

Supplemental Materials for "Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data"

Lifeng Wang, Guang Chen and Hongzhe Li

Additional plots for simulation data

Figures 1 - 10 show the true and estimated transcriptional effects for 10 simulated TFs for noise variance of 1 and 3 based on gSCAD and simple linear regressions (SLR).

Additional plots for the yeast cell-cycle data analysis

Figure 11 shows the estimated transcriptional effects of 10 randomly selected TFs that are related to the yeast cell-cycle process.

Proof of Theorem 1

In Fan and Li (2001), the oracle property of the SCAD penalized estimates for standard linear models was established, which indicates that the SCAD penalty enables consistent variable selection and parameter estimation simultaneously, as if the subset of relevant variables is already known. To study theoretical properties of gSCAD, we generalize the arguments in Fan and Li (2001) to the group selection settings assuming that the knot locations are held fixed as the sample size increases. Without loss of generality, we consider the model with $\mu = 0$ and assume that a random design where $\mathbf{z}_{ij} = (\mathbf{x}_{ij}, y_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, T$, are independently and identically distributed with

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}^* + \varepsilon_{ij},$$

$x_{ij}^t = (B_1(t_j)X_{i1}, \dots, B_{L+4}(t_j)X_{iK})$, $E\varepsilon_{ij} = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$. We further denote $\mathbf{X} = (\mathbf{x}_{11}^t, \dots, \mathbf{x}_{nT}^t)^t$. Let $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^{*t}, \dots, \boldsymbol{\beta}_K^{*t})^t$ and assume that $\boldsymbol{\beta}^*(1) = (\boldsymbol{\beta}_1^{*t}, \dots, \boldsymbol{\beta}_s^{*t})^t$ are the non-zero coefficients, and $\boldsymbol{\beta}^*(2) = (\boldsymbol{\beta}_{s+1}^{*t}, \dots, \boldsymbol{\beta}_K^{*t})^t = 0$, i.e., the first s TFs in this model are relevant to gene expression levels over time and the next $K - s$ TFs are not relevant. Let $m = nT$ and we have the following asymptotic theorem:

Theorem 1. *Assume $\Sigma = E(\mathbf{X}\mathbf{X}^t)$ is positive definite, and $\lambda_m \rightarrow 0$ and $\sqrt{m}\lambda_m \rightarrow \infty$ as $m \rightarrow \infty$. Then,*

(a.) $\hat{\boldsymbol{\beta}}^*(2) = \mathbf{0}$ with probability approaching 1.

(b.) $m^{1/2}(\Sigma(1) + \Omega)(\hat{\boldsymbol{\beta}}^*(1) - \boldsymbol{\beta}^*(1)) \rightarrow N(0, \Sigma(1))$ in distribution, where $\Sigma(1)$ is the covariance matrix Σ corresponding to $\boldsymbol{\beta}(1)$ and $\Omega = \text{diag}\left\{\frac{\partial^2 l(\|\boldsymbol{\beta}_1^*\|)}{\partial \boldsymbol{\beta}_1^* \partial \boldsymbol{\beta}_1^{*t}}|_{\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1^*}, \dots, \frac{\partial^2 l(\|\boldsymbol{\beta}_s^*\|)}{\partial \boldsymbol{\beta}_s^* \partial \boldsymbol{\beta}_s^{*t}}|_{\boldsymbol{\beta}_s^* = \boldsymbol{\beta}_s^*}\right\}$.

Proof: First we shall prove the consistency of $\hat{\boldsymbol{\beta}}^*$ and quantify its rate of convergence by showing that $\|\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*\| = O_P(m^{1/2})$ as $\lambda \rightarrow 0$.

Let $\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^* = \delta \mathbf{u}$ with δ a scalar and \mathbf{u} a vector satisfying $\|\mathbf{u}\| = 1$. Then,

$$\begin{aligned}
l(\hat{\boldsymbol{\beta}}^*) - l(\boldsymbol{\beta}^*) &\geq \|\mathbf{Y} - \mathbf{X}(\boldsymbol{\beta}^* + \delta \mathbf{u})\|_2^2 - \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*\|_2^2 + m \sum_{k=1}^s (p_\lambda(\|\boldsymbol{\beta}_k^* + \delta \mathbf{u}\|_2) - p_\lambda(\|\boldsymbol{\beta}_k^*\|_2)) \\
&= -(2\boldsymbol{\varepsilon}^t \mathbf{X} \mathbf{u})\delta + (\mathbf{u}^t \mathbf{X}^t \mathbf{X} \mathbf{u})\delta^2 \\
&\quad + m \sum_{k=1}^s \left(\frac{\partial p_\lambda(\|\boldsymbol{\beta}_k^*\|)}{\partial \boldsymbol{\beta}_k^*} \mathbf{u}_k \delta + \frac{1}{2} \mathbf{u}_k^t \frac{\partial^2 p_\lambda(\|\boldsymbol{\beta}_k^*\|)}{\partial \boldsymbol{\beta}_k^* \partial \boldsymbol{\beta}_k^{*t}} \mathbf{u}_k \delta^2 \{1 + o(1)\} \right) \\
&= (-O_P(m^{1/2}) + n \sum_{k=1}^s \frac{\partial p_\lambda(\|\boldsymbol{\beta}_k^*\|)}{\partial \boldsymbol{\beta}_k^*} \mathbf{u}_k) \delta \\
&\quad + n(\mathbf{u}^t \Sigma \mathbf{u} + \sum_{k=1}^s \frac{1}{2} \mathbf{u}_k^t \frac{\partial^2 p_\lambda(\|\boldsymbol{\beta}_k^*\|)}{\partial \boldsymbol{\beta}_k^* \partial \boldsymbol{\beta}_k^{*t}} \mathbf{u}_k) \delta^2 \{1 + o_P(1)\}. \tag{1}
\end{aligned}$$

In (1), $\mathbf{u}^t \Sigma \mathbf{u} > 0$ because Σ is positive definite, and the first order and the second order derivatives of $p_\lambda(\|\boldsymbol{\beta}^*\|)$ are equal to zero when $\lambda \leq \min_{1 \leq k \leq s} \|\boldsymbol{\beta}_k^*\|$. Also note that $l(\hat{\boldsymbol{\beta}}^*) - l(\boldsymbol{\beta}^*) \leq 0$. Therefore, $\|\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*\| = O_P(m^{1/2})$ as $\lambda \rightarrow 0$.

