Penalized Adaptive Method in Forecasting
with Large Information Set and Structure
Change

Wolfgang K. Härdle
Xinjue Li
Lenka Zbořáková

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
WISE –The Wang Yanan Institute for Studies in Economics
Xiamen University
lvb.wiwi.hu-berlin.de
irtg1792.hu-berlin.de
wise.cmu.edu.cn
Excess bond premium

Figure 1: FEDS Notes: Excess bond premium, Jan 1973 - Mar 2016 (link), shaded areas are NBER designated recessions
Motivation

Excess bond premium models

Figure 2: Real (black) vs. fitted excess bond premium for 2-year bonds by Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009), Jan 1964 - Dec 2003

Penalized Adaptive Method
Challenges

- High-dimensional data
  - Stock prices
  - Macroeconomic variables
- Dimension reduction
  - Systemic risk indicator
  - Excess bond premium modelling
- Non-stationarity

Penalized Adaptive Method
**Dimension reduction**

- Factor analysis
- Principal component analysis
- Penalized regression analysis
  - Selects important variables
  - Good interpretation of the fitted model
  - Tibshirani (1996): Lasso
  - Fan and Li (2001): SCAD penalty

Penalized Adaptive Method
Motivation

Penalized likelihood

\[ Q(\beta) = n^{-1} \sum_{i=1}^{n} l_i(\beta) - \sum_{j=1}^{p} p_\lambda(|\beta_j|) \]

with \( l_i(\cdot) \) non-penalized log-likelihood function and \( p_\lambda(\cdot) \) a penalty function with parameter \( \lambda > 0 \)

- Fan and Li (2001): Local quadratic approximation (LQA)
- Zou and Li (2008): Local linear approximation (LLA)

\[ Q(\beta) \approx n^{-1} \sum_{i=1}^{n} l_i(\beta) - \sum_{j=1}^{p} p'_\lambda(|\beta_j|) |\beta_j| \]
Window selection

- Time varying coefficients
  - Heterogeneity throughout time

- Use of rolling windows with fixed window size

- Polzehl and Spokoiny (2004, 2006): Adaptive window choice
  - Data driven choice of the longest homogeneous interval
  - Propagation & separation
  - Change point analysis
Motivation

Propagation-separation approach

- Series of nested intervals for a given time point $t$
  
  \[
  I_t^{(1)} \subset I_t^{(2)} \subset I_t^{(3)} \subset \ldots \subset I_t^{(M)}
  \]
  
  with $n^{(m)}$ observations in $I_t^{(m)}$, $m = 1, \ldots, M$

- Propagation: Extension of local model in a (nearly) homogeneous situation

- Separation: Extension is restricted to the region of local homogeneity

Penalized Adaptive Method
Outline

1. Motivation ✓
2. Penalized adaptive method
3. Real data application
4. Concluding remarks
Penalized adaptive method

Combination of SCAD penalty with adaptive window choice

- Dimension reduction
- Longest homogeneous interval detection
- Prediction based on the estimated sparse coefficients
SCAD penalty

- Linear model $Y = X\beta + \varepsilon$
  with $Y_{(n \times 1)}, X_{(n \times p)}, \beta_{(p \times 1)}, \varepsilon_{(n \times 1)} \sim iid (0, \sigma^2)$

- Fan and Li (2001): Quadratic spline function with knots at $\lambda$ and $a\lambda$ with

\[ \frac{\partial p_\lambda(\beta)}{\partial \beta} = \lambda \left\{ I(\beta \leq \lambda) + \frac{(a\lambda - \beta)^+}{(a - 1)\lambda} I(\beta > \lambda) \right\} \]

  for $a > 2$ and $\beta > 0$

- Zou and Li (2008): LLA algorithm

  Details
Hypothesis

- Recall the series of nested intervals for a given time point $t$
  $$I_t^{(1)} \subset I_t^{(2)} \subset I_t^{(3)} \subset \ldots \subset I_t^{(M)}$$

