A Regularization Approach to the Minimum Distance Estimation : Application to Structural Macroeconomic Estimation Using IRFs*

Senay SOKULLU †

September 22, 2011

Abstract

This paper considers the problem of invertibility of the covariance matrix of Impulse Response Functions in Dynamic Stochastic General Equilibrium Models, which use Impulse Response Function Matching Estimation. We propose to use a regularized inverse and present a simple simulation analysis to show that the regularized inverse performs better than the other weighting matrices adopted in the literature.

Keywords: Dynamic Stochastic General Equilibrium Models, Impulse Response Function, Tikhonov Regularization

JEL Classification: C5, E3

*Preliminary and incomplete

†University of Bristol, E-mail: *senay.sokullu@bristol.ac.uk*

1 Introduction

Starting with Rotemberg and Woodford (1997), interest for Impulse Response Function Matching estimation (IRFME) has increased as a result of its simplicity. However, in almost all of the Dynamic Stochastic General Equilibrium (DSGE) models, the IRFME suffers from the non-invertible covariance matrix of impulse response functions (IRFs). Hence, the macroeconometricians have adopted two solutions: the first one uses an identity matrix as a weighting matrix and the second one uses the inverse of the diagonal matrix of the variances of IRFs. In this paper, we propose to use a regularization scheme to invert the variance-covariance matrix of the IRFs and with a simple Monte Carlo simulation, we show that the method we propose performs better in terms of Mean Squared Errors in small samples.

The Minimum Distance Estimation (MDE) of structural VAR models, or in other words Impulse Response Function Matching Estimation (IRFME) has contributed a lot to the econometrics of DSGE models over the recent years. (See Rotemberg and Woodford, 1997; Christiano, Eichenbaum, and Evans, 2005; Giannoni and Woodford, 2004; Altig, Christiano, Eichenbaum, and Linde, 2004). The method has gained favor as it is easy to implement and as it captures the dynamics of the model by construction. It basically matches the empirical IRFs derived from a structural VAR and the theoretical IRFs obtained from the DSGE model. However, the method often suffers from the non-invertible covariance matrix of IRFs, as the inclusion of many lags leads to stochastic singularity. To solve this issue, researchers either use an identity matrix like in Rotemberg and Woodford (1997) or a diagonal matrix with the inverse variances of the IRFs on the diagonal as in Christiano, Eichenbaum, and Evans (2005). In this chapter, we address this problem and propose to use a regularized inverse of the covariance matrix of the IRFs. To the best of our knowledge, this is the first time that regularization is proposed to solve the problem of invertibility of the optimal weighting matrix in IRFME.

The use of non-optimal weighting matrix do not affect the consistency of the estimator but it affects the variance of the asymptotic distribution. As a result, variance of the asymptotic distribution of estimators obtained with non-optimal weighting matrix is different than that of optimal estimator. Thus using the variance of optimal estimator for non-optimal one leads to errors in inference. Feve, Matheron, and Sahuc (2009) present a correction for this mistake based on bootstrap methods. They propose to construct the distribution of this non-optimal estimator by bootstrapping methods. On the other hand, number of horizons to be included in the estimation of DSGE model is also closely related to the efficiency. Hall, Inoue, Nason, and Rossi (2007), propose a selection rule for the optimal horizon to get the most efficient estimation. Nonetheless, they continue to use non-optimal weighting matrix. The method presented in this chapter, leads to asymptotic distribution equal to the that of the optimal estimator. Moreover, by construction it can deal with infinite horizon and hence does not need a selection rule.

This problem in IRFME is indeed very similar to the problem of many instruments in econometrics literature. Although the increasing number of instruments improves the asymptotic efficiency of the estimator as is shown by Carrasco and Florens (2003), in finite samples, increasing number of instruments increases the bias. Moreover, it leads to non-invertible covariance matrix of moment conditions. The problem can be solved by taking a few instruments but this will result in an efficiency loss. Donald, Imbens, and Newey (2003) modify the empirical likelihood estimator for many instruments and show that it is asymptotically efficient. Donald and Newey (2001), derive the mean squared error of the estimator and propose to select the number of instruments which minimizes MSE. However, their method works best when one has the information of most relevant instruments. On the other hand, Carrasco, Florens, and Renault (2007), specify the estimation with many instruments as an inverse problem and point out the ill-posedness which is the result of non-invertible covariance matrix. Following this idea, they propose to use the regularized inverse of the covariance matrix. Moreover, Carrasco (2006) and Carrasco and Florens (2000) show that this regularized estimator is consistent and asymptotically normal. In this chapter, we follow Carrasco (2006) and Carrasco and Florens (2000) and treat the model as an ill-posed inverse problem and propose to regularize the covariance matrix of the IRFs by *Tikhonov Regularization*.

