Economics Letters 235 (2024) 111531

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Econometric identification of the attainable maximal sharpe ratio by optimal shrinkage of the cross-section of asset returns[☆,☆☆]

Yuting Chen [a,b,c], Valerio Potì [a,d,*]

[a] College of Business, University College Dublin, Blackrock, Dublin, Ireland
[b] ESC Rennes School of Business, Rennes, France
[c] Central Bank of Ireland, Dublin, Ireland
[d] Dipartimento di Economia & Management, Università degli Studi A. Moro, Bari, Italy

ARTICLE INFO

JEL classification:
G1
C5
C14

Keywords:
Regularization
Maximal Sharpe ratio
Pricing kernel volatility
Polynomial pricing kernels

ABSTRACT

In this paper, we propose using a search for the optimal regularization of GMM as a way to identify the attainable maximal Sharpe ratio in a given investment opportunity set (e.g., the economy). Regularization is achieved by imposing a bound on the volatility of a flexible specification of the candidate pricing kernel, alongside other economically motivated restrictions. In an empirical application of this methodology to US equities, our estimates of the maximal attainable Sharpe ratio in the economy are between 21 and 35 percent annually, depending on the cross-validation criterion used in the search, thus in the low region of the range of values hitherto considered, either on theoretical or on empirical grounds, by the literature.

1. Introduction and overview

There is an extensive literature that considers versions of CAPM where the inter-temporal marginal rate of substitution (IMRS) of a rational representative investor is approximated as a polynomial of the market portfolio return. This approach has enjoyed considerable empirical success (e.g., Harvey and Siddique (2000)) due to the flexibility of the pricing kernel specification. It leaves the door open, however, to the possibility of over-fitting the cross-section of asset returns, as noted by Dittmar (2002), Post et al. (2008), Poti and Wang (2010).

Poti and Wang (2010), in testing the 3 and 4-moment CAPMs, reduce the risk of over-fitting by estimating the pricing kernel under a positivity constraint and an upper bound on its volatility. The positivity constraint is to rule out arbitrage. The volatility bound is to rule out so-called "good deals" (Cochrane and Saa-Requejo, 2000). More recently, Kozak et al. (2020) follow the same approach, applying it in a more general setting to regularize the estimation of general high-dimensional candidate pricing kernel models. Neely (2022) creatively uses the pricing kernel volatility bound and its implications for asset return predictability to regularize impulse response functions that describe the transmission of monetary policy.

In this study, we follow a similar line of reasoning as Poti and Wang (2010) and Kozak et al. (2020), but we turn it on its head. Rather than imposing a volatility bound specified ex-ante on *a-priory* grounds (theory and prior evidence), we estimate it. We do so by treating it as a regularization parameter in a Ridge-like regression, and selecting it in a way that maximizes the (pseudo-)out-of-sample fit of the cross-section of asset returns.

In what follows, we first outline our proposed methodology. We then describe the data we use in an empirical application. The section with the empirical results follows. The final section offers concluding remarks and a road-map for future research.

2. Methodology

In this section, we first outline the econometric specification of the candidate pricing kernel and the estimation method, including the regularization achieved by imposing sign restrictions and a volatility upper bound on the candidate kernel. We then explain how we use this econometric specification to make inference on the volatility of the (minimum variance) kernel that prices the assets in the investment opportunity set.

☆ Both authors gratefully acknowledge funding support from the M. Smurfit School of Business at University College Dublin. Yuting Chen gratefully acknowledges funding support from ESC Rennes School of Business and the UCD China Scholarship Scheme (CSC).
☆☆ The views in the paper are the authors' only and not of the Central Bank of Ireland.
* Corresponding author.
E-mail address: valerio.poti@ucd.ie (V. Potì).

https://doi.org/10.1016/j.econlet.2024.111531
Received 6 November 2023; Received in revised form 17 December 2023; Accepted 6 January 2024
Available online 18 January 2024

Table 1
Optimal volatility bounds from cross-validation.

