# Averaging Estimators for Regressions with a Possible Structural Break


Bruce E. Hansen*

University of Wisconsin†

www.ssc.wisc.edu/~bhansen


September 2007

*Preliminary*


**Abstract**

This paper investigates selection and combination of linear regressions with a possible structural break. We investigate the form of the Mallows criterion for the structural break model, and propose an approximate feasible penalty for the structural break model. We propose an averaging estimator, which combines the estimates from the no-break model and the break model. This estimator is simple to compute, as the weights are a simple function of the ratio of the penalty to the Andrews SupF test statistic. We show that this estimator has lower squared error than conventional unrestricted and pretest estimators.



*Research supported by the National Science Foundation.



†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706

# 1 Introduction

Structural change is an important issue in time series econometrics. Applied economists routinely test their models for the presence of structural change, typically using the Andrews (1993) and Andrews-Ploberber (1994) tests. Sometimes, when the evidence supports it, a structure break model is estimated. The breakdate may be estimated formally (as recommend by Bai (1997)) or may be selected informally, but the practical effects are rather similar. This means that applied econometricians may be *de facto* using a pretest estimator: using a restricted estimator (linear regression) when the structural change test is insignificant, and using the unrestricted estimator (the structural change estimator) when the test is significant.

This practice is unfortunate because it is well known that pretest estimators generically have poor sampling properties. The squared error of pretest estimators is parameter-dependent, but can be quite high relative to unrestricted estimation.

This paper advocates the use of the Mallows information criterion, which is better suited for selection of estimators with low risk. Furthermore, advocate the use of an averaging estimators, with the weight selected to minimize the Mallows criterion.

We derive the form of the Mallows penalty term for the structural change model, and find that it is non-standard. We tabulate the penalty terms, and show how they can be used for the selection of structural break models and for averaging estimators. The Mallows averaging estimator is simple to compute, as the weights are a simple function of the ratio of the Mallows penalty to the Andrews SupF test statistic.

We investigate the asymptotic performance of the estimators using three squared error criteria. The new averaging estimator has excellent risk performance as measured by all criteria.

## 2 Model and Estimation

The model of interest is a linear time-series regression with a possible structural break. The observations are $(y_t, \mathbf{x}_t)$ for $t = 1, ..., n$, where $y_t$ is scalar and $x_t$ is an $m$ vector which may contain lagged values of $y_t$. The model is

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta}_1 1\left(t < k\right) + \mathbf{x}_t'\boldsymbol{\beta}_2 1\left(t \geq k\right) + e_t \\ E\left(e_t \mid \mathbf{x}_t\right) &= 0 \\ E\left(e_t^2 \mid \mathbf{x}_t\right) &= \sigma^2 \end{aligned} \tag{1}$$

which has parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, k, \sigma^2)$. The true breakdate $k_0$ is unknown but constrained to satisfy the restriction $k_1 \leq k_0 \leq k_2$. A potential parameter of interest is the difference in regression slopes $\boldsymbol{\theta} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$.

If there is no break in the slope coefficients, then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and the model simplifies to

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t \tag{2}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and the breakdate $k$ drops out.

When a structural break is uncertain it is common in applications to employ a two-step procedure where the first step is to test for the presence of a structural break and the second step is to estimate the model selected by the test. Let us decribe this procedure in detail.

Estimation of (1) is easiest by concentration. First, fix $k$. Then equation (1) is estimated by least-squares, which we write as

$$y_t = \mathbf{x}_t'\hat{\boldsymbol{\beta}}_1(k) 1\left(t < k\right) + \mathbf{x}_t'\hat{\boldsymbol{\beta}}_2(k) 1\left(t \geq k\right) + \hat{e}_t(k). \tag{3}$$

Let $\hat{\mathbf{e}}(k)$ denote the $n \times 1$ residuals from (3). The concentrated sum of squared errors, given $k$, is then $\hat{\mathbf{e}}(k)'\hat{\mathbf{e}}(k)$. The least-squares estimate of $k$ is then found by numerically minimizing this criterion:

$$\hat{k} = \underset{k_1 \leq k \leq k_2}{\operatorname{argmin}}\, \hat{\mathbf{e}}(k)'\hat{\mathbf{e}}(k).$$

The remaining estimates are then obtained using $\hat{k}$ :

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= \hat{\boldsymbol{\beta}}_1(\hat{k}) \\ \hat{\boldsymbol{\beta}}_2 &= \hat{\boldsymbol{\beta}}_2(\hat{k}). \end{aligned}$$

We write the fitted model as

$$y_t = \mathbf{x}_t'\hat{\boldsymbol{\beta}}_1 1(t < \hat{k}) + \mathbf{x}_t'\hat{\boldsymbol{\beta}}_2 1(t \geq \hat{k}) + \hat{e}_t. \tag{4}$$

Let $\hat{\mathbf{e}} = \hat{\mathbf{e}}(\hat{k})$ denote the vector of fitted residuals.

The no-break model (2) is also estimated by least-squares, which we write as

$$y_t = \mathbf{x}_t' \tilde{\boldsymbol{\beta}} + \tilde{e}_t. \tag{5}$$

As an estimator of the parameters in (1), we set $\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_2 = \tilde{\boldsymbol{\beta}}$. Note that the estimator of $\boldsymbol{\theta} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$ is $\tilde{\boldsymbol{\theta}} = \mathbf{0}$. Let $\tilde{\mathbf{e}}$ denote the $n \times 1$ vector of residuals from (5).