Now we prove part (a) of Theorem 1. To this end, use proof by contradiction. Suppose that there exists a constant $\delta > 0$ such that with probability at least δ , there exist a large m and a $k_0 > s$ such that $\hat{\boldsymbol{\beta}}_{k_0}^* \neq \mathbf{0}$. Let $\tilde{\boldsymbol{\beta}}^*$ be a vector constructed by replacing $\hat{\boldsymbol{\beta}}_{k_0}^*$ with $\mathbf{0}$ in $\hat{\boldsymbol{\beta}}^*$. Let $m^{1/2}\lambda \rightarrow \infty$. Then, $\lambda > \|\hat{\boldsymbol{\beta}}_{k_0}^*\|$ for sufficiently large m , and

$$\begin{aligned}
l(\hat{\boldsymbol{\beta}}^*) - l(\tilde{\boldsymbol{\beta}}^*) &= \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*\|_2^2 - \|\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}^*\|_2^2 + n p_\lambda(\|\hat{\boldsymbol{\beta}}_{k_0}^*\|) \\
&= \{-2\mathbf{Y}^t \mathbf{X}(\hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^*) + \hat{\boldsymbol{\beta}}^{*t} \mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^{*t} \mathbf{X}^t \mathbf{X} \tilde{\boldsymbol{\beta}}^*\} + m\lambda \|\boldsymbol{\beta}_{k_0}^*\|. \tag{2}
\end{aligned}$$

Note that $\|\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*\| = O(m^{-1/2})$, which implies $\|\hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^*\| = O(m^{-1/2})$ and $\|\hat{\boldsymbol{\beta}}_{k_0}^*\| = O(m^{-1/2})$. Therefore, in (2), the first term is $O(1)$, whereas the second term is $O(m^{1/2}\lambda)$, which dominates the first term. This contradicts the fact that $l(\hat{\boldsymbol{\beta}}^*) - l(\tilde{\boldsymbol{\beta}}^*) \leq 0$, which proves part (a).

For part (b), note that $\|\hat{\boldsymbol{\beta}}^*(1) - \boldsymbol{\beta}^*(1)\| = O_P(m^{-1/2})$, and $\hat{\boldsymbol{\beta}}^*(2) = \mathbf{0}$ with probability approaching 1. It can be easily verified that $\hat{\boldsymbol{\beta}}^*(1)$ is a local minimizer of $l(\boldsymbol{\beta}^*(1), \boldsymbol{\beta}^*(2))|_{\boldsymbol{\beta}^*(2)=\mathbf{0}}$ as a function

of $\boldsymbol{\beta}^*(1)$. Therefore,

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*(1)} \Big|_{\boldsymbol{\beta}^* = (\hat{\boldsymbol{\beta}}^*(1)^t, \mathbf{0}^t)^t} &= -2\mathbf{X}(1)^t \mathbf{Y} + 2\mathbf{X}(1)^t \mathbf{X}(1) \hat{\boldsymbol{\beta}}^*(1) + m \left(\sum_{k=1}^s \frac{\partial p_\lambda(\|\boldsymbol{\beta}_k^*\|)}{\partial \boldsymbol{\beta}^*(1)} \right) \\
&= -2\mathbf{X}(1)^t \boldsymbol{\varepsilon} + 2\mathbf{X}(1)^t \mathbf{X}(1) (\hat{\boldsymbol{\beta}}^*(1) - \boldsymbol{\beta}^*(1)) \\
&\quad + m \{ \mathbf{b} + \Omega (\hat{\boldsymbol{\beta}}^*(1) - \boldsymbol{\beta}^*(1)) (1 + o_P(1)) \} = 0,
\end{aligned} \tag{3}$$

where

$$\mathbf{b} = ([\partial l(\|\boldsymbol{\beta}_1^*\|) / \partial (\boldsymbol{\beta}_1^*)]^t \Big|_{\boldsymbol{\beta}_1^* = \hat{\boldsymbol{\beta}}_1^*}, \dots, [\partial l(\|\boldsymbol{\beta}_s^*\|) / \partial (\boldsymbol{\beta}_s^*)]^t \Big|_{\boldsymbol{\beta}_s^* = \hat{\boldsymbol{\beta}}_s^*})^t,$$

and

$$\Omega = \text{diag} \{ \partial^2 l(\|\boldsymbol{\beta}_1^*\|) / \partial \boldsymbol{\beta}_1^* \partial \boldsymbol{\beta}_1^{*t} \Big|_{\boldsymbol{\beta}_1^* = \hat{\boldsymbol{\beta}}_1^*}, \dots, \partial^2 l(\|\boldsymbol{\beta}_s^*\|) / \partial \boldsymbol{\beta}_s^* \partial \boldsymbol{\beta}_s^{*t} \Big|_{\boldsymbol{\beta}_s^* = \hat{\boldsymbol{\beta}}_s^*} \}.$$

By standard arguments, we have

$$m^{1/2} (\boldsymbol{\Sigma}(1) + \frac{1}{2} \boldsymbol{\Omega}) (\hat{\boldsymbol{\beta}}^*(1) - \boldsymbol{\beta}^*(1)) \rightarrow N\left(\frac{\mathbf{b}}{2}, \boldsymbol{\Sigma}(1)\right).$$

Theorem 1 then follows when $\lambda \rightarrow 0$ and $m^{1/2} \lambda \rightarrow \infty$.

