Empirical reverse engineering of the pricing kernel

Mikhail Chernov

Graduate School of Business, Columbia University, Rm 413, Uris Hall 3022 Broadway, New York, NY 10027-6902, USA

Abstract

This paper proposes an econometric procedure that allows the estimation of the pricing kernel without either any assumptions about the investors preferences or the use of the consumption data. We propose a model of equity price dynamics that allows for (i) simultaneous consideration of multiple stock prices, (ii) analytical formulas for derivatives such as futures, options and bonds, and (iii) a realistic description of all of these assets. The analytical specification of the model allows us to infer the dynamics of the pricing kernel. The model, calibrated to a comprehensive dataset including the S&P 500 index, individual equities, T-bills and gold futures, yields the conditional filter of the unobservable pricing kernel. As a result we obtain the estimate of the kernel that is positive almost surely (i.e. precludes arbitrage), consistent with the equity risk premium, the risk-free discounting, and with the observed asset prices by construction. The pricing kernel estimate involves a highly nonlinear function of the contemporaneous and lagged returns on the S&P 500 index. This contradicts typical implementations of CAPM that use a linear function of the market proxy return as the pricing kernel. Hence, the S&P 500 index does not have to coincide with the market portfolio if it is used in conjunction with nonlinear asset pricing models. We also find that our best estimate of the pricing kernel is not consistent with the standard time-separable utilities, but potentially could be cast into the stochastic habit formation framework of Campbell and Cochrane (J. Political Economy 107 (1999) 205).

Published in: ***Journal of Econometrics***

1. Introduction

Estimating the pricing kernel (PK), or stochastic discount factor, is of paramount importance to financial economics. Beginning with the seminal work of Hansen and Singleton (1982), scores of papers have refined and tested this link between consumption data and asset prices.

This paper proposes an alternative methodology for estimating the PK directly from observed asset prices that avoids potentially noisy and infrequently observed consumption. Our approach is based on no-arbitrage, and it differs from previous work, which focuses on the state-price density (SPD), because it uses multiple assets, therefore, allowing the recovery of the PK itself.[1]

Following Garman (1976), we specify a continuous-time parametric model of asset prices. The first stage of the methodology involves estimating the parameters using fixed income, commodities futures, and equity data. The second stage assumes that the PK is unobservable and backs it out from the estimated model given different combinations of the assets used at the first stage (reverse engineering).

Unlike the equilibrium-based approach, our approach ensures that the PK is inherently consistent—risk premia, and interest rate puzzles are absent. All the problems can be explained by the model misspecification that can be identified via econometric tests. The disadvantage is that, like all other no-arbitrage strategies, it lacks economic foundation. However, since daily consumption data are not available, an equilibrium-based theory would not be testable. The best we can hope for is to see whether the inferred PK can be potentially conformable to the ones designed for monthly cross-sections.

There are a number of extant methods for estimating the SPD that avoid the use of the consumption data (Aït-Sahalia and Lo, 1998; Jackwerth and Rubinstein, 1996; Rosenberg and Engle, 2002, among others). Though different methods are proposed, the common thread to this work is the selection of the same set of assets to be priced by the kernel, namely the S&P 500 index and options on the index. These methods are designed to handle single assets because they rely on the estimation of SPD, and generalizations to multiple assets are not obvious.

Moreover, the estimation procedures all involve nonparametric methods to various degrees. Nonparametric methods are capable of accurate representing the data, but valid inference using the asymptotic properties of the estimators requires more data than is typically available for this type of empirical study. While this approach may still be suitable for the unconditional estimation of the PK, it is hardly appropriate for estimating its dynamic conditional behavior.

The estimation of the multiple-asset continuous time models has so far received limited attention because empirically plausible specifications involve at least two factors, including latent ones. This complicates the empirical implementation considerably and was until recently not practically feasible. The examples of some early work on the subject are Longstaff (1989), who studied the effects of time aggregation on the empirical implications of the CAPM, and Ho et al. (1996) who estimated a continuous time model involving eight asset returns.

[1] SPD is equal to the PK multiplied by the probability density of an individual asset returns distribution.

The choice of assets used for estimating the PK at the second stage of the procedure is critical for pricing out of sample. To appraise the performance of the kernel estimated based on different combinations of assets, we use an approach introduced in Hansen and Jagannathan (1997).

Using multiple assets taken simultaneously sharpen the estimates of the common parameters pertaining to the dynamics of the PK. We find that the best PK can be described as a complicated non-linear function that depends on the contemporaneous and lagged returns of the S&P 500 index.[2] Since this function nests the CAPM model, we conclude that CAPM can be interpreted as the first order approximation to the true PK. Alternatively, in the light of the Roll (1977) critique, this non-linear function of the observable S&P 500 returns approximates a linear function of the unobservable market portfolio. Our best estimate of the PK is not consistent with a class of time-separable utilities; however, it can potentially be explained in the framework of stochastic habit formation model of Campbell and Cochrane (1999).

The paper is organized as follows. Section 2 presents and motivates the model we consider and derives the dynamics of the PK. We also briefly discuss the literature related to our approach. Section 3 describes the data, outlines the estimation strategy and results, while Section 4 evaluates the obtained estimates of the PK. The last section concludes and technical material is provided in several appendices.

2. The model

2.1. Specification

The typical asset pricing setup involves the specification of a model of asset prices under the objective probability measure and a model of the PK. These two components completely determine the pricing framework and a mapping to the risk-neutral measure via the prices of risk in particular. We, however, want to emphasize that we infer the law of the PK dynamics from the asset prices. Therefore, we start the discussion of our theoretical framework with the specification of the assets behavior under the objective and risk-neutral measures. This approach allows us to determine the PK.

We adopt the following system of stochastic differential equations (SDE) to specify the dynamics of the asset i's price $S^i(t)$:

$$\frac{\mathrm{d}S^i(t)}{S^i(t)} = (r_0 + \alpha^i U_2(t))\,\mathrm{d}t + \beta^i\sqrt{1-\rho^2}\sqrt{U_2(t)}\,\mathrm{d}W_1(t) + \beta^i\rho\sqrt{U_2(t)}\,\mathrm{d}W_2(t) + \gamma^i\sqrt{V^i(t)}\,\mathrm{d}Z^i(t), \tag{1}$$

$$\mathrm{d}U_2(t) = (\theta - \kappa U_2(t))\,\mathrm{d}t + \sqrt{U_2(t)}\,\mathrm{d}W_2(t), \tag{2}$$

$$\mathrm{d}V^i(t) = (\eta^i - \iota^i V^i(t))\,\mathrm{d}t + \sqrt{V^i(t)}\,\mathrm{d}W^i(t). \tag{3}$$

Eq. (1) implies that the stock-price process $S^i(t)$ follows a geometric Brownian motion with the drift $r_0+\alpha^i U_2(t)$ and a two-component stochastic variance ${\beta^i}^2 U_2(t)+{\gamma^i}^2 V^i(t)$.

[2] These findings parallel the work of Bansal and Viswanathan (1993), who directly specify the pricing kernel as a non-linear function of the contemporaneous monthly market returns.

The factor $U_2(t)$ is common to all assets and determines the dynamics of the drift (predictability property) as well as the variance. Eq. (2) states that U_2 follows a square-root mean-reverting process with the long-run mean θ/κ and the speed of adjustment κ. $V^i(t)$, the component which determines the asset-specific variance, follows a similar process with the dynamics described by $\eta^i/\imath^i$ and $\imath^i$, respectively. Finally, instead of considering the process for S^i, we will study $U_1^i = \log S^i$. Using Itô's lemma we replace (1) with:

$$\mathrm{d}U_1^i(t) = [r_0 + \alpha^i U_2(t) - \tfrac{1}{2}\,{\beta^i}^2 U_2(t) - \tfrac{1}{2}\,{\gamma^i}^2 V^i(t)]\,\mathrm{d}t + \beta^i\sqrt{1-\rho^2}\sqrt{U_2(t)}\,\mathrm{d}W_1(t)$$
$$+ \beta^i\rho\sqrt{U_2(t)}\,\mathrm{d}W_2(t) + \gamma^i\sqrt{V^i(t)}\,\mathrm{d}Z^i(t). \quad (4)$$

This specification modifies the Heston (1993) model in two directions. First, we have a stochastic drift in the fundamental process (1) rather than a constant to allow for autocorrelation in asset returns. Second, we add an idiosyncratic component $\sqrt{V^i(t)}\,\mathrm{d}Z^i(t)$, a variation in return attributable to specific asset characteristics. In other words, we assume that $W_1(t)$ and $W_2(t)$ represent the systematic shocks which do not span the whole security space, i.e. we have an incomplete market setup. We assume all the Wiener processes to be independent and we model the leverage effect between the systematic factors via ρ.

In order to characterize the PK we need to make assumptions about the distribution of the process under the risk-neutral measure. Since we are in an incomplete market setting, the equivalent martingale measure is not unique. However, there are considerations which help us identify the risk-neutral measure specification. In particular, since there is no risk-premium, the drift in (1) becomes equal to the risk-free interest rate that we model as $r_0 + r_1 U_2$.[3,4] Furthermore, we want the risk-neutral version of $\sqrt{V^i(t)}\,\mathrm{d}Z^i(t)$ to remain a martingale in order to preserve our structure. This specification implies that the risk premium on this term is equal to zero and, therefore, we can view it as an idiosyncratic noise.[5] Given these considerations, we assume that the market prices of risk is $\lambda_j(t) \equiv \lambda_j\sqrt{U_2(t)}$ for $W_j(t)$ and zero for $Z^i(t)$ and $W^i(t)$. It means that, according to the Girsanov theorem, the systematic factors W_j^* under the risk-neutral measure P^* are related to the actual systematic factors as follows:

$$W_j^*(t) = W_j(t) + \int_0^t \lambda_j(s)\,\mathrm{d}s, \quad j = 1, 2. \quad (5)$$

Therefore, the stock-return dynamics under the risk-neutral measure P^* evolve according to

$$\mathrm{d}U_1^i(t) = [r_0 + r_1 U_2(t) - \tfrac{1}{2}\,{\beta^i}^2 U_2(t) - \tfrac{1}{2}\,{\gamma^i}^2 V^i(t)]\,\mathrm{d}t$$

---

[3] This risk-free rate specification is along the lines of the translated CIR model considered in Pearson and Sun (1994). We will comment on the realism of the model in Section 4.3.2.

[4] This is an important difference with respect to the Heston model. Our specification with state varying drift implies that if $U_2(t) = 0$, i.e. we have no uncertainty, the required rate of return becomes equal to the risk-free interest rate that becomes equal to r_0. This is impossible in the constant drift specification in Heston (1993), since there are arbitrage opportunities in this case (see Chernov and Ghysels, 2000 for further details).

[5] This follows from the Kunita–Watanabe formula. Details are available in Chernov (2000).

$$
+\beta^i\sqrt{1-\rho^2}\sqrt{U_2(t)}\,\mathrm{d}W_1^*(t)+\beta^i\rho\sqrt{U_2(t)}\,\mathrm{d}W_2^*(t)
+\gamma^i\sqrt{V^i(t)}\,\mathrm{d}Z^i(t), \tag{6}
$$

$$
\mathrm{d}U_2(t)=(\theta-\kappa^*U_2(t))\,\mathrm{d}t+\sqrt{U_2(t)}\,\mathrm{d}W_2^*(t), \tag{7}
$$

$$
\mathrm{d}V^i(t)=(\eta^i-\iota^iV^i(t))\,\mathrm{d}t+\sqrt{V^i(t)}\,\mathrm{d}W^i(t), \tag{8}
$$

where we used the following notations:

$$
r_1=\alpha^i-\lambda_1\beta^i\sqrt{1-\rho^2}-\lambda_2\beta^i\rho, \tag{9}
$$

$$
\kappa^*=\kappa+\lambda_2. \tag{10}
$$

Now we can establish the mapping between P and P^*. The Radon–Nikodym derivative (RND) is computed as follows:

$$
\frac{\mathrm{d}P^*}{\mathrm{d}P}(t,\tau)=\exp\left\{-\frac{(\lambda_1^2+\lambda_2^2)}{2}\int_t^{t+\tau}U_2(s)\,\mathrm{d}s
-\lambda_1\int_t^{t+\tau}\sqrt{U_2(s)}\,\mathrm{d}W_1(s)-\lambda_2\int_t^{t+\tau}\sqrt{U_2(s)}\,\mathrm{d}W_2(s)\right\} \tag{11}
$$

and any asset price at time t can be computed as

$$
\pi(t)=\mathrm{E}_t\left(\mathrm{e}^{-\int_t^{t+\tau}(r_0+r_1U_2(s))\,\mathrm{d}s}\frac{\mathrm{d}P^*}{\mathrm{d}P}(t,\tau)\xi(t+\tau)\right)=\mathrm{E}_t\left(\frac{\Lambda(t+\tau)}{\Lambda(t)}\xi(t+\tau)\right) \tag{12}
$$

where the asset's payoff at time $t+\tau$ is equal to $\xi(t+\tau)$ and $\Lambda(t)$ is the PK.[6] The discrete-time models often refer to the quantity $m(t+\tau)=\Lambda(t+\tau)/\Lambda(t)$ as the PK. We will use this notation when we work in discrete time at the estimation stage. Eqs. (11) and (12) yield the dynamics of the PK

$$
\frac{\mathrm{d}\Lambda(t)}{\Lambda(t)}=-(r_0+r_1U_2(t))\,\mathrm{d}t-\lambda_1\sqrt{U_2(t)}\,\mathrm{d}W_1(t)-\lambda_2\sqrt{U_2(t)}\,\mathrm{d}W_2(t). \tag{13}
$$

We should note here that the PK is, indeed, determined by the systematic factors only—this is an assumption underlying the initial specification. $\Lambda(t)$ is, however, not the only PK possible. Any PK $\Lambda'(t)$, which satisfies

$$
\frac{\mathrm{d}\Lambda'(t)}{\Lambda'(t)}=\frac{\mathrm{d}\Lambda(t)}{\Lambda(t)}+\mathrm{d}L(t) \tag{14}
$$

where $L(t)$ is orthogonal to the space of assets payoffs, will price assets the same way as $\Lambda(t)$. Hence, we obtain the minimum variance estimate of the PK. On the other

[6] Note that under such a setup the Novikov condition is satisfied if $U_2(t)$ is bounded. The properties of the square-root process (see Chernov, 2000 for details) imply that we can bound $U_2(t)$ on any finite time interval. Hence the RND is a martingale, and we can use the Girsanov theorem about measure transformation.

hand, if we omitted a systematic factor in the specification then the kernel's variance will be too small.

