Shanghai Advanced Institute of Finance
Shanghai Jiao Tong University

Jun Pan
junpan@saif.sjtu.edu.cn

Empirical Asset Pricing, Fall 2019

Class 2: Consumption-Based Asset-Pricing Models

The empirical asset-pricing literature focuses on the estimation and evaluation of asset-pricing models. For most empiricists, the consumption-based asset-pricing models has an enormous appeal, as it establishes a direct link between the aggregate stock market returns and the fundamentals of the economy. In an endowment economy, the risk-averse representative agent makes optimal consumption and investment decisions, and his optimization in equilibrium gives rise to a powerful pricing relation that links security returns to risk-aversion and consumption growth. With data on consumption and stock returns ready, this famous first-order condition was first tested in the classic paper of Hansen and Singleton (1982), which is the starting point of our today's class. Written more as a methodological piece to introduce the generalized method of moments estimators (GMM) of Hansen (1982), this paper has all the necessary ingredients to show us the equity-premium puzzle. But it is not until Mehra and Prescott (1985) that the poor performance of the consumption-based models starts to take center stage in the asset-pricing literature. Over the span of two decades since the publication of Mehra and Prescott in 1985, solving the equity premium puzzle has been one of the most active literature in asset pricing. Among others, the two most notable developments have been the habit model of Campbell and Cochrane (1999) and the long-run risk model of Bansal and Yaron (2004). Over the past ten years, this literature has not been as active, but learning about this literature and understanding the key issues should be an integral part of your PhD training. For this reason, your first mandatory assignment is to reproduce the result of Hansen and Singleton (1982).

1 Hansen and Singleton (1982)

1.1 Model-Implied Pricing Relation

First-Order Condition: The pricing relation arises from the first-order condition for the optimal consumption and portfolio formation:

$$P_t U'(C_t) = E_t \left[\beta U'(C_{t+1}) (P_{t+1} + D_{t+1}) \right] ,$$

where $U(\cdot)$ is the representative agent's utility function, $0 < \beta < 1$ his time discount factor, and C_t his time-t consumption.

Pricing Kernel: Consider a dividend paying security with time-t dividend D_t. The first-order condition, also known as the Euler equation, links its price P_t at time t to payoff $P_{t+1} + D_{t+1}$ at time $t+1$ via the famous pricing kernel:

$$P_t = E_t \left[q_{t+1} \left(P_{t+1} + D_{t+1} \right) \right] ,$$

where

$$q_{t+1} = \beta \frac{U'(C_{t+1})}{U'(C_t)} .$$

Security Returns: For empiricists who are more used to working with returns,

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1$$

this pricing relation can be further simplified,

$$\boxed{E_t \left[q_{t+1} \left(R_{t+1} + 1 \right) \right] = 1} \tag{1}$$

1.2 Generalized Method of Moments (GMM) Estimations

Moment Conditions: Under the GMM framework, we can express the pricing relation in (1) more generally by,

$$E_t \left[h \left(x_{t+n}, b_0 \right) \right] = 0 ,$$

where x_{t+n} is a k dimensional vector of variables observed by agents and the econometrician as of date $t+n$, b_0 is an l dimensional parameter vector that is unknown to the econometrician, h is a function mapping $R^k \times R^l$ into R^m, and E_t is the expectations operator conditioned on the agent's period t information set, I_t.

Instruments: Let z_t denote a q dimensional vector of variables observable as of date t. Taking advantage of the conditional expectation, the original m dimensional moment condition can be further expanded by

$$f \left(x_{t+n}, z_t, b_0 \right) = h \left(x_{t+n}, b_0 \right) \otimes z_t .$$

And the unconditional version of the moment condition

$$E \left[f \left(x_{t+n}, z_t, b_0 \right) \right] = E \left[h \left(x_{t+n}, b_0 \right) \otimes z_t \right] = 0 ,$$

represents a set of $r = m \times q$ from which an estimator of b_0 can be constructed, provided that r is at least as large as l, the number of unknown parameters.