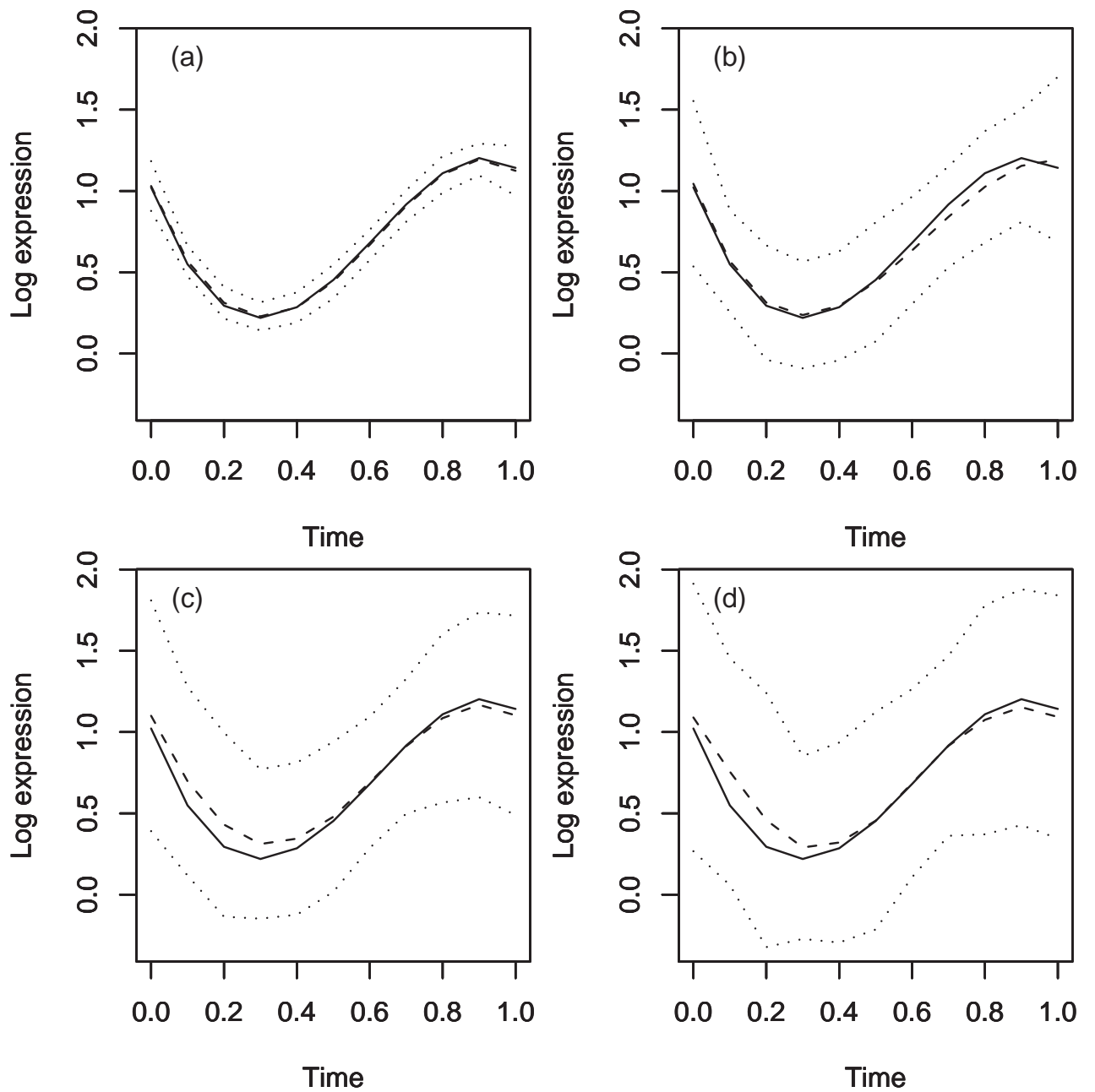


Figure 1: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 1 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

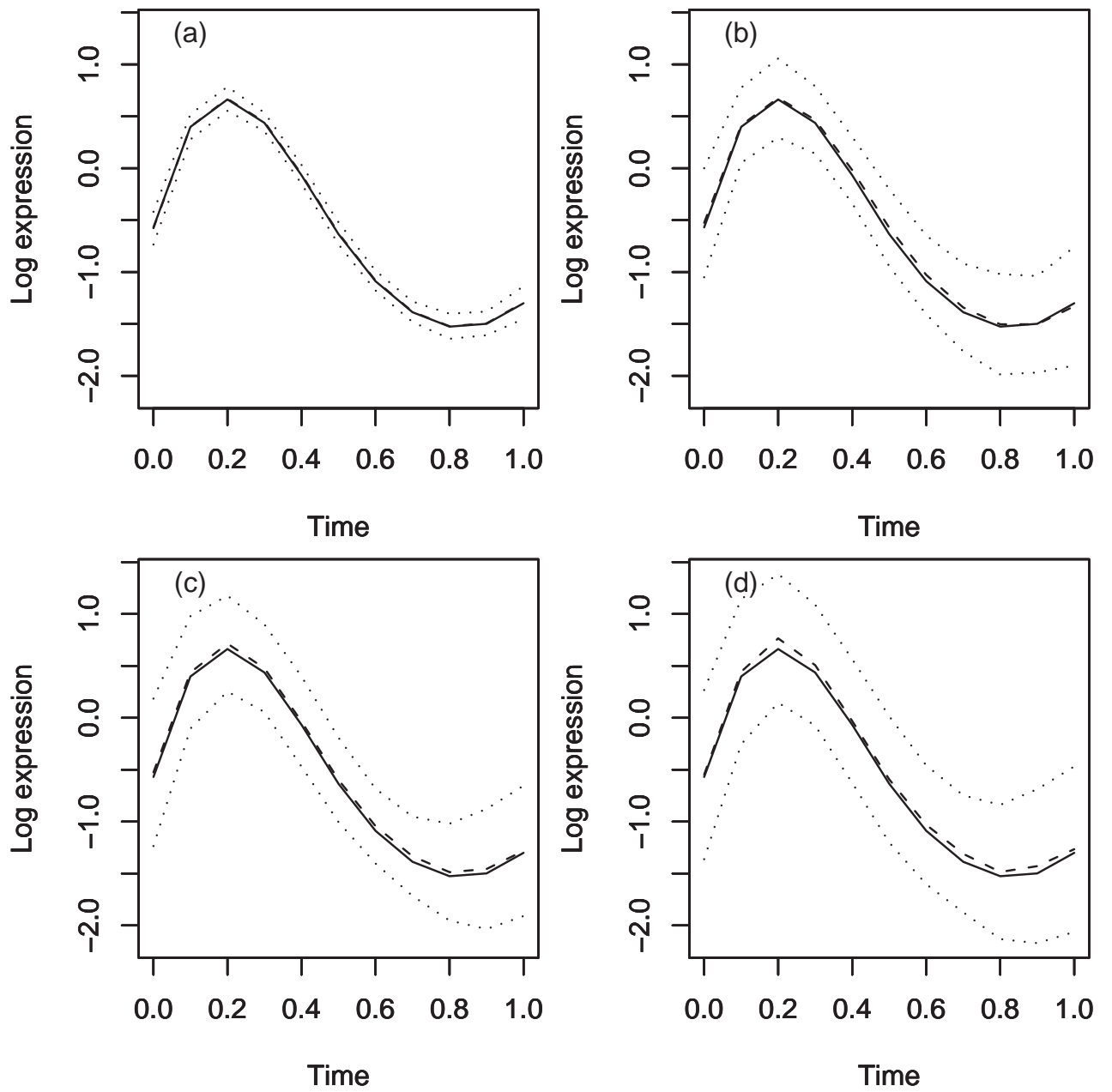


Figure 2: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 2 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