To understand the PK estimation challenge better, note that, while the objective measure parameters can be estimated from the assets return series, we have to use derivatives data to estimate λ_1 and λ_2.[7] The knowledge of the relevant parameters is not enough to obtain the value of Λ at each date t. In particular we can not solve the SDE analytically and even if we could do so, it would depend on the unobservable factor U_2.

2.2. Discussion

It will be useful to discuss some of the properties of Λ before we proceed with the estimation. According to Harrison and Pliska (1981), the existence of the equivalent martingale measure (which follows from our model construction) implies no arbitrage. Certain properties of the PK are satisfied automatically because of this feature of our model, while alternative approaches to its construction often require additional restrictions. In particular, Λ is positive almost surely. This follows from no arbitrage or can be shown formally based on the representation in (12). Furthermore, (13) implies

$$\mathrm{E}_t\left(\frac{\mathrm{d}\Lambda(t)}{\Lambda(t)}\right) = -(r_0 + r_1 U_2(t))\,\mathrm{d}t \equiv -r(t)\,\mathrm{d}t \tag{15}$$

and (together with (1) and (9)):

$$\mathrm{E}_t\left(\frac{\mathrm{d}S^i(t)}{S^i(t)}\right) = (r_0 + \alpha^i U_2(t))\,\mathrm{d}t = r(t)\,\mathrm{d}t - \mathrm{Cov}_t\left(\frac{\mathrm{d}S^i(t)}{S^i(t)}, \frac{\mathrm{d}\Lambda(t)}{\Lambda(t)}\right). \tag{16}$$

Hence asset pricing inconsistencies such as the risk-premium or risk-free rate puzzles are not a concern.

Indeed, Eqs. (15) and (16) work as restrictions on the PK model at the estimation stage. This contrasts the equilibrium based approach, where the terms involving Λ (or marginal utility in that case) are estimated based on the utility function and the use of consumption data, while the remaining components are estimated from the observed asset prices. Matching the data in such a way leads to the equity premium puzzle (Mehra and Prescott, 1985). This controversy makes it hard to detect the source of the problem, i.e. whether it is related to the utility function or the quality of the consumption data. However, in our case the discrepancy between the observed and modeled returns can be attributed only to the asset returns model misspecification.

In addition to these features, our approach separates the PK from the physical distribution of a particular asset. Let us expand the expression in (12) to illustrate this point:

$$\pi(t) = \mathrm{E}_t\left(\frac{\Lambda(t+\tau)}{\Lambda(t)}\,\xi(t+\tau)\right)$$

[7] The derivative pricing formulas for the model are derived in Appendix A.

$$= \int \xi(t+\tau) \frac{\Lambda(t+\tau)}{\Lambda(t)} \,\mathrm{pdf}\{\xi(t+\tau)\}\, \mathrm{d}\xi(t+\tau)$$

$$= \int \xi(t+\tau)\mathrm{SPD}\{\xi(t+\tau)\}\, \mathrm{d}\xi(t+\tau), \tag{17}$$

where pdf stands for the probability density function of the asset and SPD stands for the state price density of the same asset. Aït-Sahalia and Lo (1998) and Jackwerth and Rubinstein (1996) develop their methodology to estimate SPD. The drawback of such an approach is that if the SPD is estimated from one asset, say S&P 500, it is not possible to use the results to value options on, say National Semiconductor. One would have to go through the full estimation cycle again to obtain results for the latter asset. Our approach allows to keep information about preferences contained in Λ from one set of estimation results and apply it to a new task. All we have to do is estimate the physical pdf of the new asset. Rosenberg and Engle (2002) use representations in (17) to estimate the PK as the ratio of SPD to pdf. Therefore, they can not integrate information from different securities markets, as we do, since these two functions are estimated separately from the S&P 500 options and returns data respectively.

3. Data

In principle, if we use the data on the universe of all traded assets, we can get a very accurate estimate of the PK. Since it is computationally infeasible to use all the data for estimation, we have to pick some assets that would still give a reasonable kernel value. We made the following selection:

$i=0$: S&P 500 index.
$i=1$: Gold futures (COMEX ticker GC).
$i=2$: "Potomac Electric Power Co" stock (NYSE ticker POM).
$i=3$: "National Semiconductor" stock (NYSE ticker NSM).

The series $i=0$ should be able to capture the dynamics of the systematic factors very well. The series $i = 1$–3 are intended to represent assets which have behavior fairly different from the market.[8] In other words, we want these assets to have variability different from that of the market. Note that the gold futures data play a dual role here. On the one hand, they represent an asset that is typically negatively correlated with the market. On the other hand, GC is a derivative contract, hence it should facilitate estimation of the parameters related to preferences and prices of risk λ_j in particular.[9]

Our decomposition into the factors and idiosyncratic components should clearly play out here. Since, we will value all the assets simultaneously, the inclusion of such series should improve the quality of the PK. It is likely that two factors are not enough to describe the behavior of asset prices. Hence, considering such a specification along

[8] They represent the commodities, electric utilities industry and semiconductor industry respectively.
[9] The details are presented in Appendix B.

with such assets, we can attenuate the effects of the particular factors by combining all others together in the idiosyncratic term.

Our theoretical framework allows us to consider fixed-income securities simultaneously with the equities because the interest rate specification is coherent with probability measure transformation. Hence, the same PK can be applied to all assets. This is important not only because bond prices are related to the interest rate, one of the most important macro variables, but also because our specification of the short interest rate involves one of the common variables (U_2). We consider the 3-month T-bill daily prices in order to facilitate the estimation of the short rate parameters and the dynamics of U_2 (we will refer to these series as asset $i=4$).

Since we do not explicitly model a possibility of extreme movements, we consider exclusively the post 1987 crash period. Schwert (1990) provides evidence that the crash effects died out by March 1988. Hence we initiate the sample on March 1, 1988. Our sample extends to August 29, 1997, and covers nine and a half years, or 2353 daily observations. Series 0 are provided by CBOE. Series 2 and 3 come from CRSP, series 1 are provided by COMEX. Series 4 are obtained from the FRED database at the Federal Reserve Bank of St. Louis. The data represent activity in several different markets, which close at different times: series 1 are obtained at 2:30 pm, series 4—at 3:30 pm, series 0, 2, 3 are recorded at 4 pm. Thus we have a non-simultaneous price problem observed in Harvey and Whaley (1991). We will address this issue at the estimation stage (see Section 4.1).

Finally, additional options data was used for the model evaluation purposes. The details are provided in Appendix B.

Let us now take a first look at the constructed dataset. Fig. 1 reports the initial series. Panels (a) and (f) show a familiar plot of the S&P 500 level and log-returns respectively. Panel (b) reports the GC prices (series 1). The next panel in the Fig. 1 shows the prices of POM (series 2). We difference the series to obtain a stationary object (panel (h) reports the log-returns). Panels (d) and (i) report analogous information for NSM (series 3). Fig. 1, panel (e) reports the yields on the 3-month T-bills. We do not difference these series because US T-bill yields are treated as stationary in most empirical work of this kind.[10]

The second column of panels in the Fig. 1 shows returns of the series 0–3. One can see that they have quite different degrees of variability among each other. The assets, sample standard deviations and correlations are reported in Table 1. The standard deviations are quite different indicating various degrees of the idiosyncratic noise. The correlation coefficients reflect the usual findings: gold is negatively correlated with the market while stocks have positive correlation.

Our model is capable of generating this relationship. If $\beta^i \geqslant 0$, $i=0,2,3$, then, since

$$\operatorname{corr}(\mathrm{d}U_1^i, \mathrm{d}U_1^j) = \frac{\beta^i \beta^j U_2}{\sqrt{({\beta^i}^2 U_2 + {\gamma^i}^2 V^i)({\beta^j}^2 U_2 + {\gamma^j}^2 V^j)}}\,\mathrm{d}t \tag{18}$$

[10] Note that we do not use the observed yields to proxy for the risk-free interest rate. It is modeled as an unobserved factor $r_0 + r_1 U_2$.

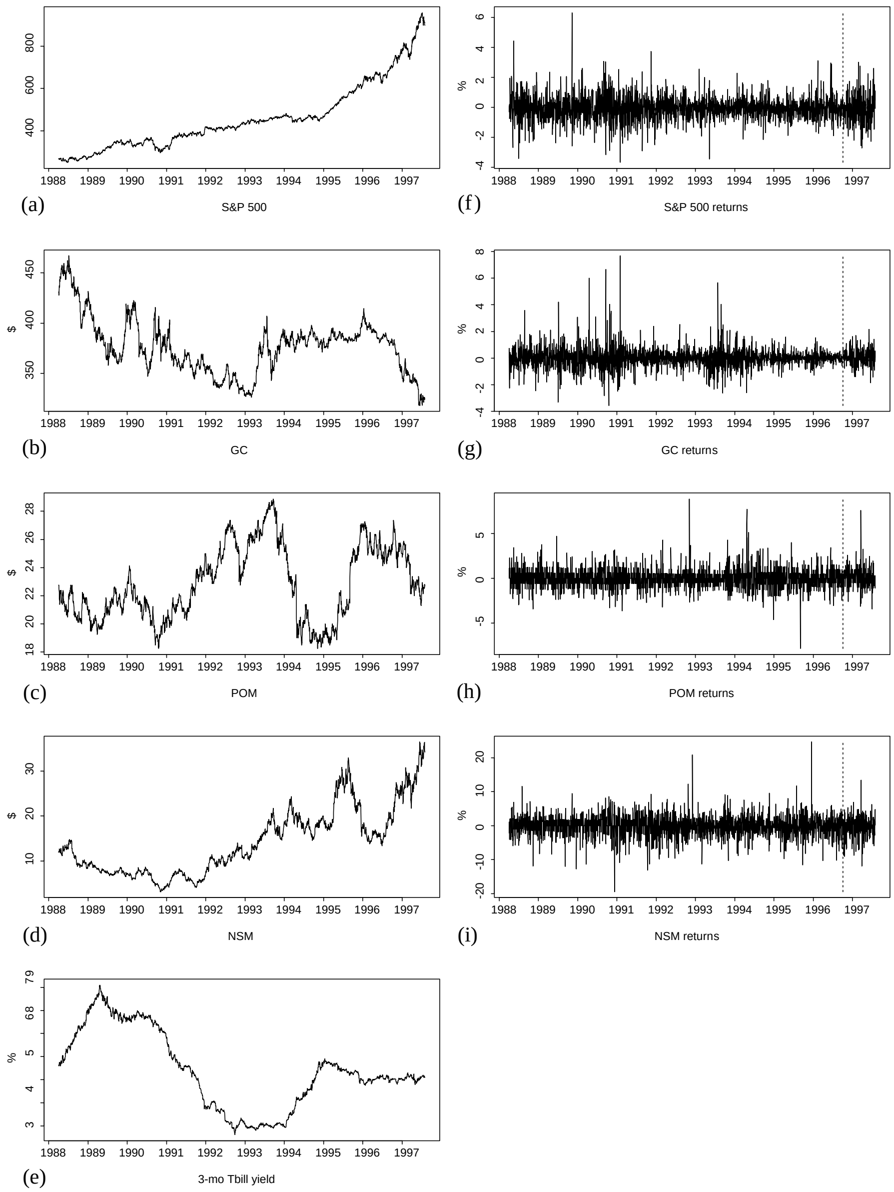

Fig. 1. The data. On this figure we plot the original data series. Panels (a)–(d) show the level of the S&P 500, the futures prices on the gold futures with the shortest possible expiration, prices of POM and prices of NSM respectively. Panels (f)–(i) depict respective returns series. Panel (e) shows bank discount yields on 3-month T-bills. The vertical dashed line separates the estimation and evaluation samples.