GMM Estimator: Let $g_0(b) = E[f(x_{t+n}, z_t, b)]$ be the population mean of the moments for any $b \in R^l$ and its sample counterpart can be constructed as

$$\boxed{g_T(b) = \frac{1}{T} \sum_{t}^{T} f(x_{t+n}, z_t, b) \ ,} \tag{2}$$

where T denotes the sample size. If the model is true, then $g_T(b)|_{b=b_0}$ should be close to zero for large values of T. Using this intuition,

$$b_T^{GMM} = \arg\min_b J_T(b) \, ,$$

where the criterion function is constructed as

$$J_T(b) = g_T(b)' \, W_T \, g_T(b) \, ,$$

where W_T is an r by r symmetric, positive definite matrix that can depend on sample information.

Two-Stage GMM: The asymptotic covariance matrix for b_T^{GMM} depends on the choice of weighting matrix W_T. It is possible to choose W_T optimally in the sense of constructing an estimator with the smallest asymptotic covariance matrix. Under the two-stage GMM of Hansen (1982), the weighting matrix W_T does not take into account of conditioning information and is calculated via a two-stage iteration of:

$$W_T^* = \left\{ R_T(0) + \sum_{j=1}^{n-1} \left(R_T(j) + R_T(j)' \right) \right\}^{-1} \ ,$$

where

$$R_T(j) = \frac{1}{T} \sum_{t=1+j}^{T} f(x_{t+n}, z_t, b_T) \ f(x_{t+n-j}, z_{t-j}, b_T)' \ .$$

To further improve the efficiency of the GMM estimator, the weighting matrix can be constructed using the conditioning information, as in Hansen (1985).

Under Hansen (1982), the resulting asymptotic covariance matrix of the two-stage GMM estimator can be computed by

$$\left(D_T' \, W_T^* \, D_T \right)^{-1} \ ,$$

where

$$D_T = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial h}{\partial b}\left(x_{t+n}, b_T\right) \otimes z_t \,.$$

To assess the goodness-of-fit and perform hypothesis testing, we go back to the sample mean of the moment condition $g_T(b)$ and the criterion function $J_T(b)$. If the model is correct, then all r components of the moment condition equal zero, and all r components of $g_T(b)|_{b=b_0}$ are close to zero for large value of T. In estimating b_T^{GMM}, l of the r moment conditions have been used, and the remaining $r-l$ can be used in testing. Indeed, T times the minimized value of the objective function $J_T(b)$ can be shown to be asymptotically distributed as a chi-square with $r-l$ degrees of freedom.

More on GMM: For a more general weighting matrix of W_0, the asymptotic covariance matrix can be computed as

$$\Omega_0 = \left(D_0' W_0 D_0\right)^{-1} D_0' W_0 \Sigma_0 W_0 D_0 \left(D_0' W_0 D_0\right)^{-1} \,,$$

where

$$\Sigma_0 = \lim_{T\to\infty} TE\left(g_T(b_0) g_T(b_0)'\right)$$

It will be a good exercise for you to prove this. Also, it will be a helpful exercise to show, as special cases of the GMM estimator, the estimations of mean, variance, skewness, and kurtosis, as well as the OLS estimator and the maximum-likelihood estimator. I understand that nowadays empiricists routinely use canned routines to calculate standard errors, but it will be a good exercise for you to go through the derivation yourself under the GMM setting. This is particularly true for the OLS regression on a panel structure, where the standard errors can be clustered by two dimensions and much of the calculation focuses on how you calculate the sample counterpart of Σ_0.

1.3 A Concrete Example

Power Utility: As a concrete example, Hansen and Singleton (1982) focuses their empirical analysis on the power utility:

$$U(c) = \frac{C^{1-\gamma}}{1-\gamma} \,,$$

where γ is the relative risk aversion coefficient. The pricing kernel for this utility specification is

$$q_{t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \,.$$

The parameter to be estimated: $b = [\beta, \gamma]'$. Mapping the notation back to that of Hansen and Singleton (1982), the relative risk aversion coefficient is their $-\alpha$.

