} \int_0^{\kappa^l} \mu(y,z) f(y|z) dy \\ \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy \end{pmatrix} = \begin{pmatrix} \kappa^l & \kappa^u \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 + E(P_t \mid Z_t, \kappa^l) - E(C_t \mid Z_t, \kappa^u) - G(z) \\ 1 - H(z) \end{pmatrix},$$

provided that $\kappa^l \neq \kappa^q$, i.e. there is more than one option. Hence the integrated risk-neutral probability masses of the lower and upper tail are identified. From the put and call prices it follows that

$$\int_0^{\kappa^l} \mu(y,z) y f(y|z) dy = \int_0^{\kappa^l} \mu(y,z) f(y|z) dy - E(P_t|Z_t, \kappa^l)$$
$$\int_{\kappa^u}^{\infty} \mu(y,z) y f(y|z) dy = E(C_t|Z_t, \kappa^l) - \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy.$$

Altogether, apart from $\{\mu_0(y,z) : y \in [\kappa^l, \kappa^u], z \in \mathcal{Z}\}$, the following four quantities are identified:

$$\mathbb{Q}(R_{t+1} \leq \kappa^l | Z_t = z) = \int_0^{\kappa^l} \mu(y,z) f(y|z) dy$$
$$\mathbb{Q}(R_{t+1} \geq \kappa^u | Z_t = z) = \int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy$$
$$E^{\mathbb{Q}} \left(R_{t+1} | R_{t+1} \leq \kappa^l, Z_t = z \right) = \frac{\int_0^{\kappa^l} \mu(y,z) y f(y|z) dy}{\int_0^{\kappa^l} \mu(y,z) f(y|z) dy}$$
$$E^{\mathbb{Q}} \left(R_{t+1} | R_{t+1} \geq \kappa^u, Z_t = z \right) = \frac{\int_{\kappa^u}^{\infty} \mu(y,z) y f(y|z) dy}{\int_{\kappa^u}^{\infty} \mu(y,z) f(y|z) dy}.$$

These quantities are of direct use for risk measurement, for example to compute the risk-neutral Value-at-Risk

$$VaR^{\mathbb{Q}}(\alpha; z) = \inf\{x \in \mathbb{R} : \mathbb{Q}(-R_{t+1} > x | Z_t = z) \leq 1 - \alpha\}, \tag{18}$$

provided $\mathbb{Q}(R_{t+1} \leq \kappa^l | Z_t = z) \leq \alpha$.[4] However, risk measures based on the risk-neutral distribution might be too conservative, as they tend to shift probability weight from gains to losses. The risk-neutral risk measures could be transformed into corresponding $\mathbb{P}$-measures, which requires knowledge of the projected pricing kernel μ in the tails. However, outside the traded range μ is only identified up to its partial expectations.

Note that without the argument z in the pricing kernel, the four quantities would be Type I (partial) integral equations which are commonly studied in the literature on nonparametric IV

[4] This model-free approach to risk management has been advocated by Aït-Sahalia and Lo (2000), among others.

estimation (e.g. Newey and Powell, 2003; Darolles et al., 2011). In this case a sufficient condition for identification is bounded completeness of the conditional density $f(y|z)$ (Blundell et al., 2007). However, the presence of z inside the unknown function of interest prevents full nonparametric identification. For any four different non-trivial functions of y, there is a linear combination with weights depending on z that can match the four conditional moments for any z. Instead, by assuming a parametric form of the pricing kernel in the tails, the four identified partial moments can be used to identify up to four parameters that can be functions of z. A popular class of models for the pricing kernel is of the exponential-affine form:

$$\mu(R_{t+1}, Z_t) = e^{-\beta_0(Z_t)+\beta_1(Z_t)r_{t,t+1}}, \quad R_{t+1} \notin [\kappa^l, \kappa^u].$$

This class has some attractive properties such as being positive, and time-aggregable when $\beta_1(z)$ is linear. Without loss of generality, we may take the middle part of the pricing kernel to also satisfy this shape but multiplied with an unknown 'correction function' of both arguments. This correction function should then be constant outside the identified interval:

$$\mu(R_{t+1}, Z_t) = e^{-\beta_0(Z_t)-\beta_1(Z_t)r_{t+1}}\alpha(R_{t+1}, Z_t), \quad R_{t+1} \notin [\kappa^l, \kappa^u], \tag{19}$$

$$\alpha(R_{t+1}, Z_t) = \begin{cases} \alpha^l(Z_t) & 0 \leq R_{t+1} \leq \kappa^l \\ \alpha^u(Z_t) & \kappa^l \leq R_{t+1} < \infty \end{cases}$$

The above specification has exactly four parameters for given Z_t. However, note that $\beta_0(z)$ is not separably identified from the level of $\alpha^l(z)$ and $\alpha^u(z)$. A useful normalization is therefore to set $\alpha^l(z) = 1$ and/or $\alpha^u(z) = 1$ for any $z \in \mathcal{Z}$. This explains the interpretation of a correction factor around the exponential-affine model. The remaining parameters for given Z_t are then overidentified by the partial moments, while the middle part is still identified by the option prices. The other advantage of the exponential-affine, or any other parametric model that can grow without bounds when $R_{t+1} \to 0$, is that it can be reasonably assumed that the unknown function lies in a compact set. This overcomes the ill-posed inverse problem and simplifies the asymptotic theory considerably, as the solution becomes a continuous mapping of the distribution of the data (Newey and Powell, 2003).

An alternative identification strategy is to assume that a subset of the conditioning variables $Z_{1t} \subsetneq Z_t$ does not affect the pricing kernel. Only one such variable would generally yield sufficient exogenous variation to identify the pricing kernel as a function of the endogenous market return. This has been used by Chen and Ludvigson (2009), under the assumption of completeness of an expected return weighted-conditional density. Even in this case, it may be useful for estimation to write the pricing kernel as

$$\mu(R_{t+1}, Z_{1t}) = e^{-\beta_0(Z_{1t})-\beta_1(Z_{1t})r_{t,t+1}}\alpha(R_{t+1}, Z_{1t}), \quad \alpha \in \Theta_\alpha, \tag{20}$$

where Θ_α is a compact functional space. The first part grows exponentially when $r_{t,t+1} \to -\infty$, so will generally not lie in a compact space, which allows separate identification of the exponential-affine component and α. Intuitively, such exogenous variation would require conditioning variables Z_{2t} that affect the distribution of the market return, but not the compensation for risk which would alter the shape of the pricing kernel. However, typical conditioning variables such as volatility factors, lagged market returns, or price-dividend ratios, that affect the conditional return distribution, are likely to influence attitudes towards risk taking and thus the pricing kernel as well. For example, fear-based models suggest that risk aversion is higher during episodes of market volatility. This would rule out identification of the market return dimension of the pricing kernel. Moreover, even when such exogenous variables exist, their instrument strength might be weak.

Some further identification results are available in related asset pricing models. Ross (2015) provides identification conditions of the pricing kernel and the objective density from option prices alone for a discrete-state Markov model, though Borovička et al. (2014) argue this actually only identifies the 'long-run' density function. Escanciano et al. (2015) identify the marginal utility function in a consumption-based model by rewriting the Euler equation as a Type II integral equation, in which form the projected pricing kernel generally cannot be written.

2.3 Estimation

Sample analogues of the conditional moments (11) can be used to estimate the projected pricing kernel using a minimum-distance criterion. Asymptotic theory for estimating an unknown function identified by conditional moments has been developed in Newey and Powell (2003), Ai and Chen (2003), and others, and is adapted here to the option pricing setting. Re-define the conditioning variables as (Z_t, X_t), where as before Z_t affects the distribution of the market return R_{t+1} and possibly the projected pricing kernel μ, while X_t are the exogenous characteristics of the option prices. The former will typically be $Z_t = (R_{t-1}, \sigma_t)$ with σ_t a proxy for conditional volatility, whereas the latter includes variables such as the moneyness ratio κ_t, the risk-free rate $r^f_{t,t+1}$), and possibly the time-to-maturity τ_t which will be one period in this section.