Table 1
Asset returns correlation matrix We report the annualized standard deviation of the assets' returns (the diagonal elements) and their correlations (off-diagonal elements). The numbers 0, 1, 2, 3 stand for S&P 500, GC, POM and NSM respectively (further details are provided in Section 2). We also report unconditional betas under the assumption that series 0 represent the market portfolio.

Series	0	1	2	3	β
0	12.67	−0.19	0.31	0.34	1.00
1		11.86	−0.09	−0.06	−0.21
2			17.54	0.07	0.43
3				49.76	1.35

the correlation between stocks will be positive. On the other hand, if we denote the futures price by $F(U_1^1,\tau)$ and apply Itô's lemma, then we find for $i \neq 1$

$$\begin{aligned}\operatorname{corr}(\mathrm{d}F(\cdot),\mathrm{d}U_1^i) &= \frac{(\beta^1\partial F/\partial U_1^1 + \rho\partial F/\partial U_2)\beta^i U_2}{\sqrt{(1/\mathrm{d}t)\operatorname{Var}(\mathrm{d}F(\cdot))({\beta^i}^2 U_2 + {\gamma^i}^2 V^i)}}\,\mathrm{d}t \\ &= \frac{(\beta^1 + \rho A_{u|_{\varphi=0}})\beta^i F(\cdot) U_2}{\sqrt{(1/\mathrm{d}t)\operatorname{Var}(\mathrm{d}F(\cdot))({\beta^i}^2 U_2 + {\gamma^i}^2 V^i)}}\,\mathrm{d}t \end{aligned} \tag{19}$$

This expression may be negative for any sign of β^1.[11]

We also report the unconditional betas of the series (assuming series 0 is a proxy for the market portfolio) in the last column of Table 1. These values also indicate that the dataset represents assets with quite different features. Overall, it seems that the selected data indeed provides a reasonable set for the PK estimation.

4. Estimating the pricing kernel

4.1. Simulated method of moments

Before we can proceed with the PK estimation, we have to know the parameter values in our models (2)–(10). The parameter vector Θ is equal to $(r_0, r_1, \lambda_1, \lambda_2, \rho, \theta, \kappa, \{\beta^i, \gamma^i, \eta^i, \imath^i\}_{i=0}^n)^\top$. Note that α^i and κ^* can be uniquely determined from Θ, (9), and (10). We also assume that the S&P 500 index is fully diversified, i.e. $\gamma^0 \equiv 0$ and we do not have to estimate η^0 and $\imath^0$.[12] The number of different assets, $n+1$, considered here is 4 (series $i = 0$–3 in Section 3). Thus, we have to estimate $7 + 4(n+1) - 3 = 20$ parameters.

Econometrically our model has all problems one could think of: no analytical expression for the likelihood function, latent state variables, non-synchronous data. Therefore, we rely on the simulated method of moments (SMM) of Duffie and Singleton (1993) as

[11] The last expression was obtained based on the results of Appendix A.
[12] Evidence in support of this assumption can be found in Chernov (2000).

the estimation method.[13] The preferred procedure in this class is the efficient method of moments (EMM) of Gallant and Tauchen (1996). The moments are constructed based on the score vector of an auxiliary model that asymptotically provides the true probability density of the data. In this case, the efficiency of EMM is close to that of maximum likelihood. If we are to implement this method in the present setup, however, we would have to estimate an auxiliary joint density of the five series under consideration. This is computationally infeasible for the current state of technology. Therefore, we will have to resort to a less efficient moment conditions.

Let us denote the five returns series that we use for estimation by r_{it}, $i = 0–4$. Then we will construct moment conditions based on the following moments:

$$
\begin{array}{ll}
\mathrm{E}(|r_{it}|), & i = 0\text{–}4 \\
\mathrm{E}(r_{it} r_{jt}), & i = 0\text{–}4,\ j = 0\text{–}4 \\
\mathrm{E}(r_{it} r_{it-1}), & i = 0\text{–}4 \\
\mathrm{E}(r_{it}^2 r_{it-1}^2), & i = 0\text{–}4 \\
\mathrm{E}(r_{0t}^2 r_{0t-1}) &
\end{array}
$$

This list amounts to $5+15+5+5+1=31$ individual moments. Our moments selection seems to represent a required minimum. We study the first and second moments, the cross-sectional variation which is captured by covariances between the series, and the intertemporal dynamics that are described by the autocovariances of the returns and their squares. The conditions based on the T-bill bank discount rate serve to identify the parameters related exclusively to the interest rate r_0 and r_1 and the parameters common with U_2, namely θ and κ. If we combine these moments with the conditions involving the absolute values of returns on the individual assets ($i = 0–3$), their second moments and the dynamics of the second moments we will be able to identify the three asset-specific parameters γ^i, η^i and ι^i. The last asset-specific parameter β^i can be identified by adding the moment conditions based on the returns autocovariances. The parameters related to the market prices of risk λ_1 and λ_2 are identified from the moments based on the T-bill and the gold futures data. Finally, the correlation coefficient ρ can be identified via the last moment that measures the changes in volatility in response to the change in the asset. Hence, we limit ourselves to the above minimal setup.[14]

It is clear that the selected moments do not span the entire distribution implied by the model. Hence, the estimation would be asymptotically more efficient if a larger set of moment conditions were used. There are two main considerations against this. Firstly, some moment conditions may force the model to fit features of the data that are not plausible within the framework of the model, i.e. the model is misspecified.[15]

[13] The alternative strategies based on the characteristic function (Chacko and Viceira, 2003; Jiang and Knight, 2002; or Singleton, 2001), or GMM (Pan, 2002) are not applicable for at least one of the three reasons.

[14] Andersen and Sørensen (1996) and Ho et al. (1996) use similar moments in a Monte-Carlo study and eight-asset application respectively.

[15] Chernov et al. (2003) show within the EMM framework that a stochastic volatility model without jumps is forced to match the tails of the distribution. As a result, the estimated parameters values are unreasonable. Namely the stochastic volatility factor wildly varies from to day contradicting the actual volatility behavior.

So, in practice, one would want to emphasize the moments which could be generated by a particular model. Secondly, Andersen and Sørensen (1996) find that adding extra moments beyond a certain set provides little additional information in finite samples. Our experimentation ranging from a just-identified system to 46 moments involving higher order and longer lags moments confirm this. Part of the reason, is that the objective function becomes so complicated that numerical convergence is harder to obtain. We also performed an additional check of our estimation strategy using the S&P 500 and T-bill series (r_0 and r_4). We estimated the sub-model corresponding to these series via the approximate maximum likelihood of Duffie et al. (2002). The obtained parameters are in line with the one obtained from our SMM estimation of the larger model.[16]

While the SMM procedure is very intuitive in theory, several important issues should be considered during the practical implementation. Most of them are related to the construction of the weighting matrix and identification. We used the diagonal continuously updated matrix.[17] Such a weighting matrix provides a substantially greater numerical stability as opposed to the optimal weighting matrix. This stability comes at a price of larger standard errors. We also verified the local identifiability of all the parameters of the model by establishing that the matrix of derivatives has full rank.

There are several issues specific to the simulated estimation that we would like to address here. In order to construct the above analytical moments we have to simulate observations from the process (2)–(4). We adopt the Euler-Maruyama order 0.5 strong discretization scheme (see Kloeden and Platen, 1995). The fact that we have to simulate allows us to incorporate the non-synchronous feature of the dataset into the estimation strategy. Namely, we simulate data points at intradaily increments to remedy this problem. We have four simulated observations within a day and we assume that one 8-hour business day is equal to $\Delta t = 1/250$ on the yearly time scale. Each time increment will be two times smaller than the previous one, and the first one in the new day will be reset back to the initial value, $\delta t = 1/4 \cdot \Delta t$ or 2 h. Then, we simulate observations at $t_1 = 12:30\text{ pm} + \delta t$, $t_2 = t_1 + 1/2 \cdot \delta t$, $t_3 = t_2 + 1/4 \cdot \delta t$, and $t_4 = t_3 + 1/8 \cdot \delta t$. This corresponds to 2:30 pm (the GC prices are reported), 3:30 pm (the T-bill yields are reported), 4:00 pm (the stock market prices are reported) and 4:15 pm (the index options prices are reported). Then we will take those observations, which correspond to the actual observation time as the end-of-the-day simulated prices.

Since the companies in the S&P 500 index pay dividends we have to take them into an account in our simulation procedure. While the historical data on the dividends is available and we could adjust the observed index level by the dividends, it is not feasible to do so with the simulated observations. Simulations are close to the real market conditions in the sense that dividends distributions are unexpected. Therefore, we assume a continuously compounded distribution rate that could be simulated. We assume a dividend rate of 2%, which is consistent with historical data.

---

[16] These results are available in Appendix D.

[17] Please refer to Den Haan and Levin (1997), Hall (2000), Hansen et al. (1996) for general treatment and to Chernov (2000) for the details of this particular application.

4.2. Reprojection

Having estimated the model parameters we are ready to proceed with the PK estimation. Note from (13) that the PK value depends on the realizations of an unobservable factor U_2. Hence we face a filtering problem. We use the reprojection methodology of Gallant and Tauchen (1998) to solve it. Denote the vector of contemporaneous and lagged observed variables by $x(t)$ and the vector of contemporaneous unobserved variables by $y(t)$. The problem at hand can be characterized as finding

$$\tilde{y}(t) = \mathrm{E}(y(t)|x(t)) = \int y\, p(y|x(t), \Theta)\, \mathrm{d}y. \tag{20}$$

In other words, we have to know the conditional probability density of $y(t)$. If we knew the analytical form of this density implied by the system dynamics, we could estimate it by $\hat{p}(y(t)|x(t)) = p(y(t)|x(t), \hat{\Theta})$. But this analytical form is not available in our case. Therefore, we estimate this density as $\hat{p}(y(t)|x(t)) = f_K(\hat{y}(t)|\hat{x}(t))$, where $\hat{y}(t), \hat{x}(t)$ are simulated from our system with parameters set equal to $\hat{\Theta}$ and f_K is the SNP density of Gallant and Tauchen (1989).[18]

The main idea behind the SNP model is to describe the observed time series as a VAR(M_u)-GARCH(M_a, M_g) model with the error terms described nonparametrically rather than by a normal or t distribution. The error term is represented by a combination of the Hermite polynomials. These polynomials are known to form an orthogonal basis in the space of squared integrable functions defined on the real line. Hence, we can represent any function by taking a linear combination of a sufficient number of such polynomials. Since they are equal to $H_n(x)\mathrm{e}^{-x^2/2}$, where $H_n(x)$ is a regular polynomial, one can interpret such a model for errors as a normal density augmented by polynomials to allow for various departures from normality. $H(K_z, K_x, M_p)$ denotes the polynomial part of the SNP model with power K_z whose coefficients are themselves polynomial functions of degree K_x in M_p lags of the time series we are modeling.

Since we use this modeling technique for the reprojection of the latent PK, the observed time series act as exogenous variables. Therefore, the reprojection SNP model will have no autoregressive part: VR(M_u)-$H(K_z, K_x, M_p)$.[19] It means, that all the heteroscedasticity in the data will have to be explained by the polynomial component. Hence we expect to see a parameterization which is much more rich than is typically encountered in the regular SNP.

Note that the above outline does not specify which observables, $x(t)$, should be used for reprojection. It means, that in essence, we can consider several different information sets. We use the following choice of $x(t)$: (i) S&P 500, (ii) S&P 500 and GC, (iii) S&P 500 and T-bill rate, (iv) T-bill rate. This seems to be an arbitrary choice of conditioning variables. However, it is similar in this respect to the choice of instrumental variables proxying for the information set in the GMM estimation of asset pricing models. The particular choice is dictated by the following considerations. We do not want to use

[18] See, for instance, Gallant and Tauchen (1998) for an intuitive description and examples of implementation.

[19] The notation VR instead of VAR emphasizes that we consider vector regression rather than autoregression.

more than two series in order to have a computationally feasible SNP density, $f_K(\cdot)$. Given this limitation, the four choices seem to be the most informative of the economic conditions and, hence, should facilitate the PK estimation.

In our particular application, $y(t) = m(t) = \Lambda(t)/\Lambda(t-1)$. We start out by simulating 100,000 observations of $n(t) = \log(\Lambda(t)/\Lambda(t-1))$ simultaneously with $U_1^0(t)$, returns on $F(U_1^1(t), \tau_t)$ and the bank discount rate based on the price of the T-bill $B(t, \tau_t)$ given the parameter vector $\hat{\Theta}$.[20] Various combinations of the last three variables will serve as different choices of the observables vector $x(t)$. Itô's formula and (13) allow us to establish the dynamics of $n(t)$ needed for the simulation scheme:

$$n(t) = \int_{t-1}^{t} d\log\Lambda(u) = -\int_{t-1}^{t}\left(r_0 + \left[r_1 + \frac{\lambda_1^2 + \lambda_2^2}{2}\right]U_2(u)\right)\mathrm{d}u$$
$$-\int_{t-1}^{t}(\lambda_1\sqrt{U_2(u)}\,\mathrm{d}W_1(u) + \lambda_2\sqrt{U_2(u)}\,\mathrm{d}W_2(u)). \quad (21)$$

The Euler–Maruyama discretization scheme is used again to obtain realizations of $n(t)$. Next, we estimate four different SNP densities $f_K(\hat{n}(t)|\hat{x}(t))$ corresponding to the choices of the exogenous variables $x(t)$ outlined above. In order to obtain the estimate of the PK we use (20), namely:

$$\tilde{m}(t) = \mathrm{E}(m(t)|x(t)) = \int m\,\hat{p}_{m(t)|x(t)}(m|x(t))\,\mathrm{d}m$$
$$= \int \mathrm{e}^n\,\hat{p}_{n(t)|x(t)}(n|x(t))\,\mathrm{d}n = \int \mathrm{e}^n f_K(n|\hat{x}(t))\,\mathrm{d}n. \quad (22)$$

This formula underlies the filtering problem that we are trying to solve: we do not attempt to forecast the PK, but we are trying to infer its value today from the contemporaneous asset prices.