The standard test of model (2) against model (1) is the SupF test of Andrews (1991). The test statistic is the standard F test

$$F_n = \frac{(\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}})}{s^2}$$

where

$$s^2 = \frac{1}{n-2m}\hat{\mathbf{e}}'\hat{\mathbf{e}} \tag{6}$$

is the bias-corrected estimator of the error variance from the full model (4). It is convenient to observe that $F_n = F_n(\hat{k})$ where

$$F_n(k) = \frac{(\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}(k)'\hat{\mathbf{e}}(k))}{s^2}$$

and that $\hat{k}$ is the argmax of $F_n(k)$.

Let $\pi = k/n$ denote the breakdate fraction. Under the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$,

$$F_n(n\pi) \rightarrow_d F(\pi) = \frac{\mathbf{W}^*(\pi)'\mathbf{W}^*(\pi)}{\pi(1-\pi)}$$

where $\mathbf{W}^*(\pi)$ is an $m$-dimensional standard Brownian bridge, and therefore

$$F_n \rightarrow_d \text{SupF} = \sup_{\pi_1 \le \pi \le \pi_2} F(\pi) \tag{7}$$

where $\pi_1 = k_1/n$ and $\pi_2 = k_2/n$. A 5% asymptotic test rejects (2) in favor of (1) if $F_n > q$ where $q$ is the 5% upper quantile of the distribution of SupF. Critical values are tabulated in Andrews (2003), and depend on $m$ and $\lambda = \pi_2(1-\pi_1)/(1-\pi_2)\pi_1$. For example, if $m = 1$, $\pi_1 = .15$ and $\pi_2 = .85$, then $q = 8.68$.

As we mentioned above, the conventional estimator of the model is to use the unrestricted estimator (4) when $F_n$ is significant, otherwise to use the restricted estimator (5). The reasoning behind this estimation strategy is fairly straightforward. A reasonable presumption is that unless there is evidence to the contrary we should use a standard linear regression model. We should use a breakdate estimator only if there is evidence of a structural break. As the most compelling evidence is a statistical test, this leads to two-step estimation.

This type of estimator is known as a pre-test estimator. In classic settings, there is an old and established literature documenting that pre-test estimators generically have poor sampling properties. We expect this problem to carry over to structural break models. A standard alternative to pre-testing is selection based on an information criterion. A recent generalization is model averaging with the weights determined by an information criterion (Hansen, 2007). However, an appropriate information criterion for the structural break model has not yet been developed.

## 3 Mallows Criterion

In this section we develop Mallows criterion appropriate for models (1) and (2). The Mallows criterion is a penalized sum of squared residuals designed to be approximately unbiased for the in-sample fit. The general approach is as follows. Write the model in vector notation as $\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}$ where $\boldsymbol{\mu}$ is the regression function. Let $\hat{\boldsymbol{\mu}} = \mathbf{P}\mathbf{y}$ be an estimator of $\boldsymbol{\mu}$ with residual vector $\hat{\mathbf{e}} = \mathbf{y} - \hat{\boldsymbol{\mu}}$. A measure of in-sample fit is $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})$ and the goal is to estimate this quantity by the sum of squared errors $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ plus a penalty. By expanding the square

$$
\begin{aligned}
\hat{\mathbf{e}}'\hat{\mathbf{e}} &= (\mathbf{e} + \boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' (\mathbf{e} + \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \\
&= (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) + \mathbf{e}'\mathbf{e} + 2\mathbf{e}' (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \\
&= (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) + \mathbf{e}'\mathbf{e} + 2\mathbf{e}' (\mathbf{I} - \mathbf{P}) \boldsymbol{\mu} - 2\mathbf{e}'\mathbf{P}\mathbf{e}.
\end{aligned}
$$

Thus the sum of squared residuals $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ equals the in-sample fit $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})$ plus three terms. The first term $\mathbf{e}'\mathbf{e}$ is independent of the estimation method and therefore does not matter. The second term $2\mathbf{e}' (\mathbf{I} - \mathbf{P}) \boldsymbol{\mu}$ has an approximate mean of zero and is therefore also ignored. The final term $2\mathbf{e}'\mathbf{P}\mathbf{e}$ has a non-zero mean which is the focus of attention. The general form of the Mallows criterion adjusts the sum of squared errors for the mean of this term:

$$
C = \hat{\mathbf{e}}'\hat{\mathbf{e}} + 2E \left( \mathbf{e}'\mathbf{P}\mathbf{e} \right).
$$

In the case of the restricted least-squares estimator (5) for model (2), $\mathbf{P} = \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ where $\mathbf{X}$ is the regressor matrix. Under standard conditions $\mathbf{e}'\mathbf{P}\mathbf{e} \to_d \sigma^2 \chi^2_m$ so $E (\mathbf{e}'\mathbf{P}\mathbf{e}) \to \sigma^2 m$. It follows that the Mallows criterion for this model takes the classic form

$$
\tilde{C} = \tilde{\mathbf{e}}'\tilde{\mathbf{e}} + 2s^2 m \tag{8}
$$

where $s^2$ is (6).