Define the infinite-dimensional parameter vector $\theta = (\beta_0, \beta_1, \alpha) \in \Theta_\beta \times \Theta_\beta \times \Theta_\alpha \equiv \Theta$, where Θ_β and Θ_α are compact spaces. The previous section implies that there is a unique $\theta_0 \in \Theta$ which minimizes the population criterion

$$Q(\theta) = E\left(\rho(Z_t, X_t, \theta)^T \Sigma_t^{-1} \rho(Z_t, X_t, \theta)\right), \tag{21}$$

for any series of positive definite weighting matrix Σ_t. This is the criterion discussed in Newey and Powell (2003) and Ai and Chen (2003). Now define the theoretical pricing function

$$P(Z_t, X_t, \theta) = E\left(m(R_{t+1}, Z_t, X_t, \theta)h(R_{t+1}, X_t) \mid Z_t, X_t\right). \tag{22}$$

Note $\rho(Z_t, X_t, \theta) = P(Z_t, X_t, \theta) - E(P_t \mid Z_t, X_t)$. Then a closely related objective is

$$Q'(\theta) = E\left((P(Z_t, X_t, \theta) - P_t)^T \Sigma_t^{-1} (P(Z_t, X_t, \theta) - P_t)\right), \tag{23}$$

which minimizes the fit between the theoretical and observed prices according to a weighted nonlinear least-squares criterion. Since the difference

$$Q'(\theta) - Q(\theta) = E\left(\epsilon_t^T \Sigma_t^{-1} \epsilon_t\right) \tag{24}$$

does not depend on the parameter vector θ, $Q'(\theta)$ is also uniquely minimized at $\theta_0 \in \Theta$. The sample analogue of $Q(\theta)$ based on first stage nonparametric estimators $\hat{\rho}(Z_t, X_t, \theta)$ and $\hat{\Sigma}_t$, yields the sample criterion

$$Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \hat{\rho}(Z_t, X_t, \theta)^T \hat{\Sigma}_t \hat{\rho}(Z_t, X_t, \theta). \tag{25}$$

The sample analogue of $Q'(\theta)$ requires a first stage nonparametric estimator of the pricing function $P(Z_t, X_t, \theta)$:

$$Q'_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (\hat{P}(Z_t, X_t, \theta) - P_t)^T \hat{\Sigma}_t (\hat{P}(Z_t, X_t, \theta) - P_t). \tag{26}$$

This criterion measures the discrepancy between observed prices P_t and estimated theoretical prices $\hat{P}$. Its minimization is therefore similar to the common practice of calibrating the parameter vector by minimizing the sum of squared pricing errors in a certain period. Calibration methods, however, often do not have a well-defined population criterion, so that estimated parameters may not be identified and as a result estimates may vary considerably with the sample that is used.[5] Linking $Q'_T(\theta)$ to sample analogues of the conditional moment restriction (11) provides a statistical framework for model selection and inference. Furthermore, conditions for consistency of the minimizer of $Q'_T(\theta)$ can be stated in terms of the pricing error directly instead of on the moment residual $g(R_{t+1}, Z_t, X_t, \theta)$. The moment residual is a nonlinear function of the market return and will generally require more strict conditions than the pricing errors.

The minimizers of the sample analogues $Q_T(\theta)$ and $Q'_T(\theta)$ need not be the same. This will depend on the first stage estimation of $\rho(Z_t, X_t, \theta)$. When the first stage estimator for $\hat{\rho}(Z_t, X_t, \theta)$

[5] When model parameters are assumed constant over time, this displays an internal inconsistency of the model. The classic example is the Black-Scholes model, which assumes constant volatility, yet estimated implied volatilities vary severely over time and are therefore treated as a time series itself.

and $\hat{P}(Z_t, X_t, \theta)$ are series estimators and θ appears linearly in ρ, $Q'_T(\theta)$ is equivalent to a two-stage least-squares estimate. This follows from the idempotency of the projection matrix of functions of (Z_t, X_t) (Newey and Powell, 2003). While the kernel weighting matrix does not satisfy this property, the next section shows that the difference vanishes for large samples under standard conditions.

The benefits of distinguishing (Z_t, X_t) become apparent when estimating the pricing function $\hat{P}(Z_t, X_t, \theta)$. Since R_{t+1} is conditionally independent of X_t given Z_t, it follows that

$$P(z, x, \theta) = \int_0^\infty m(y, z, x, \theta) h(y, x) f(y|z) dy, \tag{27}$$

where the conditional density $f(R_{t+1}|Z_t)$ does not depend on the option characteristics X_t. This suggests the natural plug-in estimator

$$\hat{P}(z, x, \theta) = \int_0^\infty m(y, z, x, \theta) h(y, x) \hat{f}(y|z) dy, \tag{28}$$

in terms of a nonparametric conditional density estimator $\hat{f}(y|z)$, such as the kernel-based estimator

$$\hat{f}_h(y|z) = \frac{\sum_{t=1}^{T-1} K_{h_y}(R_{t+1} - y) K_{h_z}(Z_t - z)}{\sum_{t=1}^{T-1} K_{h_z}(Z_t - z)} \tag{29}$$

for $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ for some kernel $K(\cdot)$ and bandwidths h_y and h_z. For the semiparametric pricing kernel (19), with $X_t = (R_t^f, \kappa_t)$ this becomes

$$\hat{P}_h^C(Z_t, X_t, \theta) = \int_0^\infty e^{-r_t^f - \beta_0(Z_t) - \beta_1(Z_t) \ln y} \alpha(y, Z_t)(y - \kappa_t)^+ \hat{f}_h(y|z) dy, \tag{30}$$

and similarly for the stock and risk free rate pricing equations in (11).

Another advantage of the pricing error-based criterion (26) is that it does not use the observation R_{T+1}, so that also the last observed price P_T can be included. The conditional density estimator then simply leaves out the last period. For the direct local estimation of the conditional moments in (25) it would not be possible to include the time-T observations.

Instead of minimizing $Q'_T(\theta)$ over the infinite dimensional functional space Θ, the method of sieves controls the complexity of the model in relation to the sample size by minimizing over an approximating finite-dimensional sieve spaces $\Theta_T \subseteq \Theta_{T+1} \subseteq \cdots \subseteq \Theta$ which become dense in Θ. This gives rise to the Sieve-Minimum Distance estimator (Ai and Chen, 2003)

$$\hat{\theta}_T = \min_{\theta \in \Theta_T} Q'_T(\theta). \tag{31}$$

Each Θ_T will depend on a finite number of parameters only. When $\hat{\theta}_T$ is a finite linear combination of a known set of functions it is a series estimator. The compactness assumption on Θ relies on the norm that is specified. Our specification follows closely that of Newey and Powell (2003), who use the Sobolev norm which sums the integrals of partial derivatives. To this end, consider a d-dimensional function g and let λ denote a $d \times 1$ vector of nonnegative integers, with $|\lambda| = \sum_{l=1}^{d} \lambda_1^d$, and $D^\lambda g(y) = \frac{\partial^{|\lambda|} g(y)}{\partial y_1^{\lambda_1} \dots \partial y_d^{\lambda_d}}$. Then for some positive integers m and p, define the unweighted Sobolev norm

$$\|g\|_{m,p} = \left\{ \sum_{|\lambda| \leq m} \int \left(D^\lambda g(z) \right)^p dz \right\}^{1/p}. \tag{32}$$

It is implicitly assumed that the random variables that enter g are standardized to have mean zero and variance one. For $p = 2$ and some positive constant B, define the functional space

$$\Theta = \left\{ g(z) : \|g\|_{m,2}^2 \leq B \right\} \tag{33}$$

Now consider the finite-dimensional series approximator $g_L(w) = \sum_{l=1}^{L} \gamma_j p_j(w)$, in terms of a set of basis functions $(p_1(w), ..., p_L(w))$. Define

$$\Lambda_L = \sum_{|\lambda| \leq m} \int D^\lambda p_L(z) D^\lambda p_L(z)^T dz \tag{34}$$

Then it holds that $g_L(w) = \sum_{l=1}^{L} \gamma_j p_j(w) \in \Theta$ if and only if $\gamma^T \Lambda_J \gamma \leq B$. Thus we can define the optimization in (31) to be over the compact finite-dimensional subspace $\Theta_T \subseteq \Theta_{T+1} \subseteq ... \subseteq \Theta$, where

$$\Theta_T = \left\{ g(w) = \sum_{l=1}^{L(T)} \gamma_j p_j(w) : \gamma^T \Lambda_{L(T)} \gamma \leq B \right\}.$$

Also define the Sobolev sup-norm

$$\|g\|_{m,\infty} = \max_{|\lambda| \leq m} \sup_z \left| D^\lambda g(z) \right|. \tag{35}$$

Then the closure $\bar{\Theta}$ of Θ with respect to the norm $\|g\|_{m,\infty}$ is compact for that norm (Gallant and Nychka, 1987; Newey and Powell, 2003). For the case of vector-valued functions, such as $\theta = (\beta_0, \beta_1, \alpha)$, define for $p \geq 1$

$$\|\theta\|_{m,p} = \|\beta_0\|_{m,p} + \|\beta_1\|_{m,p} + \|\alpha\|_{m,p}. \tag{36}$$

