A Rehabilitation of Stochastic Discount Factor Methodology

John H. Cochrane[1]

October 9, 2001

[1]University of Chicago and NBER. 1101 E. 58th St. Chicago IL 60637. 773 702 3059. john.cochrane@gsb.uchicago.edu. Data, programs and updated drafts of this paper are available at http://gsbwww.uchicago.edu/fac/john.cochrane/research/Papers/ This research is supported by a grant from the National Science Foundation administered by the NBER, and by the University of Chicago Graduate School of Business. I thank two anonymous referees for helpful comments.

Abstract

In a recent *Journal of Finance* article, Kan and Zhou (1999) find that the "Stochastic discount factor" methodology using GMM is markedly inferior to traditional maximum likelyhood even in a simple test of the static CAPM with i.i.d. normal returns. This result has gained wide attention. However, as Jagannathan and Wang (2001) point out, this result flows from a strange assumption: Kan and Zhou allow the ML estimate to know the mean market return ex-ante. I show how this information advantage explains Kan and Zhou's results. In fact, when treated symmetrically, the discount factor - GMM and traditional methodologies behave almost identically in linear i.i.d. environments.

1 Introduction

Kan and Zhou (1999) compare the "stochastic discount factor" (SDF) methodology, using GMM, to a "traditional" maximum likelihood estimate and test, applied to the static linear CAPM with i.i.d. normal returns. They conclude that the SDF methodology performs much worse than the traditional estimate and test even in this very simple environment. They summarize their results as follows

> "The accuracy of the [SDF] parameter estimation can be poor: the standard error of the estimated risk premium is often more than 40 times greater than that of the traditional methodologies... The SDF methodology is not very reliable in detecting even gross misspecifications in an asset pricing model." (p.1222)

Kan and Zhou's result has attracted a great deal of attention, in part because it is so unexpected. GMM is usually well-behaved in linear models with i.i.d. normal variables. It often reduces exactly to standard procedures (for example, OLS regression) in those environments. The stochastic discount factor expression of an asset pricing model is mathematically equivalent to its expected return - beta expression, so that part of the methodology cannot make any difference. When paired with GMM, and using the pricing errors as moments, the SDF/GMM methodology prescribes a cross-sectional regression (see Cochrane 1996, 2001). The right hand variable in the cross-sectional regression is the covariance or second moment of returns with factors, rather than the traditional betas, and the regression uses a slightly different weighting matrix, but one would not expect these minor differences to amount to much. True, when the factor is a return, ML prescribes a time-series rather than a cross-sectional regression approach, gaining one degree of freedom. But we have run cross-sectional regressions for over 30 years without any hint of gross inefficiency. In fact many authors prefer robust but even more inefficient OLS cross-sectional regressions, as in the Fama-MacBeth procedure, rather than ML's GLS cross-sectional regressions.

Kan and Zhou's results stem from a strange assumption, as pointed out by Jagannathan and Wang (2001). They allow the ML procedure to know the factor risk premium – the mean market return, in the case of the CAPM – while the GMM/SDF procedure must, as usual, estimate it. This note explains how that assumption produces their results.

If one treats the two methods symmetrically, the GMM/SDF methodology performs almost identically to traditional ML-based time-series and cross-sectional regressions when applied to standard setups, featuring linear models, and i.i.d. returns and factors. Jagannathan and Wang (2001) show this right answer by analytical examination of asymptotic distributions, and Cochrane (2001) presents a Monte Carlo analysis. Gratifyingly, all the above suppositions turn out to be correct.

The interesting question remains to be answered: How do maximum likelihood and SDF/GMM compare in highly challenging environments, with non-normal distributions, time-varying betas and time-varying factor risk premia? Maximum likelihood has well-known optimality properties when you know the exact data-generating model, but the relevant comparison is between maximum likelihood with the wrong model and GMM with potentially inefficient moments. This, interesting, comparison has not yet been investigated in an asset pricing context.

2 Kan and Zhou's result

Kan and Zhou run a Monte Carlo simulation of a test of the CAPM on size decile portfolios. They express the asset pricing model in expected return - beta form $E(R^e) = \beta\lambda$ where R^e denotes a vector of N excess returns and β denotes a conformable vector of regression coefficients of returns on the factor f. Table I collects Kan and Zhou's reported sampling variation in the estimated factor risk premium $\hat{\lambda}$. As you can see, the "ML" table entries are a factor of 40 smaller than the "SDF/GMM" entries.