Moreover, since what we are doing is *minimum distance estimation*, the problem is a bit different than the case in GMM. Because, the optimal weighting matrix is indeed given by the covariance matrix of IRFs which are the linear transformations of existing moments in the sample. Hence, when we take infinite horizon for the IRFs, this will lead to an-infinite dimensional optimal weighting matrix with finite rank which obviously simplifies the problem compared to GMM with infinite number of moment conditions. We show that, regularized IRFME is consistent and asymptotically normal. Additionally, we simulate a simple dynamic model and show that with the optimal choice of the regularization parameter, we obtain better results than the adopted techniques in DSGE models.

The paper proceeds as follows. In *Section 2* we define IRF matching estimators. In *Section 3*, we introduce the use of Tikhonov Regularization in IRFME while in *Section 4* we show its asymptotic properties. Monte Carlo simulation and its results are presented in *Section 5*. Finally in *Section 6*, we conclude.

2 IRF Matching Estimator

In this section we define the IRF matching estimator that is used to estimate structural parameters in DSGE models.

Let x_t be the $(n_x \times 1)$ vector of variables of interest at date $t = 1, 2, ..., T$. IRF matching estimator is based on the minimum distance estimator which minimizes the distance between the empirical IRFs obtained by fitting a VAR model to x_t and the theoretical IRFs implied by the structural model. Suppose that the VAR model is given by the following:

$$x_t = \Gamma_0 + \Gamma_1 x_{t-1} + \Gamma_2 x_{t-2} + ... + \Gamma_s x_{t-s} + \epsilon_t$$

where $s \geq 0$ and $\epsilon \sim iid(0, \Omega)$. Let $\hat{\varphi}_h$ denote the vector of estimated IRFs up to horizon h and $\psi_h(\theta)$ denote the IRFs obtained from the structural model up to horizon h. Then the IRF matching estimator is given by:

$$\hat{\theta}_h = \underset{\theta \in \Theta}{\operatorname{argmin}} [\hat{\varphi}_h - \psi_h(\theta)]' \hat{A} [\hat{\varphi}_h - \psi_h(\theta)] \tag{1}$$

where $\hat{A}$ is the estimated weighting matrix and it converges in probability to A_0 which is symmetric and positive definite. Under standard regularity conditions, we can show that the estimator $\hat{\theta}$ is consistent and asymptotically normal, i.e:

$$\hat{\theta} \xrightarrow{p} \theta_0$$

and

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$$

where

$$\Sigma_0 = \left\langle A_0 \frac{\partial \psi}{\partial \theta'}(\theta_0), \frac{\partial \psi}{\partial \theta'}(\theta_0) \right\rangle^{-1} \left\langle A_0 \frac{\partial \psi}{\partial \theta'}(\theta_0), W A_0 \frac{\partial \psi}{\partial \theta'}(\theta_0) \right\rangle \left\langle A_0 \frac{\partial \psi}{\partial \theta'}(\theta_0), \frac{\partial \psi}{\partial \theta'}(\theta_0) \right\rangle^{-1}$$

Moreover, as in the case of GMM, the optimal weighting matrix is given by the inverse of the covariance matrix of IRFs. (See Gourieroux, Monfort, and Vuong, 1995; Ruud, 2000). As pointed out in Feve, Matheron, and Sahuc (2009), in many applications the number of IRFs included is much larger than the number of parameters to be estimated. Thus, this give the problem of invertibility for the optimal weighting matrix. In the next section we present our method to deal with this problem of invertibility.

