Model Comparison Using the Hansen-Jagannathan Distance

Raymond Kan and Cesare Robotti

Appendix on Excess Returns

1. Asymptotic Distributions of the Parameter Estimates and of the Sample Hansen-Jagannathan Distance

The notations in this appendix follow those in Kan and Robotti ("Specification Tests of Asset Pricing Models Using Excess Returns," *Journal of Empirical Finance* 15, 2008). In that paper, we show that when excess returns are used, we should normalize the linear stochastic discount factor as

$$y_t(\lambda) = 1 - \lambda'(f_t - \mu_1). \tag{1}$$

Let r_t be a vector of excess returns on N test assets at time t. The pricing errors of the normalized SDF model on the excess returns are given by

$$e(\lambda) = E[r_t y_t(\lambda)] = \mu_2 - V_{21}\lambda, \tag{2}$$

In addition, when excess returns are used, the squared HJ-distance needs to be modified and is defined as

$$\delta^2 = \min_{\lambda} e(\lambda)' V_{22}^{-1} e(\lambda) = \mu_2' V_{22}^{-1}\mu_2 - \mu_2' V_{22}^{-1} V_{21}(V_{12}V_{22}^{-1}V_{21})^{-1} V_{12} V_{22}^{-1}\mu_2. \tag{3}$$

The λ that minimizes $e(\lambda)'V_{22}^{-1}e(\lambda)$ is given by

$$\lambda_m = (V_{12}V_{22}^{-1}V_{21})^{-1}(V_{12}V_{22}^{-1}\mu_2). \tag{4}$$

We define $e_m = \mu_2 - V_{21}\lambda_m$ and $y_t(\lambda_m) = 1 - \lambda_m'(f_t - \mu_1)$. Since there is no source of confusion, we suppress the subscript m in λ_m and e_m from now on.

The estimates of λ and δ^2 are given by

$$\hat{\lambda} = (\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}(\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2), \tag{5}$$

$$\hat{\delta}^2 = \hat{\mu}_2'\hat{V}_{22}^{-1}\hat{\mu}_2 - \hat{\mu}_2'\hat{V}_{22}^{-1}\hat{V}_{21}(\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2. \tag{6}$$

Kan and Robotti (2008) show that the asymptotic distribution of $\hat{\lambda}$ under a potentially misspecified model is given by

$$\sqrt{T}(\hat{\lambda} - \lambda) \overset{A}{\sim} N\left(0_K, \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}']\right), \tag{7}$$

where

$$h_t = HV_{12}V_{22}^{-1}(r_t - \mu_2)(1 - y_t) + H[(f_t - \mu_1) - V_{12}V_{22}^{-1}(r_t - \mu_2)]u_t + \lambda, \tag{8}$$

with $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$ and $u_t = e'V_{22}^{-1}(r_t - \mu_2)$.

The asymptotic distribution of $\hat{\delta}^2$ under the correctly specified model is given by

$$T\hat{\delta}^2 \overset{A}{\sim} \sum_{i=1}^{N-K} \xi_i x_i, \tag{9}$$

where the x_i's are independent χ_1^2 random variables, and the ξ_i's are the eigenvalues of $P'V_{22}^{-1}SV_{22}^{-1}P$, where P is an $N \times (N-K)$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}}V_{21}$,

$$S = \sum_{j=-\infty}^{\infty} E[g_t g_{t+j}'], \tag{10}$$

and $g_t = r_t y_t$. When $\delta > 0$, the asymptotic distribution of $\hat{\delta}^2$ is given by

$$\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\sim} N\left(0_K, \sum_{j=-\infty}^{\infty} E[q_t q_{t+j}']\right), \tag{11}$$

where

$$q_t = 2u_t y_t - u_t^2 + \delta^2. \tag{12}$$

2. Model Comparison

We consider two competing models. Let $x_1 = [f_1', \ f_2']'$ and $x_2 = [f_1', \ f_3']'$, where f_1 to f_3 are three sets of distinct factors, and f_i is of dimension $K_i \times 1$, $i = 1, 2, 3$. We assume that the SDF of model 1 is linear in x_1 and is given by $y_1 = 1 - \eta'(x_1 - E[x_1])$ whereas the SDF of model 2 is linear in x_2 and is given by $y_2 = 1 - \lambda'(x_2 - E[x_2])$. Let $D_1 = \mathrm{Cov}[r, x_1']$ and $D_2 = \mathrm{Cov}[r, x_2']$ and assume that both D_1 and D_2 have full column rank, so that the SDF parameters that minimize the HJ-distances of the two models are uniquely identified as