If the breakdate $k$ were known then a similar argument can be applied to the estimates (3) and we would obtain the Mallows criterion

$$
\hat{C}(k) = \hat{\mathbf{e}}(k)'\hat{\mathbf{e}}(k) + 4\hat{\sigma}^2 m.
$$

However the case with unknown breakdate is non-standard. For fixed $k$, we can write $\hat{\boldsymbol{\mu}}(k) = \mathbf{P}(k)\mathbf{y}$ where $\mathbf{P}(k)$ is the projection matrix onto the space of regressors $\mathbf{x}_t 1 (t < k)$ and $\mathbf{x}_t 1 (t \geq k)$. With estimated $\hat{k}$ *we* have $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\hat{k}) = \mathbf{P}(\hat{k})\mathbf{y}$ and thus the desired penalty is $E \left( \mathbf{e}'\mathbf{P}(\hat{k})\mathbf{e} \right)$.

To obtain an asymptotic approximation to $E\left(\mathbf{e}'\mathbf{P}(\hat{k})\mathbf{e}\right)$, we adopt the canonical framework

$$\boldsymbol{\theta} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = n^{-1/2}\sigma\mathbf{c} \tag{9}$$

and assume $\pi_0 = k_0/n$ is constant as $n \to \infty$. The local nesting (9) is the setting in which asymptotic comparisons of model (1) and (2) are interesting. Let $\mathbf{M} = E\left(\mathbf{x}_t\mathbf{x}_t'\right)$. Andrews (1993, Theorem 4) shows that when (9) holds, then

$$F_n(n\pi) \to_d F\left(\pi \mid \delta, \pi_0\right) = \frac{\left(\mathbf{W}^*(\pi) + \mathbf{M}^{1/2}\mathbf{c}\left((\pi_0 \wedge \pi) - \pi_0\pi\right)\right)'\left(\mathbf{W}^*(\pi) + \mathbf{M}^{1/2}\mathbf{c}\left((\pi_0 \wedge \pi) - \pi_0\pi\right)\right)}{\pi\left(1-\pi\right)} \tag{10}$$

which is a function only of the scalar parameters $\pi_0$ and $\delta = \mathbf{c}'\mathbf{M}\mathbf{c}$. The parameter $\delta$ indexes the magnitude of the structural change.

From (10) it follows that

$$\frac{\hat{k}}{n} \quad \to_d \quad \pi^* = \underset{\pi_1 \le \pi \le \pi_2}{\operatorname{argmax}}\, F\left(\pi \mid \delta, \pi_0\right). \tag{11}$$

This is the asymptotic distribution of the breakdate estimate under the local break assumption (9). (11) shows that the breakdate fraction is inconsistent for $\pi_0$. However, as $\delta$ gets large, the distribution $\pi^*$ concentrates about the breakdate $\pi_0$.

We now can deduce our main result.

**Theorem 1** *If (9) holds then as $n \to \infty$*

$$E\left(\mathbf{e}'\mathbf{P}(\hat{k})\mathbf{e}\right) \to m + p\left(m, \delta, \pi_0, \lambda\right). \tag{12}$$

*where $p\left(m, \delta, \pi_0, \lambda\right) = E\left(\pi^*\right)$. It follows that the correct penalty for the Mallows criterion is twice $m + p\left(m, \delta, \pi_0, \lambda\right)$. This function has the limiting values*

$$\begin{aligned} p\left(m, 0, \pi_0, \lambda\right) &= E\left(\text{SupF}\right) \\ \lim_{\delta\to\infty} p\left(m, \delta, \pi_0, \lambda\right) &= m \end{aligned}$$

*where* SupF *is the distribution in (7).*

The theorem shows that correct form of the penalty depends on the unknown values $\delta$ and $\pi_0$ and is therefore infeasible. As a practical solution, we recommend approximating $p\left(m, \delta, \pi_0, \lambda\right)$ by an average of the limiting cases $\delta = 0$ and $\delta \to \infty$ which does not depend on unknowns. This is

$$p^*\left(m, \lambda\right) = \frac{1}{2}\left(m + E\left(\text{SupF}\right)\right).$$

Using this value to approximate $p\left(m, \delta, \pi_0, \lambda\right)$, we obtain the practical Mallows criterion for the structural change model.

**Proposition 1** *The correct Mallows criterion for the structural change model is*

$$\hat{C} = \hat{\mathbf{e}}'\hat{\mathbf{e}} + 2s^2 \left(m + p\left(m, \delta, \pi_0, \lambda\right)\right)$$

*which is infeasible. A practical approximate Mallows criterion is*

$$\hat{C}^* = \hat{\mathbf{e}}'\hat{\mathbf{e}} + 2s^2 \left(m + p^*\left(m, \lambda\right)\right). \tag{13}$$

The penalty coefficients $p^*\left(m, \lambda\right)$ are displayed in Table 1 as a function of $\pi_1$.