Table I. Monte Carlo Standard Deviation of Estimated Risk Premium $\hat{\lambda}$. Simulations are calibrated to the CAPM on 10 NYSE size portfolios. T gives the sample size in months. The "ML" and "SDF" column are taken from Kan and Zhou (1999) Table I, p.1234. The third column is calculated with $\sigma(f) = 1$ as specified by Kan and Zhou. All table entries are multiplied by 100.

T	ML	SDF/GMM	$\sigma(f)/\sqrt{T}$
120	0.20	10.49	9.12
360	0.11	5.53	5.27
600	0.09	4.26	4.08
720	0.08	3.86	3.72

For a test of the CAPM with i.i.d. normal returns, the true maximum likelihood estimate of the factor risk premium is the mean of the market excess return, $\hat{\lambda} = E_T(R_t^{em})$, and its standard error is $\sigma(\hat{\lambda}) = \sigma(R^{em})/\sqrt{T}$. (See Campbell, Lo and MacKinlay 1997 or Cochrane 2001. Throughout, I use the notation $E_T = \frac{1}{T}\sum_{t=1}^{T}$ to denote sample mean.) In this traditional and simple environment, one cannot improve on the average market return as an estimator of the factor risk premium. Additional returns in a cross-section include the market premium plus idiosyncratic error, so they tell us nothing new about the market premium.

Kan and Zhou renormalize the model so that the factor f has a standard deviation of one, implicitly using a leveraged market portfolio as the factor. This is unusual, but not incorrect, since any mean-variance efficient portfolio can serve as reference return. The last column of Table I contains a calculation of $\sigma(f)/\sqrt{T}$, which is $1/\sqrt{T}$ given Kan and Zhou's normalization. Now, comparing the rows, you see that Kan and Zhou's SDF/GMM simulation almost exactly recovers the traditional standard error $\sigma/\sqrt{T}$. It is slightly inefficient in small samples, but not disastrously so. This is what we expect. GMM in linear models with i.i.d. normal errors is usually well behaved.

The surprise from Table I is that the "traditional" ML estimates seem to improve on $\sigma/\sqrt{T}$ by a factor of 40. Kan and Zhou don't find *poor* performance of GMM, they find astoundingly *good* performance of their "traditional" estimator. How can you beat maximum likelihood by a factor of 40? Obviously, you can't. There must be something unusual in the calculation of the *traditional* estimates. Jagannathan and Wang (2001) found the unusual assumption: *Kan and Zhou assume that they know the mean of the factor.* ("$E[f_t|\Phi_{t-1}] = 0$" on top of p.1223.) Giving the traditional estimate this false information advantage accounts for Kan and Zhou's results.

3 Effects of assuming that you know the factor mean

To see what happens if you assume that you know the mean of the factor, I trace what it does to estimation and testing strategies.

Case 1. Factor is a return. When the factor is a return, as for the CAPM, the asset pricing model says that the mean of the factor is the same as the factor risk premium.

$$\lambda = E(f). \tag{1}$$

Again, if you *don't know* $E(f)$, the maximum likelihood estimate of the factor risk premium is the sample mean of the factor. If you *know* the mean of the factor – if we make $E(f)$ a known quantity rather than a parameter to be estimated in maximum likelihood – then we *know* the factor risk premium λ, and its sampling variation is *zero:*

$$\hat{\lambda} = \lambda = E(f);\ \sigma(\hat{\lambda}) = 0.$$

This case shows most dramatically how assuming that you know the mean of the factor lowers the sampling variation of the estimated factor risk premium. It leaves a puzzle, though: how did Kan and Zhou get any sampling variation *at all* for the "ML" estimate in Table I? The answer is that they also ignored the restriction of the CAPM that the factor is a return.

Case 2. Factor is not a return. When the factor is *not* a return, restriction (1) does not hold, and the mean of the factor is no longer equal to the factor risk premium. In this case, standard ML (you don't know the factor mean) specifies a cross-sectional regression: run sample mean returns on the betas, and the estimated slope is the factor risk premium.

I offer three ways to see how adding the false assumption that we know the factor mean generates estimates with very small but non-zero sampling variation in this case. They increase in generality, but also in algebraic complexity, so they decrease in transparency.

