Methods for variance esitmation of high dimensional data

Xuelong Wang

2020-01-02

Contents

1 Motivation

2 Linear regression

3 GCTA method

3.1 GCTA approach

The GCTA approach estimates variances of weak effects…

3.1.1 Model assumption

GCTA approach is built on a linear model:

$$y_i = \mu_i + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i. \tag{1}$$

where y_i denotes a outcome (quantitative measurement) and x_{ij} , $j = 1, \ldots, p$ are the standardized covariates measurements for subject i. Besides we also assume the independence between the covaraites and error terms, $\epsilon_i \perp\!\!\!\perp x_{jk}$.The equation 1 may be re-expressed as

$$Y = \mu + X\beta + \epsilon. \tag{2}$$

where X is a $n \times p$ matrix with element as x_{ij}, $Y = (y_1, \dots, y_n)^T$, $\mu = (\mu_1, \dots, \mu_n)^T$, $\beta = (\beta_1, \dots, \beta_p)^T$,and $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$.

The goal here is to estimate how much variation of the outcome is accounted for by the covariates. Based on the assumptions of model 1, the variance of y could be composed into two parts,

$$Var(y_i) = Var(X_i^T\beta) + Var(\epsilon) = \sigma_\beta^2 + \sigma_\epsilon^2.$$

The σ_β^2 is the response's variability counted by covariates. Since $x_i's$ are independent, we have $\sigma_\beta^2 = \sum_{i=1}^p \beta_i$. To estimate the variance, one intuitively approach is to estimate the σ_ϵ^2 by a regression model. But it may not feasible when the number of covaraites is large. The GCTA approach uses a working random effects model to estimate σ_β^2 without knowing the error terms variance.

3.1.2 Advantages of the GCTA apporach

For the environmental study mentioned before, the GCTA method demonstrates some advantages than other variance estimation approaches.

- a working random effects model to estimate $Var(X\beta)$
- Don't need to select the casual covariates, so that could work with weak signal problem
- relatively little bias compared to other methods

3.1.3 Two more Obstacles

Although we discussed the GCTA approach could be a good tool for the environmental health analysis, there are still issues we need to tackle.

- Theoretical analysis of The GCTA approach suggests the Independence of causal covariates, but most of the environmental data are high correlated.
- In SNP studies, the number of covariates is large and the number of interactive terms is also going to be very large, which makes the interactive effect even harder to be estimated. Therefore, interaction effect usually is not considered in SNPs studies. Although in environmental studies the number of predictors is not large (within 40), directly applying the GCTA method to estimate the interactive effect still hardly guarantee good performance.

3.2 The proposed method

With those two problems in mind, we develop a new method by modifying the GCTA method for correlated covariates. The main idea is to transform the correlated covariates into uncorrelated ones. The transformation process is also called decorrelation. We consider a linear transformation so that the transformation does not change the variance structure.

3.2.1 Transformation for correlated covariates

The linear transformation is

$$Z = A^{-1}X,$$

where X are the covariates vector,A is a linear transformation operator which is a full rank square matrix. After transformation, the covariance of the new covariates Z will be

$$Var(Z) = I_p.$$

Moreover, based on the model assumed by GCTA (model 2), we have

$$Y = \mu + X^T\beta + \epsilon = Z^T A^T \beta + \epsilon = Z^T\alpha + \epsilon,$$

where $\alpha = A^T\beta$. Let's look the total effect of *X and *Z:

$$Var(X^T\beta) = Var(Z^T A^T\beta) = Var(Z^T\alpha).$$

Therefore, the Z will be the uncorrelated predictors and $Z^T\alpha$ should keep the same total cumulative effect as $X^T\beta$. If X follows a normal distribution, i.e. $X \sim N(0, \Sigma)$, then the $Z \sim N(0, I_p)$. Therefore, Z's elements are independent to each other, which is the exact condition we want for the GCTA appraoch. Although for non-normal covariates the decorrelation procedure only reduces linear association with no guarantee of independence, it still can improve the performance of GCTA method.