3 Tikhonov Regularization for IRFME

As already presented in the previous section, IRFME base on estimating parameters by minimization of a weighted distance between the IRFs obtained from a reduced form VAR and IRFs of the structural DSGE model. Furthermore, it can be seen as a GMM estimator where the moment conditions is defined by these distances. Let $f(x_t, h, \theta)$ define the distance between the IRF of structural model and IRF of reduced form model up to h:

$$f(h, \theta) = \varphi_h - \psi_h(\theta_0) \tag{2}$$

where φ_h is the IRFs coming from the reduced form VAR and $\varphi_h = \varphi[m(X)]$, m is the existing moments of process X. $\psi_h(\theta)$ is the IRFs of the structural DSGE model where θ_0 is the true value of the vector of parameters. The estimator of θ_0 is given by the following minimization problem:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left\| \hat{K}\hat{f}(x_t, h, \theta) \right\|^2 \tag{3}$$

where $\hat{K}$ converges to its true value K and $\hat{f}(.)$ is the empirical counterpart of the equation (2)[1]:

$$\hat{f}(x_t, h, \theta) = \varphi(\hat{m}) - \psi_h(\theta)$$

Note that when h is taken to be finite, K is a matrix, i.e., $K : \mathbb{R}^q \to \mathbb{R}^q$, where q depends on h and the dimension of X. When h is infinite, K is an operator on $\mathcal{E}$ to $\mathcal{E}$ where $\mathcal{E}$ is a Hilbert space. The estimator obtained in (3) is consistent for any $\hat{K} \to K$ and it is efficient if $\hat{K}^*\hat{K}$ is a consistent estimator of the inverse of covariance matrix of IRFs.[2] In cases where the number of IRFs exceeds the number of structural parameters to be estimated we have the exact problem as in the problem of many instruments in IV regression[3]. In this case $\hat{K}^*\hat{K} = \hat{W}$ becomes near singular and thus non-invertible. In other words, the estimation becomes an ill-posed inverse problem.

Inverse problems exist in many areas of econometrics, especially whenever we try to recover structural parameters from reduced form equations. IRFME is an example of this case since we try to recover the structural parameters of DSGE model from a reduced form VAR. Another example can be the structural estimation of auctions where we observe the bids and we try to recover the distribution function of valuation of bidders. Nonetheless, not every inverse problem is necessarily ill-posed. An inverse problem is well-posed when the

[1] Regarding the formulation of the estimator used in the previous section, $A = K^*K$

[2] See Gourieroux, Monfort, and Vuong (1995),Ruud (2000).

[3] See Donald and Newey (2001), Donald, Imbens, and Newey (2003), Hansen, Hausman, and Newey (2006), Carrasco and Florens (2000).

solution exists, is unique and continuous. (See Engl, Hanke, and Neubauer, 1996) In the case of IRFME, when the number of IRFs included is large compared to the number of parameters of interest, the efficient weighting matrix W becomes nearly singular, i.e., it has eigenvalues which are very close to zero. Thus the solution that depends on the W^{-1} becomes unstable and this results in an ill-posed IRFME problem. To solve this ill-posed inverse problem, following Carrasco, Florens, and Renault (2007), Carrasco (2006) and Carrasco and Florens (2000), we propose to regularize the solution using Tikhonov Regularization scheme[4]. Under this scheme, the regularized inverse of the weighting matrix is given by[5]:

$$(W^{\alpha})^{-1} = (\alpha I + W'W)^{-1}W \tag{4}$$

where $\alpha > 0$ and $\alpha \overset{T\to\infty}{\longrightarrow} 0$ is called *regularization parameter* and I is the identity matrix.

It is clear from equation (4) that we are disturbing the optimal weighting matrix by adding α. On the other hand, by regularizing the inverse of the covariance matrix, we make it more stable, thus we decrease the variance of the estimator, it can be seen as a smoothing parameter. Hence the choice of α is very important. The optimal α can be derived from the MSE, however, since the regularized estimator is consistent and has the same asymptotic distribution as the optimal estimator, we need to do a second order analysis.[6]. This choice in regularization of the problem of many instruments is clearly studied by Carrasco (2006). After obtaining an analytical expression for MSE, she derives the formula of α which minimizes it.[7] Nonetheless, in the simulations we make, we use a data based selection rule for α which is defined in *Section 5.1*.

4 Consistency of the Regularized IRFME

In this section we show that the regularized IRFME is consistent. The consistency of the regularized estimator in case of many instruments in GMM is shown before by Carrasco and Florens (2000) and Carrasco (2006), nonetheless here we prove it for the impulse response

[4]There are also other regularization schemes in inverse problems literature, but in this chapter we adopt Tikhonov Regularization as it is easy to implement. However, in some cases Tikhonov regularization is not very practical as it requires the inversion of a matrix. For those cases, Landweber-Friedman Regularization should be used which is an iterative approach and does not require the inversion of a matrix. Carrasco (2006) shows the implementation of other regularization schemes as well as the asymptotic results.