# 4 Selection and Combination

Model selection based on the Mallows Criterion (13) picks the structural change model (4) if $\hat{C}^* < \tilde{C}$, and picks the restricted model (5) if $\tilde{C} < \hat{C}^*$. Equivalently, it picks (4) if $F_n > 2p^*\left(m, \lambda\right)$. This is similar to the pre-test estimator, but replaces the critical value from Andrews' table with the value $2m^*$. Since the latter value is smaller (4.98 versus 8.68 for the case $m = 1$ and $\pi_1 = .15$), selection based on Mallows criterion tends to pick the structural change model more frequently.

In a recent paper, Hansen (2007) has argued for model combination based on Mallows weights, rather than selection. Model combination assigns a weight of $w$ to model (1) and a weight of $1 - w$ to model (2). The Mallows criterion for the weighted average is

$$C\left(w\right) = \left(\hat{\mathbf{e}}w + \tilde{\mathbf{e}}\left(1 - w\right)\right)' \left(\hat{\mathbf{e}}w + \tilde{\mathbf{e}}\left(1 - w\right)\right) + 2s^2 \left(m + p^*\left(m, \lambda\right) w\right).$$

The Mallows weight is the value in $[0, 1]$ which minimizes $C(w)$. The solution is

$$\hat{w} = \begin{cases} 0 & \text{if} \quad F_n < p^*\left(m, \lambda\right) \\ 1 - \dfrac{p^*\left(m, \lambda\right)}{F_n} & \text{if} \quad F_n \geq p^*\left(m, \lambda\right) \end{cases} \tag{14}$$

Viewed as a function of the test statistic $F_n$, the weight $\hat{w}$ as a smoothed version of the selection criterion.

Given the weight $\hat{w}$, the estimates of the model parameters are obtained as weighted averages:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1(w) &= \hat{w}\hat{\boldsymbol{\beta}}_1 + \left(1 - w\right)\tilde{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}_2(w) &= \hat{w}\hat{\boldsymbol{\beta}}_2 + \left(1 - w\right)\tilde{\boldsymbol{\beta}} \end{aligned}$$

# 5 Numerical Comparison

We now illustrate the efficiency gains achievable by Mallows averaging through a simple numerical investigation. We take model (1) with $m = 1$, normalize $Ex_t^2 = 1$ and $\sigma^2 = 1$, and set $\beta_2 - \beta_1 = \theta = n^{-1/2}\delta$. We make large sample (asymptotic) comparisons, so the environment is fully determined by the parameters $\pi_0$ and $\delta$.

We compare four estimation methods: (1) unrestricted least-squares; (2) Pre-Test estimation (based on an asymptotic 5% test); (3) Selection using the Mallows criterion (13); and (4) Averaging using the Mallows weights (MMA) (14).

The methods are compared using three criterion:

$$\begin{aligned} \text{Fit} &= E\sum_{t=1}^{n}(\mu_t - \hat{\mu}_t)^2 \\ \text{MSE}(\hat{\theta}) &= nE\left(\hat{\theta} - \theta\right)^2 \\ \text{MSE}(\hat{\beta}_2) &= nE\left(\hat{\beta}_2 - \beta_2\right)^2 \end{aligned}$$

Fit is an overall measure of performance. $\text{MSE}(\hat{\theta})$ is useful when the goal is to measure the degree of structural change. $\text{MSE}(\hat{\beta}_2)$ equals out–of-sample mean squared forecast error.

In additional to the four feasible estimation procedures, we also include the optimal weighted average estimator. The optimal weights depend on the parameters $\pi_0$ and $\delta$ and are therefore infeasible.

Figure 1 displays the calculations based on Fit. There are three panels displaying results for $\pi_0 = .15$, $\pi_0 = .30$ and $\pi_0 = .45$. The parameter $\delta$ is varied along the $x$-axis, Fit is displayed in the $y$-axis. Five lines are plotted. The long dashes are for the unrestricted least-squares estimator. The short dashes are for the PreTest estimator. The closely spaced dots are for the Selection estimator. The solid line is for the MMA estimator, and the remaining dotted line is the the optimal weighted average.

To facilitate comparisons, the final panel displays the Maximal Regret of the four estimators as a function of $\pi_0$, maximized over $\delta$. The Regret is the difference in Fit with the optimal estimator, and the Maximal Regret is the largest value of the Regret across all values of $\delta$. We plot the Maximal Regret as a function of $\pi_0$ to show how this criterion varies with the location of structural change.

Figure 2 displays similar calculations for $\text{MSE}(\hat{\theta})$. Figures 3 and 4 display the results for $\text{MSE}(\hat{\beta}_2)$. This criterion is more sensitive to $\pi_0$, so we display results for $\pi_0 = .15$, .30, .45, .55, .65, .75 and .85, with the Maximal Regret displayed in the final panel of Figure 4.

How do the estimators compare? Figure 1 displays the comparisons based on the Fit criterion. First, compare the risk of the unrestricted and pretest estimators. We see that the pretest estimator achieves much lower risk for small values of $\delta$, but at the cost of much higher risk for moderate values of $\delta$. The result is that the pretest estimator has the highest maximum risk. Second, compare the risk of these estimators with that of the MMA estimator. We see that the MMA estimator uniformly dominates the unrestricted estimator, and has much lower risk than the pretest estimator for almost all parameter values. The MMA estimator has considerably lower Regret than the other estimators.