3.2.2 Decorrelation procedure

There are many methods and algorithms for data decorrelation. One of commonly used methods is to apply the eigenvalue decomposition to the covariance matrix. Let Σ_X be the covariance matrix of X, so Σ_X is a symmetric and positive-definite. Then eigenvalue decomposition of Σ_X will be

$$Var(X) = \Sigma_X = U\Lambda U^T,$$

where X is the random vector with dim as $p \times 1$, Σ_X is $p \times p$ symmetry and p.d. matrix, Λ is a diagonal matrix with each diagonal element as the eigenvalue. If the Σ_X is full rank, then we could just take the reciprocal of each square root of eigenvalue as following.

$$\Sigma_X^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}U^T, \text{ and } \Lambda^{-\frac{1}{2}} = \begin{bmatrix} \lambda_1^{-\frac{1}{2}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p^{-\frac{1}{2}} \end{bmatrix}.$$

So that after transformation, the $\Sigma_X^{-\frac{1}{2}}X$ will have an identity covariance matrix as following:

$$Var(\Sigma_X^{-\frac{1}{2}}X) = \Sigma_X^{-\frac{1}{2}}\Sigma_X\Sigma_X^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}U^T U\Lambda^{-1}U^T U\Lambda^{-\frac{1}{2}}U^T = I_p.$$

If the Σ_X is not full rank, then we can still use the eigenvalue decomposition. But the procedure cannot guarantee the identity covariance matrix anymore. The reason is that some eigenvalues will be zero, so we can not take the reciprocal. One straightforward solution is just leave them there:

$$Var(X) = \Sigma_X = U\Lambda U^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1\Lambda_1U_1^T,$$

where U_1 is a $p \times r$ matrix with $r < p$ and in most of case r = n the sample size. Then after applying the same procedure we get following,

$$\tilde{\Sigma}_X^{-\frac{1}{2}} = U_1\Lambda_1^{-\frac{1}{2}}U_1^T,$$

Where $\tilde{\Sigma}_X^{-1}$ is the Moore Penrose inverse. After transformation the X we have,

$$Var(\tilde{\Sigma}_X^{-\frac{1}{2}}X) = \tilde{\Sigma}_X^{-\frac{1}{2}}\Sigma_X\Sigma_X^{-\frac{1}{2}} = U_1\Lambda_1^{-\frac{1}{2}}U_1^T = U_1U_1^T,$$

Where $U_1U_1^T + U_2U_2^T = I_p$ and $(U_1U_1^T)^T U_1U_1^T = U_1U_1^T$. Besides, $U_1U_1^T$ and $U_2U_2^T$ are indempotent and $rank(U_2U_2^T) + rank(U_1U_1^T) = p$.

3.3 Interactive effect

For analysing the interactive effect, we need to consider interactive terms in our model. Let's just consider a 2-way interaction model

$$y_i = \mu_i + \sum_{j=1}^{p} x_{ij}\beta_j + \sum_{l \neq k} \gamma_{lk} x_{il} x_{ik} + \epsilon_i, \tag{3}$$

where γ_{jk} denotes interactive coefficients. Anything else will be same as GCTA model 2. This model also can be expressed in the matrix form

$$Y = \mu + X\beta + X\Gamma X^T + \epsilon. \tag{4}$$

Where Γ is a $p \times p$ matrix with element as γ_{jk}. Let's also assume that $X_i \perp\!\!\!\perp \epsilon_i$, then the variance of y_i can be decompose as following

$$Var(y_i) = Var(X_i^T\beta) + Var(X_i^T\Gamma X_i) + 2Cov(X_i^T\beta, X_i^T\Gamma X_i) + Var(\epsilon_i). \tag{5}$$

After adding the interactive terms, the situation becomes complicated.

1. Besides the interactive effect there is an additional covariance term of $X_i^T\beta, X_i^T\Gamma X_i$ to deal with.
2. The main and interactive terms are bonded to be dependent, even thought all elements X are independent. Same situation for the 2-way interactive terms, they are also dependent.

As we mentioned before, independence of covariates is suggested for GCTA approach to work well, so we cannot guarantee the performance of GCTA approach.

