Supplementary Material for 'A Focused Information Criterion for Locally Misspecified Vector Autoregressive Models'

Jan Lohmeyer, Franz Palm, Hanno Reuvers, Jean-Pierre Urbain

Department of Quantitative Economics

Maastricht University SBE

The Netherlands

November 5, 2017

A The Selection Matrices and Various Definitions for the VAR Models

The selection matrices encountered in the main text are:

$$\begin{aligned}
\boldsymbol{L} &:= \boldsymbol{L}^{(1)} \otimes \boldsymbol{I}_K, \text{ with } \boldsymbol{L}^{(1)} = \begin{bmatrix} \boldsymbol{I}_{p_1} \\ \mathbf{O}_{p_2 \times p_1} \end{bmatrix} && (Kp \times Kp_1), \\
\boldsymbol{S}_0 &:= \boldsymbol{S}_0^{(1)} \otimes \boldsymbol{I}_K, \text{ with } \boldsymbol{S}_0^{(1)} = \begin{bmatrix} \mathbf{O}_{p_1 \times p_2} \\ \boldsymbol{I}_{p_2} \end{bmatrix} && (Kp \times Kp_2), \\
\boldsymbol{S}_m &:= \boldsymbol{S}_m^{(1)} \otimes \boldsymbol{I}_K, \text{ with } \boldsymbol{S}_m^{(1)} = \begin{bmatrix} \boldsymbol{I}_m \\ \mathbf{O}_{(p-m) \times m} \end{bmatrix} && (Kp \times Km), \\
\boldsymbol{\Pi}_m' &:= \boldsymbol{\Pi}_m'^{(1)} \otimes \boldsymbol{I}_K, \text{ with } \boldsymbol{\Pi}_m'^{(1)} = \begin{bmatrix} \boldsymbol{I}_{m-p_1} \\ \mathbf{O}_{(p-m) \times (m-p_1)} \end{bmatrix} && (Kp_2 \times K(m-p_1)).
\end{aligned} \tag{A.1}$$

An overview of several other definitions are:

$$\begin{aligned}
\boldsymbol{\theta}_{T,m} &= \operatorname{vec}(\boldsymbol{\Theta}_{T,m}) && (K^2 m \times 1) \text{ vector}, \\
\boldsymbol{\delta} &= \operatorname{vec}(\boldsymbol{\Delta}) && (K^2 p_2 \times 1) \text{ vector}, \\
\boldsymbol{\Omega} &= \operatorname{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{z}_{T,t-1} \boldsymbol{z}_{T,t-1}' && (Kp \times Kp) \text{ matrix}, \\
\boldsymbol{A}_m &= [\boldsymbol{S}_m' \boldsymbol{\Omega} \boldsymbol{S}_m]^{-1} \boldsymbol{S}_m' \boldsymbol{\Omega} \boldsymbol{S}_0 \left(\boldsymbol{I}_{p_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \otimes \boldsymbol{I}_K && (K^2 m \times K^2 p_2) \text{ matrix}, \\
\boldsymbol{V}_{jk} &= \left[\boldsymbol{S}_j' \boldsymbol{\Omega} \boldsymbol{S}_j \right]^{-1} \boldsymbol{S}_j' \boldsymbol{\Omega} \boldsymbol{S}_k \left[\boldsymbol{S}_k' \boldsymbol{\Omega} \boldsymbol{S}_k \right]^{-1} \otimes \boldsymbol{\Sigma} && (K^2 j \times K^2 k) \text{ matrix}, \\
\boldsymbol{P}_m &= \boldsymbol{S}_m [\boldsymbol{S}_m' \boldsymbol{\Omega} \boldsymbol{S}_m]^{-1} \boldsymbol{S}_m' \otimes \boldsymbol{I}_K && (K^2 p \times K^2 p) \text{ matrix}, \\
\boldsymbol{C}_m &= \left(\boldsymbol{S}_m [\boldsymbol{S}_m' \boldsymbol{\Omega} \boldsymbol{S}_m]^{-1} \boldsymbol{S}_m' \boldsymbol{\Omega} - \boldsymbol{I}_{Kp} \right) \boldsymbol{S}_0 \left(\boldsymbol{I}_{Kp_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \otimes \boldsymbol{I}_K && (K^2 p \times K^2 p_2) \text{ matrix}, \\
\boldsymbol{D}_\theta &= \partial \boldsymbol{\mu}(\boldsymbol{\theta}_\infty, \boldsymbol{\sigma}) / \partial \boldsymbol{\theta}' && (l \times K^2 p) \text{ vector}, \\
\boldsymbol{D}_\sigma &= \partial \boldsymbol{\mu}(\boldsymbol{\theta}_\infty, \boldsymbol{\sigma}) / \partial \boldsymbol{\sigma}' && (l \times K(K+1)/2) \text{ vector}.
\end{aligned} \tag{A.2}$$

B Impulse Responses and their Gradients for AR(3) Models

The first six impulse responses in the AR(3) model $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + u_t$ are

$$\begin{aligned} &\mu_1(\boldsymbol{\theta}) = \theta_1, && \mu_4(\boldsymbol{\theta}) = \theta_1^4 + 3\theta_1^2\theta_2 + \theta_2^2 + 2\theta_1\theta_3, \\ &\mu_2(\boldsymbol{\theta}) = \theta_1^2 + \theta_2, && \mu_5(\boldsymbol{\theta}) = \theta_1^5 + 4\theta_1^3\theta_2 + 3\theta_1\theta_2^2 + 3\theta_1^2\theta_3 + 2\theta_2\theta_3, \\ &\mu_3(\boldsymbol{\theta}) = \theta_1^3 + 2\theta_1\theta_2 + \theta_3, && \mu_6(\boldsymbol{\theta}) = \theta_1^6 + 5\theta_1^4\theta_2 + 6\theta_1^2\theta_2^2 + \theta_2^3 + 4\theta_1^3\theta_3 + 6\theta_1\theta_2\theta_3 + \theta_3^2. \end{aligned} \tag{B.1}$$

The gradient vectors corresponding to the impulse responses in Equation (B.1) are:

$$\begin{aligned} &\frac{\partial\mu_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}1\\0\\0\end{bmatrix}, && \frac{\partial\mu_4(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}4\theta_1^3 + 6\theta_1\theta_2 + 2\theta_3\\ 3\theta_1^2 + 2\theta_2\\ 2\theta_1\end{bmatrix}, \\ &\frac{\partial\mu_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}2\theta_1\\1\\0\end{bmatrix}, && \frac{\partial\mu_5(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}5\theta_1^4 + 12\theta_1^2\theta_2 + 3\theta_2^2 + 6\theta_1\theta_3\\ 4\theta_1^3 + 6\theta_1\theta_2 + 2\theta_3\\ 3\theta_1^2 + 2\theta_2\end{bmatrix}, \\ &\frac{\partial\mu_3(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}3\theta_1^2 + 2\theta_2\\ 2\theta_1\\ 1\end{bmatrix}, && \frac{\partial\mu_6(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}6\theta_1^5 + 20\theta_1^3\theta_2 + 12\theta_1\theta_2^2 + 12\theta_1^2\theta_3 + 6\theta_2\theta_3\\ 5\theta_1^4 + 12\theta_1^2\theta_2 + 3\theta_1^2 + 6\theta_1\theta_3\\ 4\theta_1^3 + 6\theta_1\theta_2 + 2\theta_3\end{bmatrix}. \end{aligned} \tag{B.2}$$

Remark 1

If the true DGP is a white noise, i.e. $\theta_1 = \theta_2 = \theta_3 = 0$*, then the gradients of the impulse responses will be zero from horizon 4 onwards. The limiting distribution should now be based on the second order delta method. This example supports Remark 3 from the paper.*

C Derivation of the Optimal Weights in the Simplified Model

The simplified model is given by $y_t = \alpha y_{t-1} + \frac{\delta}{\sqrt{T}} y_{t-2} + u_t$. We develop the elements of the matrix $\boldsymbol{\Psi}^\infty$ for the case *with bias correction*.

$$\begin{aligned} \boldsymbol{\Psi}_{1,1}^\infty &= a_1\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) - a_1 + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{\Omega}\boldsymbol{P}_1\boldsymbol{D}_\theta' \\ &= a_1\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) - a_1 + \sigma^2\boldsymbol{D}_\theta\boldsymbol{S}_1\left[\boldsymbol{S}_1'\boldsymbol{\Omega}\boldsymbol{S}_1\right]^{-1}\boldsymbol{S}_1'\boldsymbol{D}_\theta' \\ &= a_1\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) - a_1 + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta' \\ \boldsymbol{\Psi}_{1,2}^\infty &= \boldsymbol{D}_\theta\boldsymbol{P}_1(\sigma^2\boldsymbol{\Omega})\boldsymbol{P}_2\boldsymbol{D}_\theta' = \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta' \\ \boldsymbol{\Psi}_{2,2}^\infty &= \boldsymbol{D}_\theta\boldsymbol{P}_2(\sigma^2\boldsymbol{\Omega})\boldsymbol{P}_2\boldsymbol{D}_\theta' = \sigma^2\boldsymbol{D}_\theta\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\theta' \end{aligned} \tag{C.1}$$

The solution to the optimization problem will not change if we subtract a constant from every element. We subtract $\sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta'$, and denote the result by $\tilde{\boldsymbol{\Psi}}^\infty$. The resulting matrix looks like

$$\tilde{\boldsymbol{\Psi}}^\infty = \begin{bmatrix} a_1\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) - a_1 & 0 \\ 0 & \sigma^2\boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}_\theta' \end{bmatrix}. \tag{C.2}$$

The $(1,1)$-element of this matrix can become negative whenever $\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) < 1$. From this point onwards, we consider the case *without bias correction*. Now,

$$\tilde{\boldsymbol{\Psi}}^\infty = \begin{bmatrix} a_1\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta\boldsymbol{C}_1\boldsymbol{\delta})^2/a_1\right) & 0 \\ 0 & \sigma^2\boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}_\theta' \end{bmatrix} := \begin{bmatrix}\kappa_1 & 0\\ 0 & \kappa_2\end{bmatrix}, \tag{C.3}$$

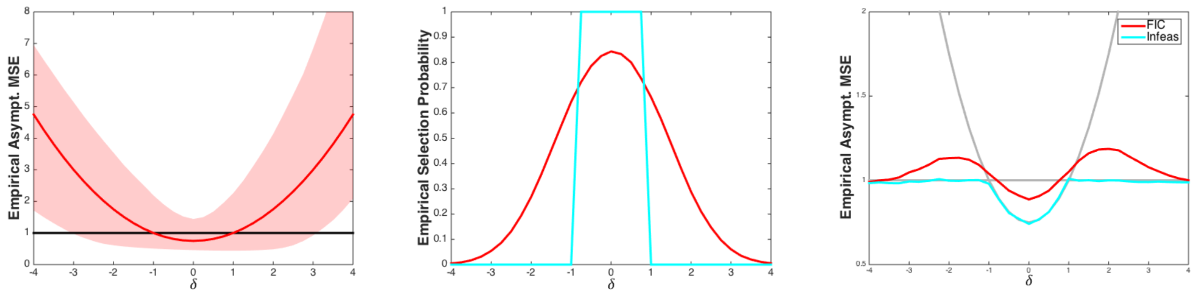
and if we assume $\boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}'_\theta > 0$ (see main text), then we have $\tilde{\boldsymbol{\Psi}}^\infty > 0$ hence unique weights. The optimal weight w^* is the minimizer of $\kappa_1 w^2 + \kappa_2(1-w)^2$ subject to $0 \leq w \leq 1$. The unconstrained solution is

$$w^* = \frac{\kappa_2}{\kappa_1 + \kappa_2} = \frac{\sigma^2 \boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}'_\theta}{a_1 \chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta \boldsymbol{C}_1 \boldsymbol{\delta})^2/a_1\right) + \sigma^2 \boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}'_\theta}. \tag{C.4}$$

This solution is both positive and in the interval $[0, 1]$ because $\kappa_1, \kappa_2 > 0$. The constraint $w \in [0, 1]$ is thus automatically satisfied. It follows that

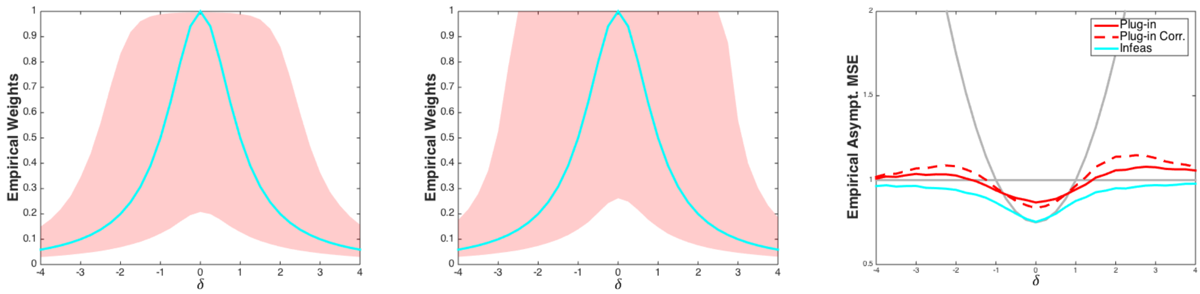
$$\begin{aligned}\Pr(w^* \leq x) &= \Pr\left(\frac{\kappa_2}{\kappa_1 + \kappa_2} \leq x\right) = \Pr\left(\kappa_1 \geq \frac{\kappa_2[1-x]}{x}\right) \\ &= \Pr\left(\chi^2_{noncentral}\left(1, (\boldsymbol{D}_\theta \boldsymbol{C}_1 \boldsymbol{\delta})^2/a_1\right) \geq \frac{\sigma^2 \boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}'_\theta [1-x]}{a_1 x}\right).\end{aligned} \tag{C.5}$$

D Further Simulation Results on the Simplified Model



(a) (b) (c)

Figure 1: (a) The asymptotic MSE of the models with one and two lags (red and black line, respectively). The area between the 5% and 95% empirical quantile of $\widehat{FIC}_1$ and $\widehat{FIC}_2$ are shaded in red and grey. (b) The empirical selection probabilities of the FIC. (c) The AMSE of the models with $m = 1$ and $m = 2$ together with the empirical MSE of the feasible FIC (red) and infeasible FIC (cyan). This figure was obtained at $T = 1000$.



(a) (b) (c)

Figure 2: The 5% and 95% empirical quantile of the weights distribution without bias correction (a) and with bias correction (b). The infeasible weights are displayed in cyan. The empirical MSE of plug-in methods is shown in (c). This figure was obtained at $T = 1000$.