$$\eta = (D_1'V_{22}^{-1}D_1)^{-1}D_1'V_{22}^{-1}\mu_2, \tag{13}$$

$$\lambda = (D_2'V_{22}^{-1}D_2)^{-1}D_2'V_{22}^{-1}\mu_2. \tag{14}$$

It follows that the pricing errors and the squared HJ-distances of the two models are given by

$$e_i = \mu_2 - D_i(D_i'V_{22}^{-1}D_i)^{-1}D_i'V_{22}^{-1}\mu_2 \qquad i = 1, 2, \tag{15}$$

$$\delta_i^2 = \mu_2'V_{22}^{-1}\mu_2 - \mu_2'V_{22}^{-1}D_i(D_i'V_{22}^{-1}D_i)^{-1}D_i'V_{22}^{-1}\mu_2 \qquad i = 1, 2. \tag{16}$$

The sample estimates are analogously defined. When $K_1 = 0$, the two models do not share a common factor. When $K_2 = 0$, the second model nests the first model as a special case. Similarly, when $K_3 = 0$, the first model nests the second model as a special case. When both $K_2 > 0$ and $K_3 > 0$, the two models are not nested.

2.1 Nested Models

Without loss of generality, we assume $K_2 = 0$, so that model 2 nests model 1 as a special case. For the nested models case, the following lemma shows that $\delta_1^2 = \delta_2^2$ implies some restrictions on the SDF parameters of model 2.

Lemma 1: $\delta_1^2 = \delta_2^2$ if and only if $\lambda_2 = 0_{K_3}$, where λ_2 is the vector of the last K_3 elements of λ.

Note that the above lemma is applicable even when the models are misspecified. In order to test the equality of HJ-distances of the two models, the above lemma suggests that one can simply perform a test of $H_0 : \lambda_2 = 0_{K_3}$ in model 2. Suppose that $\hat{V}(\hat{\lambda}_2)$ is a consistent estimator of the asymptotic variance of $\sqrt{T}(\hat{\lambda}_2 - \lambda_2)$. Then, under the null hypothesis $H_0 : \lambda_2 = 0_{K_3}$,

$$T\hat{\lambda}_2'\hat{V}(\hat{\lambda}_2)^{-1}\hat{\lambda}_2 \stackrel{A}{\sim} \chi^2_{K_3}, \tag{17}$$

which can be used for testing $H_0 : \delta_1^2 = \delta_2^2$. However, it is important to note that, in general, we cannot conduct this test using the usual standard error of $\hat{\lambda}$ which assumes that model 2 is correctly specified. Instead, we need to rely on the misspecification robust standard errors of $\hat{\lambda}_2$ based on (7)–(8) to perform the test of $H_0 : \lambda_2 = 0_{K_3}$.

Alternatively, we can derive the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ and use it for the purpose of testing $H_0 : \delta_1^2 = \delta_2^2$. The following proposition presents the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$.

Proposition 1: Partition $H_2 = (D_2'V_{22}^{-1}D_2)^{-1}$ as

$$H_2 = \begin{bmatrix} H_{2,11} & H_{2,12} \\ H_{2,21} & H_{2,22} \end{bmatrix}, \tag{18}$$

where $H_{2,22}$ is $K_3 \times K_3$. Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2$,

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \stackrel{A}{\sim} \sum_{i=1}^{K_3} \xi_i x_i, \tag{19}$$

where the x_i's are independent χ^2_1 random variables and the ξ_i's are the eigenvalues of $H_{2,22}^{-1}V(\hat{\lambda}_2)$, with $V(\hat{\lambda}_2)$ being the asymptotic variance of $\sqrt{T}(\hat{\lambda}_2 - \lambda_2)$.