To handle the covariance terms, we now focus on the situations where $Cov(X_i^T\beta, X_i^T\Gamma X_i) = 0$, so that we don't have to worry about it. For the cases where we cannot ignore the covariance term, it's hard to estimate both of the effects well. The reason is that the covariance term will be somehow mixed into both main and interactive effect estimations, so it is not easy to separate covariance part from the effects' estimation. We will discuss it latter in this paper. Let's just assume that covariates are independent and centered to each other and there is no square terms in the model 3, e.i. $\gamma_{jj} = 0$,

$$\begin{aligned}
Cov(X^T\beta, X^T\Gamma X) &= E[(X^T\beta - E(X^T\beta))(X^T\Gamma X - E(X^T\Gamma X))] \\
&= E[X^T\beta(X^T\Gamma X - E(X^T\Gamma X)) \\
&= E[X^T\beta \cdot X^T\Gamma X] \\
&= E[(\sum_{h}^{p}(x_h\beta_h))(\sum_{j}^{p}\sum_{k}^{p}\gamma_{jk}x_jx_k)] \\
&= 0
\end{aligned}$$

For the second issues, we extend our proposed approach to handle the interactive terms. Although it's impossible to make the interactive terms independent with themselves or the main terms, we still can transform them into uncorrelated. Therefore, we could combine the main and interactive term together as a larger covariate matrix

$$X_{i,t} = \begin{bmatrix} X_i \\ X_{i,inter} \end{bmatrix},$$

where $X_{i,inter} = (x_{i1}x_{i2}, \dots, x_{i(p-1)}x_{ip})^T$. Then apply the decorrelation process on the combined matrix $X_t = (X_{1t}, \dots, X_{nt})^T$. Given the independence of the covariates, simulation studies have shown that the proposed method could estimate both of the cumulative and interactive effect with little bias. This also suggests that the uncorrelation of covariates may be good enough to let the GCTA works appropriately. Therefore, we may release the condition from independent to uncorrelated covariates.

3.4 available software

3.5 Simulation study compare two GCTA and GCTA_rr

GCTA_rr is the `mixed.solve` function from `rrBLUP` r package.
Based on the following simulation results,

1. when $n < p$ case, those two methods' results are very closed to each other.
2. when $n > p$ case, in terms of effect estimation and jackknife variance estimation those two methods's reuslts are similar to each other. But for the variance corrections are quite different. That is the statistics Q of our method has a very large variance which leads to negative correction result.

3.5.1 setup

- Independent
- Normal
- $p = 100$
- $n = \{50, 75, 100, 150, 200\}$
- with interaction terms
- main effect: $Var(X^T\beta) = \{0, 8, 100\}$

3.5.2 Simulation result

3.5.3 $Var(X^T\beta) = \{0\}$

```
     n  MSE est_var est_mean NA_main GCTA_main_jack GCTA_v_jack_1
1:  50 3.40    1.83     1.32       0           0.59          9.55
2:  75 1.19    0.98     0.56       0           0.46          2.73
3: 100 1.08    0.84     0.57       0           0.44          1.35
4: 150 0.28    0.19     0.32       0          -1.09          0.86
5: 200 0.21    0.12     0.32       0          -1.60          0.78
   GCTA_v_jack_2 GCTA_v_corr
1:          9.62       -9.28
2:          2.68       -5.64
3:          1.35       -0.77
4:          0.67      -64.08
5:          0.69      -46.14
```

```
     n  MSE est_var est_mean NA_main GCTA_rr_main_jack GCTA_rr_v_jack_1
1:  50 3.40    1.83     1.32       0              0.60             9.55
2:  75 1.19    0.98     0.56       0              0.46             2.73
3: 100 1.08    0.84     0.57       0              0.44             1.35
4: 150 0.28    0.19     0.33       0             -0.17             0.62
5: 200 0.21    0.12     0.33       0              0.28             0.61
   GCTA_rr_v_jack_2 GCTA_rr_v_corr
1:             9.47         -3.560
2:             2.68         -5.643
3:             1.35         -0.770
4:             0.61         -1.204
5:             0.61         -0.041
```