The procedure outlined above will yield four alternative estimates of the PK. Section 5 will discuss our approach to the selection of the final estimate. Before we get there, Section 4.3 will discuss the results of the described above methodology implementation.

4.3. Estimation results

4.3.1. Estimating the assets dynamics model

The estimated model parameters that are obtained via SMM are reported in Table 2. Table 3 complements it by discussing characteristics of the estimated model that are more intuitive for interpretation. The first thing we notice from Table 2 is that despite the inefficiency of our estimator of the variance-covariance matrix of the moment conditions, many parameters, namely $\theta, r_0, r_1, \lambda_2$, all β's, γ^2, γ^3, all η's, and all $\imath$'s are significant. However, some of the standard errors are large and indicate potentially insignificant parameters. In order to gain further intuition about our model, we will concentrate on the entries in Table 3.

[20] The time index in τ_t indicates that every day the time to maturity will be different.

Table 2
Estimation results We calibrate the assets dynamics model

$$\frac{dS^i(t)}{S^i(t)} = (r_0 + \alpha^i U_2(t))\,dt + \beta^i\sqrt{1-\rho^2}\sqrt{U_2(t)}\,dW_1(t) + \beta^i\rho\sqrt{U_2(t)}\,dW_2(t) + \gamma^i\sqrt{V^i(t)}\,dZ^i(t)$$

$$dU_2(t) = (\theta - \kappa U_2(t))\,dt + \sqrt{U_2(t)}\,dW_2(t)$$

$$dV^i(t) = (\eta^i - \imath^i V^i(t))\,dt + \sqrt{V^i(t)}\,dW^i(t)$$

$$\alpha^i = r_1 + (\lambda_1\sqrt{1-\rho^2} + \lambda_2\rho)\beta^i$$

to the returns on the S&P 500 index, GC, POM, NSM and the T-bill bank discount rate via the Simulated Method of Moments. The S&P 500 is assumed to be fully diversified, i.e. $\gamma^0 = 0$. Panel A reports parameters common to all the assets considered, while panel B shows the asset specific parameters. Panel B also shows the relationship between the index i and an asset. The dynamics of the T-bill are completely determined by the common parameters, hence it is not mentioned in the panel B. Standard errors are reported in parentheses.

Panel A

θ	κ	r_0	r_1	λ_1	λ_2	ρ
0.7172	1.8077	0.0214	0.0894	0.5581	−0.8371	−0.4817
(0.42)	(1.07)	(0.01)	(0.03)	(0.66)	(0.28)	(0.71)

Panel B

Assets parameters	S&P 500 $i = 0$	GC $i = 1$	POM $i = 2$	NSM $i = 3$
β^i	0.2705	−0.1984	0.2850	0.9326
	(0.04)	(0.02)	(0.04)	(0.14)
γ^i	0	0.2668×10^{-4}	0.6803	0.1378
	(−)	(0.34×10^{-4})	(0.09)	(0.01)
η^i	n.a.	2.5805	1.3155	16.2226
		(0.92)	(0.13)	(3.21)
$\imath^i$	n.a.	4.1597	0.9418	5.0421
		(0.88)	(0.31)	(1.07)

All the elements of Table 3 are computed based on the values of the estimated parameters. For instance, the average value of the systematic factor is equal to the unconditional mean of U_2 that is very well known to be equal to θ/κ. Speed with which U_2 is pulled back to its mean (or its persistence) is measured by κ. Smaller values indicate a slower speed, i.e. a more persistent process. We can use the persistence to measure the half-life of the process. Namely, half of the process shock dissipates in $\log(2)/\kappa$ years. Similarly, the average idiosyncratic factor is equal to the unconditional mean of V^i, i.e. $\eta^i/\imath^i$, the pullback speed is equal to $\imath^i$ and half-life is $\log(2)/\imath^i$. The average risk-free rate is computed in the same way, based on the expression $r_0 + r_1 U_2$. The leverage effect is equal to the correlation coefficient ρ. We can also decompose the average total variance of each asset into the systematic, ${\beta^i}^2\theta/\kappa$, and idiosyncratic, ${\gamma^i}^2\eta^i/\imath^i$, components. They will allow us to assess the contribution of the common factor

Table 3
Estimation results interpretation We use the estimated parameters reported in Table 2 to compute the implied characteristics of assets that are easier to interpret.

Panel A

Average systematic factor (U_2)	Persistence of U_2	Half-life of U_2	Average risk-free rate	Leverage effect
0.3967	1.8077	0.3834	0.0569	−0.4817

Panel B

Assets characteristics	S&P 500 $i=0$	GC $i=1$	POM $i=2$	NSM $i=3$
Idiosyncratic factor (V^i)	n.a.	0.6203	1.3967	3.2173
Persistence of V^i	n.a.	4.1597	0.9418	5.0421
Half life of V^i	n.a.	0.1666	0.7359	0.1374
Systematic variance	0.0290	0.0156	0.0322	0.3451
Idiosyncratic variance	0	0.4418×10^{-9}	0.6465	0.0611
Risk premium	0.0958	−0.0702	0.1009	0.3302

U_2 to the variability of different assets. Finally, we can also compute the average risk premium $(\lambda_1\sqrt{1-\rho^2}+\lambda_2\rho)\beta^i\theta/\kappa$ of each asset.

Overall results support the intuition developed by the unconditional characteristics in the Table 1. S&P 500 and GC have very similar relatively small volatility (as measured by the sum of the systematic and idiosyncratic variances), while POM and NSM have a substantially higher variation. Interestingly, most of the variation in POM is attributed to the noise term (0.65 of idiosyncratic term vs. 0.03 of the systematic one) while variation in NSM is mostly related to the market risk (0.06 vs 0.34 respectively). Surprisingly, the GC noise component is virtually redundant. The systematic and idiosyncratic components have quite different dynamics as measured by persistence. The half-life of the systematic volatility U_2 is roughly 96 business days. The individual volatility components are ranked (from the most persistent to the least persistent) as follows: POM (184 days), GC (42 days), and NSM (34 days).

The persistence of the state variable U_2, κ, deserves a special attention. The reason is that we model stock market volatility and risk-free rate via one state variable. If we were to separate them and denote interest rate and volatility persistence by κ_r and κ_v respectively, then we would expect the following relationship with the persistence parameter of $U_2 : \kappa_r < \kappa \cdot r_1$ and $\kappa < \kappa_v$.[21] Future research should take this observation into an account when modeling stock and bond markets simultaneously.[22]

[21] For example Gallant and Tauchen (1998) find $\kappa_r \approx 0.14$ on an annual scale based on the 1962–1995 period, and Eraker et al. (2003) find $\kappa_v \approx 5.8$ based on the 1980–1999 period.
[22] See Chernov (2000) for further discussion of this issue.

The average risk-free rate is estimated to be 5.7%. We can note that it can be decomposed into the constant part (r_0) that is equal to 2.1% and the varying term ($r_1 U_2$), whose average is equal to 3.6%. Hence, we can draw an analogy with the inflation-indexed T-bonds that add a constant interest rate (around 3.5% depending on the issue date) on top of the time-varying inflation rate. Since our average stochastic component roughly corresponds to the historical inflation rate and $U_2 = 0$ corresponds to no uncertainty (no inflation, in particular) in the economy, our estimate of r_0 can be interpreted as a fixed-rate adjustment on the inflation-indexed T-bills if they were issued.

We can make several interesting observations about the average risk-premium on our assets. Note that the risk premiums of S&P 500 (the market) and POM are very close. This is not surprising if one notices that the systematic variances of these two assets are very close in magnitude as well. Hence, though POM has a much larger total variance, most of it is diversified a way in a portfolio and the systematic risk that is left is very close to that of the market. This contradicts the value of the unconditional beta (0.43) that was reported in Table 1. The risk premiums on GC and NSM support the intuition from their respective betas.

Note that our model specification implies that the investor may require not only the premium related to the variation in the asset prices ($\lambda_1 \sqrt{U_2}$), but also the premium related to the variations in volatility ($\lambda_2 \sqrt{U_2}$). We find that λ_2 is a statistically significant parameter, hence the data supports our conjecture that compensation for the uncertainty in asset returns is not enough to attract investors. Therefore, at a minimum, one needs to consider a PK with two factors.

4.3.2. Reprojecting the pricing kernel

We can now proceed with the estimation of the PK based on the estimates of the model parameters and various sets of conditioning variables. The SNP procedure combined with the Schwarz BIC criterion selected the following models of the PK (see the outline of the classification scheme in Section 4.2):

(i) VR(1)-$H(4,2,7)$, when the conditioning information set contains the S&P 500 returns, i.e. $x(t) = U_1^0(t)$;
(ii) VR(1)-$H(9,1,3)$, when the conditioning information set contains the T-bill bank discount rate, i.e. $x(t) = B(t, \tau_t)$;
(iii) VR(1)-$H(4,1,1)$, when the conditioning information set contains the S&P 500 returns and the T-bill bank discount rate, i.e. $x(t) = (U_1^0(t), B(t, \tau_t))$;
(iv) VR(1)-$H(4,2,2)$, when the conditioning information set contains the S&P 500 returns and the GC returns, i.e. $x(t) = (U_1^0(t), F(U_1^1(t), \tau_t))$.

As one can notice, the regressive part of all the models is very simple: it involves only the contemporaneous variables in the information set $x(t)$. However, the modeling of the error terms involves highly non-linear functions. Note that models (iii) and (iv) involve bivariate series, hence a more simple functional form describing the dynamics of the kernel.

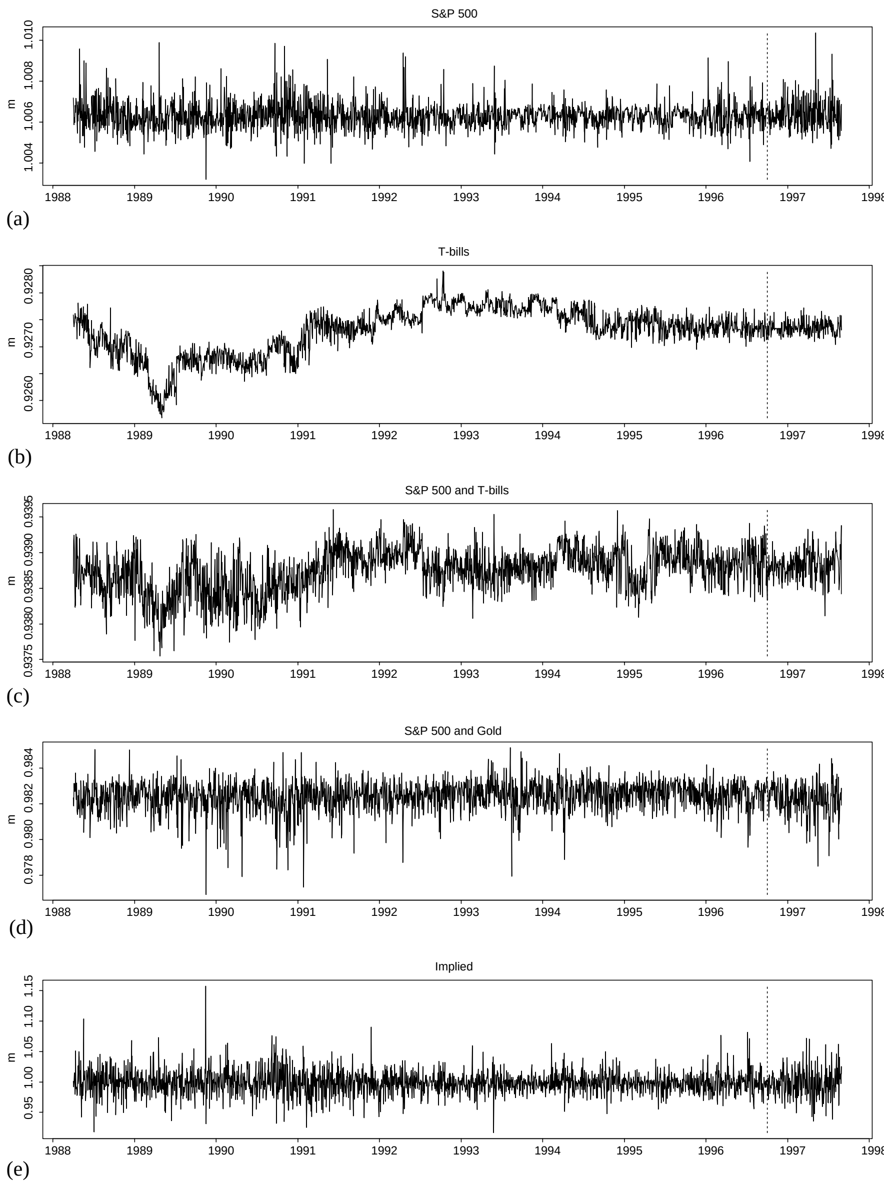

Fig. 2. Pricing kernels. We plot the time series of various PK estimators. Panels (a)–(d) show the kernels obtained from the reprojection procedure conditioned on the observable assets named in the respective panels headers. Panel (e) shows the PK implied from the T-bill prices and S&P 500 returns. The vertical dashed line separates the estimation and evaluation samples.