Note that in model (19) the functions β_0 and β_1 are supported on $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$, while α is supported on $\mathbb{R}^+ \times \mathcal{Z} \subset \mathbb{R}^{d_z+1}$. For each of these functions we can construct multivariate basis functions using tensor products of univariate basis functions $p_L(w)$ (Ai and Chen, 2003). If for each variable we use L basis functions, this leads in total to L^{d_z+1} coefficient for the constructed functions in case the endogenous market return is included, and L^{d_z} coefficient otherwise. For the exponential-affine model (19), we may want to impose smoothness at the boundary points where $\alpha^l(z) = \alpha^u(z) = 1$. This may be achieved by requiring the true function α to lie in the set

$$\Theta'_\alpha = \left\{ \alpha(y,z) : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}^+; \|\alpha\|^2_{m,2} \leq B_\alpha, D\alpha(\kappa^l) = D\alpha(\kappa^u) = 0 \right\}, \tag{37}$$

where $\mathcal{Y} = [\kappa^l, \kappa^u]$ is the identified region of $\alpha(y,z)$ as a function of y. The restriction on the first derivative at the boundary points translate into a linear restriction on the coefficients of the series estimator.

In the next section we derive conditions for the consistency result $\|\hat{\theta} - \theta_0\|_{m,\infty} \xrightarrow{p} 0$.

2.4 Consistency

We demonstrate consistency of the estimator $\hat{\theta}$ given by (31), when the projected pricing kernel has the general form $e^{-r_t^f} m(R_{t+1}, Z_t, \theta)$. Consider the following set of assumptions:

A1 For every $\theta \in \Theta$ with $\theta \neq \theta_0$, $\rho(Z_t, X_t, \theta) \neq 0$ for some (Z_t, X_t), while $\rho(Z_t, X_t, \theta_0) = 0$ a.s. for some $\theta_0 \in \Theta$

A2 $\hat{f}_h(R_{t+1}|Z_t = z) \xrightarrow{p} f(R_{t+1}|Z_t = z)$, the joint distribution function $f(R_{t+1}, Z_t)$ is continuous and bounded, $f(R_{t+1}|Z_t, X_t) = f(R_{t+1}|Z_t)$, R_{t+1} is supported on $\mathbb{R}^+$, and (Z_t, X_t) has compact and convex support $\mathcal{Z} \times \mathcal{X} \subsetneq \mathbb{R}^{d_z + d_x}$

A3 $\hat{\Sigma}_t$ and Σ_t are positive definite matrices, such that Σ_t is bounded and $\hat{\Sigma}_t \xrightarrow{a.s.} \Sigma_t$ for every t

A4 $|m(R_{t+1}, Z_t, \theta) - m(R_{t+1}, Z_t, \tilde{\theta})| \leq b(R_{t+1}, Z_t)\|\theta - \tilde{\theta}\|^v$ for some $v > 0$ with $E\left(b(R_{t+1}, Z_t)^2 R^2_{t+1}|Z_t\right) < \infty$, and $\mathrm{Var}\left(m(R_{t+1}, Z_t, \theta_0)|Z_t\right) < \infty$

A5 $(R_{t+1}, Z_t, X_t, \varepsilon_t)$ is a strong mixing stationary process, with $E\left(\varepsilon_t^2\right) < \infty$

Theorem 1. *Under assumptions A1-A5, the estimator defined by (31) satisfies*

$$\|\hat{\theta} - \theta_0\|_{m,\infty} \xrightarrow{p} 0 \tag{38}$$

when $T \to \infty$, $L \to \infty$, $L^{d_z+1}/T \to 0$, $h \to 0$, *and* $Th^{d_z} \to \infty$.

We can specialize this result to specific pricing kernels, such as the semiparametric exponential-affine form (19) with

$$m(R_{t+1}, Z_t, \theta) = e^{-\beta_0(Z_t) - \beta_1(Z_t) r_{t+1}} \alpha(R_{t+1}, Z_t).$$

Re-define the infinite-dimensional parameter space as $\Theta = \Theta_{\beta_0} \times \Theta_{\beta_1} \times \Theta'_\alpha$, where[6]

$$\begin{aligned}\Theta_{\beta_0} &= \left\{\beta(z) : \|\beta\|^2_{m,2} \leq B_0\right\} \\ \Theta_{\beta_1} &= \left\{\beta(z) : \|\beta\|^2_{m,2} \leq B_1, \beta_1 \geq 0\right\}.\end{aligned}$$

Consider the additional assumption

A6 $E\left(R_{t+1}^{-4(2\beta_1(Z_t)-1)}|Z_t\right) < \infty$ for all $\beta_1 \in \Theta_{\beta_1}$

Corollary 1.1. *Under assumptions A2-A3 and A5-A6, the estimator defined by (31) with pricing kernel (2.4) satisfies for any $\lambda| \leq m$*

$$\begin{aligned}\max_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}\left|D^\lambda\left(\hat{\alpha}(y,z) - \alpha(y,z)\right)\right| &\xrightarrow{p} 0 \\ \max_{z\in\mathcal{Z}}\left|D^\lambda\left(\hat{\beta}_i(z) - \beta_i(z)\right)\right| &\xrightarrow{p} 0 \qquad i \in \{0,1\},\end{aligned}$$

when $T \to \infty$, $L \to \infty$, $L^{d_z+1}/T \to 0$, $h \to 0$, and $Th^{d_z} \to \infty$.

2.5 Empirical likelihood formulation

The idea behind information-theoretic estimation is to simultaneously estimate the parameters and the likelihood of each period's observations subject to the theoretical moment conditions. Treating the option prices as random given the conditioning variables, the Conditional Empirical Likelihood (CEL) problem based on the conditional moment restriction (14) is defined as (Kitamura et al., 2004; Smith, 2007)

$$\begin{aligned}\min_{(p_{ts}),\theta} \quad & \sum_{t=1}^T\sum_{s=1}^T\left(\frac{p_{ts}-w_{ts}}{w_{ts}}\right)^2 w_{ts} \\ \text{s.t.} \quad & \sum_{s=1}^T p_{ts}\left(g(R_{s+1}, Z_s, X_s, \theta) - P_s\right) = 0 \\ & \sum_{s=1}^T p_{ts} = 1.\end{aligned} \tag{39}$$

The kernel weights $w_{ts} = \frac{K_h(Z_s-Z_t,X_s-X_t)}{\sum_r K_h(Z_r-Z_t,X_r-X_t)}$ compute the distance between observations according to their conditioning variables in terms of the kernel $K(\cdot)$. The implied probabilities p_{ts} are chosen to be as close as possible to the empirical probabilities w_{ts} while satisfying the pricing restrictions, which are computed using kernel smoothing of the option prices and the return realizations. The expectations over the discrete probabilities w_{ts} and a continuous kernel density

[6] We do not require β_0 to be nonnegative, as it is interpreted as the difference between the risk free rate and a subjective time discount parameter. A simplification would be to set $\beta_0 = 0$.

estimator are close when $h_y \approx 0$, as seen from

$$\sum_{s=1}^{T} p_{ts} g(y, Z_s, X_s, \theta) \approx \int g(R_{s+1}, Z_t, X_t, \theta) \hat{f}_{h,p}(y|Z_t) dy, \tag{40}$$

where

$$\hat{f}_{h,p}(y|Z_t) = \sum_{s=1}^{T} p_{ts} K_{h_y}(y - R_{s+1}). \tag{41}$$

The implied probabilities can be found in closed form as a function of θ (Antoine et al., 2007). They give rise to closed-form implied densities $\hat{f}_{h,p}(y|Z_t)$ which can be directly used for forecasting. Although the value R_{T+1} is not observed, it can be replaced by an expected value or be left out and the existing $T-1$ observations re-weighted. The criterion (39) when optimized over the implied probabilities with the approximation (40) can be shown to be equivalent to

$$\mathcal{L}'_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (\hat{P}_h(Z_t, X_t, \theta) - \hat{E}_h(P_t \mid Z_t, X_t))^T \hat{V'}_h^{-1}(Z_t, X_t; \theta)(\hat{P}_h(Z_t, X_t, \theta) - \hat{E}_h(P_t \mid Z_t, X_t)), \tag{42}$$

where

$$\hat{V'}_h(Z_t, X_t, \theta) = \widehat{\mathrm{Var}}_h \left(g(R_{t+1}, Z_t, X_t, \theta) - P_t \mid Z_t, X_t \right). \tag{43}$$

This is a continuously updating variant of the minimum-distance criterion (25). The quadratic or Euclidean loss function is computationally attractive as the implied probabilities are available in closed form, which is not the case for other common loss functions. The probability limit of $\mathcal{L}'_T(\theta)$ identifies the same θ as the minimizers of $Q(\theta)$ and $Q'(\theta)$ in Section 2.2, as the weighting matrix only matters for efficiency. The diagonals of the weighting matrix indicate that an efficient estimator gives high weight to those strike prices for which the discounted option trade gain or loss have a low variance. In particular, this weight is increasing in the covariance between the option price and its discounted payoff.