3.5.4 $Var(X^T\beta) = \{100\}$

```
     n   MSE est_var est_mean NA_main GCTA_main_jack GCTA_v_jack_1
1:  50  9247    1784       87       0             66          8795
2:  75 10077    1863       92       0            103          5170
3: 100 11839    2142      100       0             84          2072
4: 150 10953     443      103       0             31          1280
5: 200  9778     245       98       0             30           725
   GCTA_v_jack_2 GCTA_v_corr
1:          8793       -3687
2:          5109       -3122
3:          2081         194
4:          1148      -80475
5:           673      -32124
```

```
     n   MSE est_var est_mean NA_main GCTA_rr_main_jack GCTA_rr_v_jack_1
1:  50  9247    1784       87       0                66             8795
2:  75 10077    1863       92       0               103             5170
3: 100 11839    2142      100       0                84             2072
4: 150 11194     414      104       0               103              969
5: 200  9854     238       98       0                98              616
   GCTA_rr_v_jack_2 GCTA_rr_v_corr
1:             8787          -3492
2:             5109          -3124
```

3:	2081	194
4:	970	158
5:	616	220

3.5.5 $Var(X^T\beta) = \{8\}$

	n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack_1	GCTA_v_jack_2	GCTA_v_corr
1:	50	90	25.8	8.0	0	8.5	74.1	73.8	-190.67
2:	75	70	13.1	7.5	0	7.5	32.1	31.9	-25.67
3:	100	68	6.3	7.8	0	7.5	13.7	13.8	-0.97
4:	150	70	4.0	8.1	0	8.4	9.2	8.1	-502.59
5:	200	65	2.5	7.9	0	7.6	4.6	4.3	-214.51

	n	MSE	est_var	est_mean	NA_main
1:	50	24.0	24.0	8.0	0
2:	75	13.8	13.8	7.9	0
3:	100	8.6	8.6	8.1	0
4:	150	3.7	3.7	8.0	0
5:	200	2.7	2.7	8.0	0

	n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack_1	GCTA_rr_v_jack_2	GCTA_rr_v_corr
1:	50	90	25.8	8.0	0	8.5	74.1	73.6	-177.35
2:	75	70	13.1	7.5	0	7.5	32.1	31.8	-16.78
3:	100	68	6.3	7.8	0	7.5	13.7	13.8	-0.97
4:	150	70	4.1	8.1	0	8.1	6.9	6.9	1.49
5:	200	65	2.5	7.9	0	7.9	3.9	3.9	1.38

	n	MSE	est_var	est_mean	NA_main
1:	50	23.8	23.9	8.0	0
2:	75	13.7	13.7	7.9	0
3:	100	8.6	8.6	8.1	0
4:	150	3.8	3.8	8.0	0
5:	200	2.7	2.7	8.1	0

3.5.6 correlation test $

	n	MSE	est_var	est_mean	NA_main	cor_main_jack	cor_v_jack_1
1:	50	0.0131	0.0130	0.49	0	0.49	0.0127
2:	75	0.0083	0.0083	0.50	0	0.50	0.0079
3:	100	0.0057	0.0057	0.50	0	0.50	0.0059
4:	150	0.0038	0.0038	0.50	0	0.50	0.0039
5:	200	0.0030	0.0030	0.50	0	0.50	0.0029

	cor_v_jack_2	cor_v_corr
1:	0.0128	0.0120
2:	0.0079	0.0076
3:	0.0059	0.0057
4:	0.0039	0.0038
5:	0.0029	0.0029

3.6 compare the performance of delete 1 and delete d in variance estimation

The delete-d jackknife varinace estimator is

$$\sqsubseteq_{J(d)} = \frac{n-d}{d} \cdot \frac{1}{S} \sum_{S} (\hat{\theta}_s - \hat{\theta}_{s.})$$

, where $S = \binom{n}{d}$. Note that S could a very large value, so in the following simulation, only $S = 1000$ is used. In Jun Shao's another paper, he proposed an approximation of the deletel-d variance estimation. That is just select m from $S = \binom{n}{d}$ sub-samples and in that paper it recommended $m = n^{1.5}$.