Moment Conditions:

$$\boxed{E_t\left(\beta e^{-\gamma\, y_{t+1}+R_{t+1}} - 1\right) = 0\,,} \tag{3}$$

where y_t is the real per capita non-durable consumption growth as of period t,

$$y_t = \log\left(C_t^{ND}/POP_t\right) - \log\left(C_{t-1}^{ND}/POP_{t-1}\right)$$

and where R_t is the period-t real stock return,

$$R_t = \log(P_t + D_t) - \log(P_{t-1}) - \pi_t$$

and where π_t is the inflation rate

$$\pi_t = \log(P_t^{ND}) - \log(P_{t-1}^{ND})\,.$$

2 The Equity Premium Puzzle

This literature formally begins with Mehra and Prescott (1985) and, for a long time, has been considered by some as one of the holy grail in asset pricing. It focuses on a measure that is central to asset pricing – the expected return of the aggregate stock market in excess of the riskfree rate. This measure, often referred to as the equity premium, can be estimated using the historical data, and can also be inferred from standard asset-pricing models. So a comparison between the model and the reality is inevitable and Mehra and Prescott (1985) did just that. While the ingredients of this comparison are contained in some of the early papers, including Hansen and Singleton (1982, 1983), a direct head-to-head comparison of model and data was brought about in that famous paper, entitled "The Equity Premium, A Puzzle." The main observation is that the equity premium estimated from the data is too high to be justified by the standard asset-pricing models with moderate levels of risk-aversion coefficient.

2.1 A Concrete Example

Mehra and Prescott (1985) build their analysis on a regime switching model coupled with a representative agent with power utility. It is a useful exercise to reproduce their results

if you would like to practice on the regime-switching models. The simple example outlined in this section delivers the key intuition on the equity premium puzzle. Throughout the section, we will use annualized numbers and one period in the model is one year.

Model Implied: Let's build on the pricing relation implied by a representative agent with power utility:

$$E_t\left(\beta e^{-\gamma\, y_{t+1}+R_{t+1}}-1\right)=0\,,$$

where γ is his risk aversion coefficient. Applying this pricing relation to both the risky and riskfree securities, we have

$$\beta\exp\left(-\gamma\mu_y+\frac{1}{2}\gamma^2\sigma_y^2+\mu_R+\frac{1}{2}\sigma_R^2-\gamma\,\mathrm{Cov}_{yR}\right)=1 \tag{4}$$

$$\beta\exp\left(-\gamma\mu_y+\frac{1}{2}\gamma^2\sigma_y^2+r_f\right)=1\,, \tag{5}$$

where we have assumed that the consumption growth y_t and the real stock return R_t are joint normal, with μ_y and σ_y denoting the mean and standard deviation of the consumption growth y_t, μ_R and σ_R the mean and standard deviation of the real stock return R_t, Cov_{yR} denoting the covariance between the consumption growth and the real stock return, and r_f denoting the continuously compounded riskfree rate. Equations (4) and (5) imply that

$$\boxed{\mu_R+\frac{1}{2}\sigma_R^2-r_f=\gamma\,\mathrm{Cov}_{yR}\,.} \tag{6}$$

Data Estimated: Equation (6) effectively summarizes the tension that gives rise to the equity premium puzzle. On the left hand size is the equity premium implied by the data. Using Table 1 of Campbell (2003), we have $\mu_R=8.085\%$, $\sigma_R=15.645\%$, and $r_f=0.896\%$. Plugging these numbers to Equation (6), the left hand side equals 8.41%.[1]

To match this magnitude, there are two terms in Equation (6) with γ being the only free parameter. Here, the covariance term Cov_{yR} plays a central role in the pricing. Consider the extreme example of a security whose return R_t has a zero correlation with the consumption growth y_t, it will yield zero equity premium. In fact, this is the familiar intuition we get from the CAPM: only the systematic risk is priced with a premium. In this economy, the risk associated with the uncertainty in consumption growth y_t is what really matters, and the pricing kernel $q_t=\beta\exp\left(-\gamma\, y_t\right)$ dictates the pricing in this economy. To calculate the

[1]It should be noted that this number is close to but not the same as the actual equity premium, which is $\exp\left(\mu_R+\frac{1}{2}\sigma_R^2\right)-\exp\left(r_f\right)=8.86\%$. For small value of $\mu_R+\frac{1}{2}\sigma_R^2$ and r_f, $\mu_R+\frac{1}{2}\sigma_R^2-r_f$ is a good approximation.