R-squared	No Arbitrage		Arbitrage Allowed	
Model	Sign constraints	No sign constraints	Sign constraints	No sign constraints
1	0.23	0.23	0.23	0.23
2	0.23	0.23	0.23	0.23
3	0.23	0.23	0.23	0.23
4	0.23	0.23	0.23	0.23
5	0.23	0.23	0.23	0.23
6	0.21	0.23	0.21	0.23

HJ-distance	No Arbitrage		Arbitrage Allowed	
Model	Sign constraints	No sign constraints	Sign constraints	No sign constraints
1	0.35	0.29	0.35	0.27
2	0.35	0.35	0.35	0.35
3	0.27	0.25	0.27	0.25
4	0.35	0.27	0.35	0.27
5	0.19	0.27	0.19	0.27
6	0.09	0.25	0.09	0.25

Notes: This table reports the optimal volatility bounds on the pricing kernel estimated using GMM and cross-validation. The specifications from model 1 to model 6 are: Linear without HC (model 1), Quadratic without HC (model 2), Cubic without HC (model 3), Linear with HC (model 4), Quadratic with HC (model 5), and Cubic with HC (model 6). The criteria are R-squared and Hansen and Jagannathan distance. Cross-validation is conducted by slicing the whole sample (from January 1965 to July 2023) into 5 equal-length folders, estimating the pricing kernel using 4 folders, and testing it using the remaining folder. The basis assets are the 17 industry-sorted portfolios. The reported optimal volatility bound is calculated as the annualized standard deviation bound on the estimated pricing kernel.

2.1. Econometric specification

With n basis test asset payoffs x_{t+1}[1] the candidate pricing kernel can be estimated by solving the following minimization problem:

$$\begin{aligned} &\min_{m} \quad g_t' W_{n\times n} g_t \\ &s.t. \quad m_{t+1} > 0, \quad \sigma^2(m_{t+1}) \leq A \\ &g_t = E_t(m_{t+1}x_{t+1}) - p_t(x_{t+1}) \end{aligned} \tag{1}$$

Here, g_t is a vector of pricing errors, while W is the moment condition weighting matrix which, following Dittmar (2002) and Poti and Wang (2010), we specify as the second moment matrix proposed by Hansen and Jagannathan (1997). In our empirical application, x_{t+1} are returns on the tests assets and thus $p_t(x_{t+1})$ is a vector of ones.

In (1), the positivity restriction, $m_{t+1} > 0$, enforces no-arbitrage whereas the unconditional variance bound, $\sigma^2(m_{t+1}) \leq A$, enforces a no-good deal bound, as in Ross (2009), Poti and Wang (2010) and, more recently, Kozak et al. (2020).

Due to the duality between volatility of the (minimum variance) kernel that prices the assets in a given investment opportunity set and the attainable maximal Sharpe ratio (Hansen and Jagannathan, 1991), the variance bound represents a bound on the maximal squared Sharpe ratio attainable from forming portfolios of these assets (Ross, 2009).

We specify the kernel in (1), m_{t+1}, as a flexible polynomial of the return on a diversified portfolio of traded assets and of our measure of return on human capital (HC):

$$\begin{aligned} m_{t+1} = a_t + &b_{vw,1} R_{vw,t+1} + b_{vw,2} R^2_{vw,t+1} + b_{vw,3} R^3_{vw,t+1} + \\ &b_{l,1} R_{l,t+1} + b_{l,2} R^2_{l,t+1} + b_{l,3} R^3_{l,t+1} \end{aligned} \tag{2}$$

The variance bound plays the same role as the regularization parameter in Ridge-like or LASSO-type regressions (Zboňáková et al., 2019, Chernozhukov et al., 2021), where the regularization restricts the magnitude of the regression coefficients by restricting their L^1 (in the case of LASSO) or L^2 (in the case of Ridge regressions) norm. In our setting, the variance bound can be seen as restricting in the appropriate norm the magnitude of the parameters of the candidate pricing kernel, hence of the parameters of (2).

[1] We include the risk-free asset in x_{t+1}.

2.2. Estimation and inference on pricing kernel volatility

We make inference on the volatility of the (minimum variance) pricing kernel by searching for an optimal volatility bound (the square rot of the value of A in (1)) that yields the best out-of-sample pricing fit of the estimated kernel (i.e., m_{t+1} in (2)).