3.6.1 setup

- Independent
- Normal
- $p = \{100, 1000\}$
- $n = \{50, 75, 100, 150, 200, 500, 750, 1000, 1500, 2000\}$
- $d = 0.5 \times n$
- $n_{repeat} = 1000$ for delete d jackknife
- main effect: $Var(X^T\beta) = 8$

3.6.2 GCTA with p = 100

n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack	GCTA_v_jack_var	d	n_sub
50	25.6	25.8	8.0	0	8.5	74.1	8383.8	1.0	NA
75	13.2	13.1	7.5	0	7.5	32.1	685.1	1.0	NA
100	6.2	6.3	7.8	0	7.5	13.7	102.2	1.0	NA
150	4.0	4.0	8.1	0	8.4	9.2	16.4	1.0	NA
200	2.5	2.5	7.9	0	7.6	4.6	2.1	1.0	NA
50	25.6	25.8	8.0	0	45.5	41.2	365.2	0.5	NA
75	13.2	13.1	7.5	0	-177.5	27.1	99.7	0.5	NA
100	6.2	6.3	7.8	0	-237.3	18.5	38.1	0.5	NA
150	4.0	4.0	8.1	0	-13.8	9.4	7.5	0.5	NA
200	2.5	2.5	7.9	0	17.3	5.0	1.4	0.5	NA
50	25.6	25.8	8.0	0	35.1	41.1	366.6	0.5	354
75	13.2	13.1	7.5	0	-107.6	27.0	100.1	0.5	650
100	6.2	6.3	7.8	0	-237.3	18.5	38.1	0.5	1000
150	4.0	4.0	8.1	0	-20.2	9.3	7.0	0.5	1837
200	2.5	2.5	7.9	0	53.4	5.1	1.3	0.5	2828

3.6.3 GCTA with p = 1000

n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack	GCTA_v_jack_var	d
500	2.88	2.91	8.0	0	7.8	4.65	1.08	1.0
750	1.29	1.30	8.0	0	8.0	2.26	0.15	1.0
1000	0.77	0.78	8.0	0	8.0	1.28	0.04	1.0
1500	0.47	0.48	7.9	0	6.7	0.80	0.01	1.0
500	2.88	2.91	8.0	0	-79.1	6.56	1.17	0.5
750	1.29	1.30	8.0	0	-5.9	3.04	0.13	0.5
1000	0.77	0.78	8.0	0	40.8	1.71	0.05	0.5
1500	0.41	0.41	8.0	0	9.9	0.80	0.01	0.5
2000	0.31	0.31	8.0	0	25.6	0.48	0.00	0.5

3.6.4 GCTA_rr_rr with p = 100

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d	n_sub
50	25.6	25.8	8.0	0	8.5	74.1	8378.6	1.0	NA
75	13.2	13.1	7.5	0	7.5	32.1	685.3	1.0	NA
100	6.2	6.3	7.8	0	7.5	13.7	102.2	1.0	NA
150	4.1	4.1	8.1	0	8.1	6.9	8.5	1.0	NA
200	2.5	2.5	7.9	0	7.9	3.9	1.3	1.0	NA
50	25.6	25.8	8.0	0	52.5	40.6	363.1	0.5	NA
75	13.2	13.1	7.5	0	-198.0	26.6	100.2	0.5	NA
100	6.2	6.3	7.8	0	-257.6	18.1	38.6	0.5	NA
150	4.1	4.1	8.1	0	-11.9	9.3	7.5	0.5	NA
200	2.5	2.5	7.9	0	25.4	5.0	1.4	0.5	NA
50	25.6	25.8	8.0	0	35.2	40.5	363.4	0.5	354
75	13.2	13.1	7.5	0	-120.5	26.6	100.8	0.5	650
100	6.2	6.3	7.8	0	-257.6	18.1	38.6	0.5	1000
150	4.1	4.1	8.1	0	-17.0	9.3	7.1	0.5	1837
200	2.5	2.5	7.9	0	76.2	5.1	1.3	0.5	2828

3.6.5 GCTA_rr with p = 1000

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d
500	2.88	2.91	8.0	0	7.8	4.65	1.08	1.0

(continued)

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d
750	1.29	1.30	8.0	0	8.0	2.26	0.15	1.0
1000	0.77	0.78	8.0	0	8.0	1.28	0.04	1.0
1500	0.48	0.48	7.9	0	8.0	0.62	0.00	1.0
500	2.88	2.91	8.0	0	-79.1	6.56	1.17	0.5
750	1.29	1.30	8.0	0	-5.9	3.04	0.13	0.5
1000	0.77	0.78	8.0	0	40.8	1.71	0.05	0.5
1500	0.41	0.41	8.0	0	11.8	0.80	0.01	0.5
2000	0.31	0.31	8.0	0	24.4	0.48	0.00	0.5