Now take Figure 2, which displays the results for $\text{MSE}(\hat{\theta})$. We can see that the MSE of both the unrestricted least-squares estimator and the pre-test estimator are quite sensitive to the true value of $\delta$. For the unrestricted estimator, the MSE is monotonically decreasing with $\delta$, while the MSE of the pretest estimator is hump-shaped, with the highest MSE for moderate values of $\delta$. For both estimators, however, the maximum MSE (across $\delta$) is quite high. The MSE of the Mallows selection estimator is intermediate between these two estimators. The MSE of the MMA estimator is the least sensitive to the value of $\delta$ (the plot is relatively flat as a function of $\delta$) and most closely tracks the MSE of the optimal weighted average. As a result, the Maximum Regret of the MMA estimator is considerably lower than the other feasible estimators.

Similar lessons emerge from Figures 3 and 4, which display the results for MSE($\hat{\beta}_2$). The unrestricted estimator does quite well when $\pi_0$ is large but does poorly when $\pi_0$ is small. In constrast the pretest estimator does well when $\pi_0$ is small but does porrly when $\pi_0$ is large. The MSE of the MMA estimator is less sensitive to the values of $\pi_0$ and $\delta$, and as a result has low Regret.

## 6 Conclusion

Common empirical practice is to test time-series regressions for the presence of a structural break, and if a break is detected then account for this by allowing for structural change in the parameter estimates. This practice corresponds to a pre-test estimator with poor sampling properties. An estimator with better risk properties can be constructed as the weighted average of the no-break and break estimates, where the weight is selected to minimize a modified Mallows criterion and is a simple function of the Andrews SupF test statistic and a penalty term.

# References

[1] Andrews, Donald W. K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821-856.

[2] Andrews, Donald W. K. (2003), "Tests for parameter instability and structural change with unknown change point: A corrigendum," *Econometrica*, 71, 395-397.

[3] Andrews, Donald W. K. and Werner Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica,* 62, 1383-1414.

[4] Bai, Jushan (1997): "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, 79, 674-679.

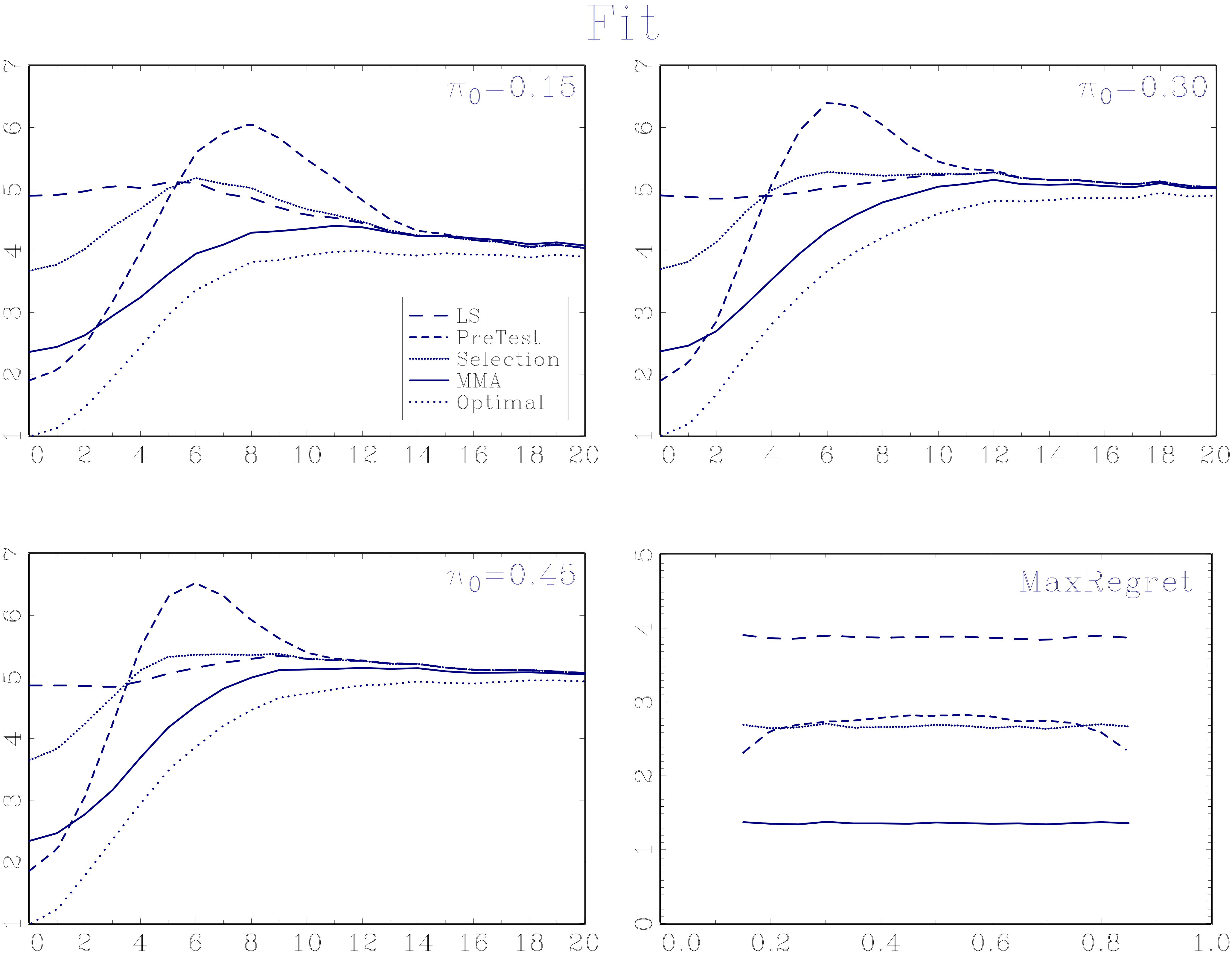
[5] Hansen, Bruce E. (2007): "Least squares model averaging," *Econometrica*, 75, 1175-1189.