[5]Note that W becomes an operator when we use $h = \infty$

[6]In next section we show that it has the same asymptotic distribution as the optimal one. Moreover, it is also proved by Carrasco (2006)and Carrasco and Florens (2000) for regularized GMM estimation with many moment conditions

[7]Though the choice of optimal α may not be the same for all inverse problems. The criterion function to be minimized by the selection of α changes from application to application. See Golub, Heath, and Wahba (1979) and Engl, Hanke, and Neubauer (1996).

function matching estimation.[8] Following assumptions are needed for the consistency:

Assumption 1 *Let X be a stationary, ergodic, random process. $\{X_t\}_{t=1}^{T}$ is an observed random sample of X.*

Assumption 2 *(i) Let φ be the functions of IRFs from the reduced form VAR. $\varphi : \mathbb{R}^r \to \mathcal{E}$, where r is the number of moments of $\{X_t\}_{t=1}^{T}$ and $\mathcal{E}$ is an Hilbert space with the inner product $\langle ., . \rangle$ that defines a norm $\|.\|$.*
(ii) Let $\psi(\theta)$ be a function of theoretical IRFs from the structural model. It is defined from $\psi_h : \mathbb{R}^k \to \mathcal{E}$, where $\theta \in \mathbb{R}^k$ is the parameters to be estimated.

Assumption 3 *The function $f : \mathcal{E} \to \mathcal{E}$ is defined as:*

$$f = \varphi(m(x)) - \psi(\theta)$$

and it takes its minimum value for $\theta = \theta_0$.

Assumption 4 *Let K be a nonrandom bounded linear operator $K : \mathcal{D}(K) \subset \mathcal{E} \mapsto \mathcal{E}$ and let $K^*K = W^{-1}$. Then $f \in \mathcal{D}(W)$ for all θ.*

Assumption 5 *Let $N(K)$ denote the null space of K, i.e., $N(K) = \{g \in \mathcal{E} | Kg = 0\}$. Then we assume that $f \in N(K)$ implies $f = 0$.*

Assumption 6 *Let $\hat{K}_n$ be a sequence of random bounded linear operators. $\hat{K}_n : \mathcal{D}(\hat{K}_n) = \mathcal{E} \mapsto \mathcal{E}$. Let $\hat{f} = \varphi(\hat{m}) - \psi(\theta)$. Then we assume that $\hat{f} \in \mathcal{D}(\hat{K}_n) \forall \theta$ and $Q_n = \left\| \hat{K}_n \hat{f} \right\|$ is a continous function of θ.*

Assumption 7 *$Q_n \to Q = \|Kf\|$ almost surely on $\theta \in \mathbb{R}^k$*

Assumption 8 *$f(\theta)$ is differentiable with respect to θ and the $(k \times k)$ matrix $\left\langle \hat{K}_n \frac{\partial \hat{f}(\bar{\theta})}{\partial \theta'}, \hat{K}_n \frac{\partial \hat{f}(\hat{\theta})}{\partial \theta'} \right\rangle$ is positive definite and symmetric.*

Assumption 9 *The inner product satisfies the following differentiation rule:*

$$\frac{\partial}{\partial \theta'} \langle u(\theta), v(\theta) \rangle = \left\langle \frac{\partial}{\partial \theta'} u(\theta), v(\theta) \right\rangle + \left\langle u(\theta), \frac{\partial}{\partial \theta'} v(\theta) \right\rangle$$

and K and $\hat{K}_n$ commute with differential operator:

$$\frac{\partial}{\partial \theta'}[Ku(\theta)] = K[\frac{\partial}{\partial \theta'} u(\theta)]$$

[8] Their results also hold for the case of continuum of moment conditions.

