True and Approximating Models and OCMT

Virtues of OCMT

1. Attempt to apply the multiple testing concept on variable selection

2. Rule out the Psuedo signals and Noise signals, while snatching the signals and Hidden signals

3. Applicable in extreme cases where # of variables are larger then the # of observations

Basic Settings and Notations

Based on the data generating process(DGP), as shown below:

$$y_t = a^{'} z_t + \sum_{i=1}^{k} \beta_i x_{it} + u_t$$

z_t is a known vector of preselected constant, linear trend, dummy variables, and stochastic variables.

x_{it} is an unknown vector of finite variables.
i.e. $i = 1, 2, \cdots, n$, n is the # of all variables and $n < \infty$

u_t is the error term.

z_t and x_{it} are assumed to be independent with u_t at time t

True Model and Approximating Model

Variables x_{it} can be split into: Signals, Hidden signals, Psuedo Signals, and Noise

k signals are in the set $S_{nt} = \{x_{it}, i = 1, 2, \cdots, n\}$

$n - k$ variables are the rest of the active set (S_{nt}), which are Psuedo signals (k^*) and Noise ($n - k - k^*$).

$\mathbf{T}$ is the matrix of obs.

If:

> The model only have Signals and Hidden Signals, the model is the *True Model*
>
> The model have all signals except Noise Signals, the model is the *Approximating Model*

OCMT and the OLS

The OCMT is a multiple staged process.

Each stage performs multiple testing and significant tests on each variable.

LS regression of y_t on z_t and the regressors in S_{nt} one at a time

For significant testings, t_i replaces the t-ratio, for $i = 1, 2, \cdots, n$

$$t_i = \frac{\mathbf{T}^{-\frac{1}{2}} x_i' \mathbf{M}_z y}{\hat{\sigma}_i \sqrt{\mathbf{T}^{-1} x_i' \mathbf{M}_z x_i}} = \frac{\mathbf{T}^{-\frac{1}{2}} x_i' \mathbf{M}_z \mu}{\hat{\sigma}_i \sqrt{\mathbf{T}^{-1} x_i' \mathbf{M}_z x_i}} + \frac{\mathbf{T}^{-\frac{1}{2}} x_i' \mathbf{M}_z u}{\hat{\sigma}_i \sqrt{\mathbf{T}^{-1} x_i' \mathbf{M}_z x_i}} = t_{i,\mu} + t_{i,u}$$

where $\mathbf{M}_z$ is the matrix of obs. on z_t
and $\mu = (\mu_1, \mu_2, \cdots, \mu_{\mathbf{T}})'$, $\mu_t = \sum_{i=1}^{k} \beta_i x_{it}$

Why $\boldsymbol{t_i}$?

As $n, \mathbf{T} \to \infty$, we rely $t_{i,u}$ to remain bounded in probability sufficiently sharply for multiple testing over extreme large n

<Case 1> - $t_{i,u}$ bounded in prob. sufficiently sharply
x_{it} not contained in $u_t \to$ Noise Signal
x_{it} contained in $u_t \to$ Hidden Signal

<Case 2> - $t_{i,u}$ NOT bounded in prob. sufficiently sharply
x_{it} net of z_t and contained in $u_t, \to$ Signal
x_{it} net of z_t and not contained in $u_t, \to$ Psuedo Signal

OCMT can identify all signal types with probability approaching 1.

Conditional Net Impact of $\boldsymbol{x}_{it}$

The OLS focuses on marginal effects of β_i

OCMT focuses on impact of x_{it} on y_t conditional on z_t

Generalizing the mean net impact coefficient by Pesaran and Smith (2014)

$$\theta_{i,\mathbf{T}}(z) = \sum_{j=1}^{k} \beta_j \sigma_{ij,\mathbf{T}}(z)$$

Psuedo? Noise? Signals?