Again, we emphasize that the misspecification robust version of $V(\hat{\lambda}_2)$ should be used to test $H_0 : \delta_1^2 = \delta_2^2$. This is because model misspecification tends to create additional sampling variation in $\hat{\delta}_1^2 - \hat{\delta}_2^2$. Without taking into account potential model misspecification, one might mistakenly reject $H_0 : \delta_1^2 = \delta_2^2$. In actual testing, we replace ξ_i with its sample counterpart $\hat{\xi}_i$, where the $\hat{\xi}_i$'s are the eigenvalues of $\hat{H}_{2,22}^{-1}\hat{V}(\hat{\lambda}_2)$, and $\hat{H}_{2,22}$ and $\hat{V}(\hat{\lambda}_2)$ are consistent estimators of $H_{2,22}$ and $V(\hat{\lambda}_2)$, respectively.

2.2 Non-Nested Models

For non-nested models, $\delta_1^2 = \delta_2^2$ can occur under two different scenarios. The first scenario is $y_1 = y_2$, which clearly implies $e_1 = e_2$ and $\delta_1^2 = \delta_2^2$. The second scenario is $y_1 \neq y_2$ (i.e., $e_1 \neq e_2$), but the aggregate pricing errors in the two models are the same — i.e., $e_1' V_{22}^{-1} e_1 = e_2' V_{22}^{-1} e_2$. It turns out that the asymptotic distributions of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ under these two scenarios are different and we have to deal with them separately.

2.2.1 Tests of Equality of Two Stochastic Discount Factors

The condition $y_1 = y_2$ imposes parametric restrictions on η and λ. Suppose we partition η and λ as $\eta = [\eta_1', \ \eta_2']'$ and $\lambda = [\lambda_1', \ \lambda_2']'$, where η_1 and λ_1 are the first K_1 elements of η and λ, respectively. At first sight, it may appear that $y_1 = y_2$ holds if and only if $\eta_1 = \lambda_1$, $\eta_2 = 0_{K_2}$ and $\lambda_2 = 0_{K_3}$. The following lemma shows that the restriction $\eta_1 = \lambda_1$ is redundant because it is implied by the other two restrictions.

Lemma 2: For non-nested models, $y_1 = y_2$ if and only if $\eta_2 = 0_{K_2}$ and $\lambda_2 = 0_{K_3}$.

Note that Lemma 2 is applicable even when the models are misspecified. It suggests that we can test $H_0 : y_1 = y_2$ by simply testing the parametric hypothesis $H_0 : \eta_2 = 0_{K_2}, \ \lambda_2 = 0_{K_3}$. Let $\psi = [\eta_2', \ \lambda_2']'$ and $\hat{\psi} = [\hat{\eta}_2', \ \hat{\lambda}_2']'$. It is easy to establish that the asymptotic distribution of $\hat{\psi}$ under potentially misspecified models is given by

$$\sqrt{T}(\hat{\psi} - \psi) \stackrel{A}{\sim} N(0_{K_2+K_3}, V(\hat{\psi})), \tag{20}$$

where

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}_{t+j}'], \tag{21}$$

with

$$\tilde{h}_t = \begin{bmatrix} H_{1b} D_1' V_{22}^{-1}(r_t - \mu_2)(1 - y_{1t}) + H_{1b}[(x_t - E[x_t]) - D_1' V_{22}^{-1}(r_t - \mu_2)]u_{1t} + \eta_2 \\ H_{2b} D_2' V_{22}^{-1}(r_t - \mu_2)(1 - y_{2t}) + H_{2b}[(x_t - E[x_t]) - D_2' V_{22}^{-1}(r_t - \mu_2)]u_{2t} + \lambda_2 \end{bmatrix}, \tag{22}$$

where $u_{1t} = e_1' V_{22}^{-1}(r_t - \mu_2)$, $u_{2t} = e_2' V_{22}^{-1}(r_t - \mu_2)$, H_{1b} is the last K_2 rows of $(D_1' V_{22}^{-1} D_1)^{-1}$, and H_{2b} is the last K_3 rows of $(D_2' V_{22}^{-1} D_2)^{-1}$.

Suppose that $\hat{V}(\hat{\psi})$ is a consistent estimator of $V(\hat{\psi})$. Then under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$,

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{23}$$

and this can be used as a statistic for testing $H_0 : y_1 = y_2$. Just like in the nested models case, it is important that we conduct this test using the robust standard error of $\hat{\psi}$ based on (21)–(22).

When $y_1 = y_2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ is given by the following proposition.