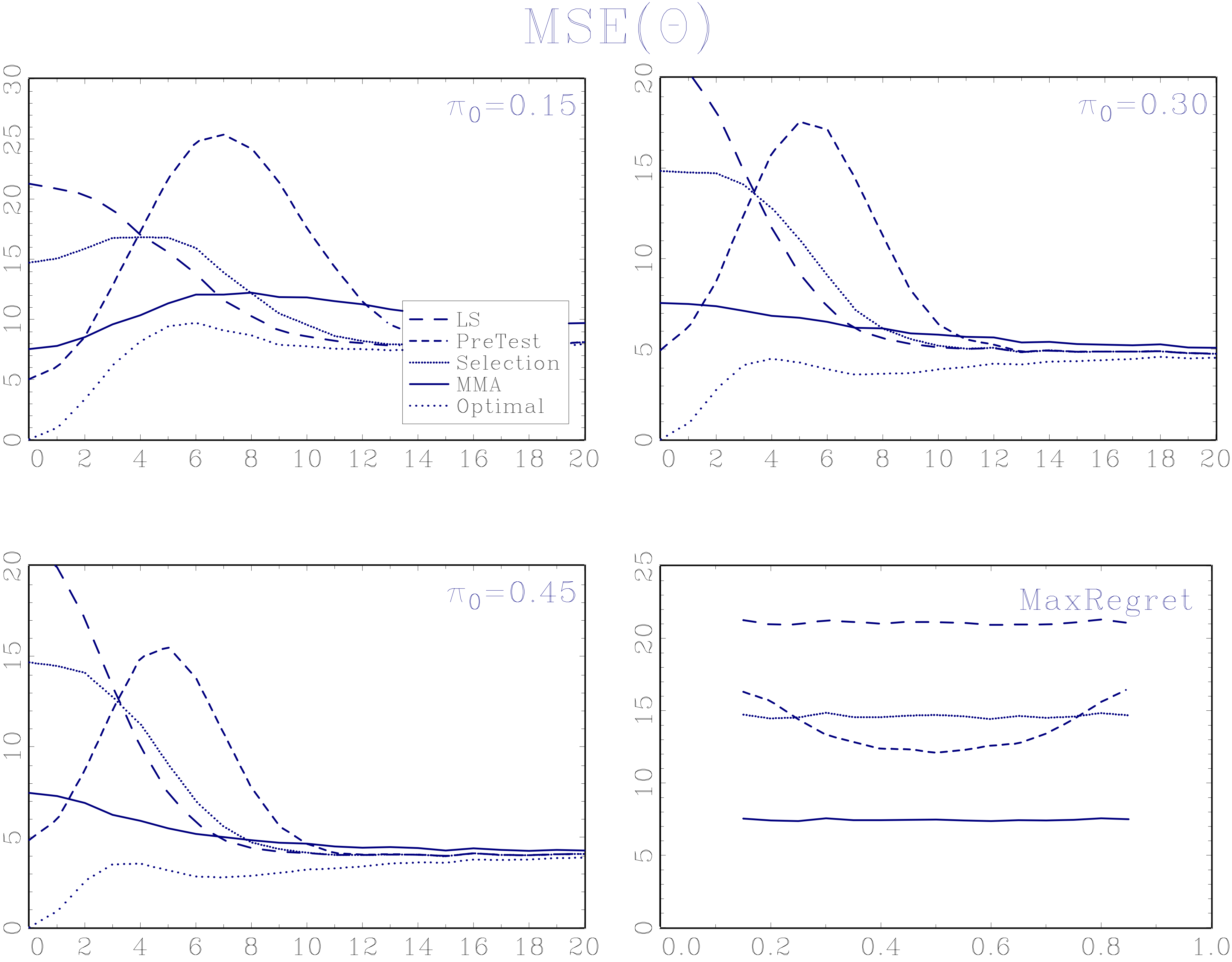
[6] Mallows, C.L. (1973): "Some comments on $C_p$," *Technometrics*, 15, 661-675.