Assumption 10 $\mathbb{E}\left\|f\right\|^4 < \infty$

Assumption 11

$$\left\|\hat{f}(\theta) - f(\theta)\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\left\|\frac{\partial \hat{f}}{\partial \theta}(\theta) - \frac{\partial f}{\partial \theta}(\theta)\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

Lemma 1 *Let $\hat{W}^{\alpha}$ denote the regularized inverse of a linear operator $W = K^*K$ by Tikhonov Regularization, i.e., $\hat{W}^{\alpha} = (\alpha I + \hat{W}^*\hat{W})^{-1}\hat{W}^*$ and W^+ denote the generalized inverse. Then:*

$$\left\|\hat{W}^{\alpha} - W^+\right\| \to 0 \quad \textit{in probability as} \quad T \to \infty, \alpha^3 T \to \infty, \alpha \to 0$$

Theorem 2 *Under the Assumptions (1) to (11), the estimator:*

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left\|\hat{f}(\theta, h)\right\|^2_{\hat{W}^{\alpha}}$$

satisfies:
(i)$\hat{\theta} \to \theta_0$ *in probability as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$
(ii)$\sqrt{T}(\hat{\theta} \to \theta_0) \sim \mathcal{N}\left(0, \left\langle W\frac{\partial f}{\partial \theta'}(\theta_0), \frac{\partial f}{\partial \theta'}(\theta_0)\right\rangle^{-1}\right)$ *as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$

Theorem (2) states that the regularized IRFME estimator is consistent and optimal.

Remark 3 *The problem of invertibility of the weighting matrix is not treated by the use of a generalized inverse in any of the aforementioned DSGE literature. Although, it would lead to the same asymptotic distribution as the regularized IRFME, in Appendix B, by a Monte Carlo simulation, we show that the estimator is not as stable as the regularized estimator.*

5 Simulation

In this section we perform a Monte Carlo simulation to compare the efficiency of the estimator we propose with the efficiency of other estimators used in IRFME literature. So we compare the estimators which are obtained from three different approaches. The first approach is the one that we propose, i.e., the covariance matrix of the IRFs is inverted by regularizing it with Tikhonov Regularization scheme. The second approach uses a diagonal weighting matrix which has the inverse of the variances of IRFs on the diagonal. Finally the third one uses an identity matrix as the weighting matrix.

5.1 The Simulation Design

We use a simple two variable one lag VAR model:

$$A^{\theta}(L)X_t = \epsilon_t$$

where

$$A^{\theta}(L) = \begin{pmatrix} (1-\gamma L) & 0 \\ (-\beta L) & (1-\gamma L) \end{pmatrix}, X_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \epsilon_t = \begin{pmatrix} v_t \\ u_t \end{pmatrix}$$

The true value of γ is chosen to be equal 0.35 while that of β is set to 3. Moreover just for simplicity we assume that the covariance matrix of ϵ is equal identity. In other words the errors are independently and identically distributed with $\mathcal{N}(0, I)$. As a result of this assumption we can get the impulse response functions by just taking the inverse of the polynomial matrix $A^{\theta}(L)$. Then the IRFs are given by:

$$A^{\theta}(L)^{-1} = \begin{pmatrix} \sum_{j=0}^{\infty} \gamma^j L^j & 0 \\ \sum_{j=0}^{\infty} (i+1)\beta\gamma^j L^{j+1} & \sum_{j=0}^{\infty} \gamma^j L^j \end{pmatrix}$$

We simulate the model 1000 times for samples of sizes 100, 200, 500 and 1000. We do several estimations with horizons equal to 2, 5 and 7. In each estimation, after we estimate the VAR model, we obtain the IRFs. Then by bootstrapping the residuals of the estimated VAR, we obtain a 100 bootstrap sample to re-estimate the model and obtain the bootstrap values of the IRFs. These bootstrap values of IRFs are then used to compute efficient weighting matrix. For each Monte Carlo replication of horizon h, the IRFME is given by:

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}} [\hat{\psi}_h - \psi_h(\theta)]' W_n [\hat{\psi}_h - \psi_h(\theta)]$$

where θ is the vector of parameters, i.e., $\theta = (\gamma, \beta)'$. Moreover, the weighting matrix W_n is the regularized inverse of the efficient weighting matrix in the first approach. It is equal to diagonal matrix with the inverse variances on the diagonal in the second approach and it is the identity matrix in the third approach.

We use a data-based selection rule to choose the regularization parameter, α. We get a grid of α and estimate the regularized model for each value on the grid. Then we choose the α that gives the minimum MSE. The grid we use is an equidistant space of 50 points in the interval $[10^{-9}, 1]$.