Finally, we can obtain the reprojections of the kernel on the observed data given the estimated SNP models. The panels (a) to (d) of the Fig. 2 plot the time series of the obtained conditional estimates based on the SNP models (i) to (iv) respectively.

We can immediately notice that the different information sets yield kernels with very different numerical values and time series properties. Since we are essentially trying to come up with a function of $x(t)$ that would approximate the kernel the best, it is natural that the resulting objects reflect the properties of the variables in $x(t)$. There are differences not only in the pattern, but in the range of the estimates as well: the range of the reprojected kernel is $(1.004-1.010)$, $(0.9260-0.9280)$, $(0.9375-0.9395)$, and $(0.9760-0.9860)$ for panels (a)–(d) respectively. The fact that they do not overlap is the evidence of how different are the probability densities $f_K(\cdot|x(t))$ implied by different information sets.

It is also clear, that the estimated kernels are affected by the model misspecification. Note that reprojected kernels involving the T-bill rate (panels (b) and (c)) have much lower values, than the ones involving the S&P 500 returns (panels (a) and (d)). Such ranking seems to be a direct consequence of the single factor that drives both interest rates and stock market volatility. The discussed above implicit relationship $\kappa_r/r_1 < \kappa < \kappa_v$ seems to be responsible for these differences in estimated kernels. Namely, the density functions $f_K(\cdot|x(t))$ are estimated from the simulated data, i.e. based on κ, while the real data can be better described by κ_r, and κ_v. Hence, when we impose the model on the data, the implicit ranking of the parameters affects the ranking of the estimates.

Moreover, the estimate based on the S&P 500 is so high, that it is greater than 1 for the whole period. These values do not imply a negative interest rate however, because the expectations, that we take in (22) are contemporaneous and, therefore, are not equal to the inverse of the gross interest rate. In other words, we estimate today's realizations of the PK, not the expectation of its tomorrow's value. There is nothing problematic with a value of the PK greater than 1. The only troublesome feature of our estimate is that it is greater than 1 for the entire period. The flip side of this feature is that T-bill based estimates are less than 1 in the same sample.

Obviously, the model misspecification and nonavailability of the full information set introduce trade-offs regarding particular estimates to be used in applications. One way to select the preferred estimate is based on the magnitude of the pricing error. However, before we do this, we explore another estimate of the kernel, which we call the implied PK.

4.3.3. Implying the pricing kernel

Observe, that the bond pricing formula in (A.2) implies that

$$\log B(t,\tau) = C(\tau)|_{\phi=0,\varphi=i} + A_u(\tau)|_{\phi=0,\varphi=i} U_2(t). \tag{23}$$

It means that if our model is perfectly specified we can invert the values of U_2 from the bond prices. However, since our model is misspecified, this operation would be similar to the implied volatility procedure frequently used in the context of the Black and Scholes (1973) model or the Heston (1993) stochastic volatility model. We will denote this implied latent factor by $\hat{U}_2$.

Since γ^0 is equal to zero, the knowledge of the index returns U_1^0 and the common latent factor $\hat{U}_2$ allow us to imply the realizations of the unobserved systematic information shocks $\mathrm{d}W_1$ and $\mathrm{d}W_2$ as a solution of a simple linear system consisting of the discretized versions of (4) and (2). We will denote these implied shocks as e_1 and e_2, respectively. If our model was ideally specified, then $e_i \sim N(0, 1/250)$. In reality, e_i's will absorb all the model errors, i.e. they will act as error terms that an econometrician adds to a model to account for misspecification. Finally, the knowledge of $\hat{U}_2$ and e_i's allows us to imply the values of the PK based on the discretized version of (13).

The implied kernel is plotted on the last panel of Fig. 2. One can notice that the implied kernel has a higher variability than the reprojected ones. This makes perfect sense, because under the null that the model is specified correctly we effectively make the full information set observable. In other words, we obtain the projection of the kernel on the time t information set, rather than a subset $x(t)$ as in the reprojection procedure of the previous section. Even if the model is misspecified, we still have more variability for the same reason. However, the values of the kernel may be biased, which does not happen with the reprojection procedure.

As a side product of our implied kernel procedure, we can look at the implied information shocks e_1 and e_2 as a simple diagnostic of the model misspecification. According to our model, these errors should be i.i.d. normal. Hence, departures from this assumption will point out the directions of the model misspecification.

We will resort to the simplest graphical techniques to evaluate the normality of random variables. Figs. 3 and 4 report such analysis. The first of these figures looks at the unconditional properties of the series. Panel (a) shows the scatter plot to check if the e_i's are uncorrelated as it is assumed in our model. The plot reveals little dependence of the two variables. Panels (b) and (c) show the QQ plots for each of the series. Panel (b) reveals that e_1 is fairly symmetric, but has heavier tails as compared to the normal distribution. Panel (c) shows that e_2 has a certain degree of negative skewness. The tails are heavy as well, but not as much as the ones of e_1. Hence we can conclude, that the implied information shocks are not normally distributed.

The normality of the e_i's may not be that important as long as the series exhibits features of white noise such as homoscedasticity and zero autocorrelation. Fig. 4 evaluates some of the time series properties of the e_i's for this reason. Panels (a) and (c), which show the simple time series plots, immediately reveal heteroscedasticity in the series. This heteroscedasticity most likely accounts for the unconditional departures from normality observed in the Fig. 3. Panels (b) and (d) show the sample autocorrelations for the series. The e_1's exhibit the desired white noise feature in this respect. However, e_2's have significantly non-zero autocorrelations that die out very slowly. This is indicative of ARMA structure in the series. This term takes on most of the misspecification in our model. It seems that the observed heteroscedasticity of the e_i's might be removed by adding a jump component. We can conclude, that our model is misspecified and, hence, the implied PK will not be consistent with the dynamics of the model.

This evidence of our model misspecification may be interpreted as a failure of the suggested methodology. However, from the empirical perspective, we have to acknowledge that any model, no matter how realistic it is, will still be misspecified. Therefore, the generic problem that any empiricist is facing, is how to make the best use out of

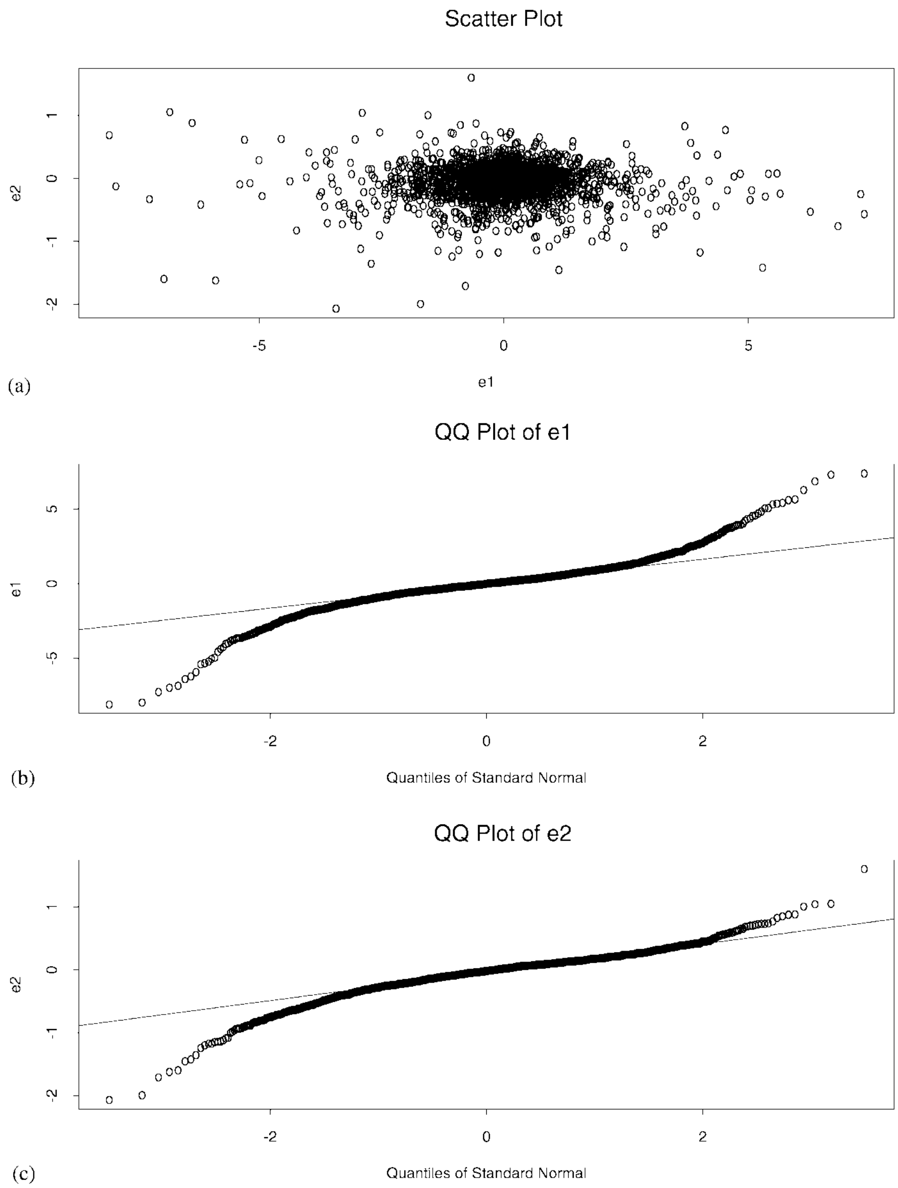

Fig. 3. Unconditional properties of the information shocks. We look at the simplest unconditional properties of the information shocks e_1 and e_2 implied from our model and the T-bill and S&P 500 data. Panel (a) shows the scatter plot, while panels (b) and (c) show the QQ plots for e_1 and e_2, respectively.

the given misspecified model. Implementation of the model that is consistent with its theoretical properties (i.e. implicit distribution of the state variables), allows to identify its pitfalls from the modeling and empirical perspectives. For instance, our diagnostics results show, that the reprojected PK is only a portion of the “true” PK (see also our

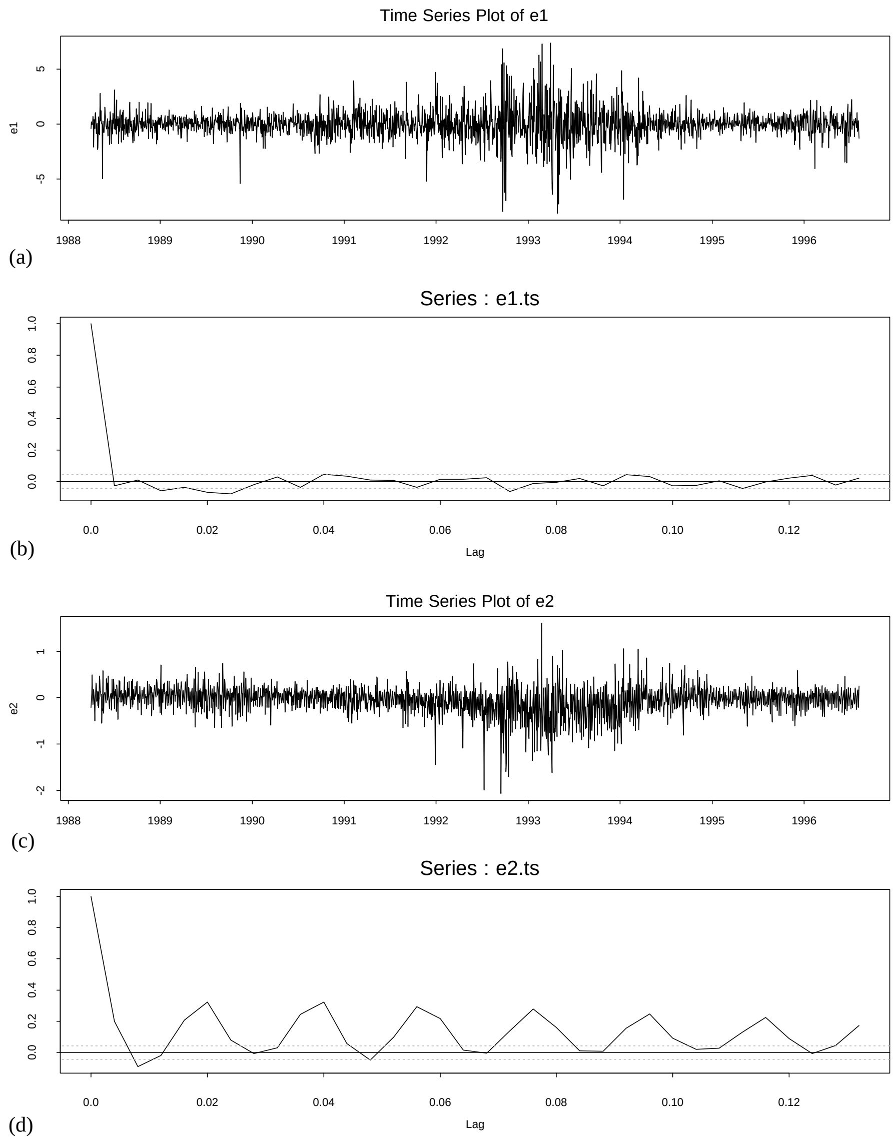

Fig. 4. Conditional properties of the information shocks. We look at the simplest conditional properties of the information shocks e_1 and e_2 implied from our model and the T-bill and S&P 500 data. Panels (a) and (c) show the times series plots, while panels (b) and (d) chart the sample autocorrelations of the respective series. Lags are measured as fractions of the sampling period, i.e. 0.01 corresponds to approximately 24 days.

discussion of (14)). Hence, we know, that empirical asset returns constructed based on this PK will represent only a part of the actual returns. Implied PK may provide the superior fit by design, but will not yield any of the insights we discussed here, because it is not consistent with the model dynamics.