2.6 Empirical pricing kernels

We apply the sieve mininium-distance estimator (31) for estimating the projecting pricing kernel implicit in option contracts written on the S&P 500 index. The option price data are obtained from OptionMetrics over the sampling period January 1996 to June 2016. The contracts have European exercise style with one month maturity. The cum-dividend returns on the S&P 500 index measure the underlying asset return, and the one month Federal Funds rate is used as a proxy for the risk-free rate. The conditioning variable is set as $Z_t = \log VIX_t$, as the VIX is an often-used proxy for market-implied volatility and risk aversion. Figure 1 shows the estimated pricing kernel $\hat{m}(R_{t+1}, Z_t)$ for the bivariate expansion with $K_R = 6$ in the return dimension and $K_Z = 2$ in the log VIX dimension. The estimator minimizes the minimum-distance criterion with continuously updated variance estimator $\hat{\Sigma}_t = \hat{\Sigma}_h(Z_t, \kappa_t, \theta)$ that allows for heteroskedasticity and spatial autocorrelation in the strike dimension. The bandwidth h_z is chosen using the nearest-neighbor method with 30 nearby observations. The pricing kernel appears fairly monotonically decreasing in the return for low and moderate levels of the VIX, in line with standard preferences that imply a higher marginal utility for lower wealth levels. However, for large values of the VIX the fitted pricing kernel appears strongly U-shaped, i.e. increases for both large positive and negative returns. This over-valuation of tail payoffs suggests the VIX contains a strong 'fear' component and does not necessarily reflect a higher probability of tail events. Regarding the dynamics, the estimate suggests that the steepness of the pricing kernels is non-monotonic: it appears to be highest during both very low and high levels of uncertainty as measured by the VIX. The U-shaped pattern is line with the non-monotonicity reported in Song and Xiu (2016) for VIX-dependent pricing kernels. This finding suggests that investors may be averse to both very high and very low levels of uncertainty, as during periods of the latter there may be fewer profitable investment opportunities.

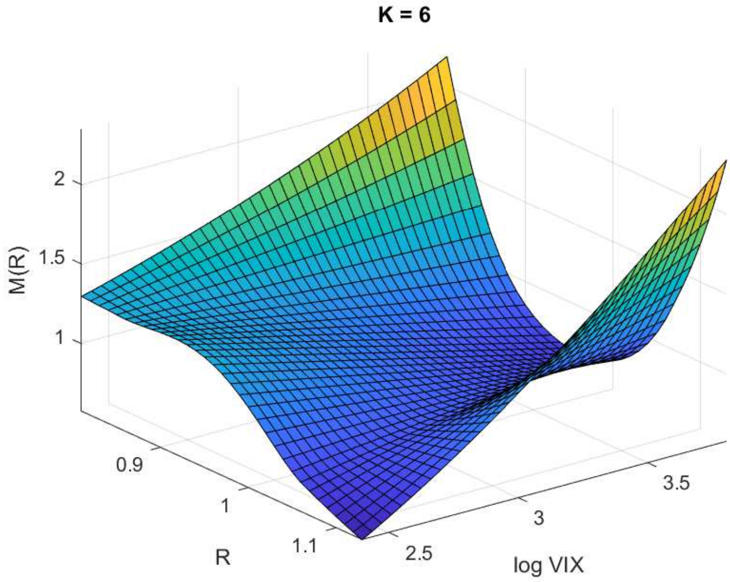

Figure 1: Fitted bivariate pricing kernel expansion with $(K_r, K_z) = 6 * 2$ on option and returns data, from January 1996 until June 2016, based on the continuous-updating estimator.

3 Market-based Density Forecasting

Option prices can be used to extract market-implied densities which incorporate the information available to investors, on top of the historical data that is available to econometricians. In particular, the risk neutral density, or state-price density, can be estimated based on the Breeden and Litzenberger (1978) result

$$f^{\mathbb{Q}}(\kappa|z) = \frac{\partial^2 E\left(C_t|Z_t = z, \kappa_t = \kappa\right)}{\partial^2 \kappa}, \qquad \kappa \in [\kappa^l, \kappa^u],\ z \in \mathcal{Z}. \tag{44}$$

The case $Z_t = t$ allows estimating the call pricing function using a cross-section or rolling window of option prices. This provides a model-free estimate that only requires a practical choice of smoothing parameters.

The main problem with using the risk-neutral density for forecasting is its bias due to the varying marginal utility of wealth over states. Typically, payoffs in poor state of the economy are more valuable to risk averse investors, which leads to overestimating the probability of losses, and therefore to conservative risk measurement. The task then is to 'correct' the risk neutral density into an unbiased estimate of the conditional 'real-world' density using a model for the pricing kernel. Several parametric models have been suggested to perform this transformation, i.e. the Generalized Beta distribution in Liu et al. (2007); Vincent-Humphreys et al. (2012), or the exponential-affine models in Hördahl and Vestin (2005); Gagliardini et al. (2011); León et al. (2012).

Section 2 discussed a flexible semiparametric model for the pricing kernel and its dynamics, and its minimum-distance estimation using conditional moments. A simple plug-in estimator of the physical density $f_t^{\mathbb{P}}(R_{t+1})$ based on equation (4) would therefore be

$$\hat{f}_t^{\mathbb{P}}(R_{t+1}) = \frac{\hat{f}_t^{\mathbb{Q}}(R_{t+1})}{\hat{\mu}(R_{t,t+\tau}, Z_t)}. \tag{45}$$

This density estimator is sensitive to small values of the estimated pricing kernel in the denominator. Moreover it is not guaranteed to be a proper non-negative density function which integrates to unity. Instead, we approach the density extraction problem using the empirical likelihood method, which simultaneously yields parameter estimates and well-behaved implied densities that automatically satisfy the pricing restrictions.

3.1 Implied densities

Suppose the dynamics of the pricing kernel are correctly specified as a function of the included state variable Z_t. In this case we can extract conditional physical densities by exactly matching the current cross-section of option prices. In particular, the following version of the CEL estimator, based on the XMM estimator proposed in Gagliardini et al. (2011), optimizes directly over implied densities instead of probabilities:[7]

$$\min_{f_1,\cdots,f_T,\theta} \sum_{t=1}^{T} \int \left(\frac{f_t(y) - \hat{f}_h(y|Z_t)}{\hat{f}_h(y|Z_t)} \right)^2 \hat{f}_h(y|Z_t)dy \tag{46}$$

$$\text{s.t.} \int g(y, Z_t, X_t, \theta) f_t(y)dy = P_t,$$

$$\int f_t(y)dy = 1. \tag{47}$$

This estimator yields closed-form estimates of the implied conditional densities $f_1, \cdots, f_T$ that incorporate the information in current option prices and can be used for market-consistent forecasting. The Lagrangian of the CEL problem (46) can be shown to equal

$$\mathcal{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (\hat{P}_h(Z_t, X_t, \theta) - P_t)^T \hat{V}^{-1}(Z_t, X_t; \theta)(\hat{P}_h(Z_t, X_t, \theta) - P_t), \tag{48}$$

where

$$\hat{V}(Z_t, X_t, \theta) = \hat{E}\left(g(R_{t+1}, Z_t, X_t, \theta) g(R_{t+1}, Z_t, X_t, \theta)^T \mid Z_t, X_t\right) \tag{49}$$
$$-\hat{E}\left(g(R_{t+1}, Z_t, X_t, \theta) \mid Z_t, X_t\right) \hat{E}\left(g(R_{t+1}, Z_t, X_t, \theta)^T \mid Z_t, X_t\right),$$

where $\hat{E}$ takes expectations with respect to the density $\hat{f}_h(R_{t+1}|Z_t)$. The population version of (48) is not exactly the same as (23), as the resulting difference term (24) now depends on the parameter θ due to the continuous updating. This implies that this CEL problem potentially identifies a different θ_0. This formulation essentially treats the price vector P_t as known given (Z_t, X_t), which is justified in the case of an observed Markovian state vector as in Gagliardini et al. (2011), but otherwise ignores the option pricing error ε_t. This is also reflected in the continuously updated variance matrix (49), which estimates the variance of the discounted next period's payoff but neglects the variance of the current pricing errors.[8]

The different identified parameters for the two CEL problems can be intuitively explained as follows. Suppose two dates t and s in the sample have the same conditioning variable $Z_s = Z_t$, but different option prices. Problem (46) attributes the pricing error to the implied densities, while (39) attributes this to a pricing error as the implied probabilities $p_{t\cdot}$ and $p_{s\cdot}$ are required to be the same. Hence the first problem interprets high price for out-of-the money puts on day t as

[7] The idea of tilting a distribution for option pricing goes back to at least Stutzer (1996), who minimizes a criterion based on the difference between the risk-neutral and historical distribution, rather than between distributions with and without option price information.