Signals: True significant variables

Noise Signals: variables that have 0 correlation with signals after filtering z_t

Psuedo Signals: variables are correlated with signals net z_t

	$\boldsymbol{\theta_i(z) \neq 0}$	$\boldsymbol{\theta_i(z) = 0}$
$\beta_i \neq 0$	Signals	Hidden Signals
$\beta_i = 0$	Psuedo Signals	Noise variables

Table 1. Simple Display of Different Variable Types

Notes on the Signals

k Signals, $\quad k^*$ Psuedo Signals, $\quad n-k-k^*$ Noise Signals

assume:

1. k is an unknown fixed constant
2. k^* can rise with n such that $\frac{k^*}{n} \to 0$ and $\frac{k^*}{\mathbf{T}} \to 0$

Suppositions of OCMT

Consider a simplified DGP:

$$y_t = a + \sum_{t=1}^{k} \beta_i x_{it} + u_t \quad \text{for } t = 1, 2, \cdots, \mathbf{T}$$

Notice that z_t contains an intercept only
the net impact coefficient $\theta_i(z)$ is now:

$$\theta_i = \sum_{j=1}^{k} \beta_j \sigma_{ij}$$

Further Assumptions on Simplified DGP

1. Only Signals and Psuedo Signals in X_{k,k^*}
2. The error term u_t is a martingale difference process, with a zero mean and finite constant variance.
3. Let x_{it} be a martingale difference process, which is weaker than iid for penalized regression.
4. Set exponential bounds for x_{it} and u_t, crucial to ensure all moments of both exists.

Further Assumptions on Simplified DGP pt.2

1. Technical Condition that is required for some parts of the OCMT theory (more related to the Appendix and Supplemental Material.)
   It is set for a more generic case of the OCMT.

2. The number of signals, k, is finite and their slope coefficient can change with $\mathbf{T}$.
   It allows the weak signal variables, whose $\beta_{i,\mathbf{T}}$ for $i = 1, 2, \cdots, k$ declines with the sample size $\mathbf{T}$, to exists.

Notes on Variables

Signals and Noise can be correlated amongst themselves.
Signals and Psuedo Signals can also be correlated.

Thus,

Signal, Psuedo Signal, and Noise can contain common factors, but Signal/Psuedo will not share factors with Noise.

If a common factor impacts both signals and a large number of remaining variables in $S_{n,t}$, condition on it to control the approximating model size.

Contrast With LASSO

1. Stochastic regressors require specific conditions
2. Max bound for all β_is
3. Assume the tuning parameters of the penalty function is known.

Process of OCMT - 1st Stage

Run n simple regressions of y_t on z_t and $x_{it}, i = 1, \cdots, n$

$$y_t = c_i + \phi_i x_{it} + u_{it}, \quad t = 1, 2, \cdots, \mathbf{T}$$

where $\phi_i = \frac{\theta_i}{\sigma_{ii}}$ and θ_i is the condition net impact of variable i

The t-ratio for ϕ_i is:

$$t_{\hat{\phi}_i,(1)} = \frac{\hat{\phi}_i}{\text{s.e.}(\hat{\phi}_i)} = \frac{x_i' \mathbf{M}_\tau y}{\hat{\sigma}_i \sqrt{x_i' \mathbf{M}_\tau x_i}}$$

Process of OCMT - 1st Stage pt.2

Critical value function:

$$c_p(n, \delta) = \Phi^{-1}(1 - \frac{p}{2cn^\delta})$$

where δ and c are positive constants
and p is the nominal size of the single testings(also known as α), $0 < p < 1$

Define an selection indicator:

$$\hat{\mathcal{J}}_{i,(1)} = I[|t_{\hat{\phi}_{i,(1)}}| > c_p(n, \delta)] \quad \text{for} \quad i = 1, 2, \cdots, n$$

variables with $\hat{\mathcal{J}}_{i,(1)} = 1$ are selected as Signals or Psuedo Signals.

The remaining active set will continue to the following stage.

Something on δ

Two values of δ

δ for the first stage and δ^* for the other stages

Required that $\delta^* > \delta$

$c_p(n, \delta)$ heavily influenced by the value of δ

Something on $c_p(n, \delta)$

$c_p(n, \delta)$ is set to control the probability of Type I Error occurence

$\delta = 1$ is the Bonferroni correction ... many other choices

Why the current $c_p(n, \delta)$?