To this end, we repeatedly estimate m_{t+1} in (2) and test its empirical admissibility by evaluating its out-of-sample fit by cross-validation. We do so $N = 100$ times, to which we refer as iterations, each time with a different choice of A from within a range of possible values.[2] In each iteration, cross-validation is conducted by slicing the whole sample (from January 1965 to July 2023) into 5 equal-length folders, estimating the pricing kernel using 4 folders at a time for the given choice of A, and testing it using the remaining folder. In each iteration, we repeat this procedure five times, using each time four different folders for estimation and a different folder for testing. Each time, within each iteration, the parameters of (2) are estimated using data from the appropriate four folders by solving the minimization problem in (1) for the given choice of A. The testing criteria are the R-squared and the Hansen and Jagannathan distance. Thus, we estimate 500 (i.e., $N \times 5 = 100 \times 5$) times the kernel[3] and record each time, hence for each of the N values of the variance bound, such a pair of cross-validation metrics. The optimal volatility bound is then the square root of the value of A for which the R-squared is maximized or the Hansen and Jagannathan distance is minimized.

The difference between our approach, as described above, and the approach of Poti and Wang (2010) and Kozak et al. (2020) is that these authors set the pricing kernel volatility bound *a-priori*, based on theoretical arguments and prior empirical literature about the curvature of economic agents' utility functions, whereas we infer the bound by finding the value that best fits the cross-section of asset

[2] In the present application, the range of possible values is from 0.01% to 100%. The range of values is obtained by computing the annualized variance corresponding to a bound expressed in terms of standard deviation of the kernel. The range of values for this standard deviation bound is from 1% to 100% with increments of 1% from one iteration to the next, for a total of 100 such iterations. The variance bound, A, is computed by multiplying by 12 the square of this standard deviation bound

[3] That is, the parametric specification for m_{t+1} in (2).