[8] Gagliardini et al. (2011) focus on using the latest cross-section of option prices and the historical time series of stock prices. The estimator described in this paper also uses the time series of option prices to identify the dynamics of the projected pricing kernel.

a high likelihood of a large negative shock happening, while the second problem merely attributes this to pricing error due to restricting the pricing kernels to be the same.

3.2 Density averaging

In fact, it is possible to interpolate between the two CEL formulations in Section 2.5. The purpose of doing so is to balance the dynamic misspecification of the pricing kernel and the physical density due to restricting the conditioning set. Given any set of observed conditioning variables Z_t, we can decompose

$$\log \frac{f_t^{\mathbb{Q}}(y)}{f^{\mathbb{Q}}(y|Z_t)} = \log \frac{f_t^{\mathbb{P}}(y)}{f^{\mathbb{P}}(y|Z_t)} + \log \frac{m_t(y)}{m(y|Z_t)},$$

where the subscript t indicates conditional on all information available at time t. A class of approximating models for the conditional physical density is

$$\log \frac{f_t^{m}(y)}{f^{\mathbb{P}}(y|Z_t)} = \pi_t(y) \log \frac{f_t^{\mathbb{Q}}(y)}{f^{\mathbb{Q}}(y|Z_t)}, \tag{50}$$

where $\pi_t(y)$ must be such that f_t^m is a proper density. Approximating the logarithms around one and rearranging yields the density averaging estimator

$$f_t^m(y) = \pi_t(y)\frac{f_t^{\mathbb{Q}}(y)}{m(y|Z_t)} + (1-\pi_t(y))f^{\mathbb{P}}(y|Z_t).$$

The first term $\frac{f_t^{\mathbb{Q}}(y)}{m(y|Z_t)}$ is a market-implied density corrected for its bias in states with similar Z_t. The second density $f^{\mathbb{P}}(y|Z_t)$ is unbiased given Z_t, but is subject to missing conditioning variables.

The parameters π and m then jointly solve the constrained maximum likelihood population problem:

$$(\pi_0, m_0) = \arg\max_{\pi,m} E\left(\log\left(\pi_t(R)\frac{f_t^{\mathbb{Q}}(R)}{m(R|Z_t)} + (1-\pi_t(R))f^{\mathbb{P}}(R|Z_t)\right)\right)$$

s.t.

$$E\left(m(R_{t+1}, Z_t)h(R_{t+1},\kappa_t) - P_t(\kappa_t)\middle|Z_t,\kappa_t\right) = 0$$

The pricing constraint assures that m_0 identifies the projected pricing kernel. The unconstrained maximum likelihood problem only identifies the projected pricing kernel when the conditioning information is correctly specified. To see this, suppose $\pi = 1$, and let $Z_t = \emptyset$. Then optimizing the unconstrained maximum likelihood over the function m yields

$$m^* = \arg\min_{m\in L^2} E\left(\log\left(\frac{f^{\mathbb{Q}}(R|S)/m(R)}{\int f^{\mathbb{Q}}(y|S)/m(y)\mathrm{d}y}\right)\right),$$

which by calculus of variation is characterized implicitly by

$$m^*(R)f(R) = \int \frac{f^{\mathbb{Q}}(R|s)}{\int \frac{f^{\mathbb{Q}}(y|s)}{m^*(y)}\mathrm{d}y} f_S(s)\mathrm{d}s.$$

When the conditioning variables are correctly specified, $S = \emptyset$, and by the law of iterated expectation the RHS implies that $m^*(R)$ equals the projected pricing kernel. In the general case with non-trivial missing conditioning variables $S \neq \emptyset$, the pricing kernel $m(R) = \frac{f^{\mathbb{Q}}(R)}{f^{\mathbb{P}}(R)}$ does not normally satisfy this equation as the denominator in the RHS distorts the expectation over S.

The 'informativeness' parameter π_0 is identified from the likelihood provided the averaging densities are not identical, which is the case with missing conditioning variables.

3.3 CEL with partially smoothed option prices

The following two-stage representation of the population problem averages the conditional moments instead of the densities:

$$(\pi_0, m_0) = \arg\max_{\pi, m} \; E\left(\log f_t(y; \pi, m)\right)$$

where

$$\begin{aligned} f_t(y; \pi, m) = \arg\max_{p_t(y)} \; & E\left(\left(\frac{p_t(y) - f_h(y|Z_t)}{f_h(y|Z_t)}\right)^2 \big| Z_t\right) \\ \text{s.t.} & \int m(y, Z_t)h(y, \kappa_t)p_t(y)\mathrm{d}y = \pi P_t(\kappa_t) + (1-\pi)E(P_t(\kappa_t)\big|Z_t) \\ & \int p_t(y)dy = 1. \end{aligned}$$

The sample analogue of this problem is a natural generalization of the two CEL problems introduced in Section 2.5. Suppose we have a balanced panel $(R^e_{t+1}, (C_{it})_i, Z_t)$ consisting of excess returns, n option prices, and conditioning variables. For a finite set of strike prices there is an infinite number of densities that solve the pricing constraint, so we need to specify a divergence measure to choose among them. The quadratic loss function leads to the problem

$$\min_{(p_t(\cdot))} \; \frac{1}{T}\sum_{t=1}^{T} \int \left(\frac{p_t(y) - \hat{f}_h(y|Z_t)}{\hat{f}_h(y|Z_t)}\right)^2 \hat{f}_h(y|Z_t)dy \tag{51}$$

$$\text{s.t.} \int m(y, Z_t)h(y, \kappa_t)p_t(y)dy = \hat{P}_t(\pi) \tag{52}$$

$$\int p_t(y)dy = 1, \qquad t = 1, \cdots, T$$

where

$$\hat{P}_t(\pi) = \pi P_t + (1-\pi)\hat{E}_h\left(P_t|Z_t\right)$$
$$\hat{f}_h(y|Z_t) = \sum_{s=1}^{T-1} w_{ts} K_{h_y}(R_{s+1} - y).$$

Let $g_t(y) = m(y, Z_t)h(y, \kappa_t)$ and $\hat{g}_t = \hat{E}_h(g_t(R_{t+1})|Z_t)$. The solution is given by

$$p_t(y) = \hat{f}_h(y|Z_t)(1 - \lambda_t^T(g_t(y) - \hat{g}_t)), \tag{53}$$

in terms of the Lagrange multipliers

$$\lambda_t = \widehat{Var}_h(g_t(R_{t+1})|Z_t)^{-1}\left(\hat{g}_t - \hat{P}_t(\pi)\right). \tag{54}$$

This leads to the continuously updated local GMM criterion

$$Q_T(m;\pi) = \frac{1}{T}\sum_{t=1}^{T} \hat{P}_t(\pi)^T \widehat{Var}_h(g_t(R_{t+1})|Z_t)^{-1} \hat{P}_t(\pi) \tag{55}$$

The criterion (55) punishes the terms $\hat{P}_h(Z_t, X_t, \theta) - \pi P_t - (1-\pi)\hat{E}_h(P_t|Z_t, X_t)$. Similar linear combinations of the current prices and their historical projections on Z_t can be derived for the variance matrix and the implied probabilities. The case $\pi = 0$ corresponds to the option prices containing no additional conditioning information beyond the variables Z_t. However, due to the overidentification of sieve parameters of the pricing kernel, they help to reduce the variability of the density estimator by incorporating the pricing restrictions. The higher $\pi \in [0,1]$, the more option prices are smoothed, and the less the implied densities will react to match them. Hence π defines a class of density forecasts with varying sensitivity to option prices. The optimal value of π can be determined empirically using the predictive likelihood.