1. Most impose restriction on the dependence structure between multiple tests
2. The # of tests is not predetermined with OCMT

Process of OCMT - Following Stages

Let $\hat{k}_{j-1}$ be the # of selected variables of $j-1$ th stage, $\quad j = 2, 3, \cdots$

Run $n - \hat{k}_{j-1}$ regressions of y_t on z_t and the already selected variables and, one at a time, the not yet seleceted variables of the active set

The t-ratio is now:

$$t_{\hat{\phi}_i,(j)} = \frac{\hat{\phi}_{i,(j)}}{\mathrm{s.e.}(\hat{\phi}_{i,(j)})} = \frac{x_i' \mathbf{M}_{(j-1)} y}{\hat{\sigma}_{i,(j)} \sqrt{x_i' \mathbf{M}_{(j-1)} x_i}}$$

Process of OCMT - Following Stages pt.2

The critical value function is now $c_p(n, \delta^*)$

Selection Indicator function is still $\hat{\mathcal{J}}_{i,(j)} = I[|t_{\hat{\phi}_{i,(j)}}| > c_p(n, \delta^*)]$

The process will continue until the stage that $\sum_{i=1}^{n} \hat{\mathcal{J}}_{i,(j)} = 0$

Notes on the Process

The # of stages is bounded in n

Snatching the Signals unravels the Hidden Signals

The OCMT estimator of β_i, denoted by $\tilde{\beta}_i$, is:

The LS estimator of the β_i in a regression of y_t on z_t and all the selected variables.

Comparison With Other Models

L_2 - Boosting

Buhlmann (2006)

1. Start with n bivariate regressions
2. 1st stage selects only the max fit (by SSR)

Other Models

Fithian, Sun, and Taylor (2014), Fithian et al. (2015), Tibshirani et al. (2016)

1. Regression models selecting variables from active set with sequences of tests
2. Only one variable selected at each stage

Comparison With Other Models Pt.2

OCMT Method

1. Not a sequencial method
   Select multiple variables in one stage (by $\hat{\mathcal{J}}_{i,(j)}$)
2. Only if Hidden Signals exists will multiple stages are needed
3. finite k signals, and # of stages $< k$ with high probability (*Proposition 1*)

Performance Criteria

Define the following:

True Positive Rates (Selecting Signals):

$$TPR_{n,\mathbf{T}} = \frac{\sum_{i=1}^{n} I(\hat{\mathcal{J}}_i = 1 \text{ and } \beta_i \neq 0)}{\sum_{i=1}^{n} I(\beta_i \neq 0)}$$

False Positive Rates (Selecting Psuedo Signals):

$$FPR_{n,\mathbf{T}} = \frac{\sum_{i=1}^{n} I(\hat{\mathcal{J}}_i = 1 \text{ and } \beta_i = 0)}{\sum_{i=1}^{n} I(\beta_i = 0)}$$

Performance Criteria Pt. 2

False Discovery Rate (Selecting Noise):

$$FDR_{n,\mathbf{T}} = \frac{\sum_{i=1}^{n} I(\hat{\mathcal{J}}_i = 1 \text{ and } \beta_i = \theta_i = 0)}{\sum_{i=1}^{n} \hat{\mathcal{J}}_i + 1}$$

Error and Coefficient Norms:

$$F_{\tilde{u}} = \mathbf{T}^{-1} \|\tilde{u}\|^2 = \mathbf{T}^{-1} \sum_{t=1}^{\mathbf{T}} \tilde{u}_t^2 \quad \text{and} \quad F_{\tilde{\beta}} = \|\tilde{\beta}_n - \beta_n\| = [\sum_{i=1}^{n} (\tilde{\beta}_i - \beta_i)^2]^{\frac{1}{2}}$$

Asymptotic Properties

Proposition 2

1. For some finite positive constants C_0 and C_1

$$\Pr[|t_x| > c_p(n,\delta) \mid \theta = 0] \; \leq \; exp[-\chi c_p^2(n,\delta)/2] + exp(-C_0\mathbf{T}^{C_1})$$

A sharp probability bound for selecting Hidden Signals.

1. Further when $\theta \neq 0$

$$\Pr[|t_x| > c_p(n,\delta) \mid \theta \neq 0] > 1 - exp(-C_2\mathbf{T}^{C_3})$$

A probability lower bound for selecting Signals.