Table 4
The pricing kernels diagnostics We report the diagnostics results based on the Hansen–Jagannathan (HJ) methodology for the candidate PKs estimates based on different combinations of samples and assets. The combination type has a name of the form "sample-#", where "sample" can be either "in" denoting the estimation period from March, 1988 to August 1996 or "out" indicating the evaluation period from September, 1996 to August, 1997. The "#" stand for the number of assets used to compute the distance. If we rely only on the assets used for estimation, then "#" is equal to 4. If we add the options prices on top of that then the "#" is equal to 8. Refer to Section 3 and Appendix C for details on the datasets. Panel A reports the optimal distance to the HJ bound statistics (OD) and the respective p-values in the parentheses. Large p-values indicate that the PK is on or inside the HJ bound. Refer to the Appendix C and, in particular, formula (C.10) for details. Panel B reports the values of the HJ distance from the estimated PKs to the set of admissible PKs.

Type	S&P 500	T-bills	S&P 500 and T-bills	S&P 500 and GC	Implied
Panel A					
in-4	1.55	2.57	2.57	1.90	0.00
	(0.11)	(0.05)	(0.05)	(0.08)	(0.50)
in-8	1.94	2.57	2.56	1.98	0.00
	(0.08)	(0.05)	(0.05)	(0.08)	(0.50)
out-4	1.15	0.99	2.29	2.38	2.00
	(0.14)	(0.16)	(0.07)	(0.06)	(0.08)
out-8	1.10	0.80	2.53	2.29	1.97
	(0.15)	(0.19)	(0.06)	(0.07)	(0.08)
Panel B					
in-4	2.90	4.33	3.96	2.99	2.35
in-8	8.94	9.50	9.34	8.97	8.81
out-4	2.82	3.03	2.97	2.83	2.55
out-8	3.63	3.79	3.74	3.63	3.41

5. Evaluation of the candidate pricing kernels

Since we have obtained five different estimates of the PK, it is important to evaluate them based on a selected criterion. We choose two evaluation tools for such an assessment both developed by Hansen and Jagannathan (1991, 1997) and are known as the HJ bounds and HJ distance (δ) respectively. The HJ bound establishes a threshold level on the second moment of an admissible PK. Violation of this bound indicates misspecification of a candidate kernel. The bound is considered to be violated if the candidate PK is outside of the bound and the distance to the bound is significantly different from zero, as judged by the optimal distance test (OD). The δ measures the distance from the candidate to the unobserved true PK and can be interpreted as the pricing error magnitude. We briefly discuss the methodology in Appendix C.

Table 4 reports the values of the OD statistic and the δ computed for each candidate PK depicted in the Fig. 2. We consider four different sets of asset returns for each estimate of the kernel. These sets vary across the sample period and the number of assets involved in the computation of the HJ distance. On the one hand we consider in- and out-of-sample pricing errors (the corresponding time periods are from March 1, 1988 to August 29, 1996 and from September 1, 1996 to August 29, 1997). On the other hand we consider the four assets used in the model estimation and then we look at these assets plus additional four series of options prices that are described in Appendix B.

Panel A of the table reports the tests of the HJ bounds violations. The OD statistic used in these tests (see formula (C.10)) involves GMM estimation based on the moment conditions featuring the second moment of the PK, $m^2(t)$. One way to compute this second moment would be just simple raising of $\tilde{m}(t)$ to the second power. However, since $\tilde{m}(t) = \mathrm{E}(m(t) \mid x(t))$, this approach will yield a biased estimate of $m^2(t)$. The unbiased estimate would be direct filtering of the second moment following the same approach as for $m(t)$ itself. In other words, we have to compute $\mathrm{E}(m^2(t) \mid x(t))$. One can notice from (22), that this quantity is simply $\int \mathrm{e}^{2n} f_K(n \mid \hat{x}(t))\,\mathrm{d}n$. We will denote it by $\tilde{m}_2(t)$ to distinguish from the second power of $\tilde{m}(t)$. Given that we already have estimated all the SNP densities $f_K(\cdot)$, it is a fairly simple matter to filter the second moments of the candidate PKs.

Overall, the OD test results judge our estimates of the PK favorably. Since Burnside (1994) provides evidence that in finite samples the test tends to overreject, even the p-values equal to 0.05 may mean a failure to reject the kernels admissibility. The most impressive results are obtained for the implied PK: it lies on the bound in sample, however it deteriorates slightly out of sample. The PK reprojected on S&P 500 also has good properties: most p-values are higher than 0.10 based on the asymptotic distribution. The performance of this kernel actually improves once we go out of sample. One explanation could be that our misspecified model may happen to recreate the out-of-sample dynamics better. The performance of other kernels seems to be worse, though the one reprojected on T-bills performs really nice out-of-sample. It is not surprising that all kernel estimates are close to the admissibility bound. As we mentioned in our discussion of Eqs. (13) and (14), our approach yields estimates of the minimum variance kernel.

Panel B of Table 4 reports the values of the HJ δ. Regardless of the sample of returns we use, we establish the following ranking of the PKs estimates (in descending order): the implied kernel, the kernel reprojected on the S&P 500 returns, the kernel reprojected on the S&P 500 and GC returns, the kernel reprojected on the S&P 500 and T-bill returns and, finally, the kernel reprojected on the T-bills. The superior performance of the implied kernel is not surprising: it is akin to superior performance of the ad-hoc implementation of the Black–Scholes model in option pricing tests. This approach uses the degrees of freedom provided by the model to adjust for whatever misspecification the model has. Every day these adjustments are different, however, which leads to a very complicated structure of the information shocks that are supposed to be white noise (see our discussion of these issues in Section 4.3.3).

If we concentrate on the more consistent reprojection procedure, the S&P 500 based filter is the best. Interestingly, the quality of the information contained in the index is so good that it is able to correct the mispricings of the T-bill based filter (compare the pricing errors of the univariate filter based exclusively on T-bills and the bivariate filter based on the T-bills and the S&P 500 index). This result is surprising in that the S&P 500 index does not seem to contain information about the market prices of risks that constitute an important component of the PK. The bivariate filters, which involve both the S&P 500 and an asset requiring the risk-neutral parameters (T-bills or the gold futures) do not improve upon the more simple filter.

We noted in Appendix C, that due to particulars of our estimation procedure, we cannot derive a test to evaluate δ. However, it is not an important drawback given the above discussion. Indeed, if we found that all δ's were not significantly different from each other, it would still indicate that we have to rely on the S&P 500 based one for the parsimony reasons. Hence, we arrive at the estimate of the PK which involves only the data on the S&P 500 returns. In this respect we come back to the CAPM, whose practical implementation often involves the same index as a proxy for the PK. The crucial difference, however, is that our estimate is based on a highly nonlinear function of the index returns while CAPM reserves to the simple linear relationship.

As we observed in Section 4.3.2, our preferred kernel is greater than 1 for the entire sample period. While this does not violate any basic principles per se, it still raises a question whether such realizations of the PK are reasonable. Except for the discussed above model misspecification, this result may be driven by the specifics of our sample. In general, except for the short contraction period in the beginning of the 1990s, the U.S. economy enjoyed “good times” (see for instance panel (a) of Fig. 1). The PK greater than 1 implies a premium put by investors on payoffs during the good time. Such a result cannot be explained in the framework of the standard time-separable utilities. Indeed, good times imply consumption growth, which in its turn implies decrease in marginal utility, and, therefore, a PK which is less than 1. Hence, the Hansen–Jagannathan diagnostics exclude this class of preferences.

Can we explain our result by different type of preferences? The stochastic habit formation model of Campbell and Cochrane (1999) can be one possible explanation. Their model implies the following PK:

$$m(t) = \varrho \left(\frac{C(t) - X(t)}{C(t-1) - X(t-1)} \right)^{-\upsilon}, \tag{24}$$

where $C(t)$ denotes the level of consumption, and $X(t)$ is the habit level. Time preference and risk aversion parameters are denoted by ϱ and υ, respectively. We can see from this expression, that $m(t)$ is greater than 1, when the surplus consumption, $C(t) - X(t)$, decreases. Such an outcome is quite plausible because we would expect the habit to grow faster than consumption during the good times.

Panel (a) of Fig. 5 shows the familiar plot of the PK reprojected on the S&P 500 returns with the beginning and the end of the NBER dated contraction cycle (dashed lines), the beginning of the evaluation sample (dotted line), and the unconditional first moment of the kernel (thick solid line). It is quite hard to interpret this plot, since the contraction period does not seem to critically differ from the expansion periods. However, one can notice higher volatility during the contraction. This observation is confirmed by panel (b) of the figure. Here we plot the discrete-time counterpart the PK's volatility, i.e. $\tilde{m}_2(t) - \tilde{m}^2(t)$. The thick solid line indicates the unconditional second moment on this panel. It is clear, that during the contraction period the volatility of the PK increases. This indicates, that the time of contraction is a good investment opportunity, since there is a potential to find portfolios with quite high Sharpe ratios.

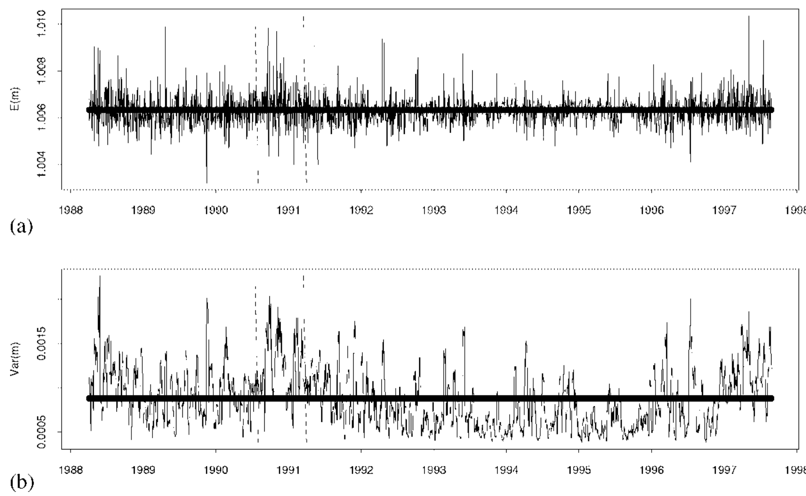

Fig. 5. The conditional expectation and variance of the pricing kernel. We plot the conditional expectation (panel (a)) and variance (panel (b)) of the PK. These moments were computed based on the reprojection procedure using the S&P 500 returns as the informations set. The dashed lines on both panels indicate the beginning and ending of the NBER dates contraction cycle. The dotted line indicates the beginning of the evaluation sample. The thick solid line shows the unconditional expectation and variance respectively.

We can also notice that the variance starts to increase in the evaluation period. It is hard to say, however, whether it is anticipation of the next contraction, or just simple out-of-sample deterioration of results.