E No Gradient Dependence in Simplified Model

For $y_t = \alpha y_{t-1} + \frac{\delta}{\sqrt{T}} y_{t-2} + u_t$, we have $\boldsymbol{\Omega} = \frac{\sigma^2}{1-\alpha^2}\left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right]$ and $\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma^2}\left[\begin{smallmatrix} 1 & -\alpha \\ -\alpha & 1 \end{smallmatrix}\right]$. The required selection matrices are $\boldsymbol{S}_1 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, $\boldsymbol{S}_0 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, and $\boldsymbol{\Pi}_1'\boldsymbol{\Pi}_1 = \mathbf{O}_{2\times 2}$. Straightforward calculations show

$$\boldsymbol{C}_1 = \left\{\begin{bmatrix} 1 & \alpha \\ 0 & 0 \end{bmatrix} - \boldsymbol{I}_2\right\}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha \\ -1 \end{bmatrix} := \boldsymbol{y} \tag{E.1}$$

and

$$\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1 = \frac{1}{\sigma^2}\begin{bmatrix} \alpha^2 & -\alpha \\ -\alpha & 1 \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} \alpha \\ -1 \end{bmatrix}\begin{bmatrix} \alpha & -1 \end{bmatrix} = \frac{1}{\sigma^2}\boldsymbol{y}\boldsymbol{y}'. \tag{E.2}$$

The elements of the weighting matrix (with bias correction in red) are now given by:

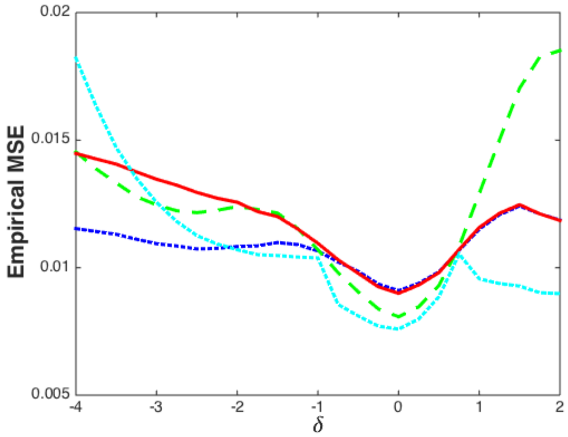
$$\begin{aligned} \boldsymbol{\Psi}_{11} &= \boldsymbol{D}_\theta\left[\boldsymbol{C}_1(\boldsymbol{\delta}\boldsymbol{\delta} - \sigma^2\boldsymbol{S}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{S}_0)'\boldsymbol{C}_1' + \sigma^2\boldsymbol{P}_1\right]\boldsymbol{D}_\theta' = (\delta^2 - 1)(\boldsymbol{D}_\theta\boldsymbol{y})^2 + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta', \\ \boldsymbol{\Psi}_{12} &= \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta', \\ \boldsymbol{\Psi}_{22} &= \sigma^2\boldsymbol{D}_\theta(\boldsymbol{\Omega}^{-1} - \boldsymbol{P}_1)\boldsymbol{D}_\theta' + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta' = (\boldsymbol{D}_\theta\boldsymbol{y})^2 + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta'. \end{aligned} \tag{E.3}$$

The weights are determined from $\boldsymbol{w}^0 = \arg\min_{\boldsymbol{w}\in\mathcal{H}} \boldsymbol{w}'\boldsymbol{\Psi}\boldsymbol{w} = \arg\min_{\boldsymbol{w}\in\mathcal{H}}(\boldsymbol{D}_\theta\boldsymbol{y})^2\boldsymbol{w}'\begin{bmatrix} (\delta^2 - 1) & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{w} + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta' = \arg\min_{0\leq w\leq 1}(\boldsymbol{D}_\theta\boldsymbol{y})^2\left[w^2(\delta^2 - 1) + (1-w)^2\right] + \sigma^2\boldsymbol{D}_\theta\boldsymbol{P}_1\boldsymbol{D}_\theta'$. This expression shows that the weights do not depend on the quantity of interest because $\boldsymbol{D}_\theta$ is no longer of importance for the optimal weight calculation.

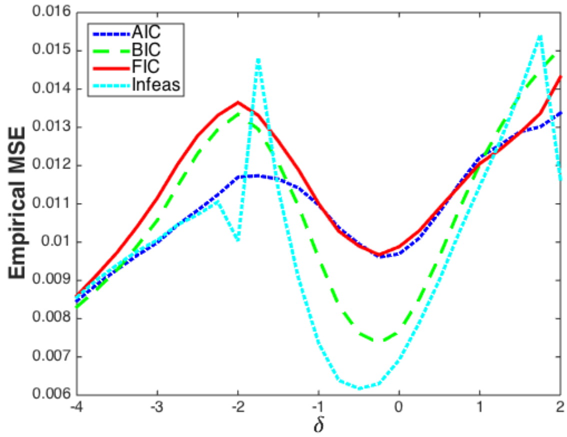
Remark 2

The diagonal elements of the weighting matrix without bias correction are also the AMSE's of the individual model. We would prefer the model with one lag if $\delta^2 < 1$. This is supported by Figure 2.

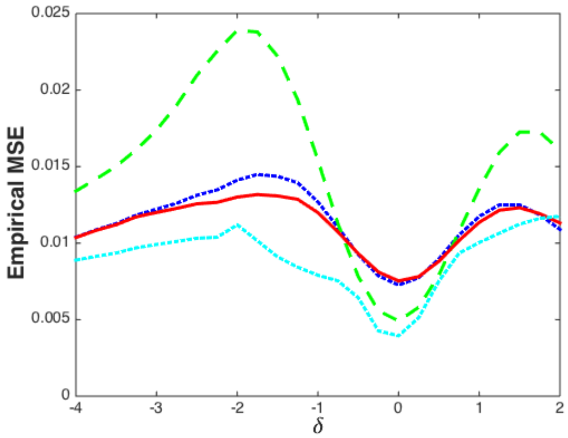
F Full Simulation Results on Misspecified AR(3) Model



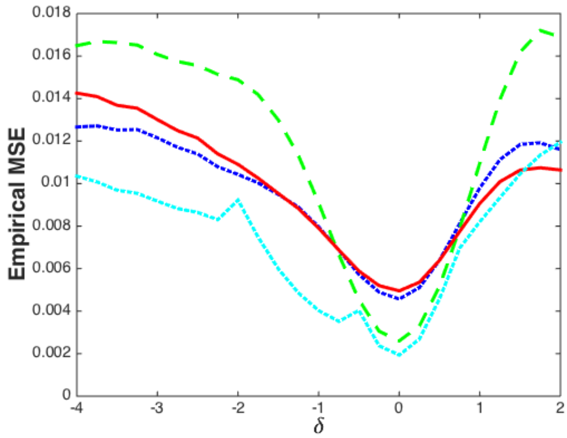
(a) $h = 1$.



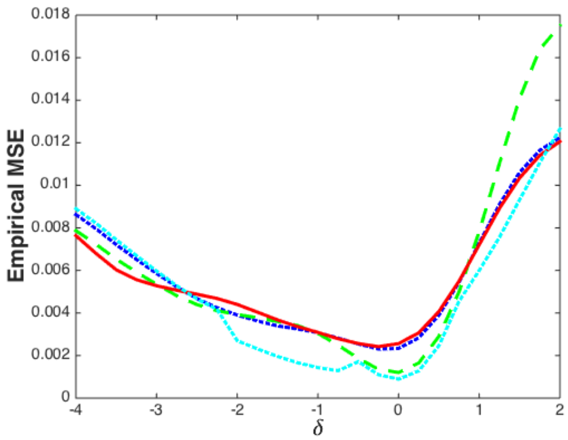
(b) $h = 2$.



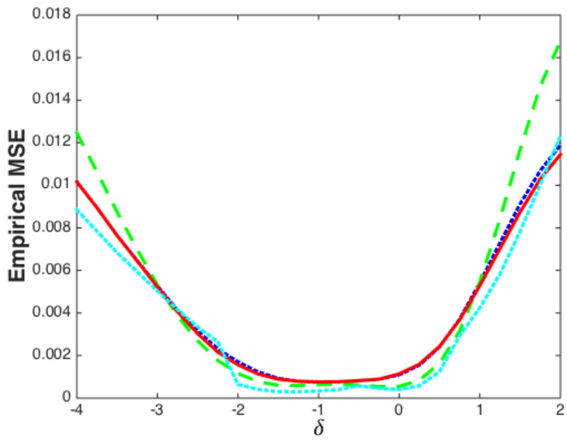
(c) $h = 3$.



(d) $h = 4$.

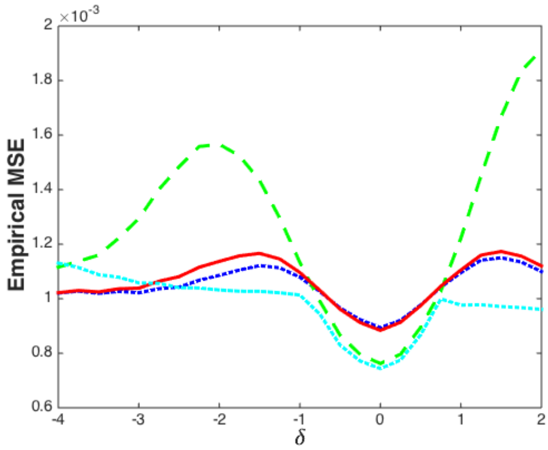


(e) $h = 5$.

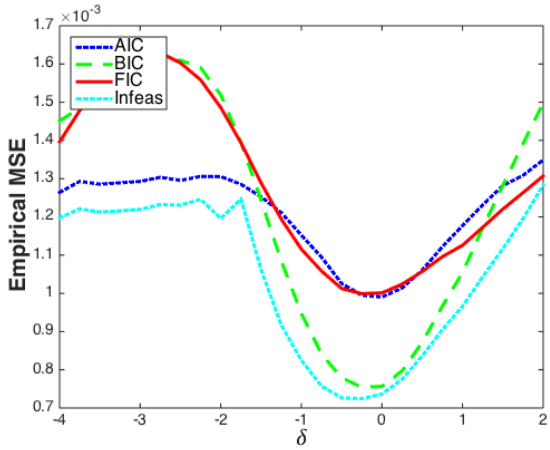


(f) $h = 6$

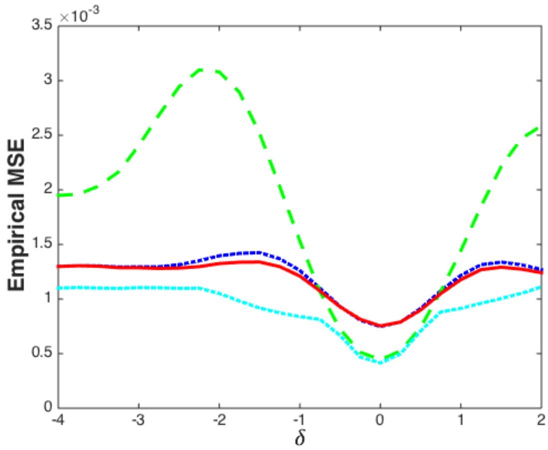
Figure 3: The empirical MSE for model selection based on Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Focused Information Criteria (FIC). 'Infeas' denotes an infeasible version of the FIC for which all quantities (and especially δ) are replaced by their true values. The DGP is $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ with $T = 100$.



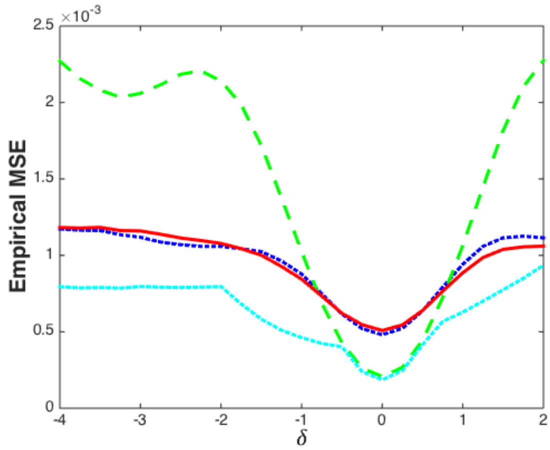
(a) $h = 1$.



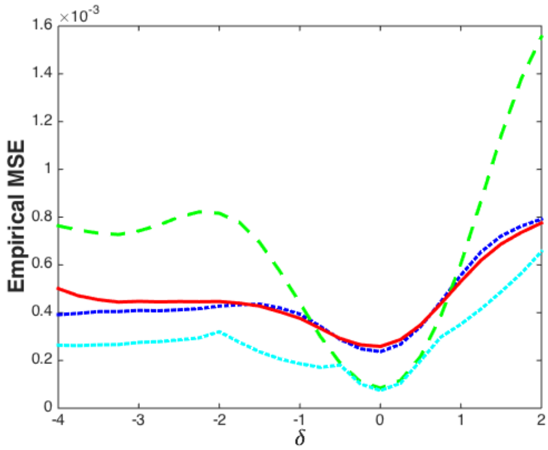
(b) $h = 2$.



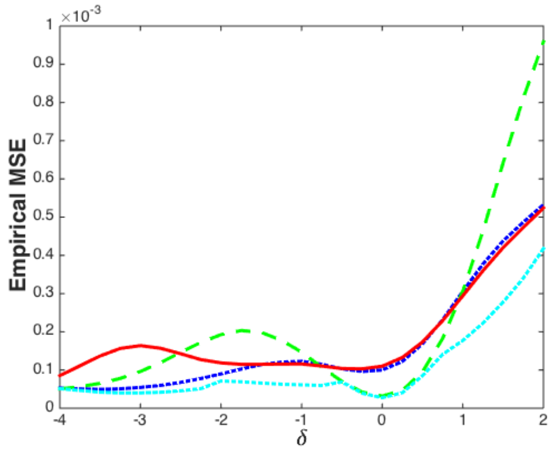
(c) $h = 3$.



(d) $h = 4$.

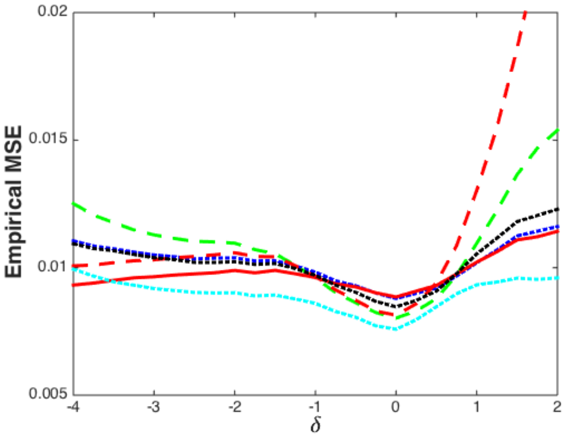


(e) $h = 5$.

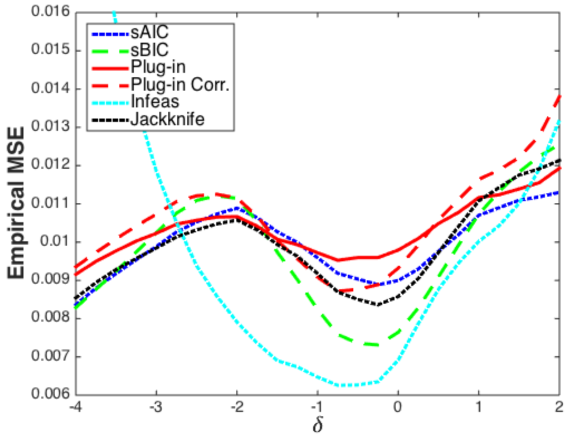


(f) $h = 6$

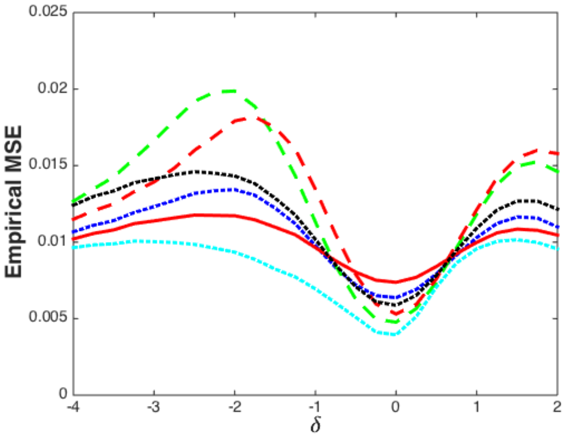
Figure 4: Idem Figure 3, but for $T = 1000$.