- $I_t^{(1)}$ homogeneous by assumption

- Hypothesis

  $$H_0 : \quad Y_t \sim \mathbb{P}_1, \quad \text{for } t \in I_t^{(m)}$$

  $$H_1 : \quad \left\{ \begin{array}{ll}
  Y_t \sim \mathbb{P}_1, & \text{for } t \in I_t^{(m-1)} \\
  Y_t \sim \mathbb{P}_2, & \text{for } t \in I_t^{(m)} \setminus I_t^{(m-1)}
  \end{array} \right.$$  

  for $m = 2, \ldots, M$ with measures $\mathbb{P}_1, \mathbb{P}_2 \in \{\mathbb{P}(\theta), \theta \in \Theta \subseteq \mathbb{R}^p\}$
Generalized likelihood ratio

- **Test statistic**
  \[
  T_t^{(m)} = \frac{n_t^{(m-1)}}{n_t^{(m)}} \max_{\beta} Q(\beta, l_t^{(m-1)}) + \frac{n_t^{(m)} - n_t^{(m-1)}}{n_t^{(m)}} \max_{\beta} Q(\beta, l_t^{(m)} \setminus l_t^{(m-1)}) - \max_{\beta} Q(\beta, l_t^{(m)})
  \]
  for \( m = 2, \ldots, M \)

- **SCAD penalty estimator**
  \[
  \tilde{\beta}_t^{(m)} = \arg\max_{\beta} Q(\beta, l_t^{(m)})
  \]

- **Adaptive estimator** \( \hat{\beta}_t^{(m)} \), for \( m = 1, \ldots, M \)
Penalized Adaptive Method

Algorithm

1. Assume $I_t^{(1)}$ homogeneous
2. Initialization $\hat{\beta}_t^{(1)} = \tilde{\beta}_t^{(1)}$
3. $m = 2$
4. While $T_t^{(m)} < \zeta_m$ and $m < M$
   
   $\hat{\beta}_t^{(m)} = \tilde{\beta}_t^{(m)}$

   $m = m + 1$

5. Final estimate $\hat{\beta}_t = \hat{\beta}_t^{(m)}$

- Critical values $\zeta_2, \ldots, \zeta_M$
- Q: How to find appropriate critical values?
Multiplier bootstrap

- Bootstrapped penalized likelihood function

\[
Q^\circ(\beta) = n^{-1} L^\circ(\beta) - \sum_{j=1}^{p} p\lambda(|\beta_j|) = n^{-1} \sum_{i=1}^{n} l_i(\beta) u_i - \sum_{j=1}^{p} p\lambda(|\beta_j|)
\]

with \( u_i \overset{iid}{\sim} (1, 1) \) for \( i = 1, \ldots, n \)

- Note \( \arg\max_{\beta} E[Q^\circ(\beta)|Y] = \arg\max_{\beta} Q(\beta) = \tilde{\beta} \)
  and \( \tilde{\beta}^\circ = \arg\max_{\beta} Q^\circ(\beta) \)
Bootstrapped test statistic

Reproduction of homogeneous situation under $H_0$

\[
T_t^{o(m)} = \frac{n_t^{(m-1)}}{n_t^{(m)}} \max_{\beta} Q^o(\beta, l_t^{(m-1)})
\]
\[
+ \frac{n_t^{(m)} - n_t^{(m-1)}}{n_t^{(m)}} \max_{\beta} Q^o(\beta, l_t^{(m)} \setminus l_t^{(m-1)})
\]
\[
- \max_{\beta} Q^o(\beta_{ts}, l_t^{(m)})
\]

with

\[
\beta_{ts} = \begin{cases} 
\beta & \text{for } l_t^{(m-1)} \\
\beta + \tilde{\beta}_{t12} & \text{for } l_t^{(m)} \setminus l_t^{(m-1)}
\end{cases}
\]

where $\tilde{\beta}_{t12} = \arg\max_{\beta} Q(\beta, l_t^{(m)} \setminus l_t^{(m-1)}) - \arg\max_{\beta} Q(\beta, l_t^{(m-1)})$
Validity of multiplier bootstrap