We would also like to obtain more intuition on the reprojection procedure by establishing a rigorous link between our PK and CAPM. Namely, observe that the PK reprojected on S&P 500 is equal to $\tilde{m}(t)$ in (22) with $x(t)$ equal to the logarithmic returns on the index, i.e. $x(t) = \log(S^0(t)/S^0(t-1)) = \log(1 + r_M(t))$, where $r_M(t)$ are the simple net returns. Let us contrast the BIC chosen SNP density f_K, which is equal to VR(1)-$H(4,2,7)$, with a VR(1) density. In the VR(1) case $l(t)\,|\,x(t) \sim N(a + bx(t), \sigma)$, where a, b, σ are the parameters of the SNP density. Then (22) implies that the reprojected PK is equal to the moment generating function of the normal distribution evaluated at 1:

$$\begin{aligned}\tilde{m}(t) &= \mathrm{MGF}_{l|x}(1) = \mathrm{e}^{a+bx(t)+\sigma^3/2} = \mathrm{e}^{a+\sigma^2/2}(1 + r_M(t))^b \\ &= A(1 + br_M(t) + o(r_M^2(t))) \approx A + Br_M(t), \qquad (25)\end{aligned}$$

where notations A and B are used for $\exp(a + \sigma^2/2)$ and Ab respectively. The last equality is based on the Taylor expansion. Hence PK reprojection based on a suboptimal probability density implies CAPM.[23]

These results are similar to the findings in Bansal and Viswanathan (1993), who also nest a linear function of the index returns (CAPM) within a nonlinear model of

[23] VR(1)-$H(4,2,7)$ is preferred to VR(1) based on the likelihood ratio test as well.

the PK. They directly impose a particular form (neural net of the length 3) on the PK as a function of the T-bill yields and market returns, treated as observable factors. The critical difference in our approach is that we assume that systematic factors are not observable and obtain the PK as a complex nonlinear function of the S&P 500 returns as a result of estimation of these unobservable factors via the observable index returns.

We conclude that we do not need to observe the market portfolio to implement asset valuation. It is very important, however, to extract information from the observable data in a efficient way. For instance, one can simply use the S&P 500 (or similar) index for the estimation of the PK as long as it happens in conjunction with a realistic model of asset returns.

Another observation we can make from Table 4 is that as soon as we introduce the options data, the pricing error increases tremendously, especially in the estimation sample. Our model specification does not require the options data to identify parameters, hence it would seem that the model should perform roughly the same with or without options. Since it did not happen, this is another indication of the model misspecification. It is clear that stochastic volatility is necessary to achieve this goal, but one should consider augmentation by a jump component as well. Also, the risk premium may be too simplistic to account for variation in the options returns. Future research should take this observation into an account to improve upon the existing assets models.

6. Conclusion

In this paper we suggest an econometric procedure that allows us to estimate the unobservable pricing kernel (PK) without either any assumptions about the investors preferences or the use of the consumption data. To this end, we propose a model of the equity's price dynamics that allowed for (i) simultaneous consideration of multiple stock prices, (ii) analytical formulas for such derivatives as futures, options and bonds, and (iii) a realistic description of all of these assets. The analytical specification of the model allows us to infer the dynamics of the PK. The model, calibrated to a comprehensive dataset including the S&P 500 index, individual equities, T-bills and gold futures, yields the conditional filter of the unobservable PK. As a result we obtained the estimate of the kernel, which is positive almost surely (i.e. precludes arbitrage), consistent with the equity risk premium, the risk-free discounting and with the observed asset prices by construction.

The main advantage of our procedure is its generality. It can be applied to a class of affine diffusions (potentially augmented by jump components). The Fourier transform methods lead to pricing of a variety of derivative products (Bakshi and Madan, 2000; Duffie et al., 2000), hence derivatives can be included in the set of assets used for estimation. It is particularly important because we can estimate parameters related to the risk-neutral probability measure. The wide variety of estimation methods allow us to successfully solve the model calibration problem. Finally, the reprojection technique of Gallant and Tauchen (1998) recovers the unobservable components of the model and, in particular, the PK. The only limitation we are aware of is the computing power.

The specifics of the model considered in this paper allow for the construction of an alternative PK estimator. We call it the implied PK because it is recovered from T-bills and S&P 500 prices in a manner very similar to the implied volatility in the Black and Scholes model. Despite the theoretical inconsistencies of the procedure, such an estimate performs the best in terms of minimizing pricing errors. The best PK obtained via the reprojection procedure involves only the information in the S&P 500 index. It is, therefore, very similar to the empirical implementation of CAPM that often involves this index as a proxy for the PK. The crucial difference, however, is that our estimate is based on a highly nonlinear function of the index returns while CAPM reserves to the simple linear relationship. Therefore, the main drawback of CAPM is not the unobservability of the market portfolio, but the linear relationship with the market proxy that it used in typical empirical applications.

Acknowledgements

This paper is based on the chapter 3 of my doctoral dissertation at the Pennsylvania State University entitled “Essays in Financial Econometrics”. I am grateful to the anonymous referee and George Tauchen, the editor, for helpful comments that substantially improved the paper. I would like to thank Torben Andersen, David Bates, Charles Cao, Ian Domowitz, Phil Dybvig, Wayne Ferson, Ron Gallant, Gordon Hanka, Lars Hansen, Frank Hatheway, Dan Houser, Eric Jacquier, Mike Johannes, Jennifer Juergens, Chris Lamoureux, Dilip Madan, Harold Mulherin, Mike Piwowar, David Robinson, Tano Santos, Dennis Sheehan, Arkady Templeman and Ramzi Zein as well as participants at the seminars at Boston College, Boston University, Charles River Associates, Columbia, Penn State, Arizona, UC Irvine, Maryland, USC, Utah, Washington, Vanderbilt, Washington University in St. Louis, at the Conference on Risk Neutral and Objective Probability Distributions at Duke University, and at the AFA 2001 Meetings in New Orleans for their comments. I am indebted to Eric Ghysels for continuous encouragement, support and guidance. Scott Byrne of COMEX, Sandy Elliott of CBOE and Kelly Hering of the Federal Reserve Bank at St. Louis provided invaluable data support. All remaining errors are my own responsibility.

Appendix A. Derivatives pricing

We consider a system which is specified as in (6)–(8) under the risk-neutral measure P^*. In this appendix we drop the superscript i to avoid clutter and since it does not cause any ambiguities. In the subsequent analysis i will denote $\sqrt{-1}$.

Both Bakshi and Madan (2000) and Duffie et al. (2000) deliver the derivatives pricing methodology which is applicable in our case. Both, the price of a bond and a futures price, can be computed via a generalized characteristic function of the state-price density:

$$f(t,\tau,\phi,\varphi) = \mathrm{E}_t^{P^*}\left[\exp\left(\mathrm{i}\varphi\int_t^{t+\tau}(r_0 + r_1 u)\,\mathrm{d}u + \mathrm{i}\phi U_{1t+\tau}\right)\right]. \tag{A.1}$$

If we know $f(t,\tau,\phi,\varphi)$, then the bond price $B(t,\tau)$ is equal to $f(t,\tau,0,i)$, while the futures price of contract on U_1 with maturity in τ periods, $F(U_{1t},\tau)$, is equal to $f(t,\tau,-i,0)$.

Following the standard approach (the details are available in Chernov, 2000) we find:

$$f(t,\tau,\phi,\varphi) = \exp\{\mathrm{i}\phi U_1 + C(\tau) + A_u(\tau)U_2 + A_v(\tau)V\}, \tag{A.2}$$

where

$$A_j = \frac{2\xi_j(\mathrm{e}^{\tau/2(A_{j+}-A_{j-})} - 1)}{A_{j+}\mathrm{e}^{\tau/2(A_{j+}-A_{j-})} - A_{j-}}, \quad j = u, v \tag{A.3}$$

$$A_{j\pm} = k_j \pm \sqrt{k_j^2 - 2\xi_j} \tag{A.4}$$

$$C = \sum_{j=u,v} 2\zeta_j \log\left(\frac{A_{j+} - A_{j-}}{A_{j+}\mathrm{e}^{\tau/2(A_{j+}-A_{j-})} - A_{j-}}\right) + \zeta_j A_{j+}\tau + \mathrm{i}r_0(\phi + \varphi)\tau \tag{A.5}$$

with

$$\zeta_u = \theta, \quad \zeta_v = \eta, \quad k_u = \kappa^* - \beta\rho\mathrm{i}\phi, \quad k_v = \imath,$$
$$\xi_u = -\tfrac{1}{2}\beta^2\phi(\phi + i) + \mathrm{i}r_1(\phi + \varphi), \quad \xi_v = \tfrac{1}{2}\gamma^2\phi(\phi - i) \tag{A.6}$$

Appendix B. Details of the dataset

B.1. Details of the T-bill data

In our model the stochastic interest rate is not observable. We can identify the relevant parameters from the T-bill prices that can be obtained from T-bill yields. Specifically, FRED reports T-bill yields on a bank discount basis. The daily yields are obtained from an arithmetic average of the secondary market quotes taken from each vendor. The vendors are the ones who collect prices from dealers and inter-dealer brokers on all actively traded Treasury issues. The quotes on T-bills auctioned the previous Monday are selected. This is done because the secondary rate refers to bills that have been in the market for at least 24 h. Therefore, the Tuesday yield is computed from quotes on a 89-days-to-maturity T-bill, …, the Friday yield is computed from quotes on a 86-days-to-maturity T-bill and, finally, the Monday yield is computed from the 83-days-to-maturity T-bill. Then the cycle starts over with the 89-day bill. This information allows to infer the average quote via the bank discount formula:

$$Y = \frac{F - B}{F}\frac{360}{t}, \tag{B.1}$$

where Y is the bank discount yield reported by FRED, F is the bill face value and B is its price; t is the time to maturity.

B.2. Details of the futures data

The futures contracts have several times to maturity and hence, in principle, several series could be constructed from the data provided by COMEX. We select short maturity contracts. The shortest maturity available is 1 month, however we do not want to include the contracts with less than 7 business days to maturity. Since only institutional features affect the behavior of time to maturity, we obtain a very regularly behaved set of maturities corresponding to the prices in panel (e). Namely, maturity bounces between 7 and 28 business days increasing from the smallest to the largest by one day and then dropping back to the smallest. Also, note that the process (1)–(3) with $i = 1$ does not describe the behavior of the futures prices. It is rather a model of the commodity price behavior (gold in our case). Commodity prices are known to exhibit mean reversion (see the discussion in Schwartz (1997)). Therefore, (1) does not seem to be appropriate as a model for a commodity price. However, Schwartz (1997) finds in a model qualitatively similar to ours that all mean reversion in the gold price can be explained by mean-reverting stochastic interest rate (the parameters of a mean-reverting convenience yield are virtually zero). Hence our model seems to be appropriate for the particular data choice.

B.3. Details of the options data

The assets described in Section 3 allow us to identify the full parameter vector, however we use additional data for the model evaluation purposes. The complimentary series are gross returns on:

- 0(a): At-the-money (ATM, moneyness closest to 1.00) medium time to maturity (closest to 50 days) puts on the S&P 500 index (CBOE ticker SPX).
- 0(b): Out-of-the-money (OTM, moneyness closest to 1.06) short time to maturity (closest to 6 days) puts on the S&P 500 index.
- 3(c): At-the-money (moneyness closest to 1.00) short time to maturity (closest to 6 days) calls on "National Semiconductor" (CBOE ticker NSM).
- 3(d): Out-of-the-money (moneyness closest to 0.94) long time to maturity (closest to 100 days) calls on "National Semiconductor"[24].

By adding these data we want to capture the cross-sectional information in the options market and see how our model explains it. The four options time-series allow us to consider all types of moneyness (OTM puts are equivalent to the ITM calls in a sense of the put-call parity), different information which may be potentially contained in the put and call trading, various time-to-maturity effects.

[24] We define moneyness as an option price divided by its strike price.

B.4. Computing gross returns of derivatives

The gross returns are computed by matching the derivatives prices with the prices of the same contract on the previous day. This way, $\xi(t+\Delta t)$ in (C.1) does not stand for the terminal payoff, say $(S(t+\tau)-K)^+$, as is typically considered but for the day $t+\Delta t$ derivative, say call, price. In the case of the futures contracts, we will understand $R(t+\Delta t)$ to be $F(S_{t+\Delta t}, \tau-\Delta t)/F(S_t, \tau)$; in the case of call options it will be $C(t+\Delta t, \tau-\Delta t, K)/C(t, \tau, K)$; in the case of puts we will, similarly, have $P(t+\Delta t, \tau-\Delta t, K)/P(t, \tau, K)$. Unfortunately, our dataset does not allow us to construct similar series for the bond prices. Ideally, we would have to consider $B(t+\Delta t, \tau-\Delta t)/B(t, \tau)$ as $R(t+\Delta t)$. However, when $\tau-\Delta t$ is equal to 89 days (Tuesday quote), we cannot find the quote on the previous day, because the Monday quote is based on the 83-days-to-maturity T-bill (see Section 3 for details). Hence we have to omit the T-bills from our HJ bound evaluations.

Appendix C. Hansen and Jagannathan methodology

This appendix briefly discusses the HJ approach to the evaluation of the PK.

The methodology was developed in a discrete time setting because it is nonparametric in its nature and we observe data only at discrete time intervals. Hence, we start with the discrete time analogue of the asset pricing formula (12):

$$\pi(t) = \mathrm{E}_t^P(m(t+\Delta t, \Theta)\xi(t+\Delta t)), \tag{C.1}$$

where we denote the discrete values of the PK by $m(t, \Theta)$ and emphasize its dependence on the model parameters vector Θ. We can rewrite this equation in the returns form:

$$1 = \mathrm{E}_t^P(m(t+\Delta t, \Theta)R(t+\Delta t)). \tag{C.2}$$

This form of the asset pricing equation immediately yields itself to the estimation of Θ via GMM because returns $R(t+\Delta t)$ are stationary and population moment conditions are readily available. The GMM estimate is equal to:

$$\hat{\Theta} = \operatorname{argmin} T \cdot g_T^\top(\Theta) W_T^{-1} g_T(\Theta), \tag{C.3}$$

where T is the sample size and $g_T(\Theta) = (1/T)\sum_t (1^\top - m(t+\Delta t, \Theta)R^\top(t+\Delta t))$ is the sample moment based on a vector of returns. The optimal weighting matrix W_T is equal to the consistent estimator of the asymptotic covariance of the pricing errors.