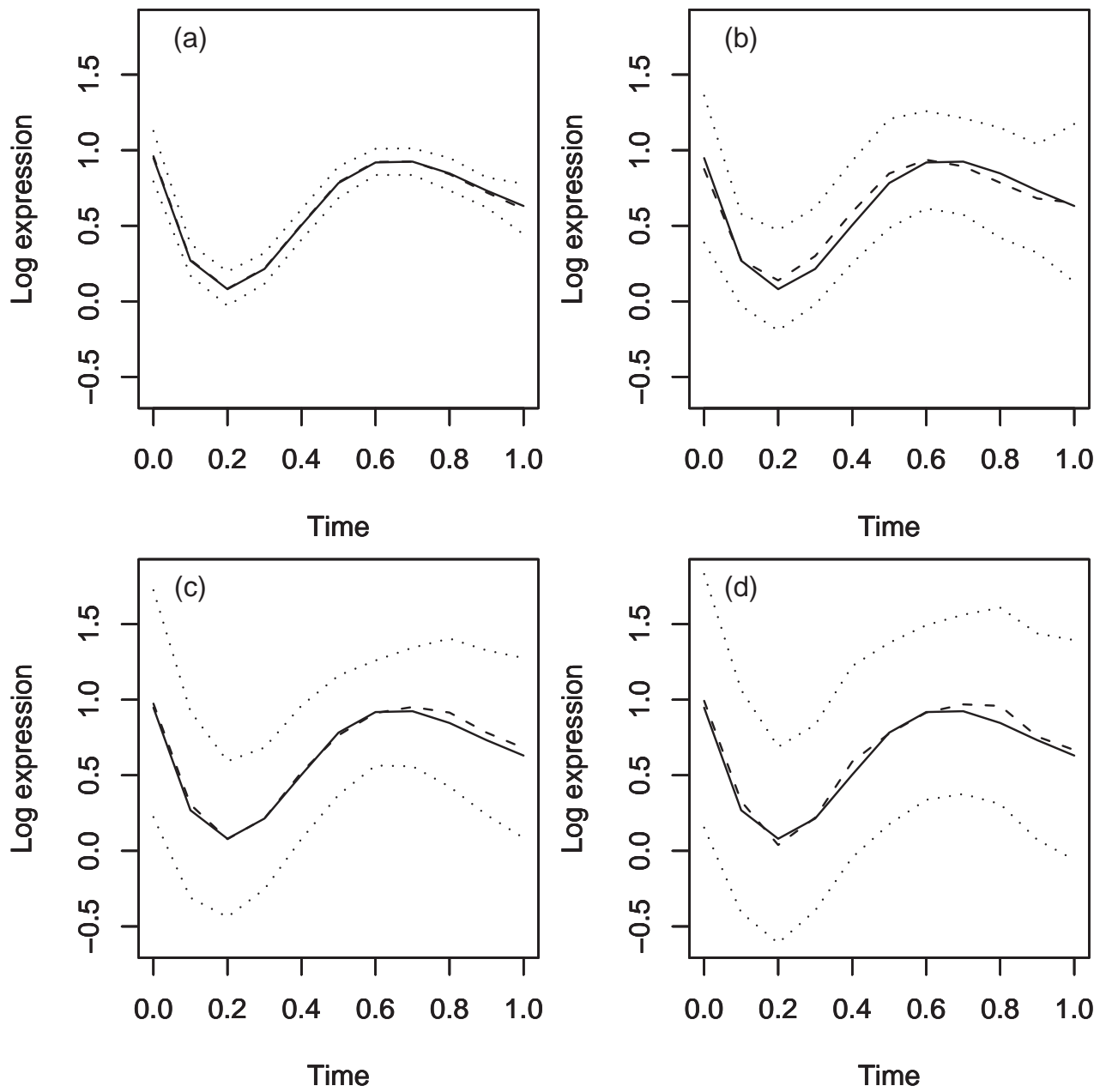


Figure 3: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 3 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

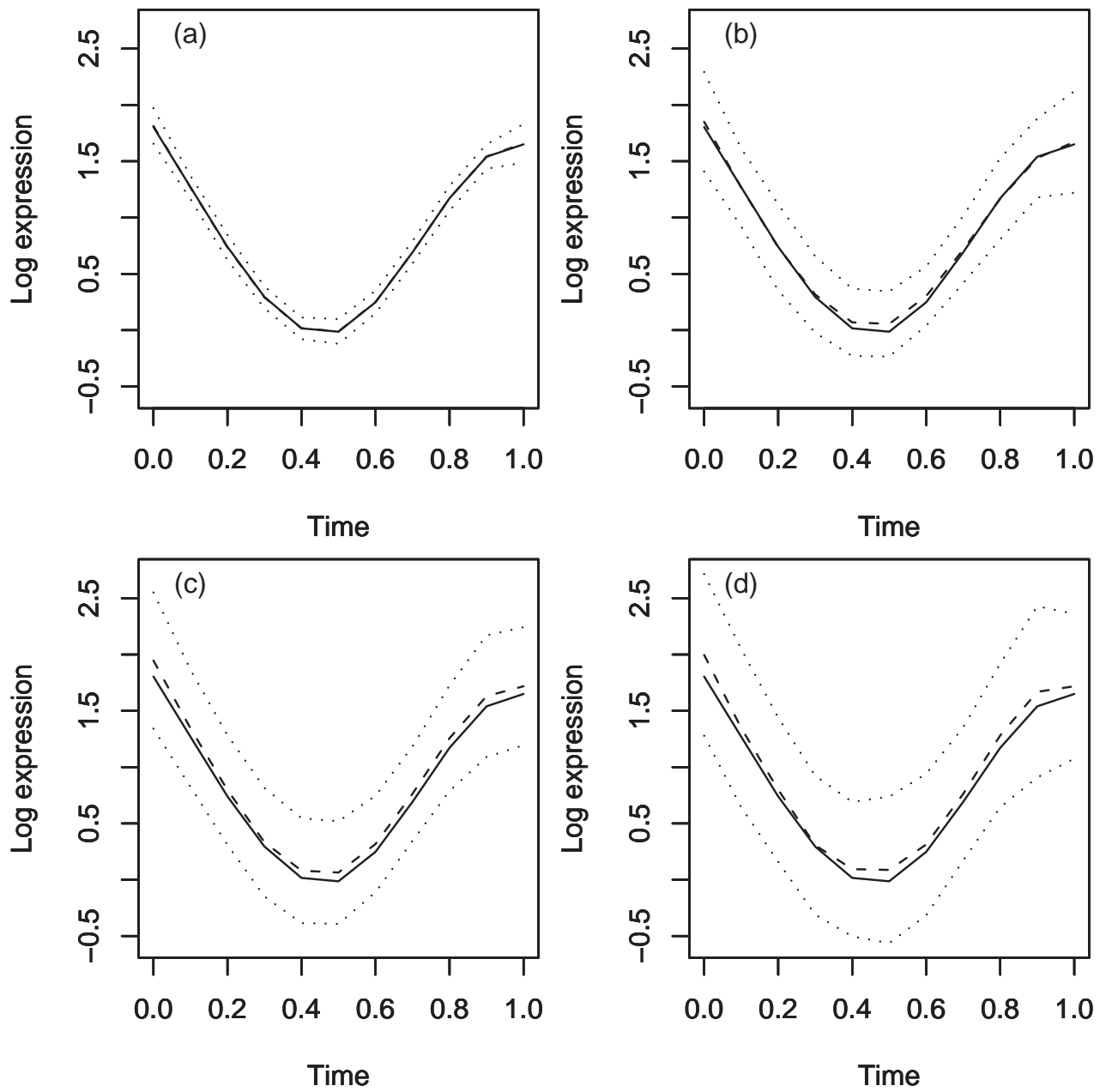


Figure 4: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 4 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