Under $H_0$

$$\mathcal{L} \left( T_t^{(m)} \right) \approx \mathcal{L} \left( T_t^{o(m)} \right | Y)$$

Critical values $\zeta_m$ for $(1 - \alpha)\%$ confidence level approximated by

$$\zeta_{t\alpha}^{o(m)} = \inf \left\{ z \geq 0 : P \left( T_t^{o(m)} > z | Y \right) \leq \alpha \right\}$$

for $m = 2, \ldots, M$
Model definition

- $y_t^{(k)}$ log yield of $k$-year discount bond at time $t$, $k = 2, \ldots, 5$
- $f_t^{(k)}$ log forward rate for loans between time $t + k - 1$ and $t + k$ specified at time $t$
- Excess log returns $r_{x_{t+1}}^{(k)}$ of $k$-year bonds

\[
rx_{t+1}^{(k)} = \beta_0^{(k)} + \beta_1^{(k)\top} f_t + \beta_2^{(k)\top} M_t + \varepsilon_{t+1}^{(k)}
\]

with $f_t = (y_t^{(1)}, f_t^{(2)}, \ldots, f_t^{(5)})$ and $M_t$ vector of macro variables

- Cochrane and Piazzesi (2005): Single forward factor (CP1F)
- Ludvigson and Ng (2009): 5 macro factors with single forward factor (LN5F) or 6 macro factors (LN6F)
Model settings

- Constant increment $n_t^{(m)} - n_t^{(m-1)} = 48, n_t^{(1)} = 48$
- Dimension $p = 36$
- Multipliers $u_i \sim \text{Pois}(1)$ for $i = 1, \ldots, 1000$
- Confidence level $(1 - \alpha) = 99\%$

Penalized Adaptive Method
Out-of-sample fit

Figure 3: Real values (black) and 1-year ahead predictions (CP1F, LN6F, PAM) of 2-year bonds, Dec 2001 - Dec 2011
Out-of-sample fit performance

<table>
<thead>
<tr>
<th></th>
<th>$r_{x_{t+1}}^{(2)}$</th>
<th>$r_{x_{t+1}}^{(3)}$</th>
<th>$r_{x_{t+1}}^{(4)}$</th>
<th>$r_{x_{t+1}}^{(5)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSPE</td>
<td>MAPE</td>
<td>RMSPE</td>
<td>MAPE</td>
</tr>
<tr>
<td>CP</td>
<td>0.008</td>
<td>0.007</td>
<td>0.50</td>
<td>0.43</td>
</tr>
<tr>
<td>CP1F</td>
<td>0.008</td>
<td>0.006</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>LN5F</td>
<td>0.008</td>
<td>0.006</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>LN6F</td>
<td>0.006</td>
<td>0.005</td>
<td>0.67</td>
<td>0.60</td>
</tr>
<tr>
<td>PAM</td>
<td>0.004</td>
<td>0.003</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CP</td>
<td>0.021</td>
<td>0.017</td>
<td>0.57</td>
<td>0.59</td>
</tr>
<tr>
<td>CP1F</td>
<td>0.021</td>
<td>0.018</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>LN5F</td>
<td>0.021</td>
<td>0.018</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>LN6F</td>
<td>0.021</td>
<td>0.017</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>PAM</td>
<td>0.025</td>
<td>0.021</td>
<td>0.64</td>
<td>0.62</td>
</tr>
<tr>
<td>CP</td>
<td>0.026</td>
<td>0.022</td>
<td>0.62</td>
<td>0.59</td>
</tr>
</tbody>
</table>

**Table 1:** Forecasting performance of PAM, Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009) models. PAM: Penalized Adaptive Method.
Conclusion