Start with the standard time-series regression specification

$$R^e_t = a + \beta f_t + \varepsilon_t;\ \ \varepsilon_t\ i.i.d.,\ E(\varepsilon_t\varepsilon_t') = \Sigma. \tag{2}$$

The asset pricing model $E(R^e) = \beta\lambda$ implies a restriction on the intercepts a of this regression. We can impose this restriction by writing the time-series regression (2) as

$$R^e_t = \beta\lambda + \beta\,[f_t - E(f)] + \varepsilon_t. \tag{3}$$

Taking the sample average of (3), we obtain

$$E_T\left(R_t^e\right) = \beta\lambda + \beta\left[E_T\left(f_t\right) - E(f)\right] + E_T\left(\varepsilon_t\right). \tag{4}$$

This is the starting point for all our λ estimates.

a) A simple example shows the point most clearly. Suppose there is one asset, and its beta is one. If we *do not know* the mean of the factor $E(f)$, we will estimate it by its sample mean, so the second term on the right hand side of (4) is zero. Then, our estimate of λ will be

$$\hat{\lambda}_{\text{don't know}} = E_T(R^e).$$

The standard error of this estimate is

$$\sigma^2(\hat{\lambda}_{\text{don't know}}) = \frac{\sigma^2(R^e)}{T} = \frac{\sigma^2(f) + \sigma^2(\varepsilon)}{T} \tag{5}$$

If we *know* the true mean of the factor $E(f)$, we will use the known value rather than estimate it as a parameter. Now, we estimate the factor risk premium from (4) by

$$\hat{\lambda}_{\text{know}} = E_T(R_t^e) - \left[E_T\left(f_t\right) - E(f)\right].$$

Substituting for R_t^e from (4),

$$\hat{\lambda}_{\text{know}} = E_T\left(\varepsilon\right) + E(f)$$

so

$$\sigma^2(\hat{\lambda}_{\text{know}}) = \frac{\sigma^2(\varepsilon)}{T} \tag{6}$$

Now, compare (5) with (6). If you know the mean of the factor, the standard deviation of the factor risk premium is driven by the *residual* variance. If you don't know the mean of the factor, the standard deviation of the factor risk premium is driven by the *return* variance, the residual variance plus the factor variance.

To compare the two standard errors, we can write

$$\frac{\sigma^2(\hat{\lambda}_{\text{know}})}{\sigma^2(\hat{\lambda}_{\text{don't know}})} = \frac{\sigma^2(\varepsilon)}{\sigma^2(f) + \sigma^2(\varepsilon)} = 1 - R^2 \tag{7}$$

where R^2 is the R^2 of the time-series regression (3) for the single asset under consideration. The R^2 of size portfolios on the market is quite high. For example, the CRSP large portfolio has market model regression $R^2 = 0.984$ in the full 1926-1998 sample. Therefore, we expect that $\sigma^2(\hat{\lambda}_{\text{know}})$ is much less than $\sigma^2(\hat{\lambda}_{\text{don't know}})$, but not zero.

b) Many returns. The same point applies to the case with many returns, requiring only a little more algebra.

If we *do not know* the mean of the factor, ML will estimate it as the sample mean, again setting the second term on the right hand side of (4) to zero, and leaving us with

$$E_T\left(R_t^e\right) = \beta\lambda + E_T\left(\varepsilon_t\right). \tag{8}$$

ML now estimates λ with a cross sectional regression. Σ/T is the covariance matrix of the error term in (8), so we get a GLS cross-sectional regression,

$$\hat{\lambda}_{\text{don't know}} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}E_T(R^e)$$

We can find the sampling variation of $\hat{\lambda}_{\text{don't know}}$ with known betas using the standard regression derivation. Substituting for $E_T(R^e)$ from (8),

$$\hat{\lambda}_{\text{don't know}} = \lambda + (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\left(\beta\left[E_T(f_t) - E(f)\right] + E_T(\varepsilon_t)\right).$$

Thus,

$$\sigma^2(\hat{\lambda}_{\text{don't know}}) = \frac{1}{T}\left[\sigma^2(f) + (\beta'\Sigma^{-1}\beta)^{-1}\right] \tag{9}$$

a natural analogue to (5).