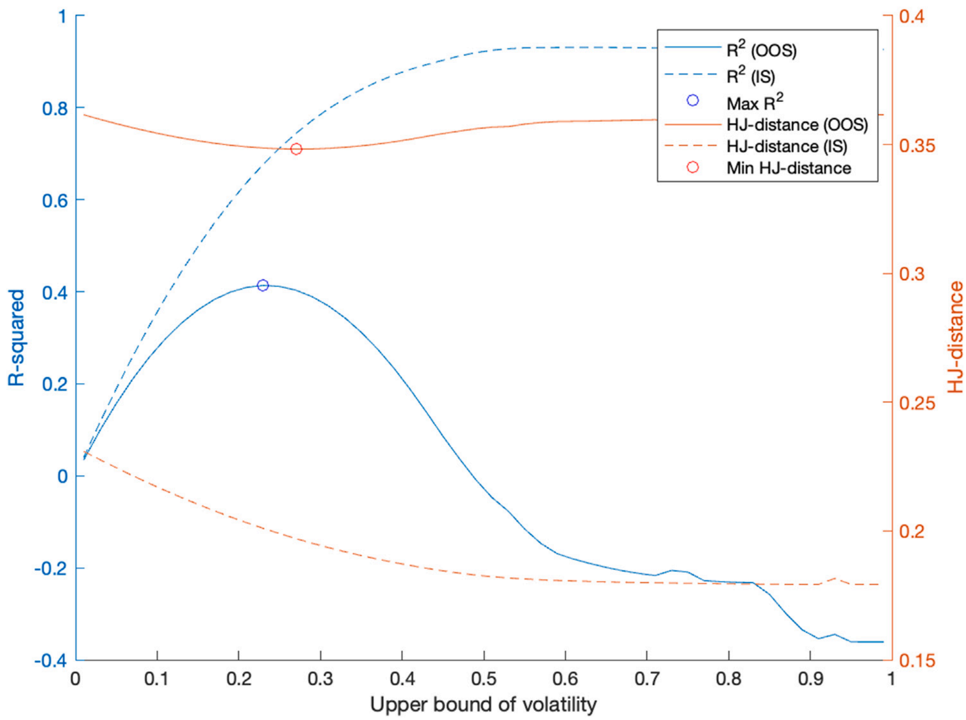

Fig. 1. This figure plots the performance of the estimated pricing kernel in cross-validation under different volatility bounds. The pricing kernel specification used here is a cubic model without human capital, with positivity and volatility constraints on the kernel and sign constraints (as in Poti and Wang (2010)) on its coefficients. The criteria are the R-squared and Hansen and Jagannathan distance. Cross-validation is conducted by slicing the whole sample (from January 1965 to July 2023) into 5 equal-length folders, estimating the pricing kernel using 4 folders, and testing it using the remaining folder. The basis assets are the 17 industry-sorted portfolios. On the horizontal axis is the upper bound of the pricing kernel's volatility, measured as the annualized standard deviation. The blue lines and the red lines represent the R-squared and HJ-distance, respectively, with the dashed lines representing their in-sample (IS) versions and the solid lines representing the out-of-sample (OOS) counter-parties. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

returns in cross-validation tests. That is, whereas Poti and Wang (2010) and Kozak et al. (2020) specify A in the variance restriction in (1), we estimate it by searching for the value of A that maximizes a criterion of cross-validation fit. For a sufficiently wide investment opportunity set, including our choice of test assets, the restriction is binding. The corresponding volatility bound, then, coincides with the volatility of the estimated pricing kernel and provides an estimate of the attainable maximal Sharpe ratio in the economy, a fundamental economic quantity and key parameter in many asset pricing models.[4]

It should be noted that, while (2) is inspired by the higher moment CAPM (i.e., the 3 and 4-moment CAPM), due to the empirical success of these specifications (Harvey and Siddique, 2000, Dittmar, 2002, Anghel et al., 2023), it is not intended as the IMRS implied by any of these asset pricing models, but rather as a specification flexible enough to stand a good chance to fit the cross-section of asset returns. Thus, (1) is not intended to yield estimates of versions of the higher moment CAPM. It is intended instead to provide estimates of a reduced form of the true yet unspecified model of the pricing kernel.

3. Data

Our sample period is from January 1965 to July 2023, for a total of 703 monthly observations. We use the CRSP value-weighted portfolio return to proxy for $R_{vw,t+1}$, NIPA data to construct a proxy for $R_{l,t+1}$, and the 17 portfolios of CRSP industry-sorted stocks kindly provided by Kenneth French on his website. The 1-month US Government Treasury Bill is used as a proxy for the risk-less asset.

[4] If the test assets payoffs span a payoff space that approximates sufficiently well the payoff space of the investment opportunity set in the economy

4. Optimal volatility bounds

We report in Table 1 the optimal volatility bounds estimated using the approach described in Section 2. As noted, (2) is intended as a flexible reduced form of the true yet unspecified model of the pricing kernel. The estimated reduced form model, however, might well represent a good approximation for the higher moment CAPM. In this case, imposing appropriate sign restrictions on the coefficients of (2) should not deteriorate the out-of-sample pricing performance of the estimated kernel. Therefore, in Table 1, we report the optimal volatility bound with and without restrictions on the signs of the coefficients of (2) consistent with the 3 and 4-moment CAPM, as in Poti and Wang (2010).

As shown in Table 1, with nearly all of the model specifications, the optimal volatility upper bound that maximizes the out-of-sample R-squared is 23%. When we use as the criterion the HJ-distance, the optimal bound ranges between 9% and 35% and is centered around 27%. The optimal bound is never slack. Thus, it coincides with the volatility of the estimated kernel. Therefore, if we assume that our n test asset payoffs (i.e., the 17 industry portfolios in our application) approximate well the span of the investment opportunity set in the economy, then we can interpret the optimal bound as an estimate of the maximal Sharpe ratio attainable in the economy. In this case, the results imply that the volatility upper bounds (implied by upper bounds to relative risk aversion) often considered by the literature (e.g., Ross (2009)) are too high.

When the search criterion is the Hansen and Jagannathan (1997) distance, the optimal pricing kernel volatility bound and, therefore, the estimated maximal attainable Sharpe ratio are somewhat less stable. With the exception of very flexible specifications (for which we possibly experienced estimation convergence problems), they are also somewhat higher. Nonetheless, the estimate remains in the low region of the range of values considered, either on theoretical or on empirical grounds, by the literature (e.g., Cochrane and Saa-Requejo (2000), Cochrane (2005), Ross (2009)).