Proposition 2: Let $H_1 = (D_1' V_{22}^{-1} D_1)^{-1}$ and $H_2 = (D_2' V_{22}^{-1} D_2)^{-1}$, and partition them as

$$H_1 = \begin{bmatrix} H_{1,11} & H_{1,12} \\ H_{1,21} & H_{1,22} \end{bmatrix}, \qquad H_2 = \begin{bmatrix} H_{2,11} & H_{2,12} \\ H_{2,21} & H_{2,22} \end{bmatrix}, \tag{24}$$

where $H_{1,11}$ and $H_{2,11}$ are of dimension $K_1 \times K_1$. Under the null hypothesis $H_0 : y_1 = y_2$, we have

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} \sum_{i=1}^{K_2+K_3} \xi_i x_i, \tag{25}$$

where the x_i's are independent χ^2_1 random variables and the ξ_i's are the eigenvalues of

$$\begin{bmatrix} -H_{1,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & H_{2,22}^{-1} \end{bmatrix} V(\hat{\psi}). \tag{26}$$

Note that (25) allows us to construct a test of $H_0 : y_1 = y_2$ using $\hat{\delta}_1^2 - \hat{\delta}_2^2$. However, it should be pointed out that unlike the Wald test in (23), there are cases in which $y_1 \neq y_2$ but yet (25) fails to reject $H_0 : y_1 = y_2$ with probability one as T goes to infinity.

Before moving on to the case of $y_1 \neq y_2$, a couple of remarks are in order. The first remark is that we can think of the results of the nested models case as a special case of testing $H_0 : y_1 = y_2$

with $K_2 = 0$. The only difference is that the ξ_i's in Proposition 1 are all positive, whereas some of the ξ_i's in Proposition 2 are negative. As a result, we need to perform a two-sided test for the non-nested models case when we use (25) to test $H_0 : y_1 = y_2$. The second remark is more subtle. Unlike (17) and (19), which are tests of $H_0 : \delta_1^2 = \delta_2^2$ for the nested models case, (23) and (25) for the non-nested models case are only tests of $H_0 : y_1 = y_2$. They should not be interpreted as pure tests of $H_0 : \delta_1^2 = \delta_2^2$. This is because $y_1 = y_2$ is a sufficient but not a necessary condition for $\delta_1^2 = \delta_2^2$. We can have $\delta_1^2 = \delta_2^2$ even when $y_1 \neq y_2$, and these cases are taken up in the next subsection.

2.2.2 Tests of Equality of the HJ-Distances of Two Distinct Stochastic Discount Factors

For two distinct non-nested SDFs (i.e., $y_1 \neq y_2$ with positive probability), the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2$ depends on whether (1) both models are correctly specified, or (2) both models are misspecified.

The first case is a little peculiar and it requires some explanation. In the likelihood ratio setting of Vuong (1989), we cannot have two distinct non-nested models that are both correctly specified. One may wonder how two distinct SDFs can be both correctly specified. Two asset pricing models are considered to be correctly specified when they both produce zero pricing errors. This occurs when the vector 1_N is in the span of D_1 as well as in the span of D_2. A simple example of this is when the first model is the correctly specified model and the second model has $f_3 = f_2 + \epsilon$, where ϵ is a vector of pure measurement errors with mean zero and independent of the returns. In this case, $D_2 = E[Rx_2'] = E[Rx_1'] = D_1$ and the second model also produces zero pricing errors even though $y_1 \neq y_2$.

The following proposition presents a simple chi-squared test for testing if both models 1 and 2 are correctly specified.

Proposition 3: Let $n_1 = N - K_1 - K_2$ and $n_2 = N - K_1 - K_3$. Also, let P_1 be an $N \times n_1$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}} D_1$ and P_2 be an $N \times n_2$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}} D_2$. Define

$$g_t(\theta) = \begin{bmatrix} g_{1t}(\eta) \\ g_{2t}(\lambda) \end{bmatrix} = \begin{bmatrix} r_t y_{1t}(\eta) \\ r_t y_{2t}(\lambda) \end{bmatrix}, \tag{27}$$

where $\theta = [\eta',\ \lambda']'$, and

$$S = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)'] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \tag{28}$$