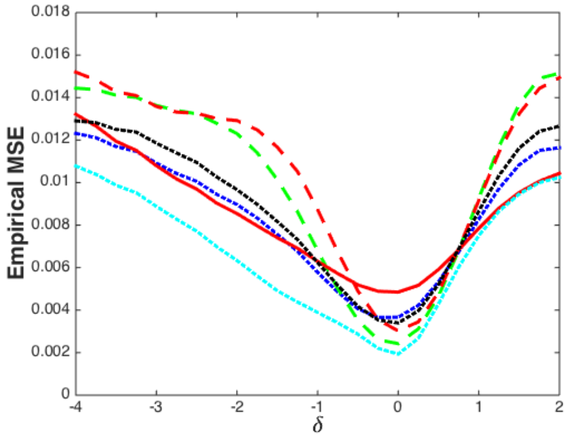
(a) $h = 1$.



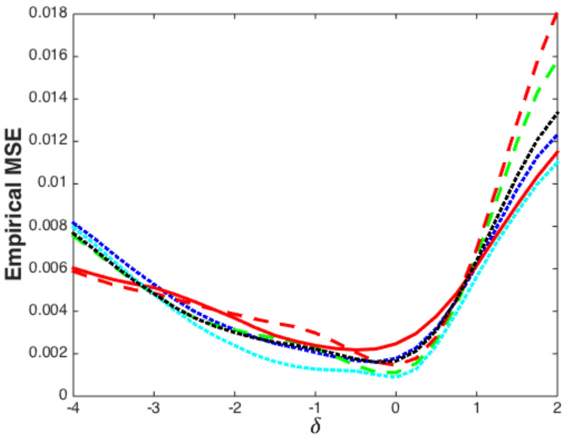
(b) $h = 2$.



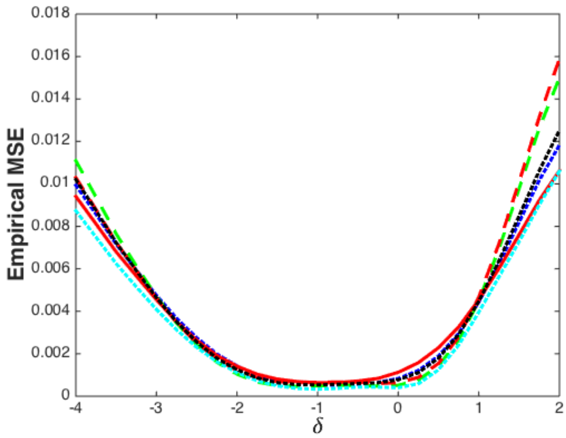
(c) $h = 3$.



(d) $h = 4$.

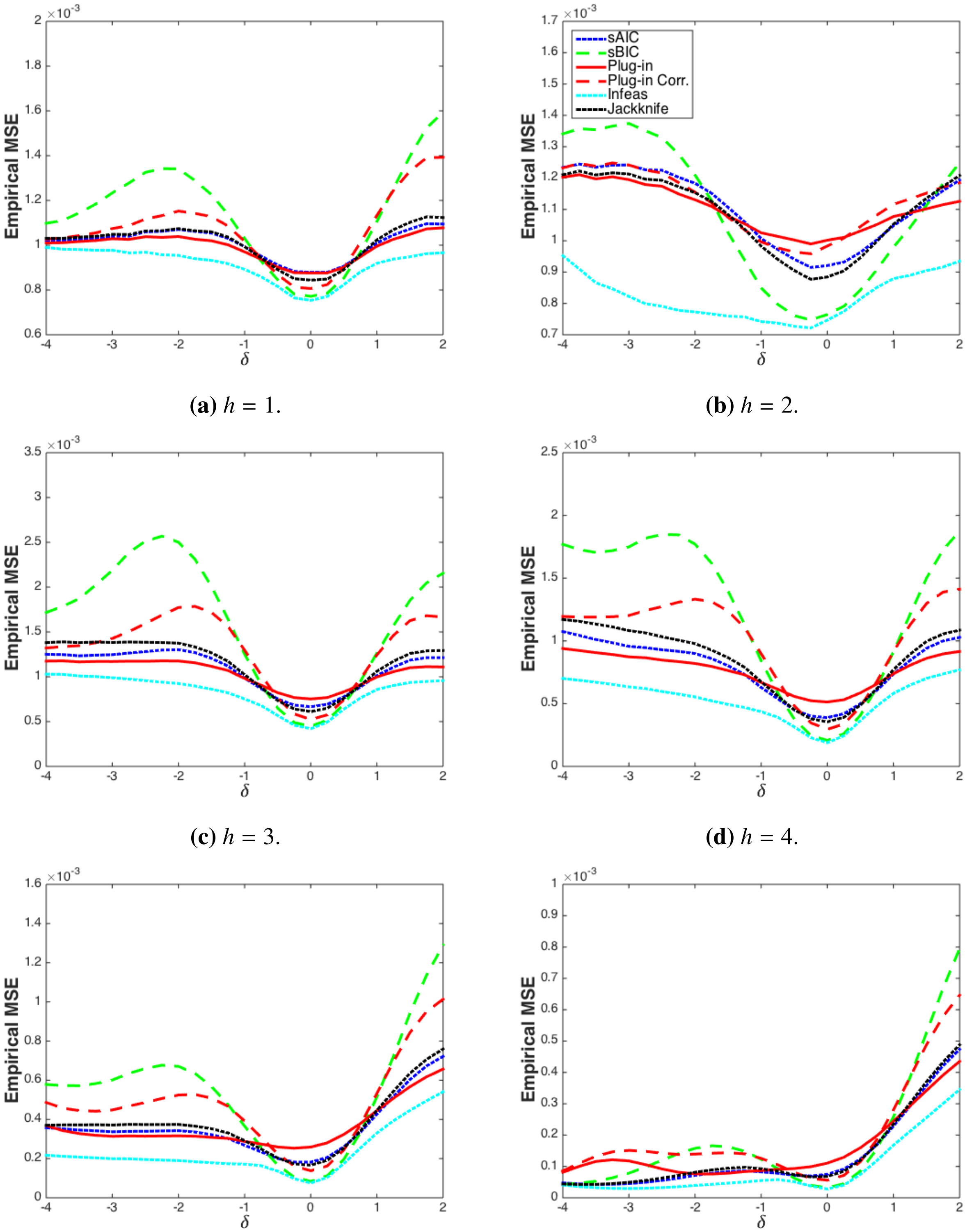


(e) $h = 5$.



(f) $h = 6$

Figure 5: The empirical MSE for model averaging based on smoothed AIC (sAIC), smoothed BIC (sBIC), the plug-in average without bias correction (Plug-in), the plug-in average with bias correction (Plug-in Corr.), the infeasible plug-in average (Infeas) and the average with Jackknife weights (Jackknife). The DGP is $y_{T,t} = 0.5 y_{T,t-1} + \frac{\delta}{\sqrt{T}} y_{T,t-2} + \frac{\delta}{2\sqrt{T}} y_{T,t-3} + u_t$ with $T = 100$.



(a) $h = 1$.

(b) $h = 2$.

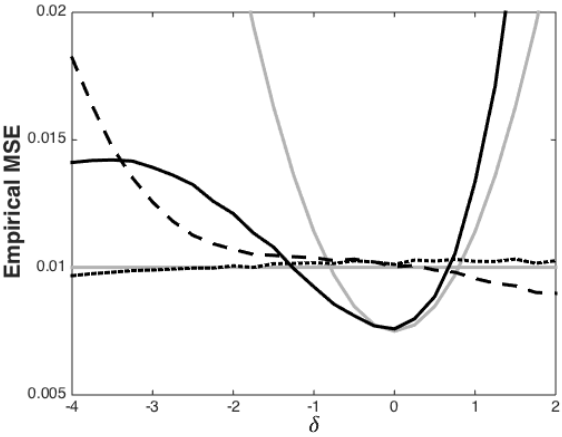
(c) $h = 3$.

(d) $h = 4$.

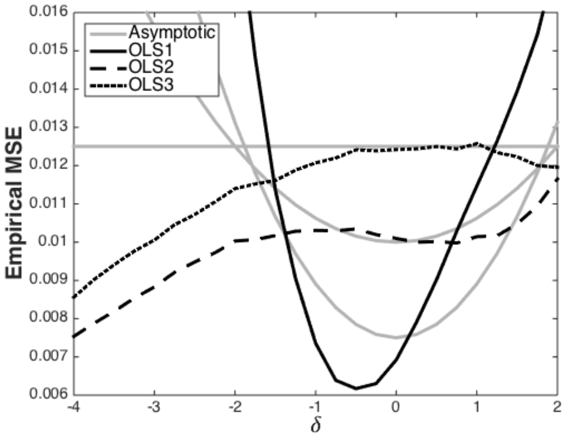
(e) $h = 5$.

(f) $h = 6$

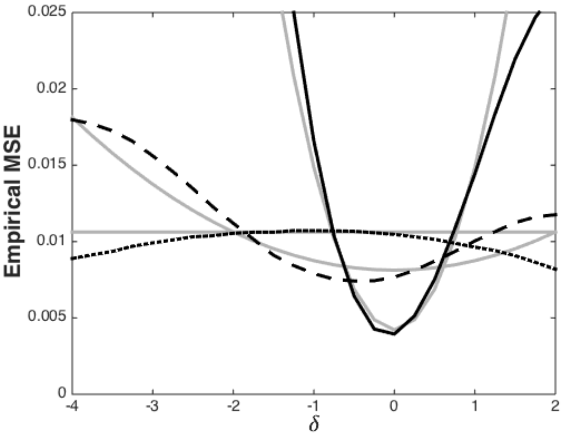
Figure 6: Idem Figure 5, but for $T = 1000$.



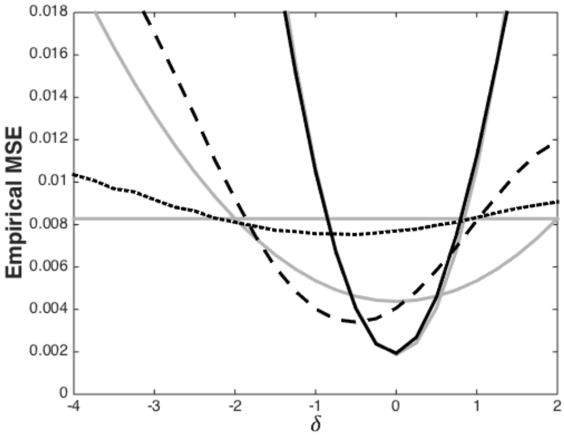
(a) $h = 1$



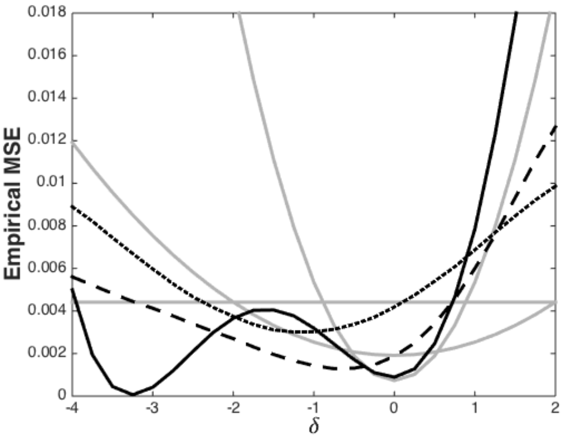
(b) $h = 2$.



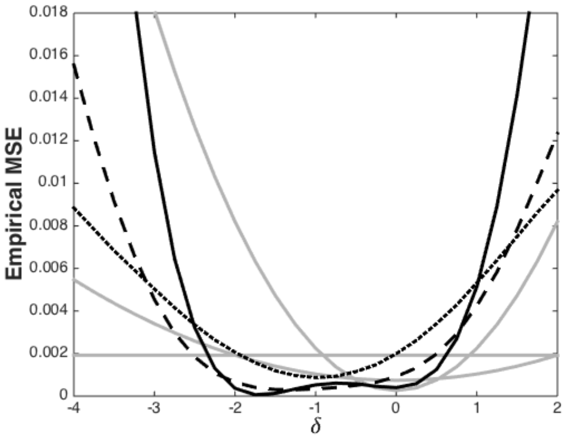
(c) $h = 3$



(d) $h = 4$.

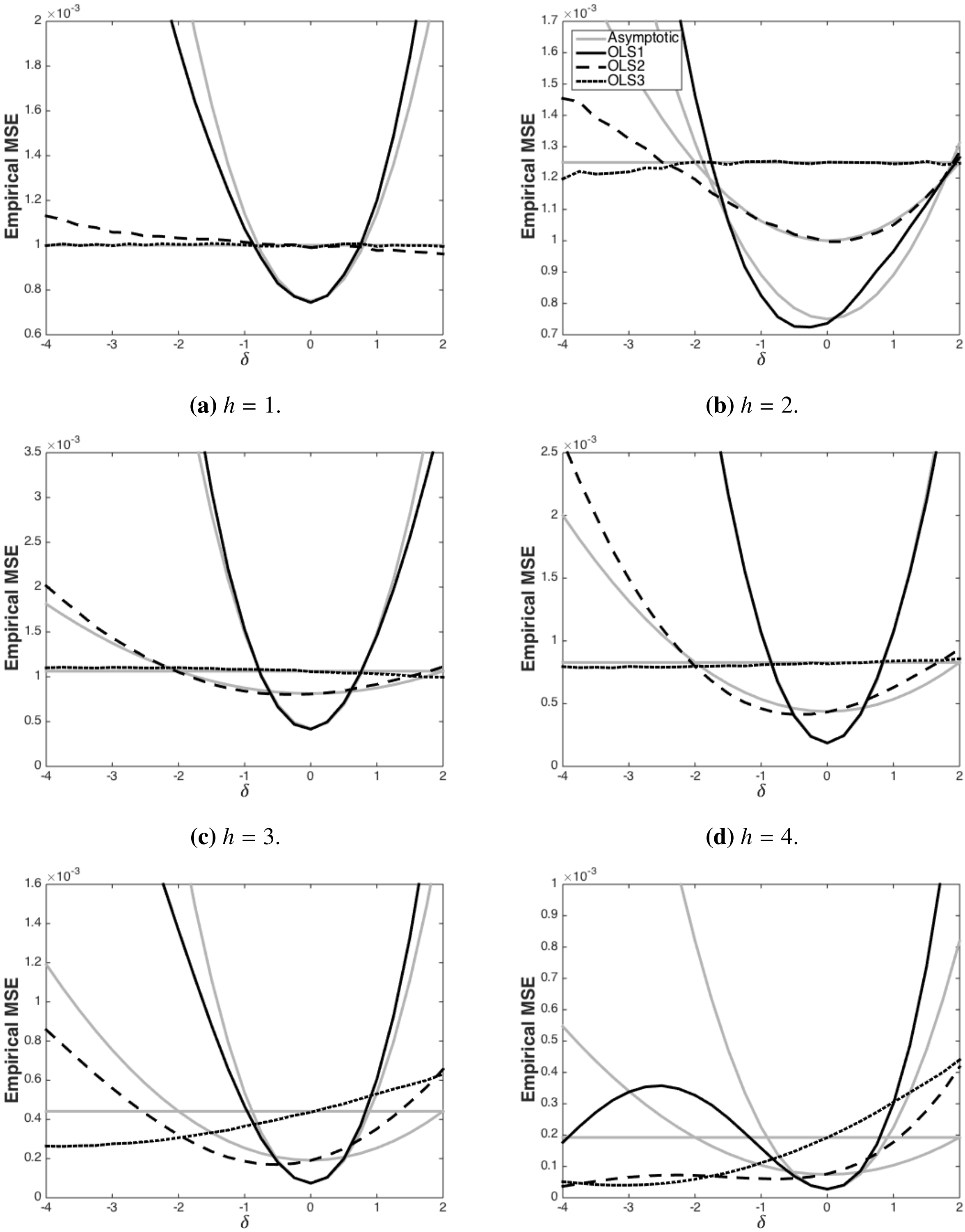


(e) $h = 5$



(f) $h = 6$

Figure 7: The empirical MSE of the OLS estimator of the model with 1 lag (OLS1), 2 lags (OLS2) and the full model with 3 lags (OLS3) for $T = 100$. Gray lines show the asymptotic MSE approximations as provided by the delta method.