magnitude of Cov_{yR}, we get $\sigma_y = 1.073\%$ from Table 2 of Campbell (2003) and the correlation between consumption growth and stock return $\rho = 34\%$ from Table 3 of Campbell (2003). This gives us $\text{Cov}_{yR} = 5.707 \times 10^{-4}$. Going back to Equation (6), we know that

$$\boxed{\gamma = \frac{\mu_R + \frac{1}{2}\sigma_R^2 - r_f}{\text{Cov}_{yR}} = \frac{8.86\%}{5.707 \times 10^{-4}} = 147}$$

For researchers back in the 1980s, a risk aversion coefficient of 147 is extremely high. So the puzzle is 8.86% is too large. Hence the equity premium puzzle. Of course, it could also be that the consumption is too smooth: $\sigma_y = 1.073\%$, much smaller than $\sigma_R = 15.645\%$.

The Riskfree Rate Puzzle: Suppose we are willing to accept the fact that γ is close to 150. There is another binding constraint that arises out of the pricing relation for the riskfree security in Equation (5):

$$r_f = -\log(\beta) + \gamma\mu_y - \frac{1}{2}\gamma^2\sigma_y^2 \,.$$

A direct impact of $\gamma = 147$ is an astronomical riskfree rate:

$$r_f = 147 \times 1.864\% - \frac{1}{2} \times 147^2 \times (1.075\%)^2 = 149.7\% \,.$$

where we have set $\beta = 1$ for simplicity. In reality, the riskfree rate is 0.896%.

Excess Volatility: Another puzzle revealed by our simple example is the excess volatility puzzle. Assuming the consumption growth to be i.i.d. normal, we can derive the equilibrium price dynamics for the aggregate stock market (as a claim to the aggregate endowment). It can be shown that the stock return volatility equals the consumption volatility. In practice, we see that consumption volatility is around 1% while the stock return volatility is around 15%. See Shiller (1981) and LeRoy and Porter (1981) if you are interested in this topic.

2.2 Campbell and Cochrane (1999)

This is an important paper in the literature of consumption-based asset pricing models. It provides a concise summary of the relevant and important empirical issues documented in the existing literature and show that such issues can be addressed and resolved by a modified pricing kernel. In a way, this is backward engineering, but by linking this pricing kernel to habit formation, this exercise provides valuable insights towards a better understanding of what gives rise to stock returns.

The Big Picture: Empirical observations on the link between asset markets and macroe-

conomics indicate countercyclical expected returns: equity risk premia are higher at business cycle troughs than they are at peaks. Moreover, price-dividend ratios are procyclical, with depressed stock market prices at business cycle troughs, and are predictive of stock returns. Against this backdrop, the time-varying stock return variance does not move one for one with estimates of conditional mean returns, indicating that the slope of the conditional mean-variance frontier, a measure of the price of risk, changes through time with a business cycle pattern.

These empirical observations are challenging for macroeconomics as standard business cycle models fail to reproduce the level, variation, and cyclical comovement of equity premia. For asset-pricing empiricists who study the markets, the pressing question: what are the fundamental sources of risk that drive expected returns?

What They Do: A simple modification of the standard representative-agent consumption-based asset pricing model via slow-moving habit. As consumption declines toward the habit in a business cycle trough, the curvature of the utility function rises, so risky asset prices fall and expected returns rise.

Unlike the long-run risk model of Bansal and Yaron (2004), they keep the dynamics of the consumption growth simple: i.i.d. lognormal, with the same mean and standard deviation as postwar consumption growth. By avoiding exogenous variation in the probability distribution of consumption, their focus is on how the pricing kernel, via slow-moving habit, can drive the time-variation of expected returns. At a conceptual level, this is very different from the approach of Bansal and Yaron (2004), who rely on an exogenous latent state variable, the long-run mean of the consumption growth, to explain the time-variation of expected returns. Personally, I very much prefer the former approach. As a modeling device, it might be convenient to shift the burden to a state variable that cannot be observed, but empirically there is not much evidence in the consumption data to justify such a modification of the consumption dynamics. Intuitively, it also makes sense that the effective risk aversion increases at business cycle troughs.