Another common distance measure is the relative entropy

$$\min_{p_t} \frac{1}{T}\sum_{t=1}^{T}\int p_t(y)\log\frac{\hat{f}_h(y|Z_t)}{p_t(y)}dy \qquad s.t. \;\; (52)$$

which has solution

$$p_t(y) = \hat{f}_h(y|Z_t)\exp(\lambda_t^T u_t(y) + \mu_t).$$

While convenient for computing the log-likelihood, the multipliers are not analytically available which makes the entropy criterion less computationally attractive than the least squares criterion.

The pricing kernel m and tuning parameter π can be chosen in a two-step approach. First, for given π, minimize $Q_T(m;\pi)$ over the sieve space to find $\hat{m}$ and compute the implied densities.

Second, maximize the log-likelihood $\mathcal{L}_T(\pi) = \frac{1}{T}\sum_{t=1}^{T} \log p_t(R;\pi,\hat{m})$ over π. When $n \to \infty$, $p_t(R;\pi,\hat{m}) \xrightarrow{p} f_t^m(y;\pi)$. This two-step approach does not require an initial estimate of $f_t^{\mathbb{Q}}$ from the cross-section of option prices, nor computing ratios of estimated densities.

Alternatively, we can estimate $\pi(\kappa)$ as a function of the option moneyness using the local regression of $g_t(R_{t+1,\kappa})$ on $C_t(\kappa)$. By de-centering both payoff and prices by their conditional expectation given Z_t, the estimated regression coefficients can be interpreted as the informative fraction of unexplained variation in option prices. Any heterogeneity in $\pi(\kappa)$ reflects that option traders' predictive ability may differ among regions of the return distribution, or that the size of risk aversion or demand shocks varies with moneyness levels. Figure 2 shows the estimated informativeness coefficients $\hat{\pi}(\kappa)$ as a function of moneyness, for the case of a flat pricing kernel $(K = 1)$ and for the sixth order polynomial from Figure 1. The graph reveals that put option prices tend to be less informative about their own payoffs than call option prices. This could be explained by large time-varying preferences for insurance against left-tail market risks. Moreover, at-the-money options are more informative than out-of-the money options, suggesting informed option traders tend to prefer the former, whereas uninformed or 'gambling' traders tend to prefer the latter. Finally, the effect of state-dependent discounting payoffs increases the informativeness of option prices at all moneyness levels. This is in line with a rational risk averse investor being more sensitive to information about returns in high utility states.

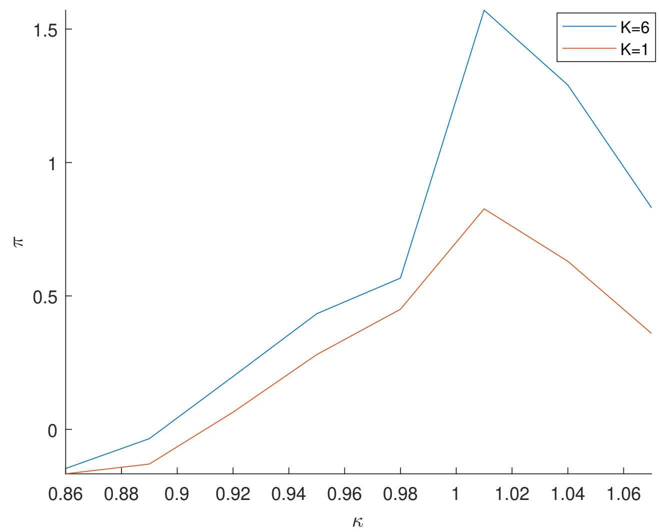

Figure 2: Estimated local informativeness parameter for varying moneyness, for a flat pricing kernel $(K = 1)$ and a sixth order bivariate polynomial.

3.4 Empirical implied predictive densities

Figure 3 shows the implied conditional density estimates for four week excess returns on the S&P 500 index, with the VIX as the conditioning variable, based on the optimization problem (51) with full option price smoothing $\pi = 0$. The implied densities are well-behaved, and feature a left skewness and relatively fat left tail. The scale of the densities is larger when the VIX increases, suggesting the VIX at least partially predicts future uncertainty rather than only the steepness of the pricing kernel. Figure 4 shows the estimated Lagrange multipliers (54) for fitting the option

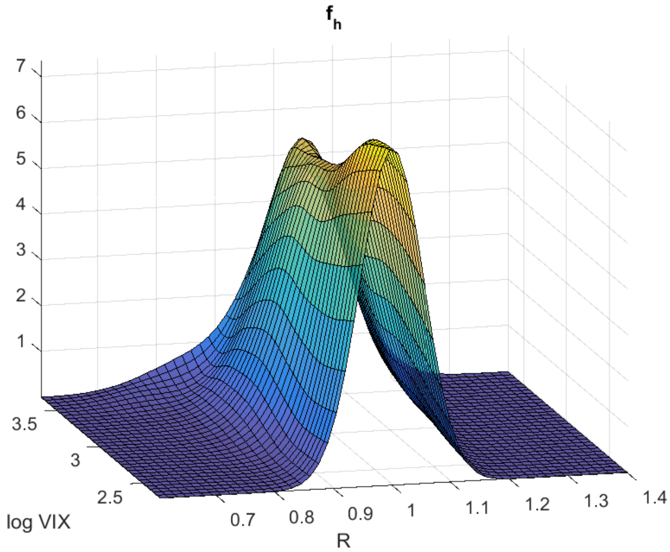

Figure 3: Implied conditional density estimates for four week excess returns on the S&P 500 index given $Z_t = \log VIX_t$, based on $\pi = 0$ and a $K = 5$ bivariate expansion of the projected pricing kernel.

price constraints in (51) using the $K = 5$ bivariate pricing kernel expansion. The largest values of the multipliers are in the left tail of the return distribution. This suggest that for the empirical pricing kernel the main challenge is to rationalize the prices of far out-of-the-money put option with the historical distribution of the returns. This problem is largest when the VIX is either well above or below its mean.

Finally, Figure 5 shows the implied predictive densities that, given the fitted pricing kernel, match the informative part of option prices based on the estimated informativeness function $\hat{\pi}(\kappa)$ from Figure 2. The predictive densities reflect the heterogeneity in informativeness across moneyness. In particular, the left tail is relatively stable, while the peaks of the density vary more considerably over time. The densities can be seen to have the largest peaks in the pre-crisis years 2004-2007, reflecting the low market volatility during this period, and vice versa for the financial crisis years 2008-2010.

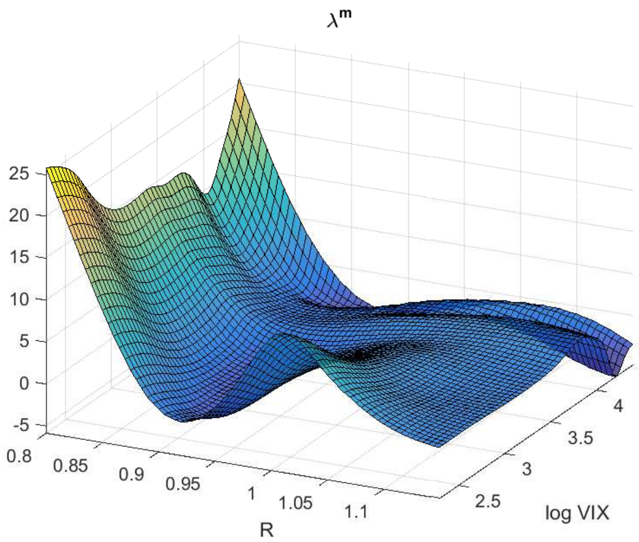

Figure 4: Estimated Lagrange multipliers for fitting the option price constraints using a bivariate pricing kernel expansion with $K = 5$.

3.5 Option-Implied Expected Returns

The projected pricing kernel also allows computing time-varying expected excess return on the market, while using information from option prices. The pricing restrictions for the options help to identify the potentially nonlinear shape of the projected pricing kernel, which in turn can be used to compute expected returns from the pricing restriction on the underlying asset. In particular, the stock pricing equation (5) minus the bond pricing equation (6) yields

$$
\begin{aligned}
&0 = E\left(e^{-r_t^f} m(R_{t+1}, Z_t, \theta)(R_{t+1} - R_t^f) \mid Z_t, r_t^f\right) \\
\Leftrightarrow\ \ & E\left(R_{t+1} - R_t^f \mid Z_t, r_t^f\right) = -e^{-r_t^f} \operatorname{Cov}(m(R_{t+1}, Z_t, \theta), R_{t+1} \mid Z_t).
\end{aligned}
$$

The time-variation in the risk premium comes from both time-varying risk aversion as well as the time-varying distribution. A plug-in estimate of the risk premia can be computed using the estimated pricing kernel and the estimated conditional densities from (53). The pricing kernel is generally decreasing with the market return, which leads to positive risk premia.