(a) $h = 1$.

(b) $h = 2$.

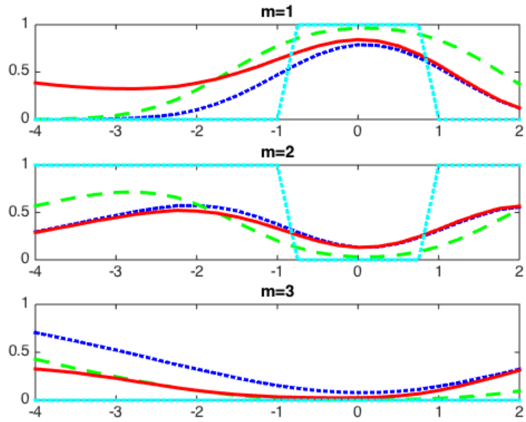
(c) $h = 3$.

(d) $h = 4$.

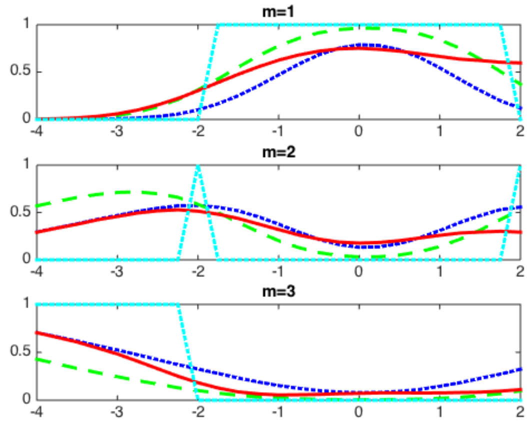
(e) $h = 5$.

(f) $h = 6$

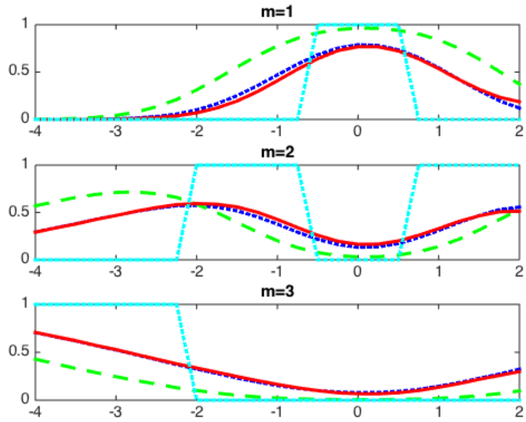
Figure 8: Idem Figure 7, but $T = 1000$.



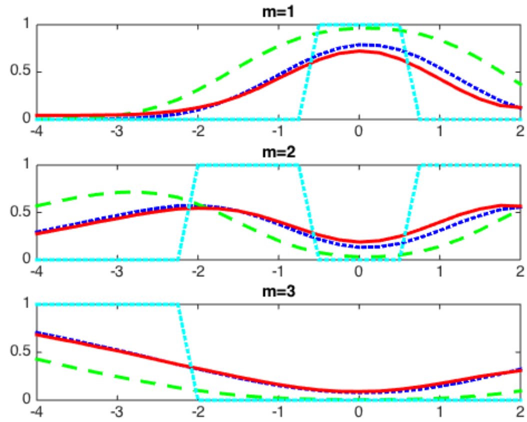
(a) $h = 1$.



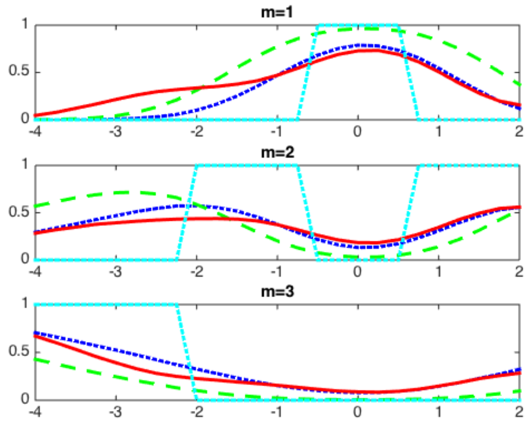
(b) $h = 2$.



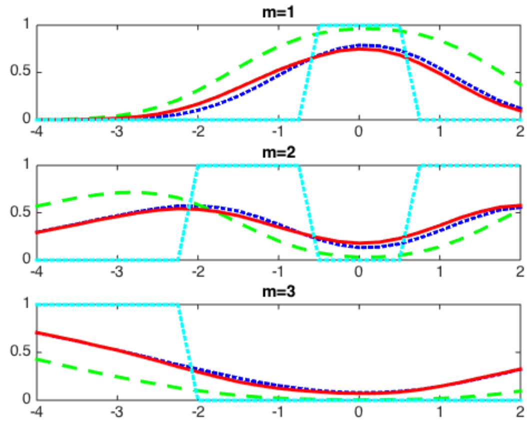
(c) $h = 3$.



(d) $h = 4$.

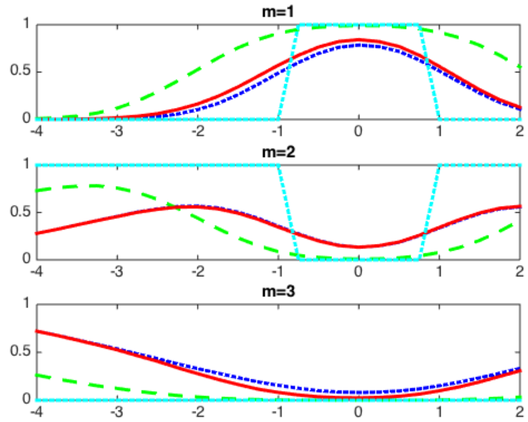


(e) $h = 5$.

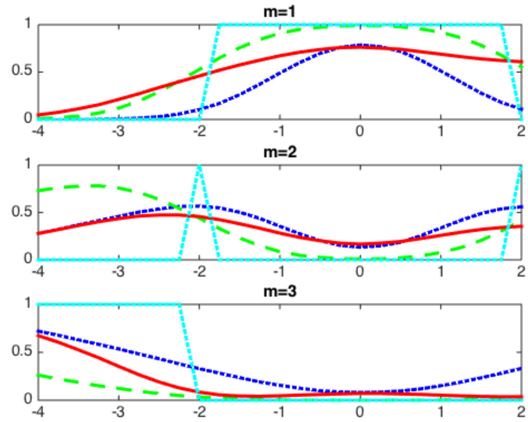


(f) $h = 6$

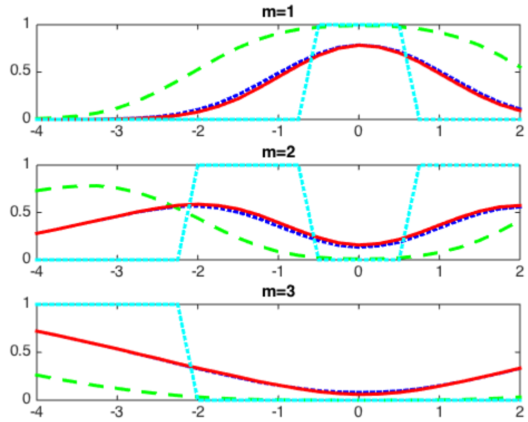
Figure 9: The selection frequencies of the various models for AIC (blue), BIC (green), FIC (red) and the infeasible estimator (cyan) as a function of the amount of misspecification, δ. The data was generated by $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ with a sample size of $T = 100$.



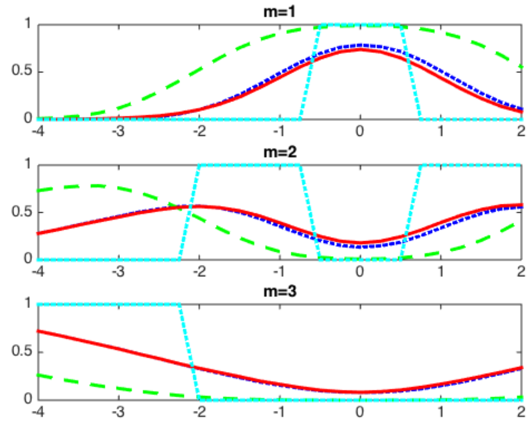
(a) $h = 1$.



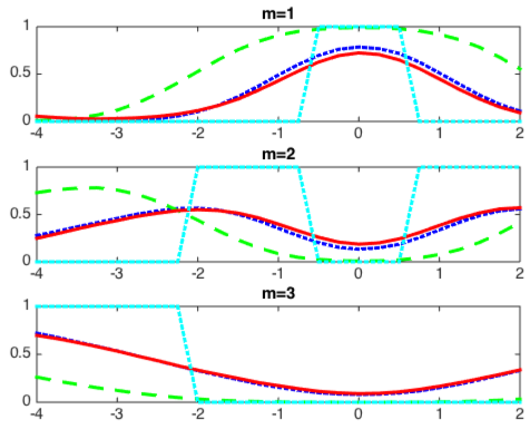
(b) $h = 2$.



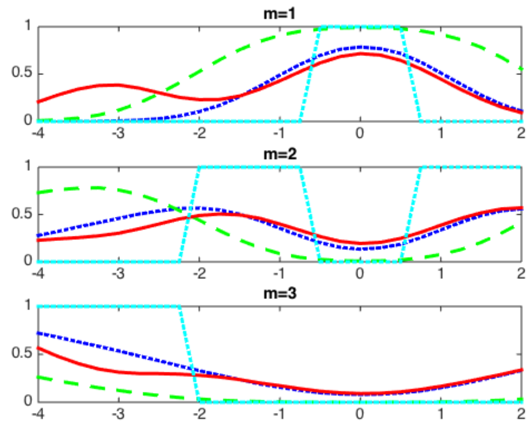
(c) $h = 3$.



(d) $h = 4$.

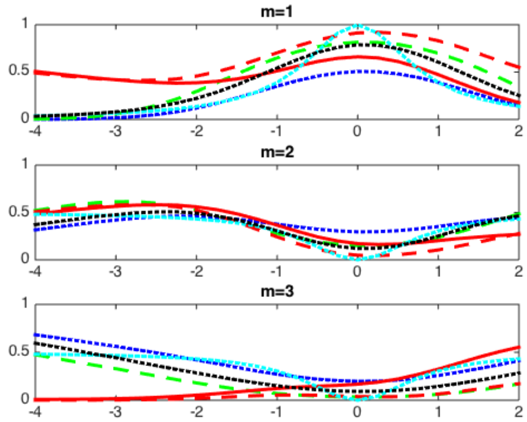


(e) $h = 5$.

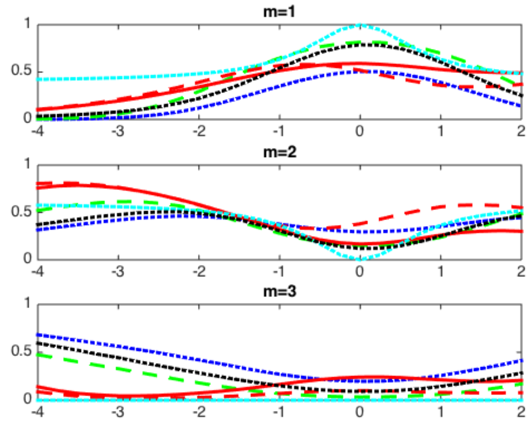


(f) $h = 6$

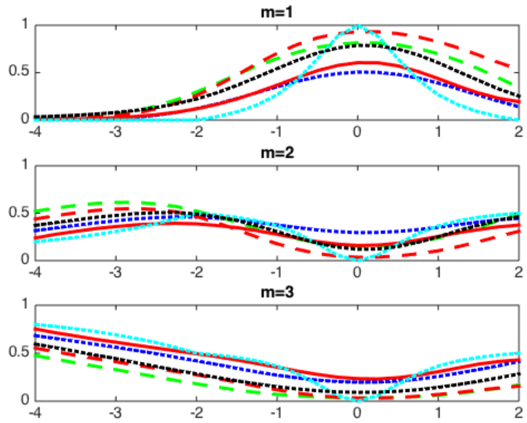
Figure 10: Idem Figure 9, but for $T = 1000$.



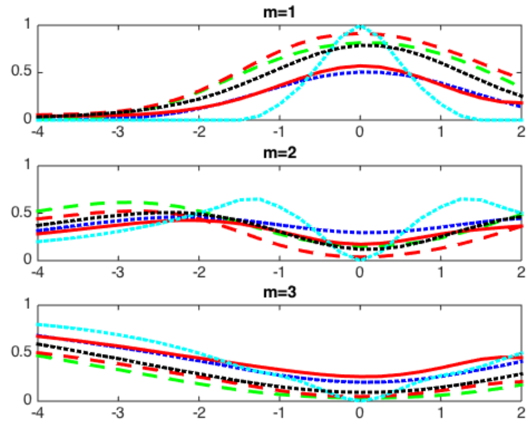
(a) $h = 1$.



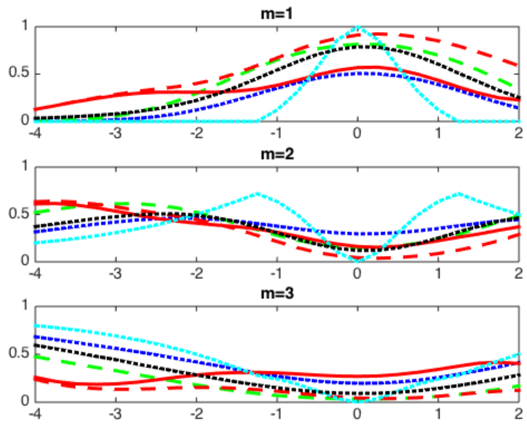
(b) $h = 2$.



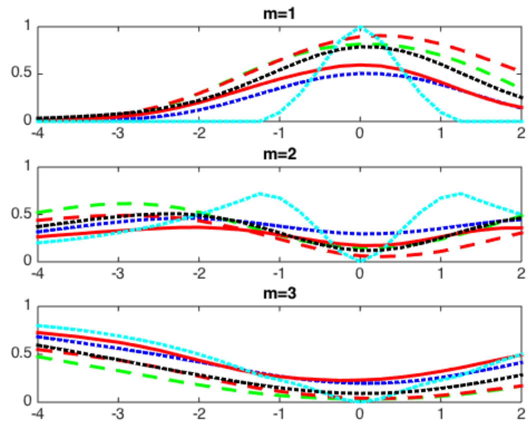
(c) $h = 3$.