Hansen and Jagannathan (1997) show, that if we define δ to be the distance from the candidate PK m to the set of admissible PKs $\mathscr{A}$:

$$\delta^2 = \min_{a \in \mathscr{A}} \mathrm{E}(m-a)^2 \tag{C.4}$$

then the δ is equal to the minimand in (C.3) with $W_T = (1/T)\sum_t R^\top(t)R(t)$. The fact that W_T is not the optimal GMM weighting matrix implies that δ does not reward PKs with higher sampling errors and, therefore, it can serve as a measure of comparison of different PKs based on the same set of asset returns.

On the other hand, if we obtain the parameters estimate $\hat{\Theta}$ via (C.3), the distribution of the test statistic based on δ is no longer χ^2. It is rather a sum of χ_1^2 distributions weighted by the positive eigenvalues of a transformation of the score of the pricing errors (see Jagannathan and Wang, 1996, for details). However, in our application, $\hat{\Theta}$ is obtained via minimization of a different function (see the moment conditions in Section 4.1). Hence, we can evaluate δ's corresponding to different candidate PKs only informally by computing the minimand in (C.3) and comparing the obtained numerical values.

Having estimated Θ via (C.3) based on the available data, one may want to see if the obtained PK is admissible. Hansen and Jagannathan (1991) provide insights into this approach. If we project m on the space spanned by the returns of the selected assets and the constant 1:

$$m = \tilde{R}^\top \beta + u, \quad \tilde{R}^\top = (1 \ R^\top) \tag{C.5}$$

(i.e. perform regression of m on R), then (C.2) and the fact that u is orthogonal to R and has a non-negative variance allows us to impose a lower bound on the second moment of m:

$$\mathrm{E}(m^2) \geqslant (\mathrm{E}(m) \ \imath^\top)[\mathrm{E}(\tilde{R}\tilde{R}^\top)]^{-1}(E(m) \ \imath^\top)^\top, \tag{C.6}$$

where $\imath$ is a vector of 1's the length of which is conformable with the one of R.

This bound can be used as the informal check of the particular model of interest. One can always compute $\mathrm{E}(m)$ and $\mathrm{E}(m^2)$ given the model and its estimated parameters. If this pair violates (C.6), then a researcher may make some conclusion about the inadequacy of the modeled PK. Burnside (1994) and Cecchetti et al. (1994) note that, since all the objects involved in the bound evaluation ($\mathrm{E}(m), \mathrm{E}(m^2)$ and the bound itself) are random, this informal procedure may be affected by the sample variation. Hence a formal statistical test is required to assess the significance of the bound violation. We adopt the tests described in Burnside (1994) to evaluate our PK.

Burnside (1994) discusses and evaluates several tests. Based on his simulation evidence, the optimal distance (OD) test seems to be the most robust in terms of the relationship between the small sample and asymptotic properties. So we choose this test as our evaluation tool.

The OD test computes the shortest distance, ζ, from the model implied point ($\mathrm{E}(m)$, $\mathrm{E}(m^2)$) to the volatility bound in (C.6) under $H_0 : \zeta = 0$. One starts out by computing

$$\hat{\zeta} = \frac{1}{T}\sum_t m(t)^2 - \left(\frac{1}{T}\sum_t m(t) \ \imath^\top\right)\left[\frac{1}{T}\sum_t \tilde{R}\tilde{R}^\top\right]^{-1}\left(\frac{1}{T}\sum_t m(t) \ \imath^\top\right)^\top. \tag{C.7}$$

If $\hat{\zeta}$ is negative, one has to perform the second step. Namely, form the following moment conditions:

$$\mathrm{E}\{m - v\} = 0, \tag{C.8}$$

$$\mathrm{E}\left\{m^2 - (v \ \imath^\top)\left[\frac{1}{T}\sum_t \tilde{R}\tilde{R}^\top\right]^{-1}(v \ \imath^\top)^\top\right\} = 0. \tag{C.9}$$

These moment conditions exploit the H_0 and estimate the kernel mean, v that corresponds to the shortest distance between the objects being compared. Then,

$$\text{OD} = \begin{cases} 0 & \text{if } \hat{\zeta} \geqslant 0 \\ J & \text{if } \hat{\zeta} < 0 \end{cases} \tag{C.10}$$

where J is the Hansen (1982) J-statistic (the optimal value of the objective function) based on the above moment conditions. OD is distributed χ_1^2 with probability 0.5 under the H_0.

Appendix D. Approximate maximum likelihood estimation of the model for the S&P 500 and T-bills

As a robustness check of the estimation method, we perform an approximate maximum likelihood (ML) estimation of the bivariate sub-model of (1)–(3), (5) using the S&P 500 and T-bills data. Duffie et al. (2002) propose a very simple and accurate approximation to the likelihood for the special case of affine diffusions, which our model belongs to. In our context this approximation works as follows. We observe the level of S&P 500, S^0, and the T-bill yield, Y. The state vector consists of U_1^0, and U_2. Observables are related to the factors via the following relations:

$$U_1^0(t) = \log S^0(t), \tag{D.1}$$

$$\begin{aligned} U_2(t) &= \frac{1}{A_u(\tau)} \log B(t,\tau) - \frac{C(\tau)}{A_u(\tau)} \\ &= \frac{1}{A_u(\tau)} \log \left[F\left(1 - Y(t,\tau)\frac{\tau}{360}\right)\right] - \frac{C(\tau)}{A_u(\tau)}, \end{aligned} \tag{D.2}$$

where the last equation is obtained from (23) and (B.1). The absolute value of the determinant of the Jacobian of the inverse relationship is equal to

$$D(t) = \frac{360}{\tau}\frac{A_u(\tau)}{F}\exp(C(\tau) + A_u(\tau)U_2(t)). \tag{D.3}$$

Therefore, the likelihood function of the data S^0, Y takes form:

$$p_\Theta(S^0, Y) = \prod_{t=1}^{T} p_\Theta(U_1^0(t), U_2(t) | U_1^0(t-1), U_2(t-1))\,\frac{1}{D(t)}. \tag{D.4}$$

Notice, that

$$\begin{aligned} & p_\Theta(U_1^0(t+\delta t), U_2(t+\delta t) | U_1^0(t), U_2(t)) \\ &\quad = p_\Theta(U_2(t+\delta t) | U_1^0(t), U_2(t))\, p_\Theta(U_1^0(t+\delta t) | U_1^0(t), U_2(t), U_2(t+\delta t)) \\ &\quad = p_\Theta(U_2(t+\delta t) | U_2(t)) \\ &\qquad \times \mathrm{E}[p_\Theta(U_1^0(t+\delta t) | U_1^0(t), U_2(t), \{U_2(s),\ s \in [t, t+\delta t]\}) | \\ &\qquad\quad U_1^0(t), U_2(t), U_2(t+\delta t)], \end{aligned} \tag{D.5}$$

Table 5
Approximate maximum likelihood estimation results We calibrate the assets dynamics model

$$\frac{dS^0(t)}{S^0(t)} = (r_0 + \alpha^0 U_2(t))\,dt + \beta^0\sqrt{1-\rho^2}\sqrt{U_2(t)}\,dW_1(t) + \beta^0\rho\sqrt{U_2(t)}\,dW_2(t)$$

$$dU_2(t) = (\theta - \kappa U_2(t))\,dt + \sqrt{U_2(t)}\,dW_2(t)$$

$$\alpha^0 = r_1 + (\lambda_1\sqrt{1-\rho^2} + \lambda_2\rho)\beta^0$$

to the returns on the S&P 500 index and the T-bill bank discount rate via the approximate maximum likelihood. Standard errors are reported in parentheses.

θ	κ	r_0	r_1	λ_1	λ_2	ρ	β^0
0.5158	1.7466	0.0252	0.1566	0.5580	−0.9091	−0.4816	0.5441
(0.21)	(0.73)	(0.01)	(0.06)	(0.23)	(0.38)	(0.20)	(0.22)

where the last equality is due to the special structure of our model, i.e. U_1^0 does not feedback into U_2. The density in the first term is non-central chi-squared, i.e. it is known in analytic form. The density in the second term is Gaussian and is approximated via a Gaussian density that does not depend on the entire path of U_2 between t and $t+\delta t$ and has mean and variance:

$$m_{\delta t} = r_0\delta t + U_1^0(t) + \frac{\alpha}{2}\,\delta t(U_2(t) + U_2(t+\delta t)), \tag{D.6}$$

$$V_{\delta t} = \frac{{\beta^0}^2}{2}\,\delta t(U_2(t) + U_2(t+\delta t)). \tag{D.7}$$

This is the first order approximation that is shown to be very accurate in Duffie et al. (2002). Given this approximation, the expectation in the last term of (D.5) can be computed analytically.

From, this point the implementation of ML is straightforward. The Table 5 provides the estimation results. Most of the estimated parameters are in line with the SMM parameters, and most standard errors are smaller as expected.

All parameters which are substantially different are related to the level of the volatility. The long-run volatility value, $\beta^0\sqrt{\theta/\kappa}$, is equal to 29% according to the MLE and to 17% according to SMM, which is more reasonable. It appears that the full dataset contains more information about volatility. This conjecture is supported by smaller SMM standard errors. Statistically, none of the parameters are significantly different from each other.

References

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial prices. Journal of Finance 53, 499–548.

Andersen, T., Sørensen, B., 1996. GMM estimation of a stochastic volatility model: a Monte Carlo study. Journal of Business and Economic Statistics 14, 328–352.

Bakshi, G., Madan, D., 2000. Spanning and derivative-security valuation. Journal of Financial Economics 55, 205–238.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: a new approach. Journal of Finance 48, 1231–1262.

Black, F., Scholes, M.S., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Burnside, C., 1994. Hansen-Jagannathan bounds as classical tests of asset-pricing models. Journal of Business and Economic Statistics 12, 57–79.

Campbell, J., Cochrane, J., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Cecchetti, S., Lam, P.-S., Mark, N., 1994. Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns. Journal of Finance 49, 123–152.

Chacko, G., Viceira, L., 2003. Spectral GMM estimation of continuous-time processes. Journal of Econometrics, this issue.

Chernov, M., 2000. Essays in financial econometrics. Ph.D. Thesis, Pennsylvania State University.

Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. Journal of Financial Economics 56, 407–458.

Chernov, M., Gallant, R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. Journal of Econometrics, this issue

Den Haan, W., Levin, A., 1997. A practitioner's guide to robust covariance matrix estimation. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 15. Elsevier Science BV, Amsterdam.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and option pricing for affine jump-diffusions. Econometrica 68, 1343–1376.

Duffie, D., Pedersen, L.H., Singleton, K., 2002. Modeling sovereign yield spreads: a case study of Russian debt. Journal of Finance, Forthcoming.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in equity index voltality and returns. Journal of Finance, forthcoming.

Gallant, A.R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. Econometrica 57, 1091–1120.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gallant, R., Tauchen, G., 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of American Statistical Association 93, 10–24.

Garman, M., 1976. A general theory of asset valuation under diffusion state processes. Working Paper No. 50, University of California, Berkeley.

Hall, A., 2000. Covariance matrix estimation and the power of the overidentifying restrictions test. Econometrica 68, 1517–1527.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. Journal of Political Economy 99, 225–262.

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

Hansen, L.P., Singleton, K., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269–1286.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Harrison, M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11, 215–260.

Harvey, C., Whaley, R., 1991. S&P 100 index option volatility. Journal of Finance 46, 1551–1561.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

Ho, M., Perraudin, W., Sørensen, B., 1996. A continuous-time arbitrage-pricing model with stochastic volatility and jumps. Journal of Business and Economic Statistics 14, 31–43.

Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51, 1611–1632.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Jiang, G., Knight, J., 2002. Estimation of continuous-time processes via the empirical characteristic function. Journal of Business and Economics Statistics 20, 198–212.

Kloeden, P.E., Platen, E., 1995. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.

Longstaff, F., 1989. Temporal aggregation and the continuous-time capital asset pricing model. Journal of Finance 44, 871–887.

Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145–161.

Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.

Pearson, N., Sun, T.-S., 1994. Exploiting the conditional density in estimating the term structure: an application to the Cox, Ingersoll, and Ross model. Journal of Finance 49, 1279–1304.

Roll, R., 1977. A critique of the asset pricing theory's tests, part I: on past and potential testability of the theory. Journal of Financial Economics 4, 129–176.

Rosenberg, J., Engle, R., 2002. Empirical pricing kernels. Journal of Financial Economics 64, 341–372.

Schwartz, E., 1997. The stochastic behaviour of commodity prices: implications for valuation and hedging. Journal of Finance 52, 923–973.

Schwert, G.W., 1990. Stock volatility and the crash of '87. Review of Financial Studies 3, 77–102.

Singleton, K., 2001. Estimation of affine asset pricing models using the empirical characteristic function. Journal of Econometrics 102, 111–141.