Table 1: Penalty Coefficients $p^*(m, \lambda)$

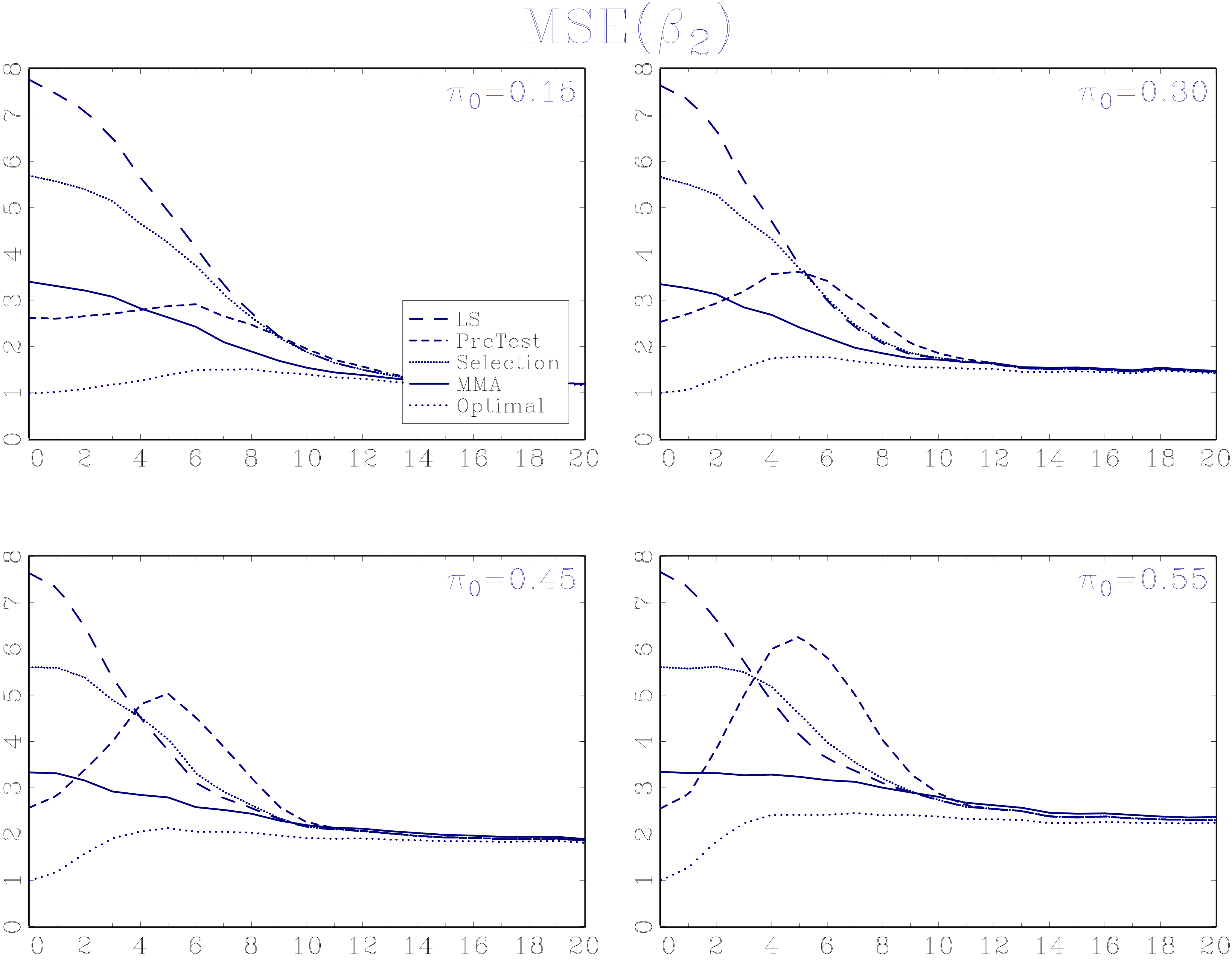
| | $\pi_1 = 1 - \pi_2$ | | | | | |
|---|---|---|---|---|---|---|
| $m$ | .01 | .05 | .10 | .15 | .20 | .25 |
| 1 | 3.28 | 2.9 | 2.67 | 2.49 | 2.33 | 2.18 |
| 2 | 5.02 | 4.56 | 4.27 | 4.05 | 3.85 | 3.66 |
| 3 | 6.56 | 3.06 | 5.73 | 5.47 | 5.24 | 5.01 |
| 4 | 8.00 | 7.45 | 7.08 | 6.80 | 6.55 | 6.30 |
| 5 | 9.37 | 8.78 | 8.40 | 8.09 | 7.81 | 7.54 |
| 6 | 10.7 | 10.1 | 9.69 | 9.37 | 9.07 | 8.77 |
| 7 | 12.0 | 11.4 | 10.9 | 10.6 | 10.3 | 9.98 |
| 8 | 13.3 | 12.6 | 12.2 | 11.8 | 11.5 | 11.2 |
| 9 | 14.6 | 13.9 | 13.4 | 13.0 | 12.7 | 12.3 |
| 10 | 15.8 | 15.1 | 14.6 | 14.2 | 13.9 | 13.5 |
| 11 | 17.1 | 16.3 | 15.8 | 15.4 | 15.0 | 14.7 |
| 12 | 18.3 | 17.5 | 17.0 | 16.6 | 16.2 | 15.8 |
| 13 | 19.6 | 18.8 | 18.2 | 17.8 | 17.4 | 17.0 |
| 14 | 20.8 | 19.9 | 19.4 | 18.9 | 18.5 | 18.1 |
| 15 | 21.9 | 21.1 | 20.5 | 20.0 | 19.6 | 19.2 |
| 16 | 23.2 | 22.3 | 21.7 | 21.2 | 20.8 | 20.4 |
| 17 | 24.3 | 23.5 | 22.9 | 22.4 | 21.9 | 21.5 |
| 18 | 25.5 | 24.6 | 24.0 | 23.5 | 23.0 | 22.6 |
| 19 | 26..7 | 25.8 | 25.2 | 24.6 | 24.2 | 23.7 |
| 20 | 27.9 | 26.9 | 26.3 | 25.8 | 25.3 | 24.8 |