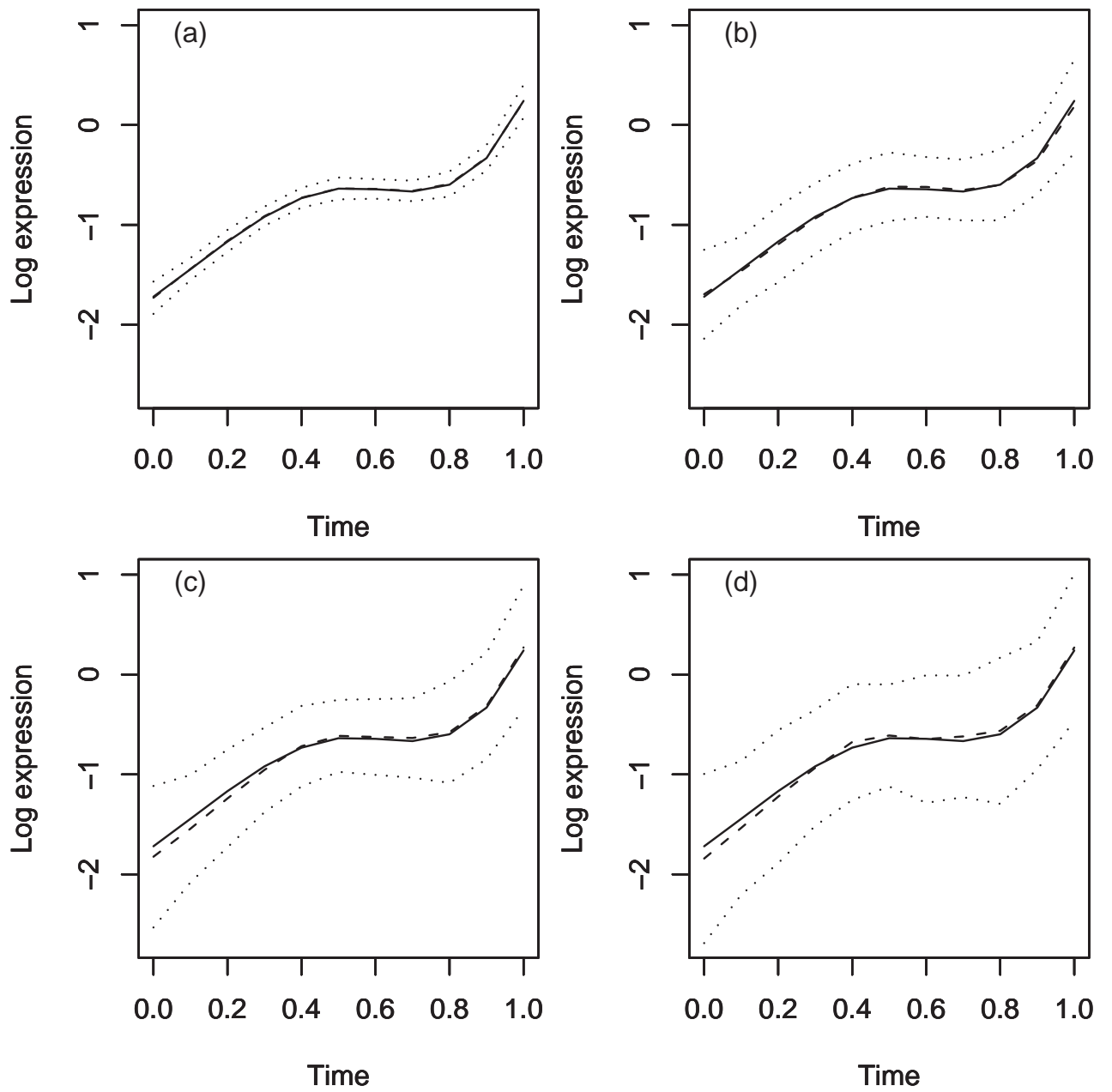


Figure 5: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 5 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

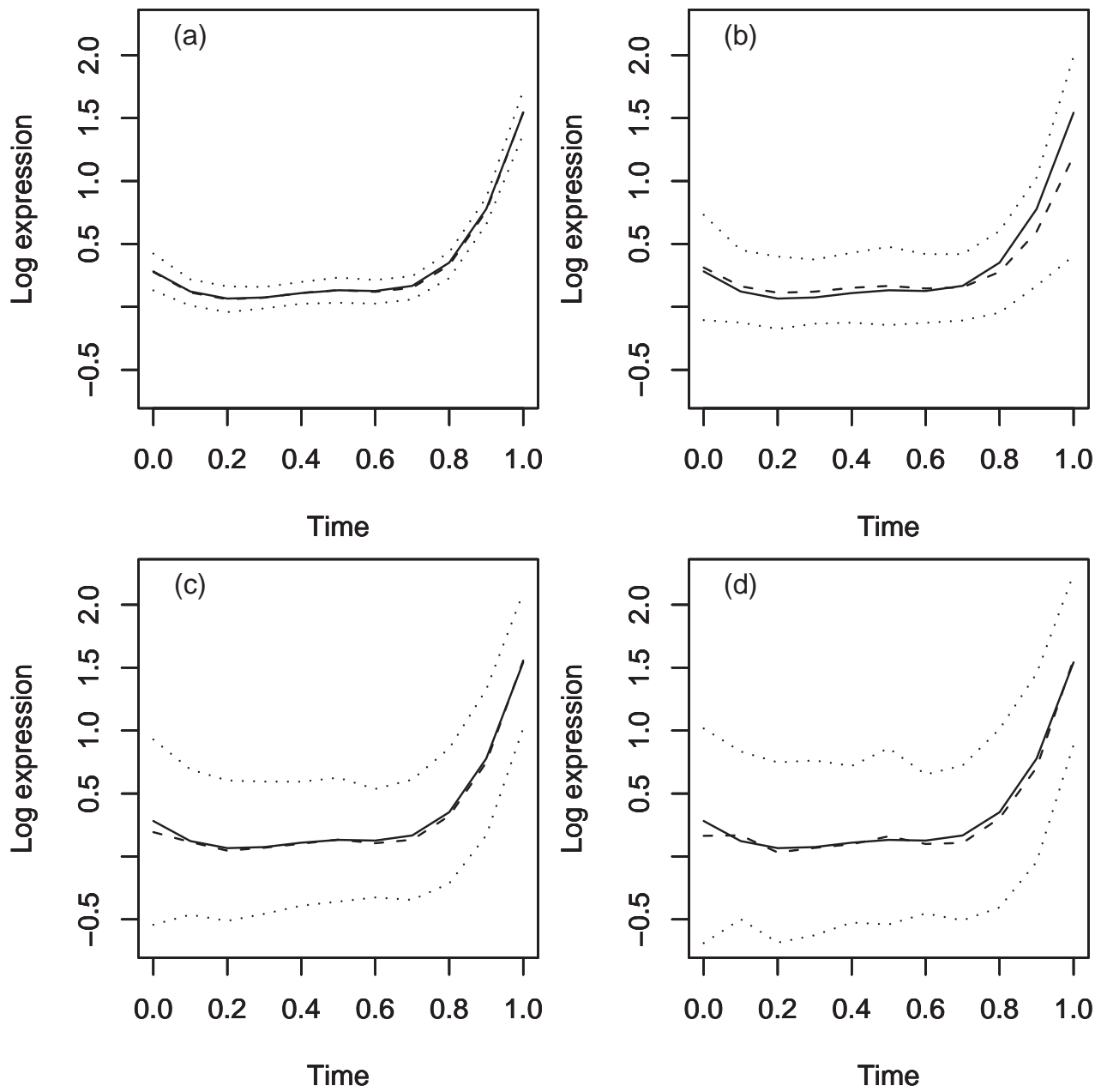


Figure 6: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 6 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