Penalized adaptive method

- Fully data-driven method
- Capturing non-stationarity and effective dimension reduction
- Improved performance in excess bond returns modelling

Outlook

- Inference for $p > n$ case
- Extension beyond linear models
- Optimal penalty parameter selection
Penalized Adaptive Method in Forecasting with Large Information Set and Structure Change

Wolfgang K. Härdle
Xinjue Li
Lenka Zboňáková

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
WISE –The Wang Yanan Institute for Studies in Economics
Xiamen University
lvb.wiwi.hu-berlin.de
irtg1792.hu-berlin.de
wise.cmu.edu.cn
References

Chernozhukov, V., Chetverikov, D. and Kato, K.
Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors

Cochrane, J. H. and Piazzesi, M.
Bond risk premia

Fan, J. and Li, R.
Variable selection via nonconcave penalized likelihood and its oracle properties

Penalized Adaptive Method
References

Favara, G., Gilchrist, S., Lewis, F. L. and Zakrajšek, E.  
Recession risk and the excess bond premium  

Kim, Y., Choi, H. and Oh, H. S.  
Smoothly clipped absolute deviation in high dimensions  

Ludvigson, S. C. and Ng, S.  
Macro factors in bond risk premia  
References

Polzehl, J. and Spokoiny, V. G.
Spatially adaptive regression estimation: Propagation-separation approach
WIAS-Preprint No. 998, 2004

Polzehl, J. and Spokoiny, V. G.
Propagation-separation approach for local likelihood estimation

Spokoiny, V. G. and Zhilova, M.
Bootstrap confidence sets under model misspecification
References

Suvorikova, A., Spokoiny, V. G. and Buzun, N.  
*Multiscale parametric approach for change point detection*  
Information Technology and Systems 2015: 979-996, 2015

Tibshirani, R.  
*Regression shrinkage and selection via the Lasso*  

Zou, H. and Li, R.  
*One-step sparse estimates in nonconcave penalized likelihood models*  

Penalized Adaptive Method
LLA algorithm

- Zou and Li (2008): for $p < n$ set $\beta^{(0)}$ as unpenalized MLE
- Kim et al. (2008): for $p > n$ set $\beta^{(0)}$ as LASSO estimator
- Algorithm
  1. Set initial value $\beta^{(0)}$
  2. For $k = 0, 1, \ldots$, repeatedly solve

$$
\beta^{(k+1)} = \arg \max_{\beta} \left\{ \sum_{i=1}^{n} l_i(\beta) - n \sum_{j=1}^{p} p'_{\lambda}(|\beta_j^{(k)}|)|\beta_j| \right\}
$$

until convergence
LLA estimator

- Continuous
- Unbiased for large parameters
- Oracle properties
  - Consistency in variable selection
  - Asymptotic normality

under condition:

$$\sqrt{n}\lambda_n \to \infty \text{ and } \lambda_n \to 0$$
## Macroeconomic variables I

<table>
<thead>
<tr>
<th>Description</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Personal Income</td>
<td>$\Delta \log$</td>
</tr>
<tr>
<td>2. Real Consumption</td>
<td>$\Delta \log$</td>
</tr>
<tr>
<td>3. Industrial Production Index (Total)</td>
<td>$\Delta \log$</td>
</tr>
<tr>
<td>4. NAPM Production Index (Percent)</td>
<td>–</td>
</tr>
<tr>
<td>5. Civilian Labor Force: Employed, Total</td>
<td>$\Delta \log$</td>
</tr>
<tr>
<td>6. Unemployment Rate: All workers, 16 years &amp; over (Percent)</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>7. NAPM Employment Index (Percent)</td>
<td>–</td>
</tr>
<tr>
<td>8. Money Stock M1</td>
<td>$\Delta^2 \log$</td>
</tr>
<tr>
<td>9. Money Stock M2</td>
<td>$\Delta^2 \log$</td>
</tr>
<tr>
<td>10. Money Stock M3</td>
<td>$\Delta^2 \log$</td>
</tr>
<tr>
<td>11. S&amp;P500 Common Stock Price Index: Composite</td>
<td>$\Delta \log$</td>
</tr>
<tr>
<td>12. Interest Rate: Federal Funds (% p.a.)</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>13. Commercial Paper Rate</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>14. Interest Rate: US Treasury Bill, Sec Mkt, 3-m (% p.a.)</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>15. Interest Rate: US Treasury Bill, Sec Mkt, 3-m (% p.a.)</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>16. Interest Rate: US Treasury Const Maturities, 1-y (% p.a.)</td>
<td>$\Delta$</td>
</tr>
</tbody>
</table>