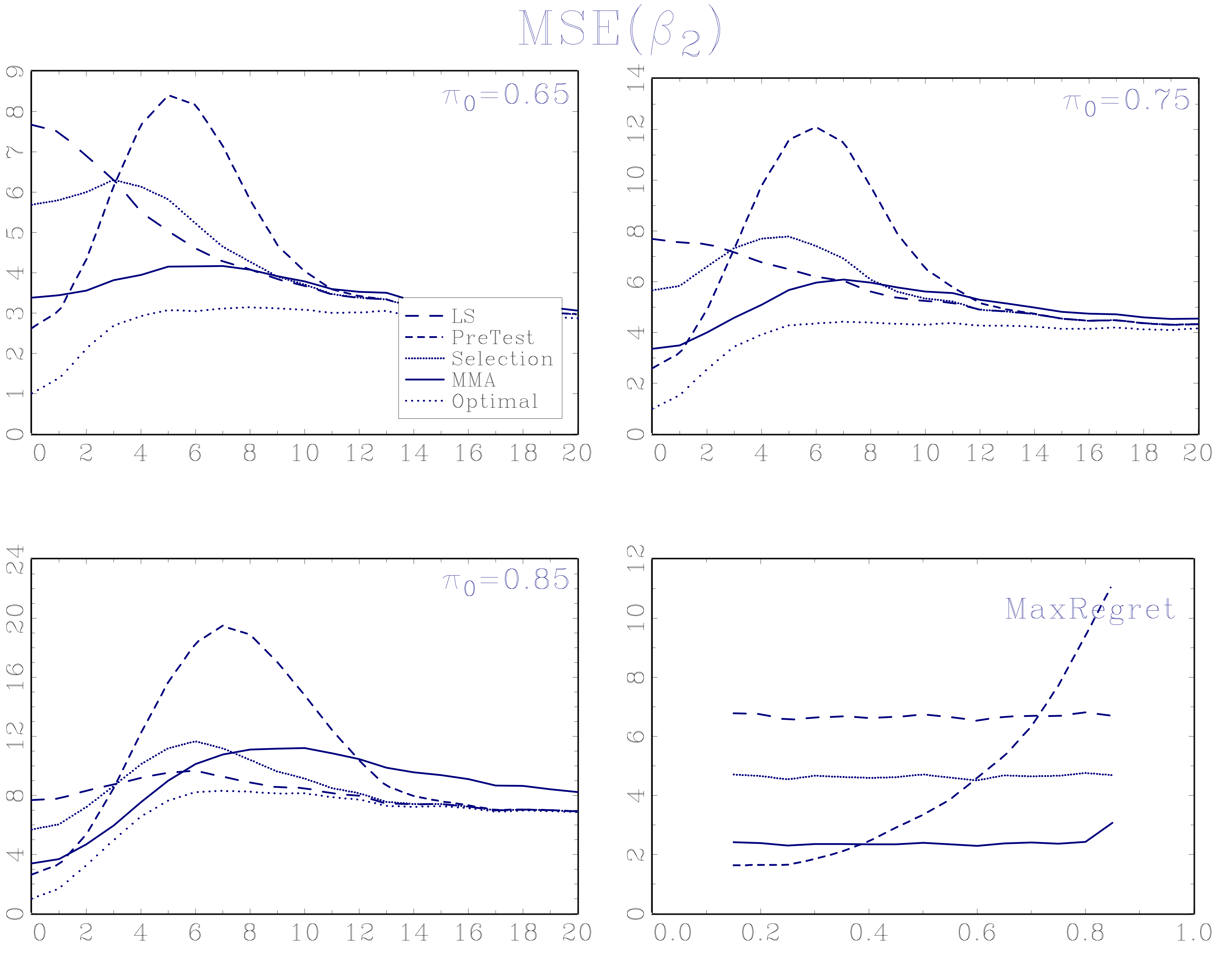
Note: $\lambda = (1 - \pi_1)^2 / \pi_1^2$


Fit
π0=0.15
π0=0.30
π0=0.45
MaxRegret
LS
PreTest
Selection
MMA
Optimal



MSE(Θ)
π0=0.15
π0=0.30
π0=0.45
MaxRegret
LS
PreTest
Selection
MMA
Optimal


MSE(β2)
π0=0.15
π0=0.30
π0=0.45
π0=0.55
LS
PreTest
Selection
MMA
Optimal
0 1 2 3 4 5 6 7 8
0 2 4 6 8 10 12 14 16 18 20



MSE($\beta_2$)
$\pi_0$=0.65
$\pi_0$=0.75
$\pi_0$=0.85
MaxRegret
LS
PreTest
Selection
MMA
Optimal