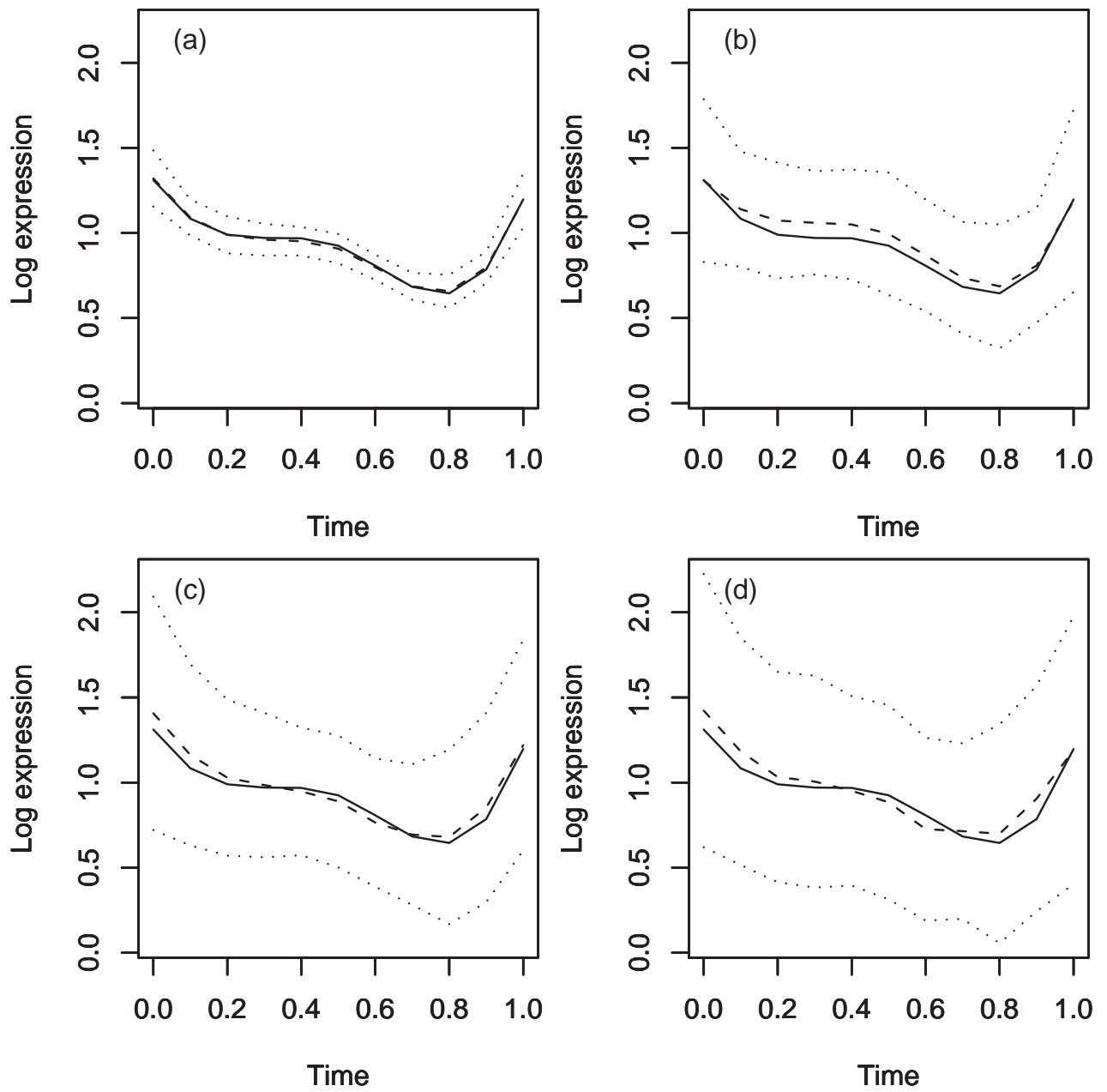


Figure 7: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 7 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