Table 2: List of macroeconomic variables from Ludvigson and Ng (2009)
# Macroeconomic variables II

<table>
<thead>
<tr>
<th>Description</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>17. Interest Rate: US Treasury Const Maturities, 5-y (% p.a.)</td>
<td>∆</td>
</tr>
<tr>
<td>18. Interest Rate: US Treasury Const Maturities, 10-y (% p.a.)</td>
<td>∆</td>
</tr>
<tr>
<td>19. Bond Yield: Moody’s Aaa Corporate (% p.a.)</td>
<td>∆</td>
</tr>
<tr>
<td>20. Bond Yield: Moody’s Baa Corporate (% p.a.)</td>
<td>∆</td>
</tr>
<tr>
<td>21. cp90 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>22. fygm3 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>23. fygm6 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>24. fygt1 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>25. fygt5 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>26. fygt10 - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>27. fyaaac - fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>28. fybaac- fyff Spread</td>
<td>–</td>
</tr>
<tr>
<td>29. Spot Market Price Index: all commodities</td>
<td>Δ² log</td>
</tr>
<tr>
<td>30. NAPM Commodity Prices Index (Percent)</td>
<td>–</td>
</tr>
<tr>
<td>31. CPI-U: All items</td>
<td>Δ² log</td>
</tr>
</tbody>
</table>

Table 3: List of macroeconomic variables from Ludvigson and Ng (2009)
Simulation settings

- 1000 scenarios
- Design matrix $X(n \times p)$
  \[
  \{X_i\}_{i=1}^n \sim \mathcal{N}_p(0, \Sigma),
  \]
  \[n = 100, 200, 400, \quad p = 10, \quad q = \|\beta\|_0 = 3, 5\]
- Covariance matrix $\Sigma(p \times p)$
  \[\sigma_{ij} = 0.5|i-j|\]
  \[i, j = 1, \ldots, p\]
- $b = 1000$ bootstrap samples
- $u_i \sim \text{Exp}(1), \text{Pois}(1)$ or from a bounded distribution