Asymptotic Properties Pt.2

Further on Proposition 2,

$$TPR_{n,\mathbf{T}} = k^{-1}\sum_{i=1}^{k}\Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta) \mid \theta_i \neq 0]$$

$$TPR_{i,\mathbf{T}} \;\geq\; 1 - exp(-C_2\mathbf{T}^{C_3}) = 1 + O[exp(-C_2 n^{C_3\kappa_1})]$$

Meaning:

$$TPR_{n,\mathbf{T}} \xrightarrow{p} 1 \text{ , for any } \kappa_1 > 0$$

Asymptotic Properties Pt.3

Also, In a similar sense,
for any $0 < \varkappa < 1$:

$$FPR_{n,\mathbf{T}} = (n-k)^{-1} \sum_{i=k+1}^{n} \Pr[|t_{\hat{\phi}_{i,(1)}}| > c_p(n,\delta) \mid \beta_i = 0]$$

$$= \frac{k^*}{n-k} + O\{exp[-\varkappa c_p^2(n,\delta) \ / \ 2]\} + O[exp(-C_0 \mathbf{T}^{C_1})]$$

$$+ O[exp(-C_2 \mathbf{T}^{C_3})/(n-k)]$$

Asymptotic Properties Pt.4

Theorem 1

Under the assumptions on simplified DGP:

(a) The probability that number of stages($\hat{P}$) exceeds k

$$\Pr(\hat{P} > k) = O(n^{1-\varkappa\delta^*}) + O(n^{1-\kappa_1/3-\varkappa\delta}) + O[exp(-n^{C_0\kappa_1})]$$

(b) The probability of selecting the approximating model($\mathcal{A}_0$)

$$\Pr(\mathcal{A}_0) = 1 + O(n^{1-\varkappa\delta}) + O(n^{2-\varkappa\delta^*}) + O(n^{1-\kappa_1/3-\varkappa\delta}) + O[exp(-n^{C_0\kappa_1})]$$

Asymptotic Properties Pt.5

Theorem 1

(c) If $\delta > 1 - \kappa_1/3$, then $TPR_{n,\mathbf{T}} \xrightarrow{p} 1$

If $\delta > min(0, 1 - \kappa_1/3)$ and $\delta^* > 1$, then $FPR_{n,\mathbf{T}} \xrightarrow{p} 0$

If $\delta > max 1, 2 - \kappa_1/3$, then $FDR_{n,\mathbf{T}} \xrightarrow{p} 0$

Asymptotic Properties Pt.6

Additional on Assumption 6

Weak signals can be picked with no cost

The power of OCMT in selecting signals rises with ratio $\sqrt{\mathbf{T}}\left|\theta_{i,(j)}\right| \ / \ \sigma_{e_i,(\mathbf{T})}\sigma_{x_i,(\mathbf{T})}$

If said ratio is low, a large $\mathbf{T}$ is required for selecting variables

Extending OCMT - Finding True Model

Tweaks in $FDR_{n,\mathbf{T}}$ (falsely picking Psuedo and Noise):

$$FDR^*_{n,\mathbf{T}} = \frac{\sum_{i=1}^{n} I(\hat{\mathcal{J}}_i = 1 \text{ and } \beta_i = 0)}{\sum_{i=1}^{n} \hat{\mathcal{J}}_i + 1}$$

And a post-OCMT selection is required to rule out Psuedo Signals
(EX. Schwarz information criterion)

Extending OCMT - Norms of Error and Coefficient

Additional regularity condition *Assumption 7* is needed

Combined with the 6 assumptions of simplified DGP, we get:

$$F_{\tilde{u}} = \mathbf{T}^{-1} \|\tilde{u}\|^2 = \sigma^2 + O_p(n^{-\kappa_1 \ / \ 2}) + O(n^{3\epsilon - 3\kappa_1 \ / \ 2})$$

and

$$F_{\tilde{\beta}} = \|\tilde{\beta}_n - \beta_n\| = O_p(n^{5\epsilon \ / \ 2 - \kappa_1}) + O_p(n^{\epsilon - \kappa_1 \ / \ 2})$$

Consistency for Coefficient Norm

Including Psuedo Signals doesn't affect the consistency

1. $\Pr(\text{pick Signals}) \xrightarrow{p} 1$ then $\beta_{n-k} \xrightarrow{\mathbf{T}\to\infty} 0$
2. Restricting the rate of k^* rising with n (*Theorem 2*),
   then $\beta_{k^*} \xrightarrow{\mathbf{T}\to\infty} 0$ and $Var(\tilde{\beta}_{k^*})$ is controlled