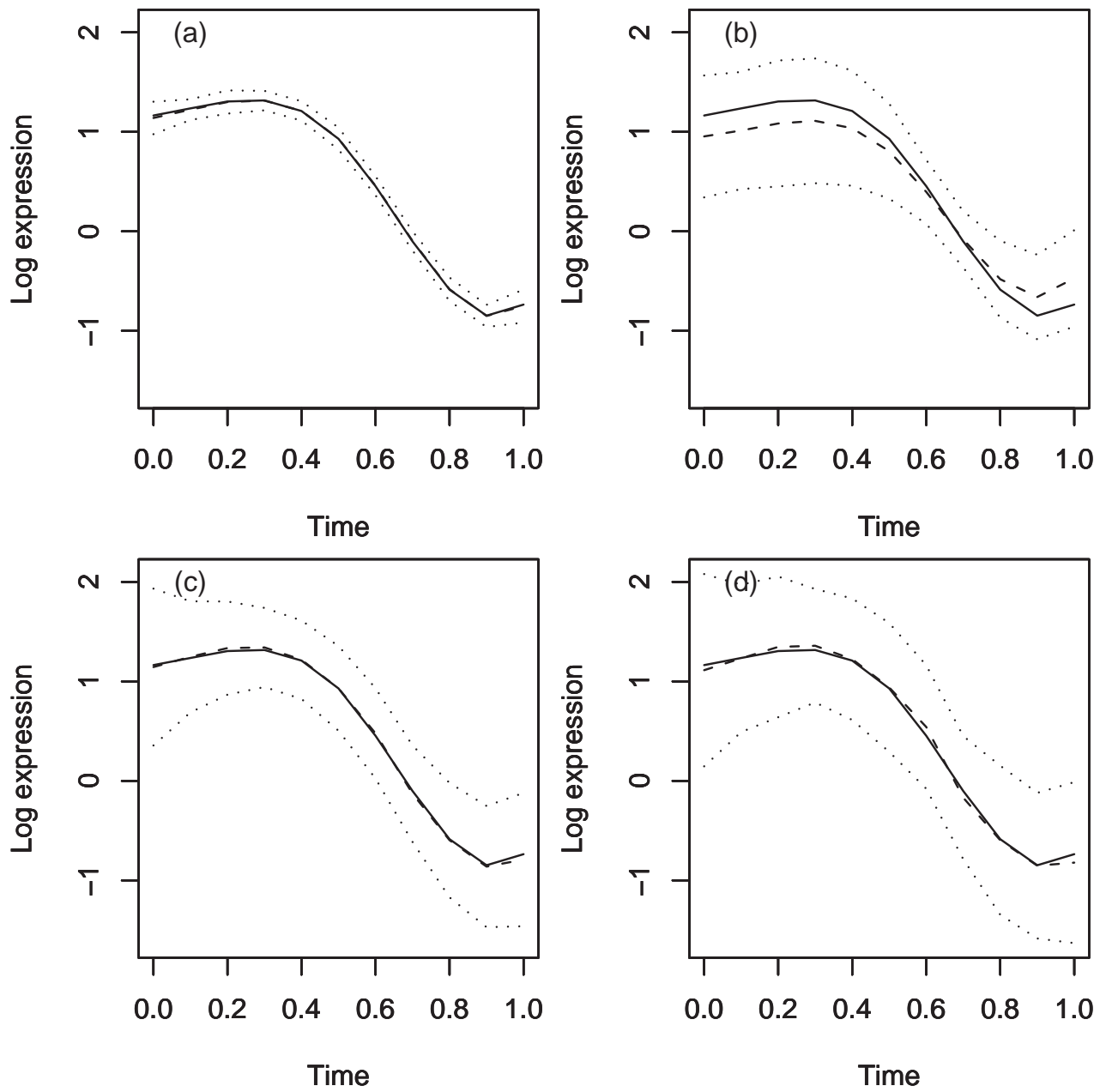


Figure 8: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 8 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

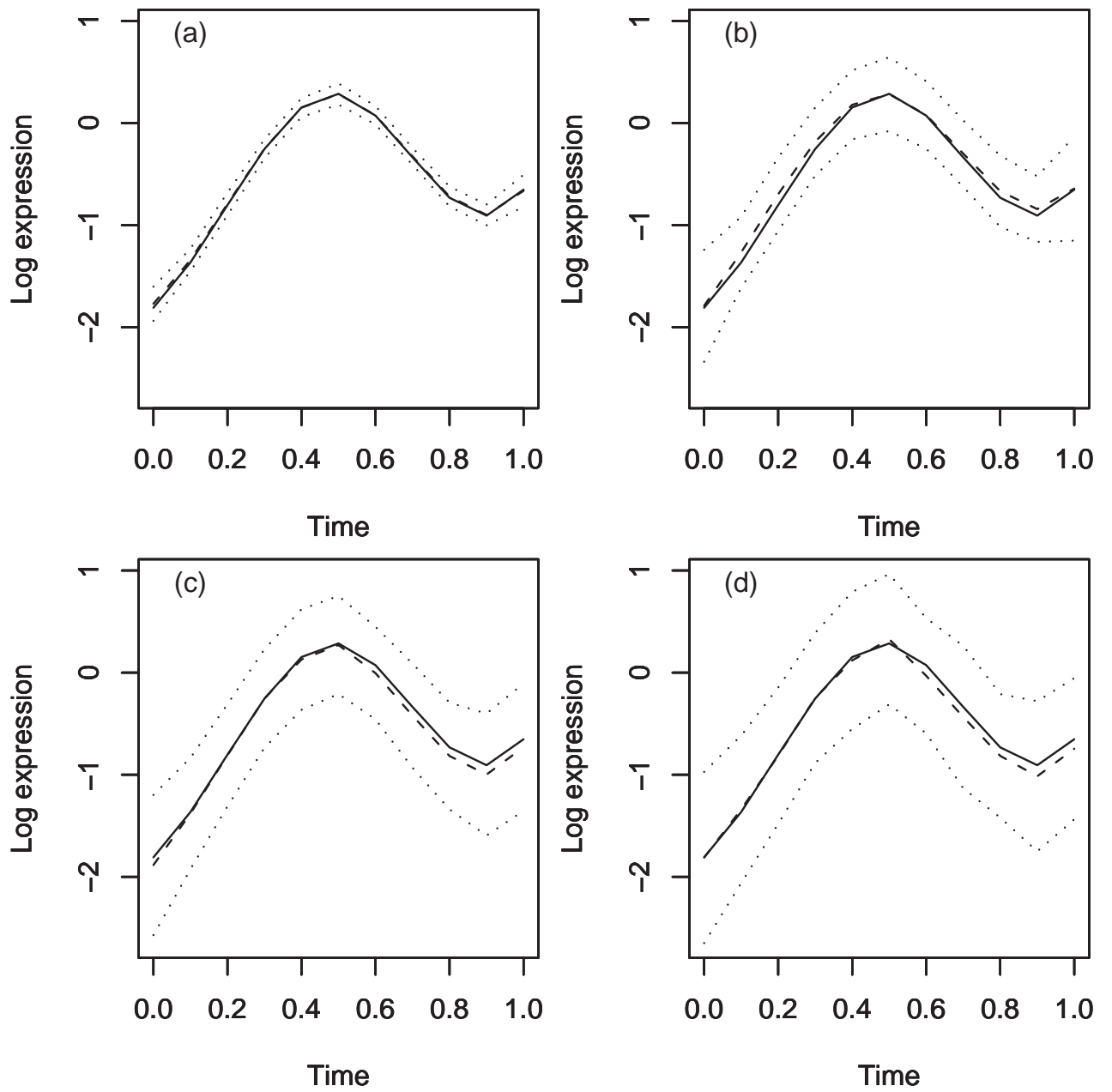


Figure 9: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 9 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