(d) $h = 4$.

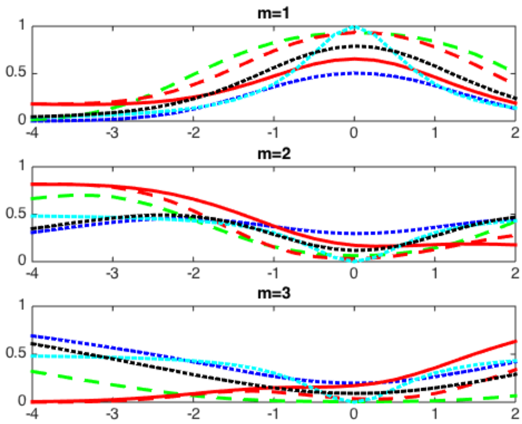


(e) $h = 5$.

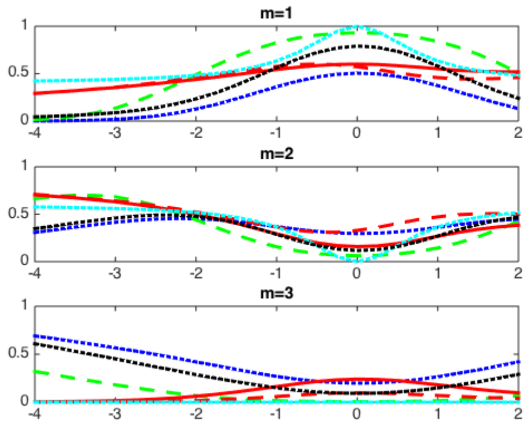


(f) $h = 6$

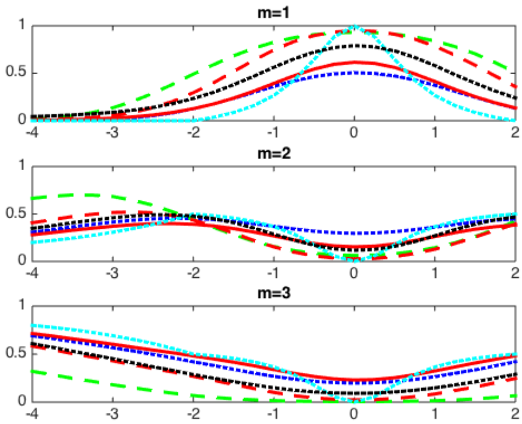
Figure 11: The empirical weights on the various models for sAIC (blue), sBIC (green), FIC (red) and the infeasible estimator (cyan) as a function of the amount of misspecification, δ. The data was generated by $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ at a sample size of $T = 100$.



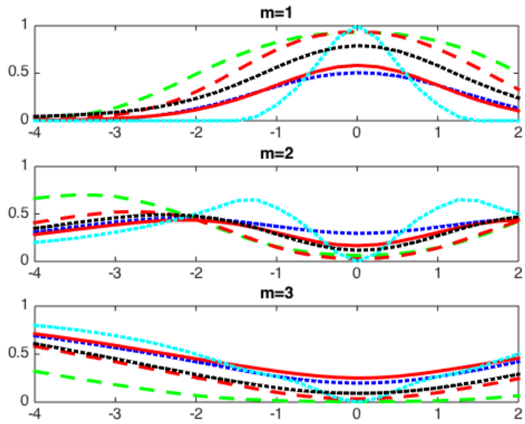
(a) $h = 1$.



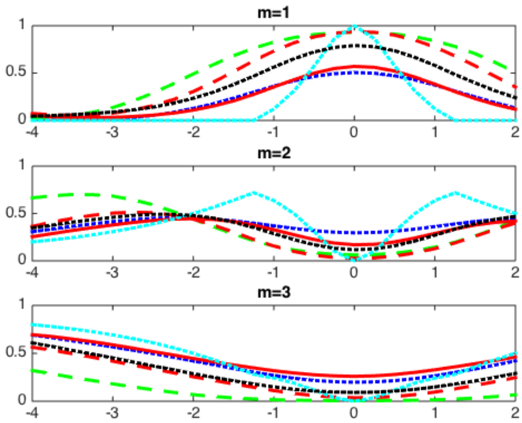
(b) $h = 2$.



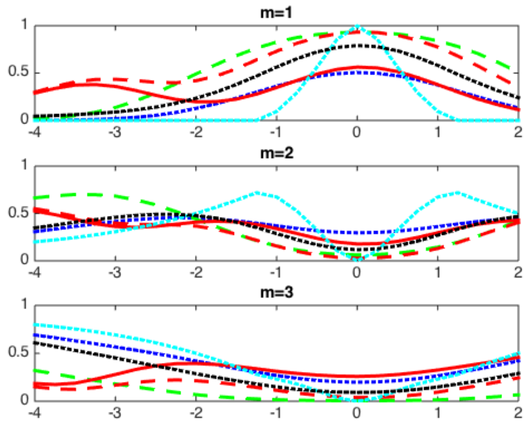
(c) $h = 3$.



(d) $h = 4$.

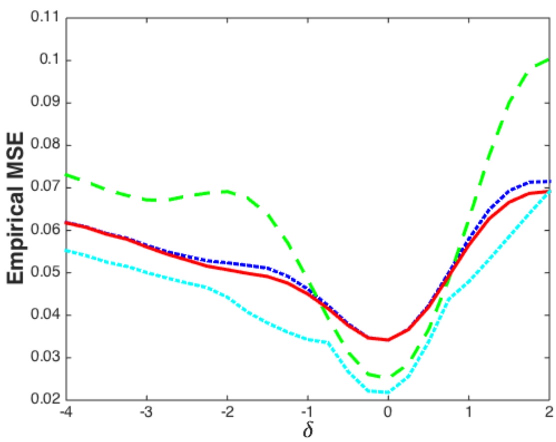


(e) $h = 5$.

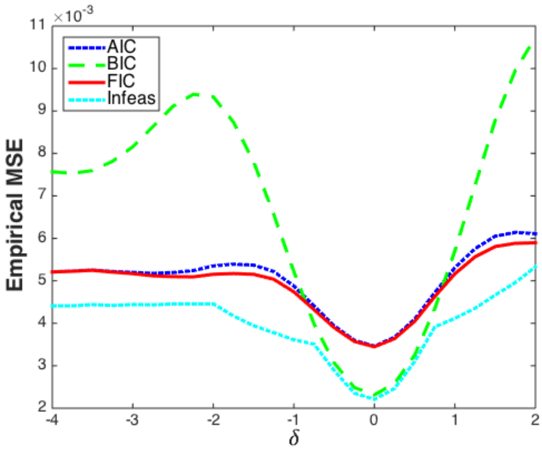


(f) $h = 6$

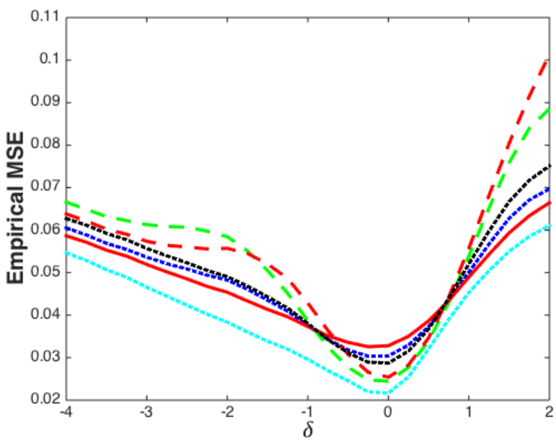
Figure 12: Idem Figure 11, but for $T = 1000$.



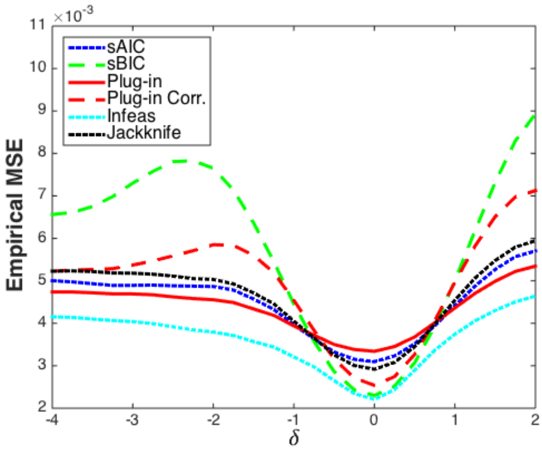
(a) Model Selection, $T = 100$.



(b) Model Selection, $T = 1000$.



(c) Model Averaging, $T = 100$.



(d) Model Averaging, $T = 1000$.

Figure 13: Simulation results for a multiple quantities of interest. Model selection/averaging is based on the impulse responses from horizon 1 up to 6. The trace of the AMSE matrix is used to map the AMSE matrix to a scalar, i.e. model selection/averaging is based on the sum of the asymptotic mean-squared errors. Therefore, we also report the sum of the empirical MSE of the first six horizons on the vertical axis.

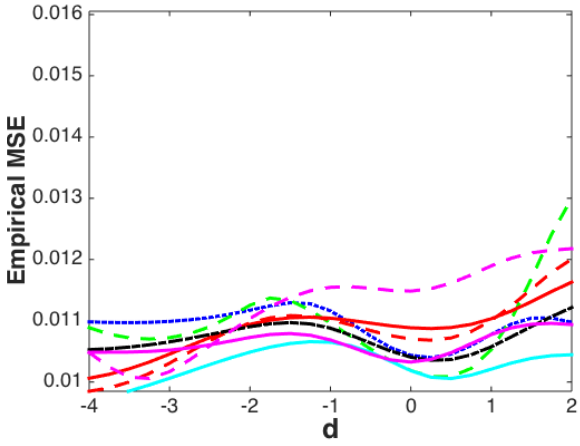
Table 1: The empirical coverage of 90% confidence intervals for the horizons: two, three and six. For several choices of δ and T in the DGP $y_{T,t} = 0.5 y_{T,t-1} + \frac{\delta}{\sqrt{T}} y_{T,t-2} + \frac{\delta}{2\sqrt{T}} y_{T,t-3} + u_t$.

	$h = 2$				$h = 3$				$h = 6$			
δ	$T = 100$	$T = 250$	$T = 500$	$T = 1000$	$T = 100$	$T = 250$	$T = 500$	$T = 1000$	$T = 100$	$T = 250$	$T = 500$	$T = 1000$
-4.0	75.76	81.70	83.04	84.08	76.54	81.13	81.99	82.31	75.85	58.81	42.44	2.04
-3.5	78.29	82.37	83.41	84.23	78.39	81.64	82.19	82.35	72.32	52.63	50.53	5.51
-3.0	80.06	82.92	83.72	84.40	79.82	82.08	82.42	82.42	67.85	47.84	6.19	11.01
-2.5	81.35	83.36	83.96	84.61	81.04	82.46	82.58	82.47	62.63	45.29	8.13	17.37
-2.0	82.33	83.79	84.16	84.76	81.92	82.64	82.54	82.46	55.30	12.54	14.52	23.33
-1.5	83.06	84.18	84.38	84.89	82.34	82.64	82.41	82.31	34.87	15.49	21.80	27.80
-1.0	83.59	84.44	84.50	84.99	82.23	82.38	82.13	82.07	24.76	23.12	27.72	31.39
-0.5	83.87	84.55	84.60	85.02	81.73	82.04	81.83	81.79	34.06	35.34	37.24	38.45
0.0	83.92	84.59	84.61	85.00	81.01	81.52	81.39	81.43	53.66	46.20	44.24	43.45
0.5	83.66	84.51	84.57	84.97	79.85	80.68	80.82	81.05	61.93	55.10	50.98	48.32
1.0	83.20	84.33	84.51	84.94	78.11	79.67	80.13	80.58	68.96	62.81	57.59	53.38
1.5	82.46	84.02	84.38	84.89	75.73	78.40	79.42	80.09	73.56	68.80	63.00	57.78
2.0	80.60	83.49	84.20	84.82	71.70	76.94	78.61	79.59	76.69	73.19	67.54	61.70

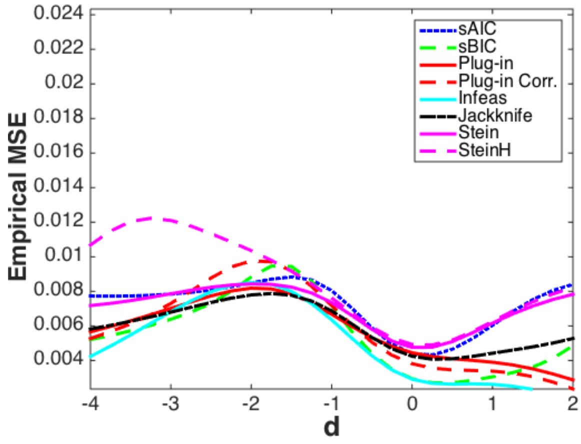
Table 2: The empirical coverage of the confidence regions as a function of sample size T and misspecification parameter δ.

	$h = \{2, 3\}$				$h = \{2, 6\}$			
δ	$T = 100$	$T = 250$	$T = 500$	$T = 1000$	$T = 100$	$T = 250$	$T = 500$	$T = 1000$
-4.0	89.71	89.94	90.01	90.08	87.14	86.47	71.25	9.87
-3.5	89.62	89.99	90.10	90.10	86.24	81.75	57.42	23.71
-3.0	89.59	90.03	90.11	90.11	84.15	73.27	41.74	56.54
-2.5	89.71	90.05	90.15	90.10	78.95	65.37	40.47	86.12
-2.0	89.64	90.05	90.07	90.02	74.91	56.85	68.57	93.01
-1.5	89.61	90.04	90.04	90.00	74.50	68.40	91.13	91.69
-1.0	89.59	90.01	90.01	90.02	76.13	89.96	90.41	90.21
-0.5	89.51	89.97	89.95	89.97	90.30	87.67	88.80	89.48
0.0	89.41	89.86	89.89	89.93	83.80	86.91	88.31	89.16
0.5	89.35	89.78	89.83	89.90	84.45	87.32	88.33	89.08
1.0	89.17	89.74	89.77	89.87	85.99	87.91	88.56	89.13
1.5	89.00	89.68	89.75	89.86	86.97	88.48	88.87	89.29
2.0	88.54	89.75	89.79	89.91	87.29	88.83	89.16	89.46

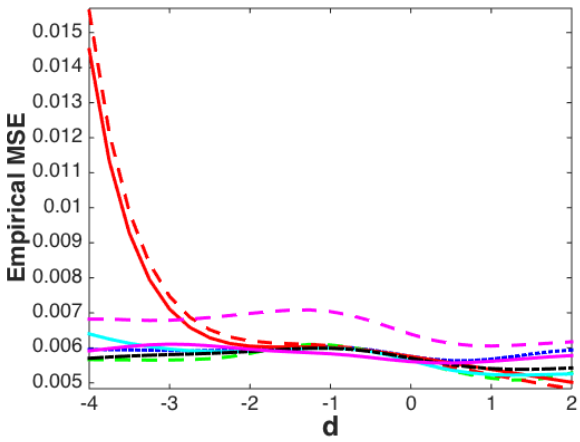
G Full Simulation Results on Misspecified VAR model



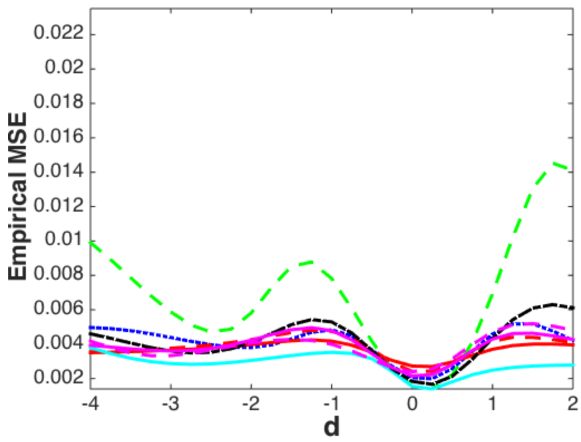
(a) $h = 1$, $1 \to 1$.