Fig. 1 visualizes the out-of-sample pricing performance of the estimated pricing kernels under different volatility bounds, for the cubic model without human capital, with the positive constraint on the pricing kernel and the sign constraints on the moment coefficients, which we treat as the benchmark case. The in-sample R-squared (HJ-distance) increases (decreases) with the volatility bound, while the out-of-sample R-squared (HJ-distance) increases (decreases) until a maximum (minimum) point. This implies that, when we impose an upper bound to the volatility of the pricing kernel, the in-sample pricing performance deteriorates, but the out-of-sample performance improves. This is typical of successful instances of regularization of predictive models. There is a critical point, below which the out-of-sample performance starts dropping. According to our cross-validation tests, the critical point is around 25%, which can be seen as an estimate of the true volatility upper bound and thus as an estimate of the maximal Sharpe ratio attainable in the economy (Poti and Wang, 2010).

Tests with different basis assets, including the 30 industry portfolios and the momentum portfolios, and different number of folders for cross-validation, all indicate a consistent pattern.

The difference in the values of the estimated pricing kernel volatility depending on the cross-validation criterion is not surprising. This is because the pricing kernel volatility that attains the highest (out-of-sample) R-squared is the volatility of the (least volatile) pricing kernel that best prices (out-of-sample) the average asset. The pricing kernel volatility that attains the lowest (out-of-sample) Hansen and Jagannathan distance, instead, is the volatility of the (least volatile) pricing kernel that best prices (out-of-sample) the asset that is most difficult to price (i.e., the pricing kernel with the smallest pricing error for the most mispriced asset). Thus, we should believe that the optimal volatility bound under the Hansen and Jagannathan criterion provides an estimate of the economy maximal Sharpe ratio if we believe that the economy pricing kernel should price well even the most mis-priced assets, rather than admitting the possibility that such assets are priced in a segregated segment of the capital market, hence that markets are imperfectly integrated.

5. Conclusion

We proposed using a search for the optimal regularization of GMM as a way to identify the attainable maximal Sharpe ratio from a given investment opportunity set (e.g., the economy), obtaining estimates of this quantity that are considerably lower than often hitherto conjectured in the asset pricing literature.

In future work, we will experiment with even more flexible conditional specifications of the kernel, explicitly allowing for time-varying prices of the factor risks by allowing for conditional time-variation of the parameters of the pricing kernel in (2), at the same time expanding the test asset payoff space with portfolios of the basis assets traded on the basis of conditioning information carried by predictive variables.

Data availability

Data will be made available on request.

References

Anghel, D.G., Caraiani, P., Rosu, A., Rosu, I., 2023. Asset pricing with systematic skewness: Two decades later. Crit. Finance Rev. 12 (1–4), 309–354.

Chernozhukov, V., Hardle, W.K., Huang, C., Wang, W., 2021. LASSO-driven inference in time and space. Ann. Statist. 49.

Cochrane, J., 2005. Asset Pricing. Princeton University Press, Princeton.

Cochrane, J.H., Saa-Requejo, J., 2000. Beyond arbitrage: Good-deal asset price bounds in incomplete markets. J. Polit. Econ. 108, 79–119.

Dittmar, R.F., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. J. Finance 57 (1), 369–403.

Hansen, L., Jagannathan, R., 1991. Restrictions on intertemporal marginal rates of substitution implied by asset returns. J. Polit. Econ. 99 (2), 225.

Hansen, L.P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. J. Finance 52 (2), 557–590.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. J. Finance 55 (3), 1263–1295.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. J. Financ. Econ. 135 (2), 271–292.

Neely, C.J., 2022. How persistent are unconventional monetary policy effects? J. Int. Money Finance 126, 102653.

Post, T., Van Vliet, P., Levy, H., 2008. Risk aversion and skewness preference. J. Bank. Financ. 32 (7), 1178–1187.

Poti, V., Wang, D., 2010. The coskewness puzzle. J. Bank. Financ. 34 (8), 1827–1838.

Ross, S.A., 2009. Neoclassical Finance. Princeton University Press.

Zboňáková, L., Monti, R.P., Hardle, W.K., 2019. Towards the interpretation of time-varying regularization parameters in streaming penalized regression models. Pattern Recognit. Lett. 125, 542–548.