Penalized Adaptive Method
# Bootstrapped quantiles coverage probability

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>q</th>
<th>$\mathcal{L}(\xi_j)$</th>
<th>90 %</th>
<th>95 %</th>
<th>97.5 %</th>
<th>99 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10</td>
<td>3</td>
<td>Bounded</td>
<td>81.2</td>
<td>86.5</td>
<td>91.0</td>
<td>94.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>80.3</td>
<td>86.5</td>
<td>90.5</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>87.9</td>
<td>92.4</td>
<td>95.3</td>
<td>97.8</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>5</td>
<td>Bounded</td>
<td>77.9</td>
<td>83.7</td>
<td>87.6</td>
<td>93.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>76.7</td>
<td>83.0</td>
<td>88.3</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>85.7</td>
<td>90.5</td>
<td>94.8</td>
<td>97.5</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>3</td>
<td>Bounded</td>
<td>90.7</td>
<td>94.8</td>
<td>97.3</td>
<td>98.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>90.4</td>
<td>95.2</td>
<td>97.3</td>
<td>98.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>92.9</td>
<td>96.5</td>
<td>98.3</td>
<td>99.2</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>5</td>
<td>Bounded</td>
<td>86.7</td>
<td>92.1</td>
<td>96.0</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>85.9</td>
<td>91.9</td>
<td>96.0</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>90.4</td>
<td>94.8</td>
<td>97.4</td>
<td>99.2</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td>3</td>
<td>Bounded</td>
<td>97.1</td>
<td>98.6</td>
<td>99.6</td>
<td>99.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>97.2</td>
<td>98.6</td>
<td>99.5</td>
<td>99.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>97.7</td>
<td>98.8</td>
<td>99.6</td>
<td>99.8</td>
</tr>
<tr>
<td>400</td>
<td>10</td>
<td>5</td>
<td>Bounded</td>
<td>94.3</td>
<td>97.5</td>
<td>98.5</td>
<td>99.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exp(1)</td>
<td>94.4</td>
<td>97.6</td>
<td>98.5</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pois(1)</td>
<td>95.2</td>
<td>98.1</td>
<td>98.5</td>
<td>99.4</td>
</tr>
</tbody>
</table>

Table 4: Empirical coverage probabilities

Penalized Adaptive Method
Bounded distribution

Random variable $Z$ with values in the interval $[0, 4]$ with a probability density function defined as

$$f(z) = \begin{cases} 
\frac{3}{14} & \text{if } 0 \leq z \leq 1 \\
\frac{1}{12} & \text{if } 1 < z \leq 4
\end{cases} \quad (1)$$
Change points detection

- 500 scenarios of \( n = 500 \) observations
- Design matrix \( X_{(n \times p)} \) as before, \( p = 10 \)
- Number of intervals \( M = 10, 5 \), \( n^{(m+1)} - n^{(m)} = 100, 50 \) for \( m = 1, \ldots, M - 1 \), \( n^{(1)} = 100, 50 \)
- Change point simulation

\[
\beta^*_i = \begin{cases} 
(1, 1, 1, 1, 1, 0, \ldots, 0) & \text{if } i < i_{cp} \\
(1, 0.8, 0.6, 0.4, 0.2, 0, \ldots, 0) & \text{if } i \geq i_{cp}
\end{cases}
\]

where \( i_{cp} \) denotes observation with a change point
- \( b = 1000, u_i \sim \text{Exp}(1), \text{Pois}(1) \) or from a bounded distribution
## Change points detection summary

<table>
<thead>
<tr>
<th></th>
<th>$M = 10$</th>
<th>$i_{cp}$</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded</td>
<td>Corr</td>
<td>99.6</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>99.6</td>
<td>84.6</td>
<td>62.6</td>
<td>42.8</td>
<td></td>
</tr>
<tr>
<td>Exp(1)</td>
<td>Corr</td>
<td>99.6</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>99.6</td>
<td>84.6</td>
<td>62.6</td>
<td>42.8</td>
<td></td>
</tr>
<tr>
<td>Pois(1)</td>
<td>Corr</td>
<td>99.4</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>99.4</td>
<td>91.8</td>
<td>76.8</td>
<td>59.8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$M = 5$</th>
<th>$i_{cp}$</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>91.6</td>
<td>83.0</td>
<td></td>
</tr>
<tr>
<td>Exp(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>91.2</td>
<td>82.4</td>
<td></td>
</tr>
<tr>
<td>Pois(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>94.0</td>
<td>87.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Percentage of correctly identified change points PAMsimCP Penalized Adaptive Method
Change points detection

- 500 scenarios of $n = 500$ observations
- Design matrix $X_{(n \times p)}$ as before, $p = 10$
- Number of intervals $M = 10, 5$, $n^{(m+1)} - n^{(m)} = 100, 50$ for $m = 1, \ldots, M - 1$, $n^{(1)} = 100, 50$
- Change point simulation

$$\beta_i^* = \begin{cases} 
(1, 1, 1, 1, 1, 0, \ldots, 0) & \text{if } i < i_{cp} \\
(1, 1, 1, 0, \ldots, 0) & \text{if } i \geq i_{cp}
\end{cases}$$

where $i_{cp}$ denotes observation with a change point
- $b = 1000$, $u_i \sim \text{Exp}(1)$, Pois(1) or from a bounded distribution