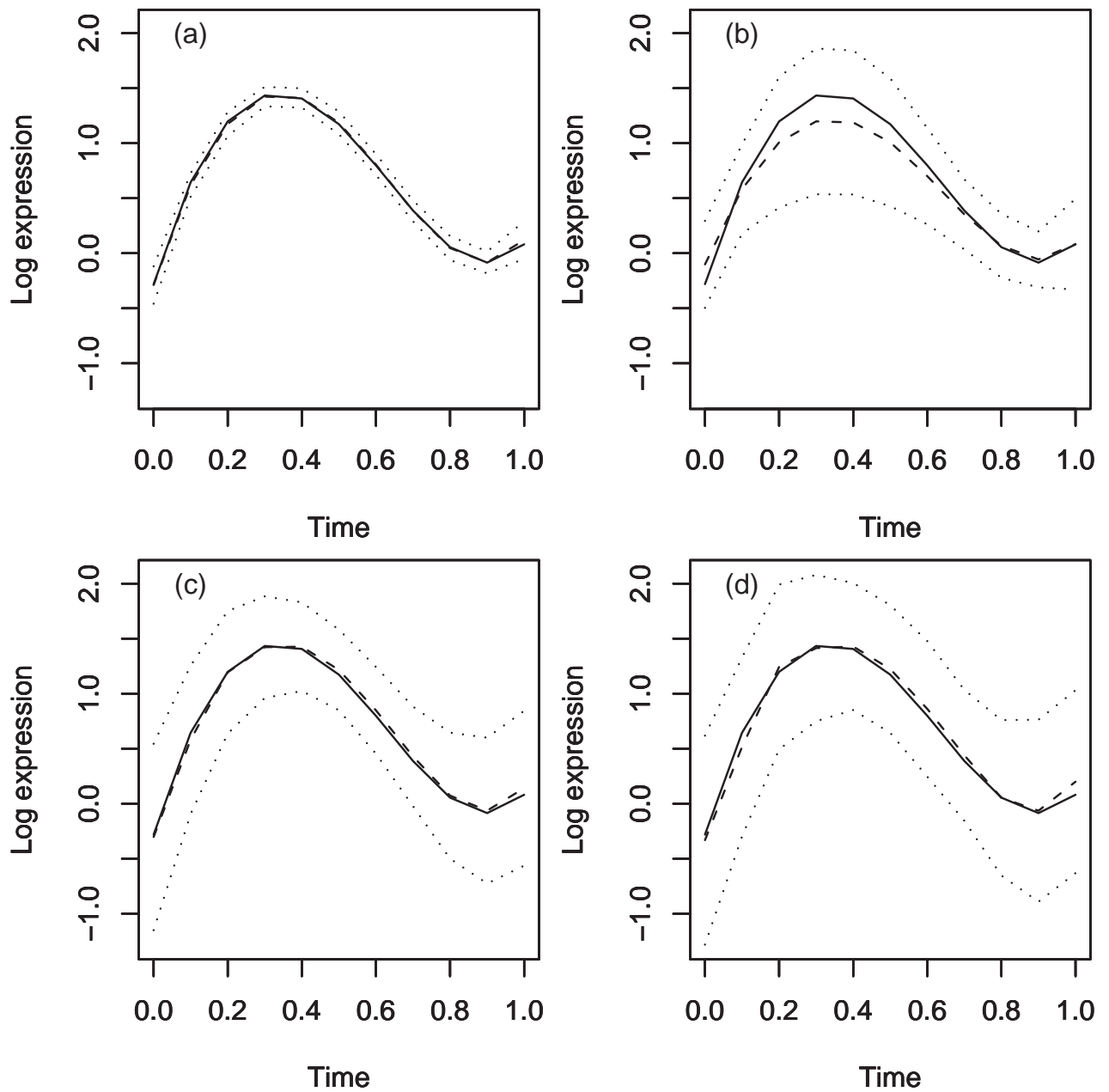


Figure 10: True (solid line) and mean (± 1 SE) of the estimated (dashed and dotted lines) time-dependent transcriptional effects for transcriptional factor 10 using the gSCAD (plots (a) and (b)) and simple linear regression (plots (c) and (d)) for noise variance of 1 (left) and 3 (right).

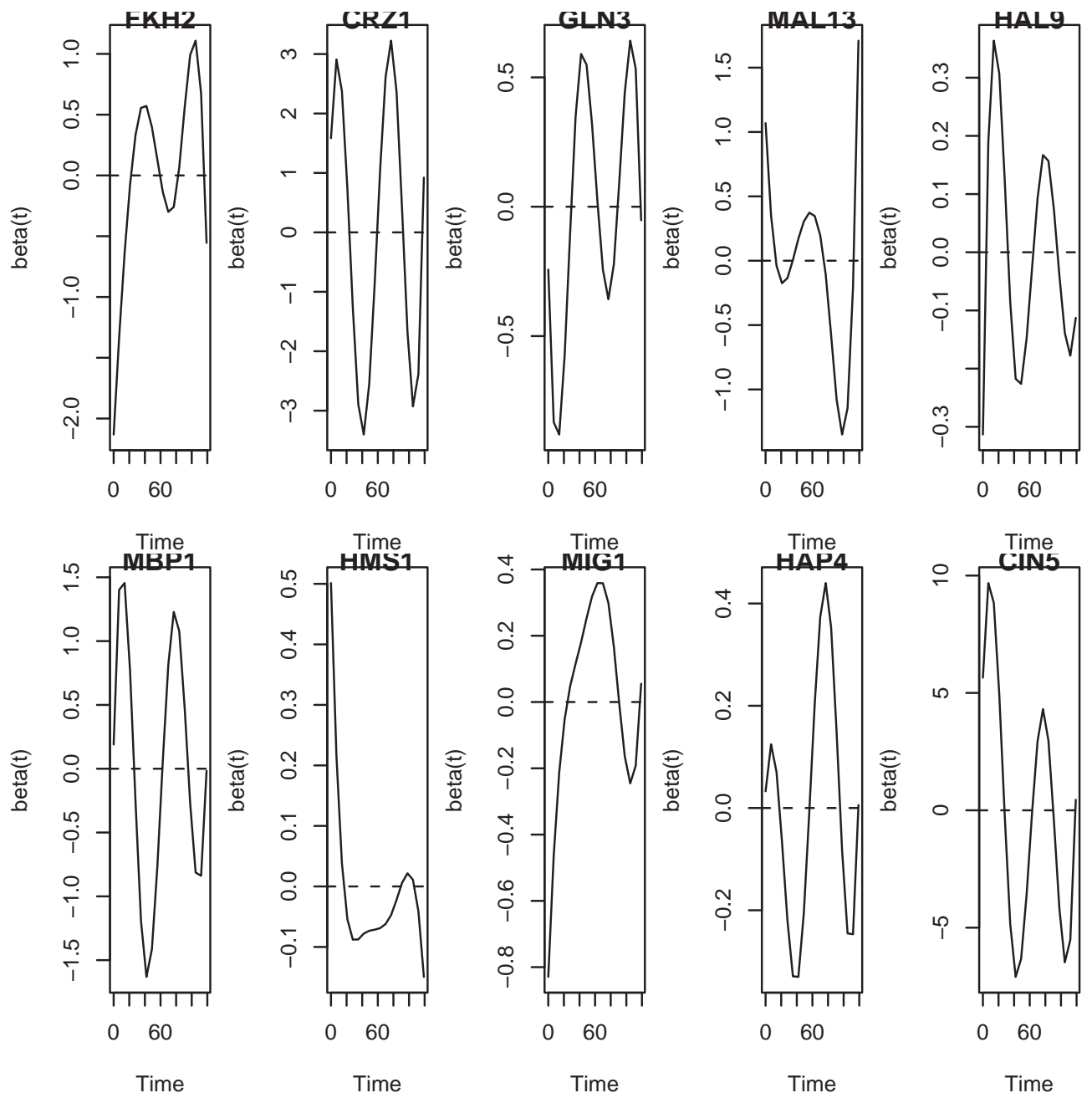


Figure 11: *Estimated transcriptional effects for 10 randomly selected TFs identified by the *gSCAD* procedure.*