(b) $h = 1$, $2 \to 1$.

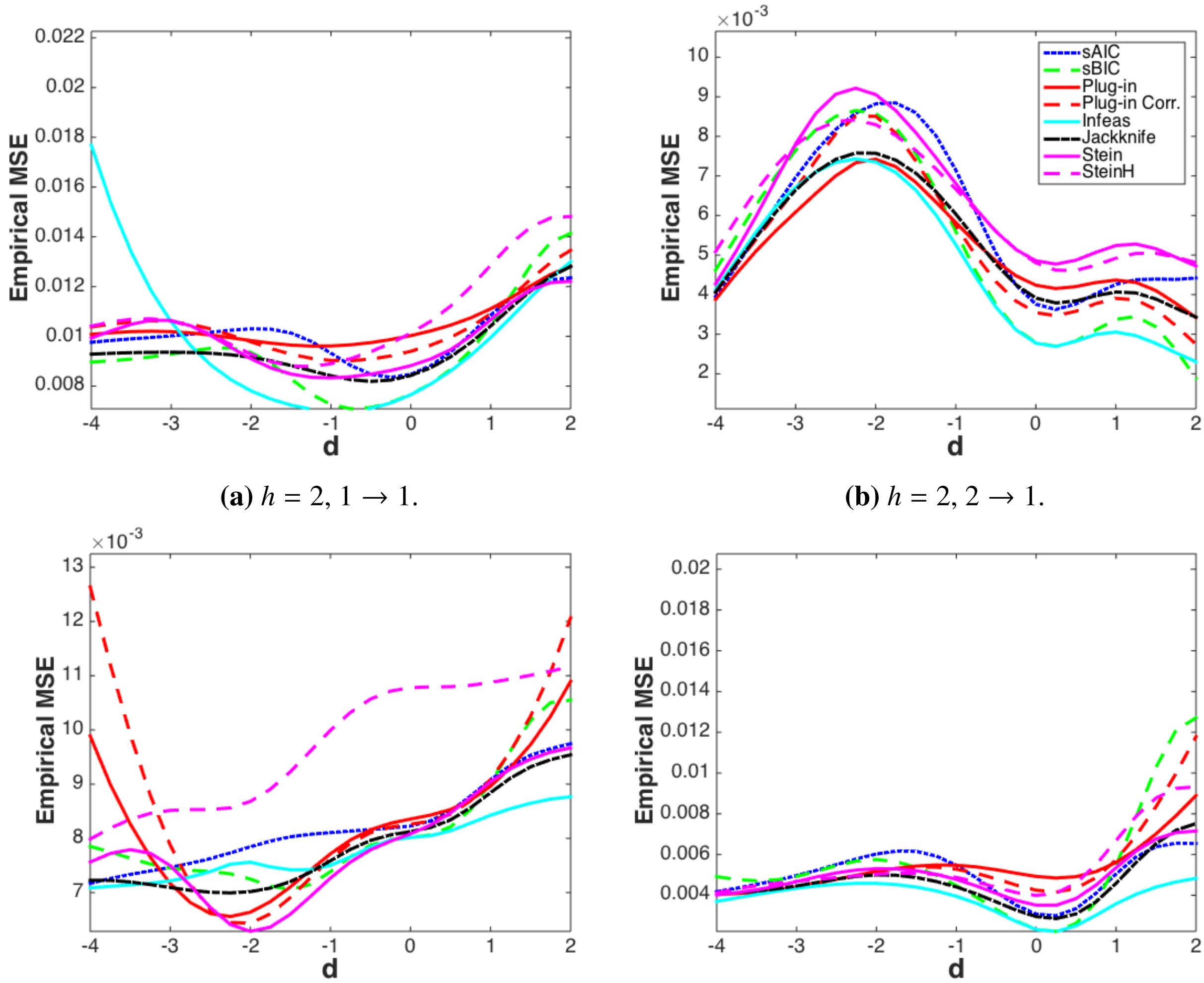


(c) $h = 1$, $1 \to 2$.



(d) $h = 1$, $2 \to 2$.

Figure 14: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



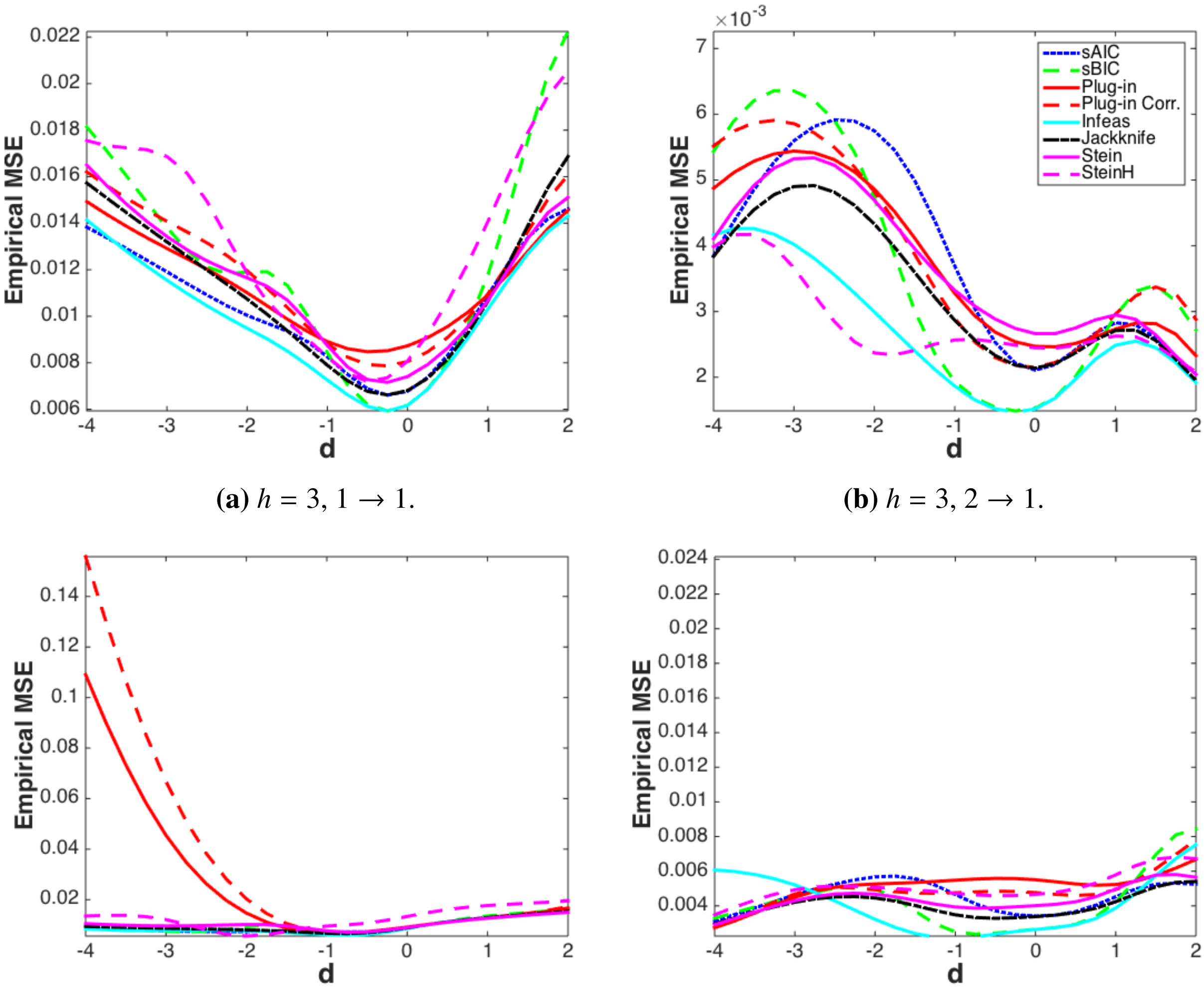
(a) $h = 2$, $1 \to 1$.

(b) $h = 2$, $2 \to 1$.

(c) $h = 2$, $1 \to 2$.

(d) $h = 2$, $2 \to 2$.

Figure 15: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



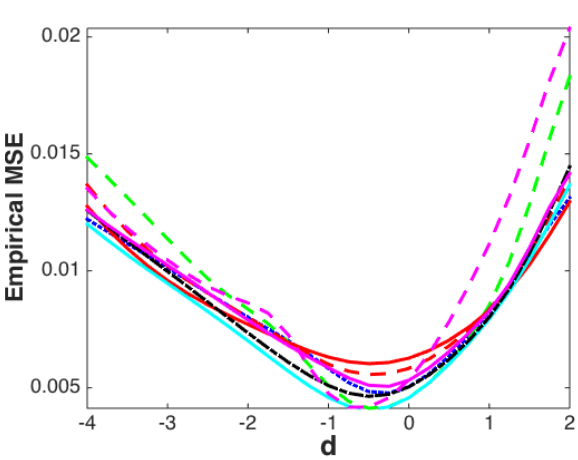
(a) $h = 3,\ 1 \to 1$.

(b) $h = 3,\ 2 \to 1$.

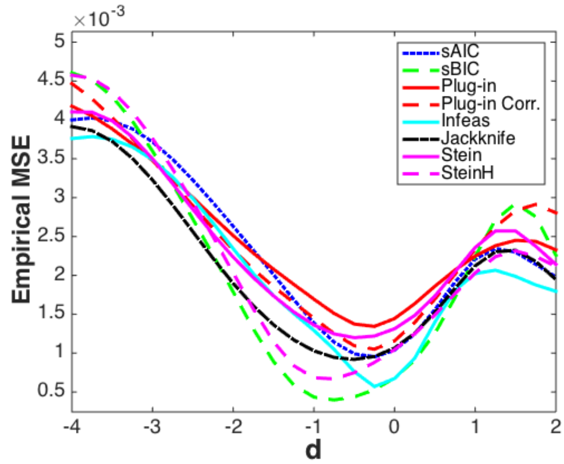
(c) $h = 3,\ 1 \to 2$.

(d) $h = 3,\ 2 \to 2$.

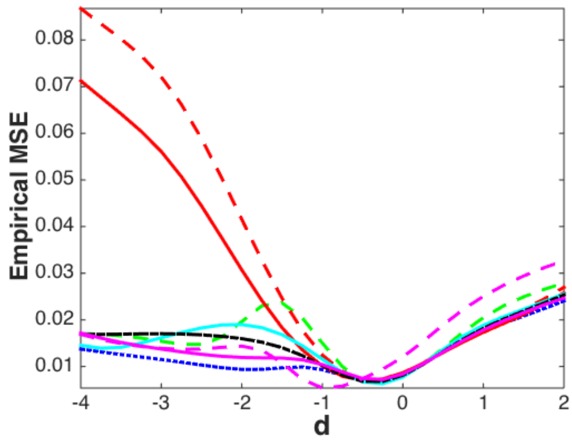
Figure 16: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



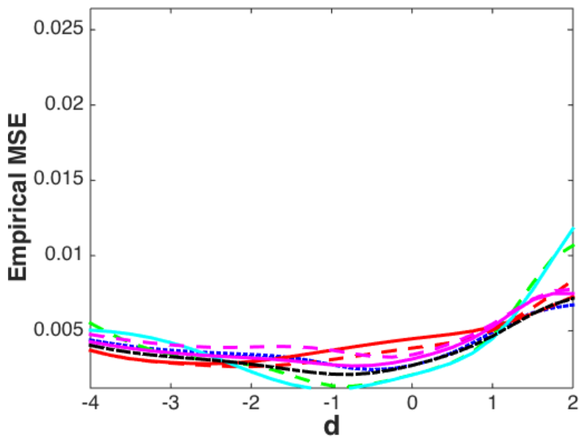
(a) $h = 4$, $1 \rightarrow 1$.



(b) $h = 4$, $2 \rightarrow 1$.

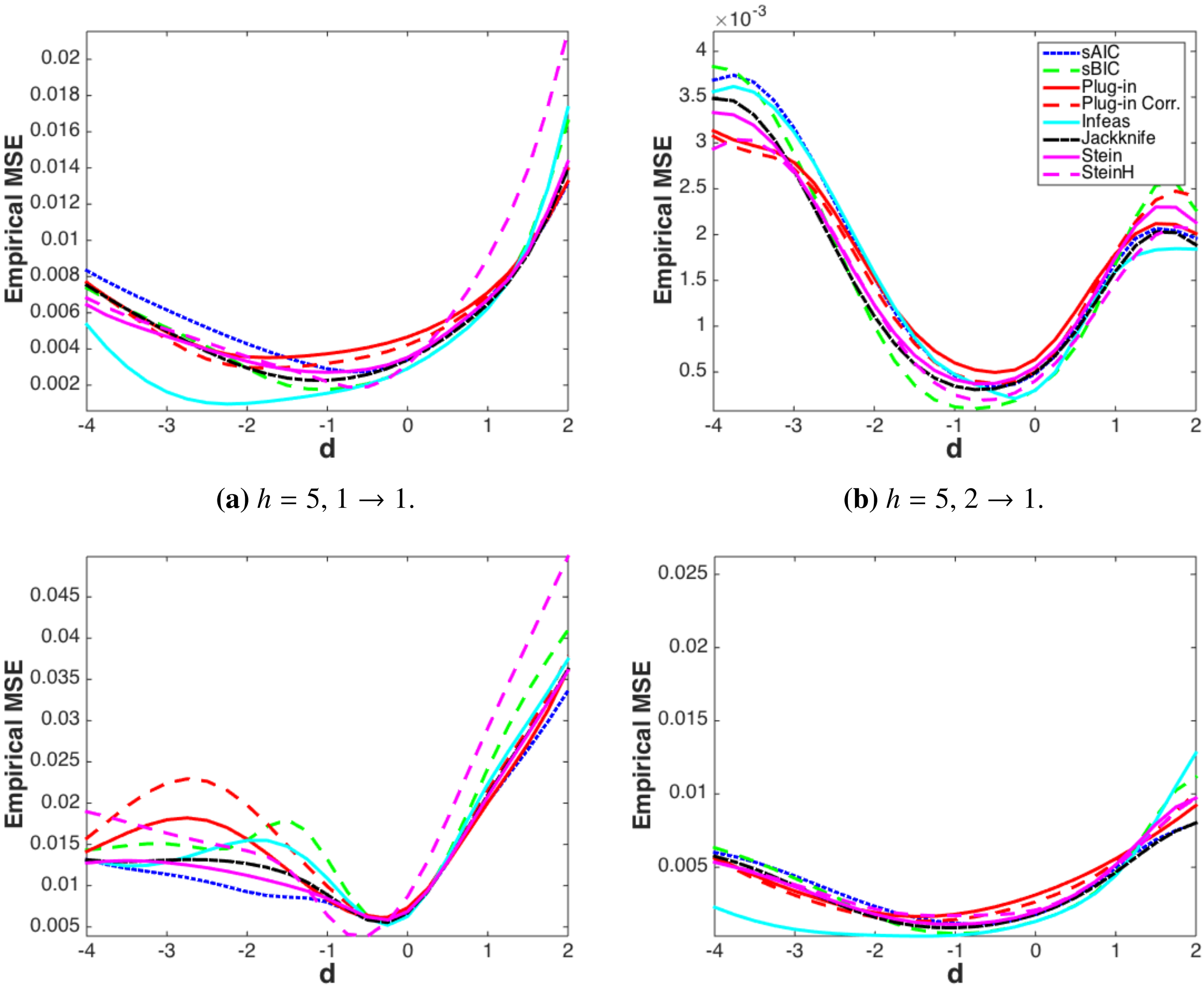


(c) $h = 4$, $1 \rightarrow 2$.