Penalized Adaptive Method
## Change points detection summary

<table>
<thead>
<tr>
<th>$M = 10$</th>
<th>$i_{cp}$</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded</td>
<td>Corr</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>85.4</td>
<td>64.0</td>
<td>41.8</td>
</tr>
<tr>
<td>Exp(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>84.4</td>
<td>63.0</td>
<td>41.0</td>
</tr>
<tr>
<td>Pois(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>91.8</td>
<td>77.8</td>
<td>58.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M = 5$</th>
<th>$i_{cp}$</th>
<th>100</th>
<th>200</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bounded</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>91.0</td>
<td>82.4</td>
</tr>
<tr>
<td>Exp(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>91.4</td>
<td>82.0</td>
</tr>
<tr>
<td>Pois(1)</td>
<td>Corr</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>1stCorr</td>
<td>100.0</td>
<td>94.0</td>
<td>87.8</td>
</tr>
</tbody>
</table>

Table 6: Percentage of correctly identified change points

Penalized Adaptive Method
Out-of-sample fit for 3-year bonds

Figure 4: Real values (black) and 1-year ahead predictions (CP1F, LN6F, PAM) of 3-year bonds, Dec 2001 - Dec 2011

Penalized Adaptive Method
Appendix

Out-of-sample fit for 4-year bonds

Figure 5: Real values (black) and 1-year ahead predictions (CP1F, LN6F, PAM) of 4-year bonds, Dec 2001 - Dec 2011
Out-of-sample fit for 5-year bonds

Figure 6: Real values (black) and 1-year ahead predictions (CP1F, LN6F, PAM) of 5-year bonds, Dec 2001 - Dec 2011
Appendix

**In-sample fit for 2-year bonds**

![Graph showing in-sample fit for 2-year bonds]

**Figure 7:** Real excess bond premium (black) and fitted CP1F, LN6F, PAM for 2-year bonds, Jan 1961 - Dec 2011. PAMinsam Penalized Adaptive Method.
In-sample fit for 3-year bonds

Figure 8: Real excess bond premium (black) and fitted CP1F, LN6F, PAM for 3-year bonds, Jan 1961 - Dec 2011
PAMinsam
Penalized Adaptive Method
In-sample fit for 4-year bonds

Figure 9: Real excess bond premium (black) and fitted CP1F, LN6F, PAM for 4-year bonds, Jan 1961 - Dec 2011

Penalized Adaptive Method
In-sample fit for 5-year bonds

Figure 10: Real excess bond premium (black) and fitted CP1F, LN6F, PAM for 5-year bonds, Jan 1961 - Dec 2011