3.6.6 cor with n = 200

n	MSE	est_var	est_mean	NA_main	cor_main_jack	cor_v_jack	d
50	0.01252	0.01265	0.50001	0	0.50432	0.01229	1.0
75	0.00774	0.00782	0.50050	0	0.50323	0.00815	1.0
100	0.00607	0.00613	0.50148	0	0.50334	0.00582	1.0
150	0.00383	0.00385	0.49584	0	0.49709	0.00391	1.0
200	0.00281	0.00284	0.49930	0	0.50027	0.00288	1.0
50	0.01252	0.01265	0.50001	0	5.32154	0.01282	0.5
75	0.00774	0.00782	0.50050	0	3.26213	0.00844	0.5
100	0.00607	0.00613	0.50148	0	2.50378	0.00595	0.5
150	0.00383	0.00385	0.49584	0	1.59064	0.00396	0.5
200	0.00281	0.00284	0.49930	0	1.46439	0.00293	0.5

3.6.7 median with n = 200

n	MSE	est_var	est_mean	NA_main	median_main_jack	median_v_jack	d
50	0.03138	0.03135	-0.00775	0	-0.00775	0.06818	1.0
75	0.02211	0.02212	-0.00212	0	0.05228	0.03113	1.0
100	0.01523	0.01523	-0.00378	0	-0.00378	0.02720	1.0
150	0.01072	0.01072	-0.00279	0	-0.00279	0.01885	1.0
200	0.00804	0.00804	-0.00051	0	-0.00051	0.01614	1.0
50	0.03138	0.03135	-0.00775	0	5.04459	0.03477	0.5
75	0.02211	0.02212	-0.00212	0	8.44376	0.02248	0.5
100	0.01523	0.01523	-0.00378	0	2.68868	0.01587	0.5
150	0.01072	0.01072	-0.00279	0	-1.47581	0.01110	0.5
200	0.00804	0.00804	-0.00051	0	3.96797	0.00827	0.5

4 EigmPrism method

5 Variance estimation in hihg-dimensional linear models

5.1 Model assumption

5.2 Signal Esitmation for $\Sigma = I$

$$E\left(\frac{1}{n}\|y\|^2\right) = \tau^2 + \sigma^2, \quad E\left(\frac{1}{n^2}\left\|X^{\mathrm{T}}y\right\|^2\right) = \frac{d+n+1}{n}\tau^2 + \frac{d}{n}\sigma^2$$

After some linear algebra, we have the corresponding estimator is

$$\hat{\sigma}^2 = \frac{d+n+1}{n(n+1)}\|y\|^2 - \frac{1}{n(n+1)}\left\|X^{\mathrm{T}}y\right\|^2, \quad \hat{\tau}^2 = -\frac{d}{n(n+1)}\|y\|^2 + \frac{1}{n(n+1)}\left\|X^{\mathrm{T}}y\right\|^2$$

Under some standard condition the estimators have asymptotic normality.

$$\psi_1^2 = 2\left\{\frac{d}{n}\left(\sigma^2+\tau^2\right)^2+\sigma^4+\tau^4\right\}$$
$$\psi_2^2 = 2\left\{\left(1+\frac{d}{n}\right)\left(\sigma^2+\tau^2\right)^2-\sigma^4+3\tau^4\right\}$$
$$\psi_0^2 = \frac{2}{\left(\sigma^2+\tau^2\right)^2}\left\{\left(1+\frac{d}{n}\right)\left(\sigma^2+\tau^2\right)^2-\sigma^4\right\}$$

If $d/n \to \rho \in [0,\infty)$, then
$n^{1/2}\left(\frac{\hat{\sigma}^2-\sigma^2}{\psi_1}\right), n^{1/2}\left(\frac{\hat{\tau}^2-\tau^2}{\psi_2}\right), n^{1/2}\left(\frac{\hat{r}^2-r^2}{\psi_0}\right) \to N(0,1)$
in distribution.