Pricing Kernel: Under their setting, the pricing kernel becomes

$$\boxed{M_{t+1} = \delta \left(\frac{S_{t+1}}{S_t} \frac{C_{t+1}}{C_t} \right)^{-\gamma},} \tag{7}$$

where δ is the time discount factor, C is the consumption, and S is the surplus consumption ratio:

$$S_t = \frac{C_t - X_t}{C_t},$$

where X is the level of habit that enters the utility function via

$$E \sum_{t=0}^{\infty} \delta^t \frac{(C_t - X_t)^{1-\gamma} - 1}{1-\gamma} .$$

This specification has a key implication: the local curvature of the utility function becomes

$$\eta_t = -\frac{CU_{cc}(C_t, X_t)}{U_c(C_t, X_t)} = \frac{\gamma}{S_t} .$$

Instead of building X explicitly has a moving average of past consumption levels, Campbell and Cochrane (1999) choose to model the surplus consumption ratio directly,

$$s_{t+1} = (1-\phi)\,\bar{s} + \phi s_t + \lambda\,(s_t)\,(c_{t+1} - c_t - g) \ ,$$

where lowercase c and S are the logs of the upper case C and S, and constant $\bar{s}$ is the long-run mean of s_t, ϕ controls the rate of mean reversion, and g is a constant parameter. In choosing the specification of the sensitivity function $\lambda(s_t)$, there is much discussion and engineering, which I am not a huge fan of. So let's focus on the intuition. If we think of the surplus ratio as how relaxed the agent is with respect to his habit level, then as s approaches to zero, he will be extremely sensitive to shocks in consumption. For this reason, $lambda(s)$ goes to infinity as s approaches zero. Likewise, for large value of s, the agent is comfortably above his habit level with $\lambda(s)$ close to zero.

Asset Pricing Implications: The surplus ratio s is the only state variable in the model to drive the dynamics of asset prices, including the expected return, volatility, and the Sharpe ratio. By construction, it is closely related to the history of the consumption growth: negative consumption shocks deplete the level of the consumption surplus ratio, and, at low levels of s, the depletion is further amplified via the construction of the sensitivity function $\lambda(s)$.

The price-dividend ratio and its connection to the surplus consumption ratio play an central role in the asset pricing implications of this model. When consumption is low relative to habit in a recession, the surplus ratio is low, the curvature of the utility function is high (γ/S_t), and prices are depressed relative to dividends. Indeed, in the model, the price-dividend ratio is nearly linear in the surplus ratio, making the price-dividend ratio a directly measurable proxy of the state variable s. Once this link is established, the implications for the expected returns, volatility, and Sharpe ratios are straightforward.

Appendix: Notes on GMM and Econometrics

A. Deriving the asymptotic variance-covariance matrix

In our notes, $H_T(\beta)$ is $g_T(b)$, the moment condition, and β is b, the parameter vector.

two emphases : ① $d_0 = E\left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_0}\right)$

NBER

NATIONAL BUREAU OF ECONOMIC RESEARCH

② Σ_0, $= \lim_{T\to\infty} E\left(H_T(\beta_0) * H_T(\beta_0)^T\right)$

Sketch of proof. Hamilton Appendix 14. A

Mean-value theorem : for some β^* between β_0 and β_T

$$\underset{(r\times 1)}{H_T(\beta_T) - H_T(\beta_0)} = \underset{(r\times q)}{\frac{\partial H_T}{\partial \beta}\Big|_{\beta^*}} \underset{(q\times 1)}{(\beta_T - \beta_0)}$$

$$\left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_T}\right)^T W_T H_T(\beta_T) - \left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_T}\right)^T W_T H_T(\beta_0)$$

$$= \left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_T}\right)^T W_T \left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta^*}\right)(\beta_T - \beta_0)$$

$$\beta_T - \beta_0 = -\left[\left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_T}\right)^T W_T \left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta^*}\right)\right]^{-1} \left(\frac{\partial H_T}{\partial \beta}\Big|_{\beta_T}\right)^T W_T H_T(\beta_0)$$