Penalized Adaptive Method
# In-sample fit performance

<table>
<thead>
<tr>
<th></th>
<th>Jan 1964 - Dec 2003</th>
<th>Jan 1961 - Dec 2011</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAE</td>
<td>$R^2$</td>
<td>$R_{adj}^2$</td>
<td>RMSE</td>
<td>MAE</td>
<td>$R^2$</td>
<td>$R_{adj}^2$</td>
</tr>
<tr>
<td>$r_x^{(2)}_{t+1}$</td>
<td>CP</td>
<td>0.007</td>
<td>0.005</td>
<td>0.322</td>
<td>0.315</td>
<td>0.007</td>
<td>0.005</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>CP1F</td>
<td>0.007</td>
<td>0.005</td>
<td>0.318</td>
<td>0.316</td>
<td>0.007</td>
<td>0.005</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>LN5F</td>
<td>0.007</td>
<td>0.005</td>
<td>0.365</td>
<td>0.357</td>
<td>0.006</td>
<td>0.004</td>
<td>0.377</td>
</tr>
<tr>
<td></td>
<td>LN6F</td>
<td>0.005</td>
<td>0.004</td>
<td>0.579</td>
<td>0.574</td>
<td>0.005</td>
<td>0.004</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>PAM</td>
<td>0.001</td>
<td>0.001</td>
<td>0.980</td>
<td>0.979</td>
<td>0.001</td>
<td>0.001</td>
<td>0.978</td>
</tr>
<tr>
<td>$r_x^{(3)}_{t+1}$</td>
<td>CP</td>
<td>0.012</td>
<td>0.010</td>
<td>0.340</td>
<td>0.333</td>
<td>0.012</td>
<td>0.010</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>CP1F</td>
<td>0.012</td>
<td>0.010</td>
<td>0.338</td>
<td>0.336</td>
<td>0.012</td>
<td>0.010</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>LN5F</td>
<td>0.012</td>
<td>0.009</td>
<td>0.385</td>
<td>0.377</td>
<td>0.011</td>
<td>0.008</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>LN6F</td>
<td>0.010</td>
<td>0.008</td>
<td>0.532</td>
<td>0.526</td>
<td>0.010</td>
<td>0.008</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>PAM</td>
<td>0.003</td>
<td>0.002</td>
<td>0.970</td>
<td>0.970</td>
<td>0.002</td>
<td>0.002</td>
<td>0.970</td>
</tr>
<tr>
<td>$r_x^{(4)}_{t+1}$</td>
<td>CP</td>
<td>0.017</td>
<td>0.013</td>
<td>0.370</td>
<td>0.363</td>
<td>0.017</td>
<td>0.013</td>
<td>0.253</td>
</tr>
<tr>
<td></td>
<td>CP1F</td>
<td>0.017</td>
<td>0.013</td>
<td>0.369</td>
<td>0.368</td>
<td>0.017</td>
<td>0.013</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td>LN5F</td>
<td>0.016</td>
<td>0.013</td>
<td>0.414</td>
<td>0.407</td>
<td>0.015</td>
<td>0.012</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>LN6F</td>
<td>0.015</td>
<td>0.012</td>
<td>0.486</td>
<td>0.479</td>
<td>0.015</td>
<td>0.011</td>
<td>0.420</td>
</tr>
<tr>
<td></td>
<td>PAM</td>
<td>0.004</td>
<td>0.003</td>
<td>0.968</td>
<td>0.967</td>
<td>0.003</td>
<td>0.003</td>
<td>0.967</td>
</tr>
<tr>
<td>$r_x^{(5)}_{t+1}$</td>
<td>CP</td>
<td>0.021</td>
<td>0.016</td>
<td>0.344</td>
<td>0.337</td>
<td>0.021</td>
<td>0.016</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>CP1F</td>
<td>0.021</td>
<td>0.016</td>
<td>0.344</td>
<td>0.343</td>
<td>0.021</td>
<td>0.016</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>LN5F</td>
<td>0.020</td>
<td>0.016</td>
<td>0.386</td>
<td>0.378</td>
<td>0.019</td>
<td>0.015</td>
<td>0.368</td>
</tr>
<tr>
<td></td>
<td>LN6F</td>
<td>0.019</td>
<td>0.015</td>
<td>0.461</td>
<td>0.454</td>
<td>0.018</td>
<td>0.014</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>PAM</td>
<td>0.005</td>
<td>0.003</td>
<td>0.965</td>
<td>0.964</td>
<td>0.005</td>
<td>0.003</td>
<td>0.962</td>
</tr>
</tbody>
</table>

Table 7: Fitted PAM and models of Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009) PAMinsam
Penalized Adaptive Method
In-sample fit summary

- Different models for different times to maturity
- Average length of homogeneous intervals 5.8 years
- Average size of active sets 13.5
- Both forward rates and macro variables selected
- Evidence of time variation