Once you decide to use this cross-sectional regression, neither the estimate or standard error contain $E(f)$. Therefore, they are unaffected by the assumption that you know the mean of the factor. The GMM/SDF estimate is also a cross-sectional regression, which is why it is unaffected by Kan and Zhou's assumption that they know the mean of the factor.

If we *do know* the mean of the factor, ML will again use that knowledge rather than estimate a known quantity, and will prescribe a different regression. Rewrite (4) as

$$E_T(R_t^e) - \beta\left[E_T(f_t) - E(f)\right] = \beta\lambda + E_T(\varepsilon_t).$$

The cross-sectional regression will therefore be

$$\hat{\lambda}_{\text{know}} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}\left\{E_T(R^e) - \beta\left[E_T(f_t) - E(f)\right]\right\}$$

Again, we find the sampling variance of $\hat{\lambda}_{\text{know}}$ by substituting for $E_T(R^e)$ from (4)

$$\hat{\lambda}_{\text{know}} = \lambda + (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}E_T(\varepsilon_t)$$

$$\sigma^2\left(\hat{\lambda}_{\text{know}}\right) = \frac{1}{T}(\beta'\Sigma^{-1}\beta)^{-1} \tag{10}$$

a natural analogue to (6).

The ratio of the do and don't variance is

$$\frac{\sigma^2\left(\hat{\lambda}_{\text{know}}\right)}{\sigma^2(\hat{\lambda}_{\text{don't know}})} = \frac{(\beta'\Sigma^{-1}\beta)^{-1}}{\sigma^2(f) + (\beta'\Sigma^{-1}\beta)^{-1}} = 1 - R^2_{\text{max}},$$

a natural analogue to (7). R^2_{max} is the time-series regression R^2 of the portfolio $(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}R^e$. This portfolio has maximum R^2 in time-series regressions of all portfolios of the original test assets R^e. The 10 size portfolios very nearly span the value-weighted market return. Hence, the maximum-R^2 portfolio has an R^2 very near one, and we expect $\sigma^2\left(\hat{\lambda}_{\text{know}}\right) << \sigma^2(\hat{\lambda}_{\text{don't know}})$.

c) Corrections for estimated betas. Formulas (9) and (10) treat betas as fixed. Shanken (1992) derives the correct asymptotic distribution of factor risk premia estimated from cross-sectional regressions, including the fact that betas are estimated, and Campbell Lo and MacKinlay (1997) and Cochrane (2001) present textbook treatments. The answer is

$$\sigma^2(\hat{\lambda}_{\text{don't know}}) = \frac{1}{T}\left[(\beta'\Sigma^1\beta)^{-1}\left(1+\frac{\lambda^2}{\sigma^2(f)}\right)+\sigma^2(f)\right]. \quad (11)$$

This differs from (9) by presence of the term $\frac{\lambda^2}{\sigma^2(f)}$. In monthly data, this term is typically small. For the CAPM, this term is the squared market Sharpe ratio, about $0.5^2/12 = 0.02$.

The correct asymptotic distribution for $\hat{\lambda}_{\text{ know}}$ is[1]

$$\sigma^2(\hat{\lambda}_{\text{know}}) = \frac{1}{T}(\beta'\Sigma^{-1}\beta)^{-1}\left(1+\frac{\lambda^2}{\sigma^2(f)}\right).$$

Again you see the same small correction, and the crucial difference that $\sigma^2(f)$ is missing from the $\sigma^2(\hat{\lambda}_{\text{know}})$.

[1] To derive this expression, just follow any standard derivation of (11) with $E_T(R_t^e) - \beta\left[E_T(f_t) - E(f)\right]$ in place of $E_T(R_t^e)$. The algebra is straightforward, tedious, and available at
http://gsbwww.uchicago.edu/fac/john.cochrane/research/Papers/

4 References

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay 1997, *The Econometrics of Financial Markets* Princeton NJ: Princeton University Press.

Cochrane, John H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model" *Journal of Political Economy* 104, 572-621.

Cochrane, John H., 2001, *Asset Pricing*, Princeton NJ: Princeton University Press.

Jagannathan, Ravi, and Zhenyu Wang, 2001, "Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods" NBER Working Paper 8098.

Kan, Raymond and Guofu Zhou, 1999, "A Critique of Stochastic Discount Factor Methodology," *Journal of Finance* LIV, 1221-1248.

Shanken, Jay, 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies* 5, 1-33