Figure 6 shows the realized returns and implied expected returns from the CEL problem (51) with $\pi = 1$, conditional on the VIX. The graph also shows a simple kernel regression estimator of the conditional mean return, which appears wiggly for high levels of the VIX, as there are few data points in this area. However, the CEL implied conditional means are more stable, as they use knowledge of the estimated pricing kernel from option prices to extrapolate through regions with few observations. Importantly, when the option pricing restrictions are imposed, it is found that high levels of the VIX positively predict returns, whereas the unrestricted kernel estimate

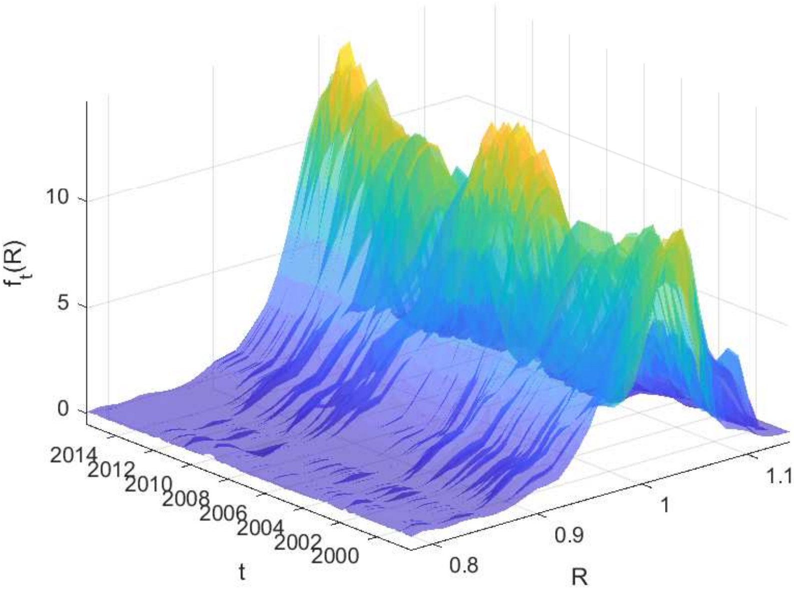

Figure 5: Implied conditional density estimates for four week excess returns on the S&P 500 index given $Z_t = \log VIX_t$, based on the estimated informativeness function $\hat{\pi}(\kappa)$ and a $(K_R, K_Z) = (6, 2)$ bivariate expansion of the projected pricing kernel.

surprisingly predicts a lower return in this case.

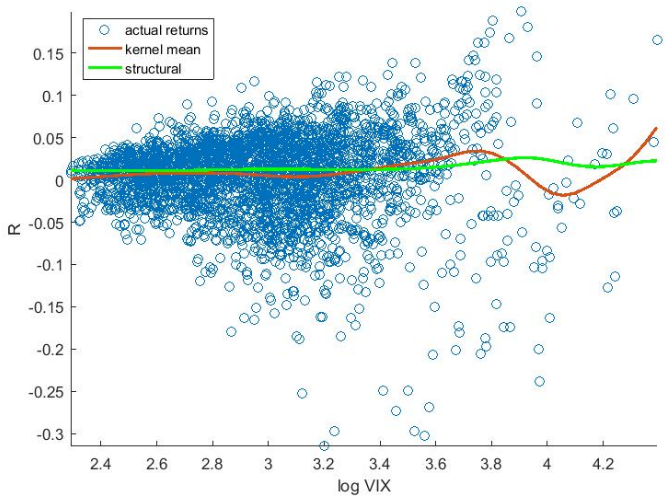

Figure 6: Actual returns and implied expected returns by fitting a bivariate pricing kernel expansion with $K = 5 * 5$ on S&P 500 options, returns, and VIX data, from January 1996 until June 2016.

4 Conclusion

This paper studies the estimation of nonlinear dynamic pricing kernels from option prices, and proposes a method to incorporate this information for density forecasting and computing expected returns. It studies the nonparametric identification and estimation of projected pricing kernels from conditional moment restrictions that come from the pricing on options, the underlying asset, and a risk-free bond. It avoids the need to compute ratios of estimated risk-neutral and physical densities, leading to stable estimates even in regions with few probability mass. Moreover, the paper proposes a conditional empirical likelihood framework to extract implied densities that make use of the pricing restrictions and the forward-looking information in option prices, and proposes a implied density combination to measure the relative importance of changes in the physical return distribution and changes in the pricing kernel. The nonlinear dynamic pricing kernels can be used to understand return predictability, and provide model-free quantities that can be compared against those implied by structural asset pricing models.

A Appendix

A.1 Proofs

Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). This requires checking that (i) there is unique θ_0 that minimizes $Q'(\theta)$ on Θ, (ii) $Q'_T(\theta)$ and $Q'(\theta)$ are continuous, $Q'(\theta)$ is compact, and $\max_{\theta\in\Theta}|Q'_T(\theta) - Q'(\theta)| \xrightarrow{p} 0$, (iii) Θ_T are compact subsets of Θ such that for any $\theta \in \Theta$ there exists a $\tilde{\theta}_T \in \Theta_T$ such that $\tilde{\theta}_T \xrightarrow{p} \theta$.

The identification condition (i) follows from Section 2.2, together with the positive definiteness of Σ_t.

The compact subset condition in (iii) hold by construction of Θ_T and Θ. Moreover for any $\theta \in \Theta$ we can find a series approximator $\theta_T \in \Theta_T$ that satisfies $\|\theta_T - \theta\| \to 0$.

For (ii), continuity of $Q'_T(\theta)$ follows from continuity of the exponential function. The remaining conditions of continuity of $Q'(\theta)$ and uniform convergence can be proved using Lemma A2 in Newey and Powell (2003). This requires pointwise convergence $Q'_T(\theta) - Q'(\theta) \xrightarrow{p} 0$ as well as the stochastic equicontinuity condition that there is a $v > 0$ and $B_n = O_p(1)$ such that for all $\theta, \tilde{\theta} \in \Theta$, $\|Q'_T(\theta) - Q'_T(\tilde{\theta})\| \leq B_n\|\theta - \tilde{\theta}\|^v$. For pointwise convergence, denote $\epsilon_t(\theta) = P(Z_t, X_t, \theta) - P_t$ and $\hat{\epsilon}_t(\theta) = \hat{P}(Z_t, X_t, \theta) - P_t$, and write

$$Q'_T(\theta) - Q'(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right)$$
$$+ \frac{1}{T}\sum_{t=1}^{T}\epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta) - E\left(\epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right).$$

The first term converges to zero in probability under consistent first stage nonparametric estimation, while the second term does so by the weak law of large numbers. The elements of the summation in the first term can be decomposed as

$$\begin{aligned}\hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta) &= \left(\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\right)^T \hat{\Sigma}_t\hat{\epsilon}_t(\theta) \\ &\quad + \epsilon_t(\theta)^T\left(\hat{\Sigma}_t - \Sigma_t\right)\hat{\epsilon}_t(\theta) \\ &\quad + \epsilon_t(\theta)^T\Sigma_t\left(\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\right)\end{aligned}$$

Since $\hat{\Sigma}_t \xrightarrow{p} \Sigma_t$ with Σ_t bounded,

$$\begin{aligned}E\left(\|\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\|_2^2\right) &= E\left(\|\hat{P}(Z_t, X_t, \theta) - P(Z_t, X_t, \theta)\|_2^2\right) \\ &= E\left(\left\| \int_0^\infty e^{-r_t^f} m(y, Z_t, \theta)h(y, X_t)\left(\hat{f}_h(y|Z_t) - f(y|Z_t)\right) dy\right\|_2^2\right) \\ &\to 0\end{aligned}$$

by consistency of $\hat{f}_h(y|z)$, and

$$\begin{aligned}E\left(\|\epsilon_t(\theta)\|_2^2\right) &\le E\left(\|\epsilon_t(\theta) - \epsilon_t(\theta_0)\|_2^2\right) + E\left(\|\epsilon_t(\theta_0)\|_2^2\right) \\ &= E\left(\|P(Z_t, X_t, \theta) - P(Z_t, X_t, \theta_0)\|_2^2\right) + E\left(\|\epsilon_t\|_2^2\right) \\ &\le C_1\|\theta - \theta_0)\|^{2v} + C_2 = O(1),\end{aligned}$$

using compactness of Θ in the last step, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\epsilon}_t(\theta)^T\hat{\Sigma}_t\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)^T\Sigma_t\epsilon_t(\theta)\right) \xrightarrow{p} 0.$$