β^* between β_0 and β_T, $\beta_T \xrightarrow{p} \beta_0$, so $\beta^* \xrightarrow{p} \beta_0$

$$\sqrt{T}(\beta_T - \beta_0) \xrightarrow{p} -\underbrace{(d_0^T W_0 d_0)^{-1} d_0^T W_0}_{(q\times r)} \sqrt{T} H_T(\beta_0)$$

$$\sqrt{T}(\beta_T - \beta_0) \Rightarrow N(0, \Omega_0)$$

$$\Omega_0 = (d_0^T W_0 d_0)^{-1} d_0^T W_0 \Sigma_0 W_0 d_0 (d_0^T W_0 d_0)^{-1}$$

B. GMM and Estimating Moments

NBER

NATIONAL BUREAU OF ECONOMIC RESEARCH

Example:

1) A. Mean

$\mu_0 = E(R_t)$ $\qquad f(R_t, \mu) = R_t - \mu.$

$H_T(\mu) = \frac{1}{T}\sum R_t - \mu.$

$\mu_T = \frac{1}{T}\sum R_t$ $\qquad W_0$: N/A. just one moment condition.

$d_0 : \frac{\partial H_T}{\partial \mu} = -1$

$\sqrt{T} H_T(\mu_0) \Rightarrow N(0, \Sigma_0)$

$\Sigma_0 = \lim_{T\to\infty} T E\left(H_T(\mu_0) H_T(\mu_0)^T\right).$

$= \lim_{T\to\infty} \frac{1}{T} E\left(\sum(R_t - \mu_0) \sum(R_t - \mu_0)\right).$

$= var(R_t) + \ldots\ldots$

$\Omega_0 = \Sigma_0$

$\sqrt{T}(\mu_T - \mu_0) \Rightarrow N(0, \Omega_0)$

$std(\mu_T) = \left(\frac{1}{T}\Omega_0\right)^{1/2} = \left(\frac{var(R_t)}{T}\right)^{1/2}$

B. mean variance, skew. kurt.

C. GMM and OLS Regression

panel regression and cluster by time and firm

NBER
NATIONAL BUREAU OF ECONOMIC RESEARCH

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 X X_t^T + \sum_{v=1}^{q}\left(1-\frac{v}{q+1}\right)\cdot \frac{1}{T}\sum_{t=v+1}^{T}\left(X_t u_t u_{t-v} X_{t-v} + X_{t-v} u_{t-v} u_t X_t\right)$$

Example

2) OLS.

$$Y = X^T\beta + u$$

$(1\times n)\ (n\times 1)$

$$f(Y, X, \beta) = X(Y - X^T\beta) \qquad n\times 1$$

exactly identified

$$H_T = \frac{1}{T}\sum_t X(Y - X^T\beta)$$

$$\frac{\partial H_T}{\partial \beta} = -\frac{1}{T}\sum_t XX^T \quad d_0 = -E(XX^T)$$

Newey-West

$$\Sigma_0 = \lim_{T\to\infty} T E(HH^T)$$

$$= \lim_{T\to\infty} \frac{1}{T} E\left(\sum X(Y - X^T\beta)\left(\sum X(Y - X^T\beta)\right)^T\right)$$

White

$$= \lim_{T\to\infty} \frac{1}{T} E\left(\sum X(Y - X^T\beta)^2 X^T\right)$$

$$\Sigma_0 = \frac{1}{T}\sum (Y_t - X_t^T\beta)^2 X_t X_t^T$$

Heteroskedastic consistent.

OLS

$$= \frac{1}{T}\sum XX^T E(Y - X^T\beta)^2$$

$\sqrt{T}(\mu_T - \mu_0)$

$\Rightarrow N(0, \Omega_0)$

$$\Omega_0 = \left[\cancel{\left(\frac{1}{T}\sum XX^T\right)^T \left(\frac{1}{T}\sum XX^T\right)}\right]^{-1} \cancel{\frac{1}{T}\sum XX^T \frac{1}{T}\sum XX^T}$$

$$d_0^{-1}\Sigma_0(d_0^T)^{-1} = \left(\frac{1}{T}\sum XX^T\right)^{-1} \frac{1}{T}\sum XX^T \sigma^2$$

$$= \left(\frac{1}{T}\sum XX^T\right)^{-1}\sigma^2$$