(d) $h = 4$, $2 \rightarrow 2$.

Figure 17: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \rightarrow j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



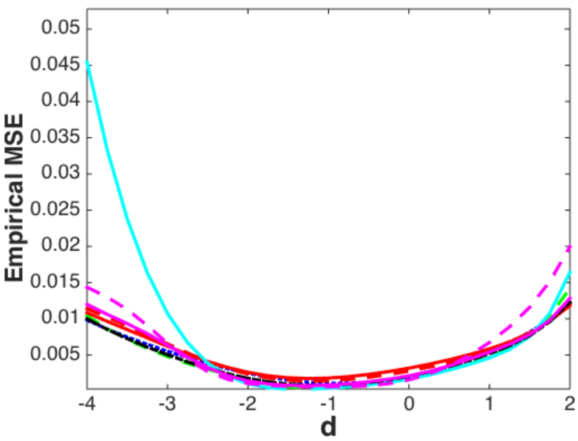
(a) $h = 5$, $1 \to 1$.

(b) $h = 5$, $2 \to 1$.

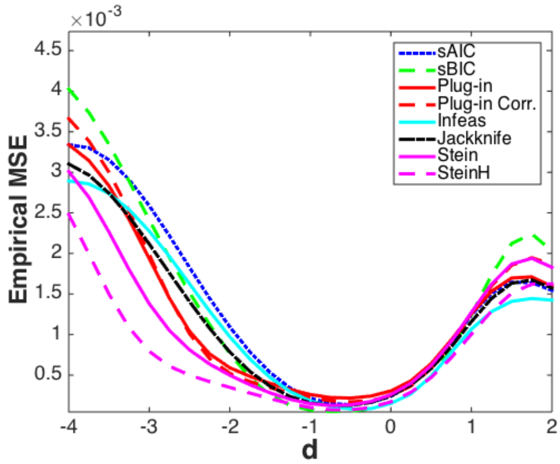
(c) $h = 5$, $1 \to 2$.

(d) $h = 5$, $2 \to 2$.

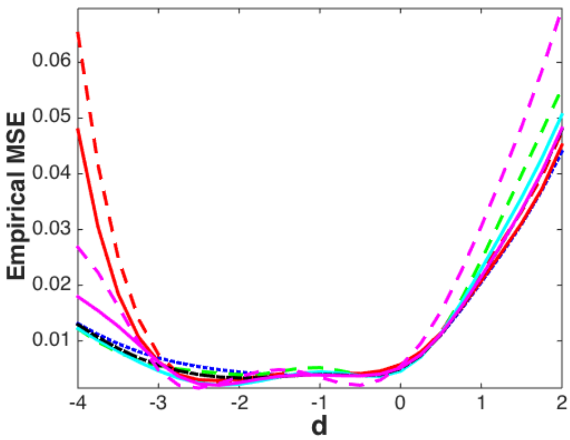
Figure 18: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



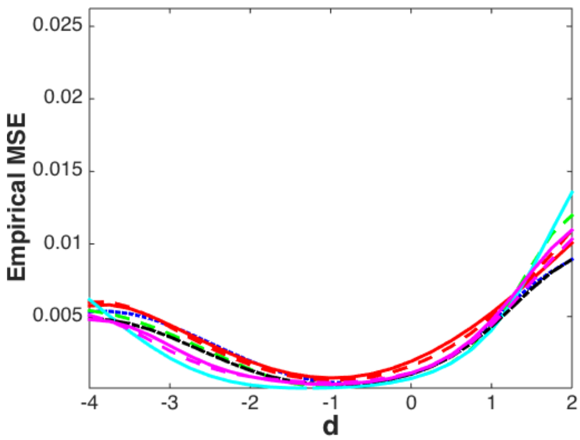
(a) $h = 6$, $1 \to 1$.



(b) $h = 6$, $2 \to 1$.

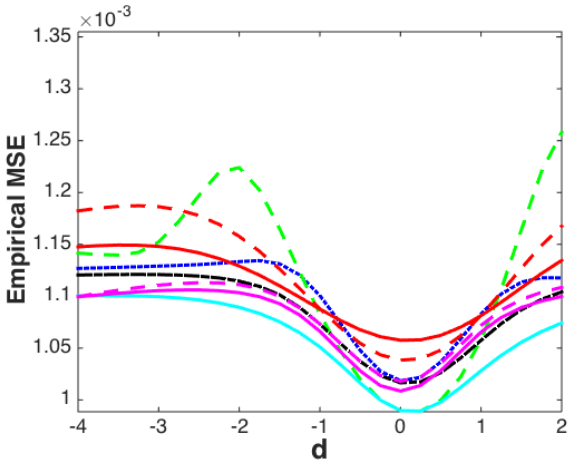


(c) $h = 6$, $1 \to 2$.

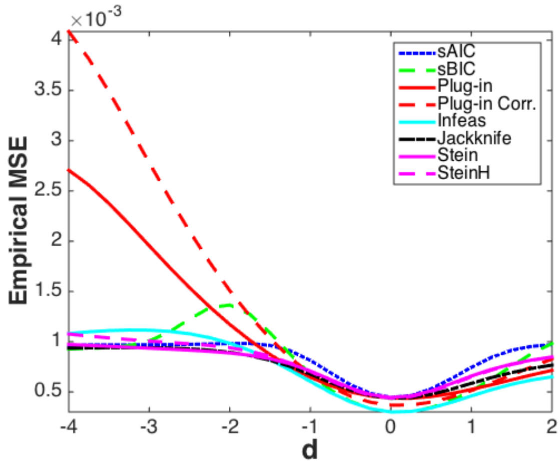


(d) $h = 6$, $2 \to 2$.

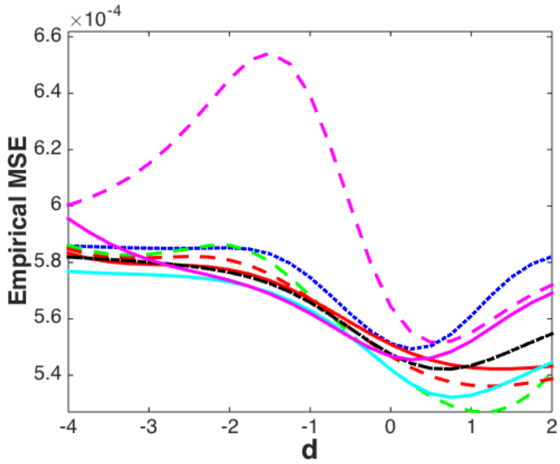
Figure 19: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.



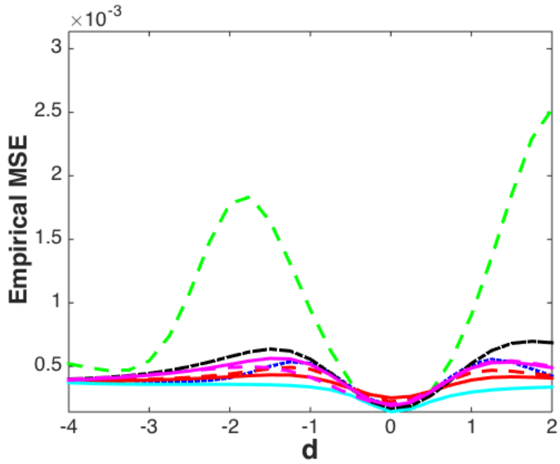
(a) $h = 1$, $1 \to 1$.



(b) $h = 1$, $2 \to 1$.

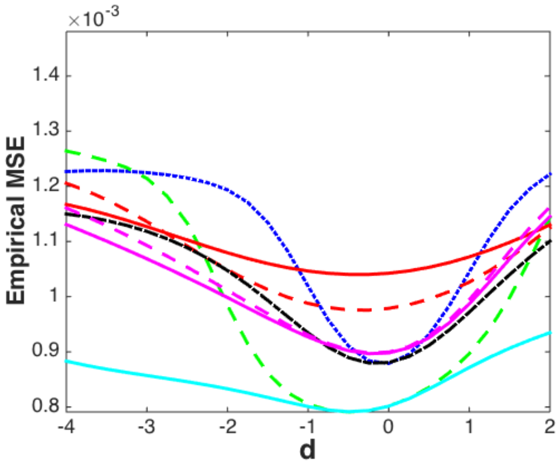


(c) $h = 1$, $1 \to 2$.

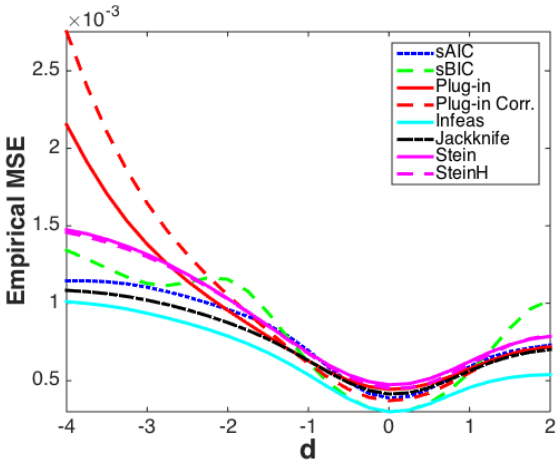


(d) $h = 1$, $2 \to 2$.

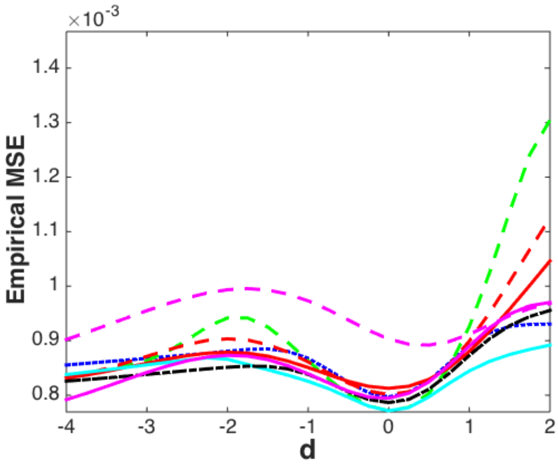
Figure 20: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.



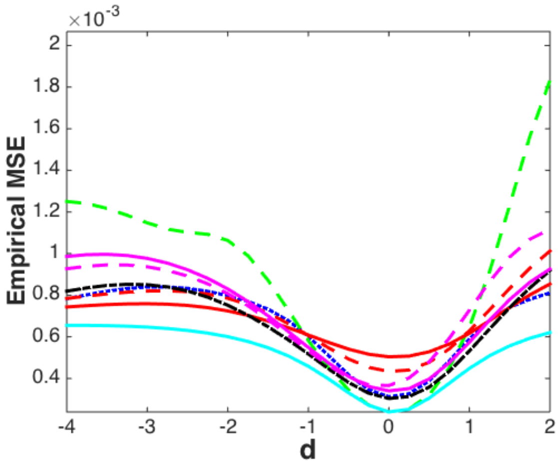
(a) $h = 2$, $1 \to 1$.



(b) $h = 2$, $2 \to 1$.

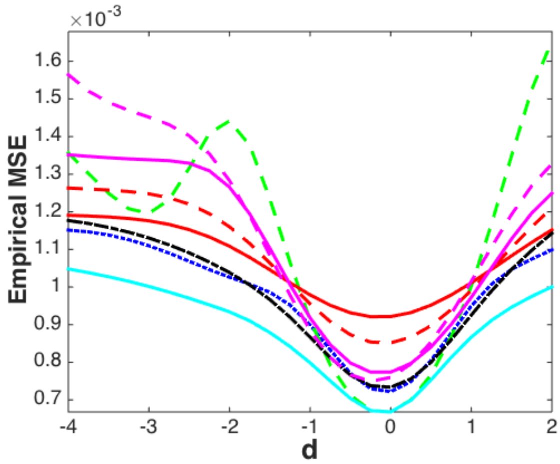


(c) $h = 2$, $1 \to 2$.

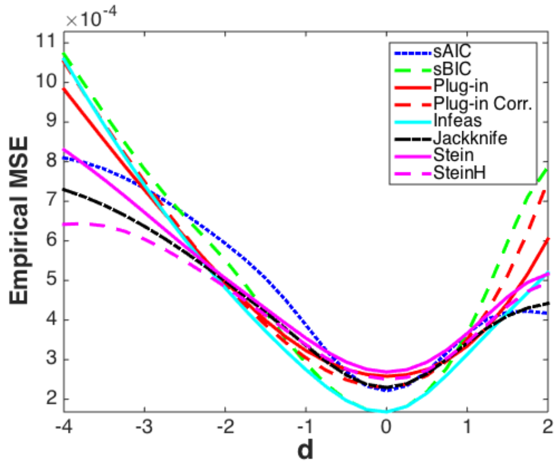


(d) $h = 2$, $2 \to 2$.

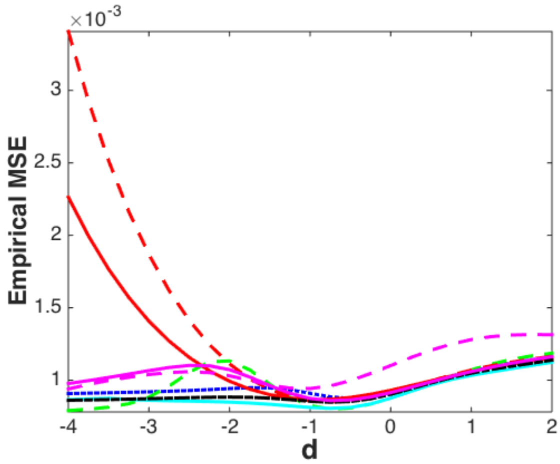
Figure 21: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.



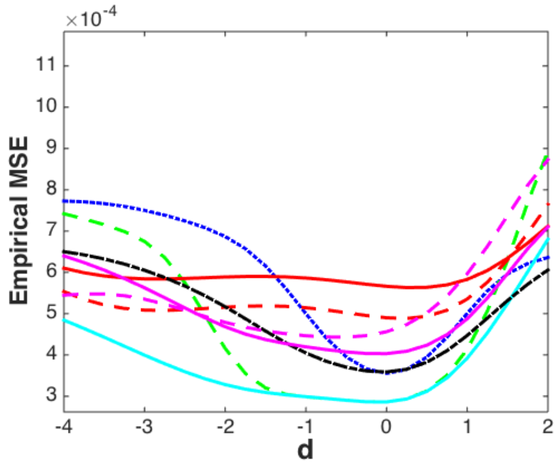
(a) $h = 3$, $1 \to 1$.