We prove the stochastic equicontinuity condition for the case $\Sigma_t = I$, which can be generalized to other p.d. matrices. Let $\epsilon(\theta) = (\epsilon_1(\theta), ..., \epsilon_T(\theta))^T$ and similarly define $\hat{\epsilon}(\theta)$. For any $\theta, \tilde{\theta} \in \Theta$,

$$\begin{aligned}|Q'_T(\theta) - Q'_T(\tilde{\theta})| &= \frac{1}{T}|\hat{\epsilon}(\theta)^T\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})^T\hat{\epsilon}(\tilde{\theta})| \\ &\le \frac{1}{T}\|\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})\|_2^2 + \frac{2}{T}\|\hat{\epsilon}(\theta)\|_2\|\hat{\epsilon}(\theta) - \hat{\epsilon}(\tilde{\theta})\|_2,\end{aligned}$$

where

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,X_t,\theta)-\hat{P}(Z_t,X_t,\tilde{\theta})\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\int_0^\infty e^{-r_t^f}\left(m(y,Z_t,\theta)-m(y,Z_t,\tilde{\theta})\right)h(y,x)\hat{f}_h(y|Z_t)dy\right)^2 \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\int_0^\infty b(y,Z_t)^2h(y,x)^2\hat{f}_h(y|Z_t)dy\|\theta-\tilde{\theta}\|^{2v} \\
&\equiv B_T\|\theta-\tilde{\theta}\|^{2v} \\
&\leq B_T C\|\theta-\tilde{\theta}\|^{v},
\end{aligned}$$

and where the last step follows by compactness of Θ. Also

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta)\|_2\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2 &\leq \frac{1}{T}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\theta_0)\|_2\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2+\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta_0)\|_2\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta)-\hat{\epsilon}(\tilde{\theta})\|_2 \\
&\leq \left(B_TC+\frac{1}{\sqrt{T}}\|\hat{\epsilon}(\theta_0)\|_2\sqrt{B_T}\right)\|\theta-\tilde{\theta}\| \\
&\equiv \tilde{B}_T\|\theta-\tilde{\theta}\|.
\end{aligned}$$

Here

$$\begin{aligned}
\frac{1}{T}\|\hat{\epsilon}(\theta_0)\|_2^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,X_t,\theta_0)-P_t\right)^2 \\
&\leq \frac{2}{T}\sum_{t=1}^{T}\left(\hat{P}(Z_t,X_t,\theta_0)-P(Z_t,X_t,\theta_0)\right)^2+\frac{2}{T}\sum_{t=1}^{T}\epsilon_t^2 \\
&= O_p(1),
\end{aligned}$$

by consistency of $\hat{f}_h(y|z)$, continuity and boundedness of $f(y,z)$, and $E(\epsilon_t^2)<\infty$. Since $h^C(R_{t+1},X_t)=(R_{t+1}-\kappa_t)^+<R_{t+1}$,

$$E\left(b(R_{t+1},Z_t)^2h(R_{t+1},X_t)^2|Z_t,X_t\right)\leq E\left(b(R_{t+1},Z_t)^2R_{t+1}^2|Z_t\right)<\infty$$

Similarly it follows that $B_T=O_p(1)$, so that $\tilde{B}_T=O_p(1)$, which proves the stochastic equicontinuity condition with $v=1$. This implies $Q'(\theta)$ is continuous and the pointwise convergence becomes uniform over Θ. This proves the three conditions. □

Proof of Corollary 1. The identification condition A1 follows from section 2.2. It remains to check the Hölder condition A4. For any $\theta, \tilde{\theta} \in \Theta$,

$$\begin{aligned}
|m(y,z,x,\theta) - m(y,z,x,\tilde{\theta})| &= e^{-rf} |e^{-\beta_0(z)-\beta_1(z)\ln y}\alpha(y,z) - e^{-\tilde{\beta}_0(z)-\tilde{\beta}_1(z)\ln y}\tilde{\alpha}(y,z)| \\
&\leq e^{-rf} \left|e^{-\beta_0(z)-\tilde{\beta}_1(z)\ln y} + e^{-\beta_0(z)}\alpha(z) + e^{-\tilde{\beta}_1(z)\ln y}\tilde{\alpha}(z)\right| \\
&\left(\left|e^{-\beta_0(z)} - e^{-\tilde{\beta}_0(z)}\right| - \left|e^{\beta_1(z)\ln y} - e^{-\tilde{\beta}_1(z)\ln y}\right| - \left|\alpha(y,z) - \tilde{\alpha}(y,z)\right|\right) \\
&\leq e^{-rf} \left|e^{-\beta_0(z)-\tilde{\beta}_1(z)\ln y} + e^{-\beta_0(z)}\alpha(z) + e^{-\tilde{\beta}_1(z)\ln y}\tilde{\alpha}(z)\right| \\
&\left(1 + e^{-\check{\beta}_0(z)} + |\ln y| e^{-\check{\beta}_1(z)\ln y}\right) \left(\left|\beta_0(z) - \tilde{\beta}_0(z)\right| + \left|\beta_1(z) - \tilde{\beta}_1(z)\right| + \left|\alpha(y,z) - \tilde{\alpha}(y\right.\right. \\
&\leq C e^{-rf - 2B_1 \ln y} \left(\|\beta_0 - \tilde{\beta}_0\|_{m,\infty} + \|\beta_1 - \tilde{\beta}_1\|_{m,\infty} + \|\alpha - \tilde{\alpha}\|_{m,\infty}\right) \\
&= C|\ln y| e^{-rf - 2B_1 \ln y} \|\theta - \tilde{\theta}\|,
\end{aligned}$$

where the last inequality uses boundedness of the functions in θ and $\tilde{\theta}$. Since

$$E\left(|\ln R_{t+1}|^2 e^{-4(2B_1 - 2r_{t+1}}|Z_t\right) \leq E\left(|\ln R_{t+1}|^2|Z_t\right)^{\frac{1}{2}} E\left(e^{-4(2B_1-1)r_{t+1}}|Z_t\right)^{\frac{1}{2}},$$

condition A4 follows with $b(R_{t+1}, Z_t) = |r_{t+1}| e^{-2(2B_1-1)r_{t+1}}$.

□

References

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

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1-2):9–51.

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, 138(2):461–487.

Bansal, R. and Viswanathan, S. (1993). No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, 48(4):1231–1262.

Beare, B. K. and Schmidt, L. D. (2014). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

Borovička, J., Hansen, L. P., and Scheinkman, J. A. (2014). Misspecified recovery. Technical report, National Bureau of Economic Research.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.

Chapman, D. A. (1997). Approximating the asset pricing kernel. *Journal of Finance*, pages 1383–1410.

Chen, X. and Ludvigson, S. C. (2009). Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics*, 24(7):1057–1093.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565.

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., Srisuma, S., et al. (2015). Nonparametric euler equation identification and estimation. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

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

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390.

Giacomini, E. and Härdle, W. (2008). Dynamic semiparametric factor models in pricing kernels estimation. pages 181–187.

Grith, M., Härdle, W. K., and Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. pages 277–305.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Härdle, W. K., Okhrin, Y., and Wang, W. (2014). Uniform confidence bands for pricing kernels. *Journal of Financial Econometrics*.

Hördahl, P. and Vestin, D. (2005). Interpreting implied risk-neutral densities: the role of risk premia. *Review of Finance*, 9(1):97–137.

Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72(6):1667–1714.

León, Á., Mencía, J., and Sentana, E. (2012). Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business & Economic Statistics*.

Linn, M., Shive, S., and Shumway, T. (2017). Pricing kernel monotonicity and conditional information. *The Review of Financial Studies*, 31(2):493–531.

Liu, X., Shackleton, M. B., Taylor, S. J., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Ludvigson, S. C. (2011). Advances in consumption-based asset pricing: Empirical tests. Technical report, National Bureau of Economic Research.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, pages 1565–1578.

Rosenberg, J. and Engle, R. F. (2000). Empirical pricing kernels. *Journal of Financial Economics*, (99-014).

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics*, 138(2):430–460.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190(1):176–196.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *The Journal of Finance*, 51(5):1633–1652.

Vincent-Humphreys, D., Noss, J., et al. (2012). Estimating probability distributions of future asset prices: empirical transformations from option-implied risk-neutral to real-world density functions.