When $y_1 \neq y_2$ and under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 = 0$,

$$T \begin{bmatrix} \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ \hat{P}_2' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_2 \end{bmatrix}' \begin{bmatrix} \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{11} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_1 & \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{12} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_2 \\ \hat{P}_2' \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{21} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_1 & \hat{P}_2' \hat{V}_{22}^{-\frac{1}{2}} \hat{S}_{22} \hat{V}_{22}^{-\frac{1}{2}} \hat{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}_1' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_1 \\ \hat{P}_2' \hat{V}_{22}^{-\frac{1}{2}} \hat{e}_2 \end{bmatrix} \overset{A}{\sim} \chi^2_{n_1+n_2}, \tag{29}$$

where $\hat{e}_1$ and $\hat{e}_2$ are the sample pricing errors of models 1 and 2, and $\hat{P}_1$, $\hat{P}_2$, $\hat{S}$ are consistent estimators of P_1, P_2, and S, respectively.

When $y_1 \neq y_2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both models are correctly specified is given in the following proposition.

Proposition 4: Using the notation in Proposition 3, when $y_1 \neq y_2$ and under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 = 0$,

$$T(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} \sum_{i=1}^{n_1+n_2} \xi_i x_i, \tag{30}$$

where the x_i's are independent χ_1^2 random variables and the ξ_i's are the eigenvalues of

$$\begin{bmatrix} P_1' V_{22}^{-\frac{1}{2}} S_{11} V_{22}^{-\frac{1}{2}} P_1 & P_1' V_{22}^{-\frac{1}{2}} S_{12} V_{22}^{-\frac{1}{2}} P_2 \\ -P_2' V_{22}^{-\frac{1}{2}} S_{21} V_{22}^{-\frac{1}{2}} P_1 & -P_2' V_{22}^{-\frac{1}{2}} S_{22} V_{22}^{-\frac{1}{2}} P_2 \end{bmatrix}. \tag{31}$$

Note that the ξ_i's are not all positive, because $\hat{\delta}_1^2 - \hat{\delta}_2^2$ can be negative. Therefore, we need to perform a two-sided test of $H_0 : \delta_1^2 = \delta_2^2$ instead of a one-sided test, as in the nested models case.

Finally, similar to the asymptotic distribution of $\hat{\delta}^2$, the asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ changes when the models are misspecified. Consequently, we cannot use Proposition 4 to test $H_0 : \delta_1^2 = \delta_2^2$ when the models are misspecified. Proposition 5 presents the appropriate asymptotic distribution of $\hat{\delta}_1^2 - \hat{\delta}_2^2$ when both non-nested models are misspecified and $y_1 \neq y_2$.

Proposition 5: Suppose $y_1 \neq y_2$. Let $d_t = q_{1t} - q_{2t}$, where

$$\begin{aligned} q_{1t} &= 2u_{1t}y_{1t} - u_{1t}^2 + \delta_1^2, \\ q_{2t} &= 2u_{2t}y_{2t} - u_{2t}^2 + \delta_2^2, \end{aligned}$$

with $u_{1t} = e_1' V_{22}^{-1}(r_t - \mu_2)$ and $u_{2t} = e_2' V_{22}^{-1}(r_t - \mu_2)$. When $\delta_1 \neq 0$ and $\delta_2 \neq 0$,

$$\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2 - (\delta_1^2 - \delta_2^2)) \overset{A}{\sim} N(0, v_d), \tag{32}$$

where

$$v_d = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]. \tag{33}$$

Under the null hypothesis $H_0 : \delta_1^2 = \delta_2^2 \neq 0$,

$$\sqrt{T}(\hat{\delta}_1^2 - \hat{\delta}_2^2) \overset{A}{\sim} N(0, v_d) \tag{34}$$

and d_t can be simplified to

$$d_t = 2u_{1t}y_{1t} - u_{1t}^2 - 2u_{2t}y_{2t} + u_{2t}^2. \tag{35}$$

The expression of d_t in (35) reveals that there are situations in which one cannot use the normal test in Proposition 5 to test $H_0 : \delta_1^2 = \delta_2^2$. This can happen when (1) $y_{1t} = y_{2t}$, which implies $u_{1t} = u_{2t}$ and hence $d_t = 0$; or (2) $y_{1t} \neq y_{2t}$ but both models are correctly specified—i.e., $u_{1t} = u_{2t} = 0$, which also leads to $d_t = 0$.