(b) $h = 3$, $2 \to 1$.

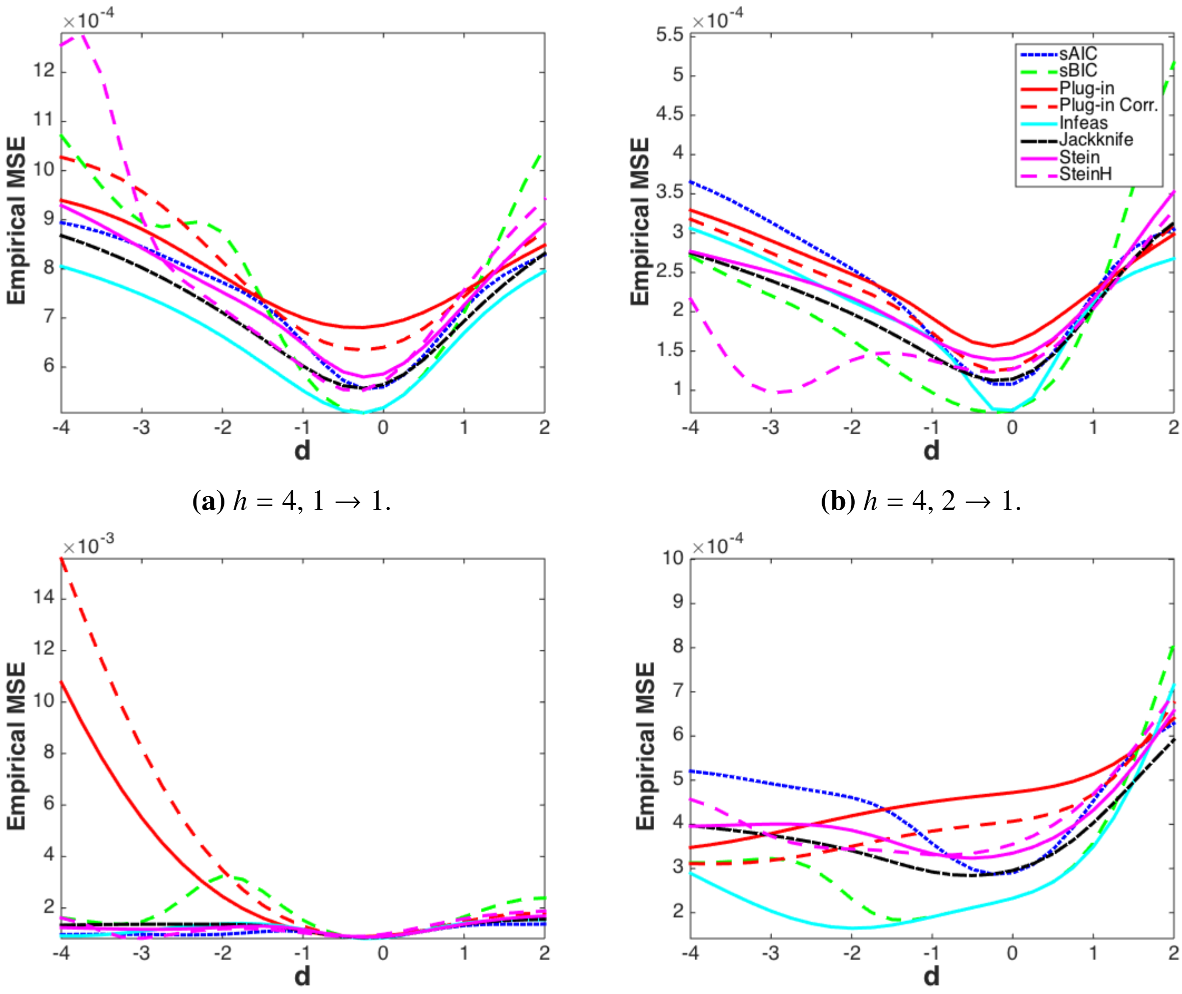


(c) $h = 3$, $1 \to 2$.



(d) $h = 3$, $2 \to 2$.

Figure 22: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.



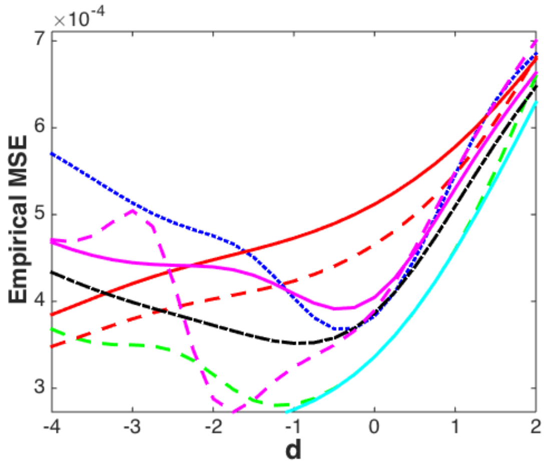
(a) $h = 4$, $1 \rightarrow 1$.

(b) $h = 4$, $2 \rightarrow 1$.

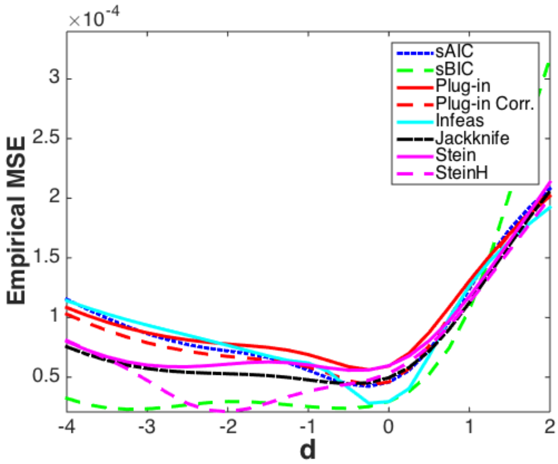
(c) $h = 4$, $1 \rightarrow 2$.

(d) $h = 4$, $2 \rightarrow 2$.

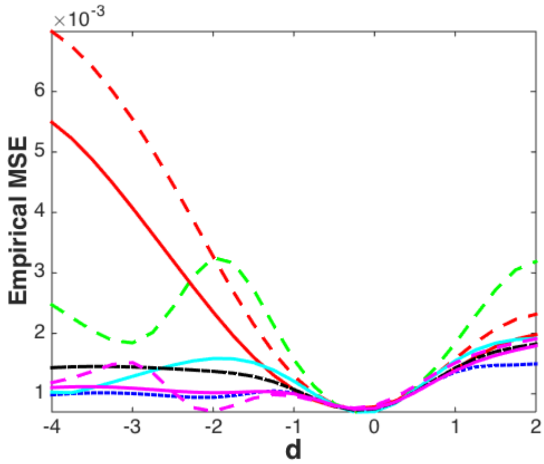
Figure 23: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \rightarrow j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.



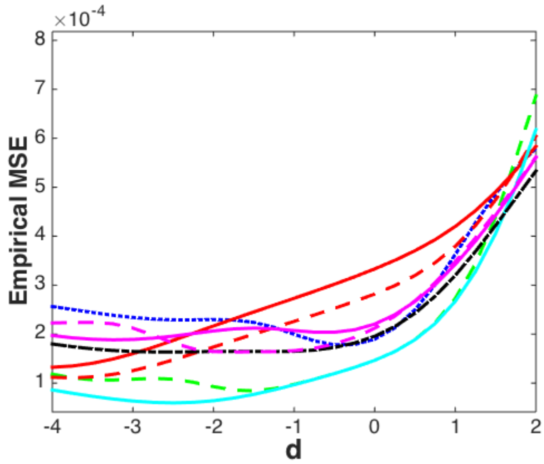
(a) $h = 5$, $1 \to 1$.



(b) $h = 5$, $2 \to 1$.

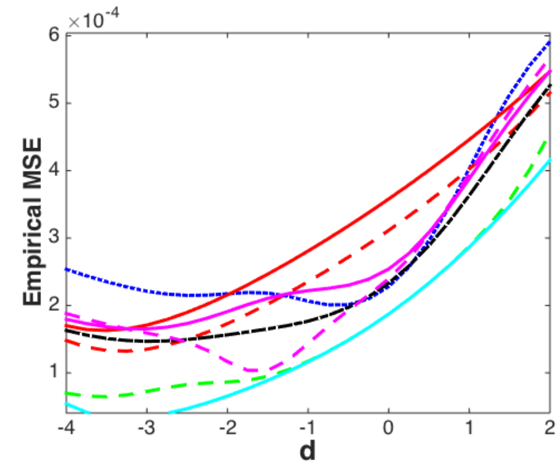


(c) $h = 5$, $1 \to 2$.

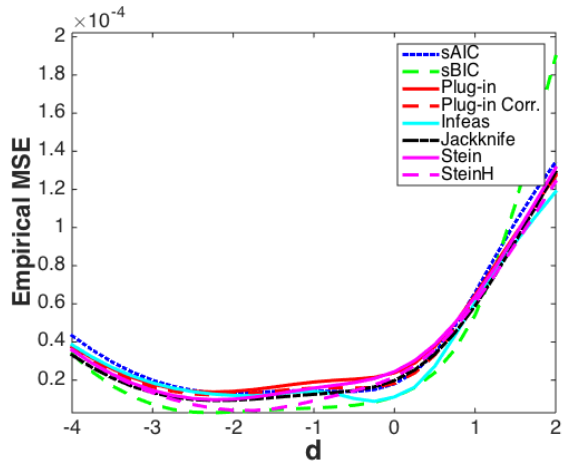


(d) $h = 5$, $2 \to 2$.

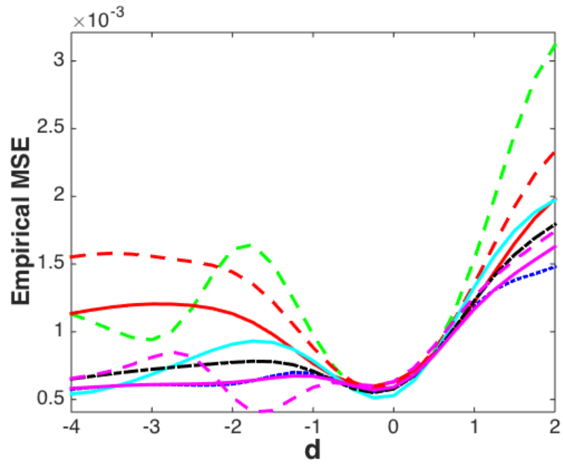
Figure 24: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.



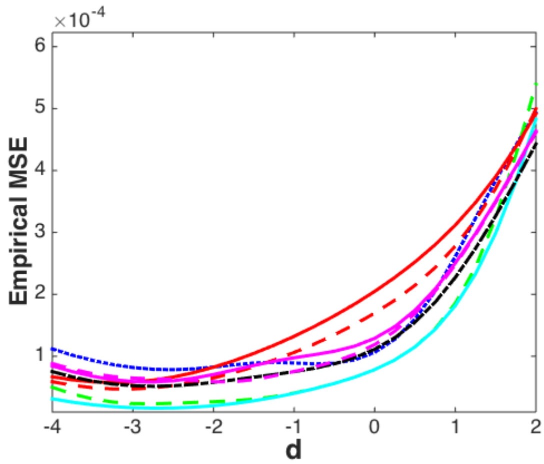
(a) $h = 6$, $1 \to 1$.



(b) $h = 6$, $2 \to 1$.



(c) $h = 6$, $1 \to 2$.



(d) $h = 6$, $2 \to 2$.

Figure 25: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 1: (a) The asymptotic MSE of the models with one and two lags (red and black line, respectively). The area between the 5% and 95% empirical quantile of $\widehat{FIC}_1$ and $\widehat{FIC}_2$ are shaded in red and grey. (b) The empirical selection probabilities of the FIC. (c) The AMSE of the models with $m = 1$ and $m = 2$ together with the empirical MSE of the feasible FIC (red) and infeasible FIC (cyan). This figure was obtained at $T = 1000$.

Figure 2: The 5% and 95% empirical quantile of the weights distribution without bias correction (a) and with bias correction (b). The infeasible weights are displayed in cyan. The empirical MSE of plug-in methods is shown in (c). This figure was obtained at $T = 1000$.

Figure 3: The empirical MSE for model selection based on Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Focused Information Criteria (FIC). 'Infeas' denotes an infeasible version of the FIC for which all quantities (and especially δ) are replaced by their true values. The DGP is $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ with $T = 100$.

Figure 4: Idem Figure 3, but for $T = 1000$.

Figure 5: The empirical MSE for model averaging based on smoothed AIC (sAIC), smoothed BIC (sBIC), the plug-in average without bias correction (Plug-in), the plug-in average with bias correction (Plug-in Corr.), the infeasible plug-in average (Infeas) and the average with Jackknife weights (Jackknife). The DGP is $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ with $T = 100$.

Figure 6: Idem Figure 5, but for $T = 1000$.

Figure 7: The empirical MSE of the OLS estimator of the model with 1 lag (OLS1), 2 lags (OLS2) and the full model with 3 lags (OLS3) for $T = 100$. Gray lines show the asymptotic MSE approximations as provided by the delta method.

Figure 8: Idem Figure 7, but $T = 1000$.

Figure 9: The selection frequencies of the various models for AIC (blue), BIC (green), FIC (red) and the infeasible estimator (cyan) as a function of the amount of misspecification, δ. The data was generated by $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ with a sample size of $T = 100$.

Figure 10: Idem Figure 9, but for $T = 1000$.

Figure 11: The empirical weights on the various models for sAIC (blue), sBIC (green), FIC (red) and the infeasible estimator (cyan) as a function of the amount of misspecification, δ. The data was generated by $y_{T,t} = 0.5y_{T,t-1} + \frac{\delta}{\sqrt{T}}y_{T,t-2} + \frac{\delta}{2\sqrt{T}}y_{T,t-3} + u_t$ at a sample size of $T = 100$.

Figure 12: Idem Figure 11, but for $T = 1000$.

Figure 13: Simulation results for a multiple quantities of interest. Model selection/averaging is based on the impulse responses from horizon 1 up to 6. The trace of the AMSE matrix is used to map the AMSE matrix to a scalar, i.e. model selection/averaging is based on the sum of the asymptotic mean-squared errors. Therefore, we also report the sum of the empirical MSE of the first six horizons on the vertical axis.

Figure 14: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 15: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 16: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 17: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 18: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 19: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 100$. See Equation (3.2) of the paper for the DGP.

Figure 20: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 21: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 22: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 23: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 24: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Figure 25: The empirical MSE of the impulse response estimator for several averaging methods. We use the notation $i \to j$ to indicate the impulse response of variable j to a structural shock in variable i. The sample size is $T = 1000$. See Equation (3.2) of the paper for the DGP.

Table 1: The empirical coverage of 90% confidence intervals for the horizons: two, three and six. For several choices of δ and T in the DGP $y_{T,t} = 0.5 y_{T,t-1} + \frac{\delta}{\sqrt{T}} y_{T,t-2} + \frac{\delta}{2\sqrt{T}} y_{T,t-3} + u_t$.

Table 2: The empirical coverage of the confidence regions as a function of sample size T and misspecification parameter δ.