5.2 Results

The results are presented in Tables (4.2), (4.3) and (4.4). We present the results obtained with optimal regularization parameter. Moreover, the bias, variance and MSE are the norm of the bias, variance and MSE of the estimated parameters. First of all, it should be noted that for all simulations, the estimation which is done with regularized weighting matrix (now on *regularized estimation*) performs best. More concretely, for all sample sizes and all horizons, the *regularized estimation* always has the smallest bias and variance and thus the smallest MSE.

For the horizons equal to 5 and 7, the second best estimator is given by the estimation done with identity weighting matrix (now on *identity*) and the worst is given by the estimation done with diagonal weighting matrix with inverse variances of IRFs on the diagonal (now on *diagonal*). For horizon equal to 2, as we have 8 moment conditions to estimate 2 parameters, the problem is not that severe and all the estimators perform more or less the same.

Moreover, for all horizons, MSEs are decreasing as sample size is increases, as expected. On the other hand, the performance of the estimators for given sample sizes depends on the selected horizon. For *regularized estimation*, the performance of the estimator do not vary with the horizon length for large samples. For sample sizes equal to 500 and 1000, the empirical MSEs are same for all horizons and they are slightly different for $T = 200$. Whereas, the MSE decreases with horizon for sample size of 100. This result suggest that, for small samples, choice of number of horizons to be included matters as it may bring more information, while in large samples, the number horizons included does not have significant effect under the *regularized estimation*.

For *diagonal* and *identity* estimators, increasing the number of horizon for a given sample size increases the MSE. As these estimators can not carry the information from extra horizons to weighting matrix, increased number of moment conditions increases the bias and the variance of the estimators. Thus, we can conclude that, with *regularized estimation* the choice of number of horizons is not a big issue if the sample is sufficiently large. However, for other approaches, the number of horizons included does matter and the Redundant Impulse Response Selection Criterion (RIRSC) of Hall, Inoue, Nason, and Rossi (2007) can be used to select the optimal number of horizons.

Note that all the results we present here holds with the regularized estimator with the optimal regularization parameter, α. Regularization introduces a bias to the estimation, at the same it decreases the variance. Hence α should be selected in such a way that it balances these two effects. In the simulations, we select the α that minimizes the empirical MSE.[9] Furthermore, α depends on the sample size and on the degree of ill-posedness of the problem. For the problems, that are slightly ill-posed, we do not need a strong regularization, thus the optimal α need not to be big. However, for severely ill-posed inverse problems, we need strong regularization which means a larger α. In Table (4.5), we present values of optimal α used in each simulation. For all sample sizes, the minimum α is used with horizon $h = 5$. It suggests that, the ill-posedness is more severe with $h = 2$ and $h = 7$. It makes sense for $h = 7$, as the covariance matrix of IRFs becomes more collinear. For the case where $h = 2$, the need for stronger regularization parameter can stem from the fact that using small number of IRFs also makes the efficient weighting matrix more collinear as the information they carry do not vary a lot. Finally, for given h, the optimal α decreases as sample size increases.

[9] Figures (1), (2), (3) and (4) show the selection rule for h=2 and for samples of sizes 100, 200, 500 and 1000, respectively.

6 Conclusion

In this chapter, we propose to use regularization for the invertibility problem of efficient weighting matrix in IRFME. After defining the regularized estimator, we show its asymptotic properties and we analyze its small sample performance with Monte Carlo simulations.

Contributions of the chapter are manyfold. First of all, it is first time that the estimation with *Tikhonov regularization* is proposed for IRFME. Secondly, we show that the estimation done with regularized weighting matrix perform best compared to the estimations with diagonal weighting matrix and identity weighting matrix for all number of horizons and for all sample sizes. Thirdly, we show that for sufficiently large samples, the regularized estimation is not affected by the choice of number of horizon, i.e., the regularization parameters adapts itself such that the increasing the length of the horizon do not improve or deteriorate the performance of the estimator. Thus, selection of optimal horizon may not be crucial issue with regularized estimation.

Although the simulations show that the regularized estimation perform better, we believe that estimation of a DSGE model with *regularized IRFME* would worth studying. Especially, estimation of the model in Christiano, Eichenbaum, and Evans (2005) can be very interesting as it has already been used as a benchmark to make comparisons by different papers. On the other hand, our approach can be compared with other estimation techniques used in DSGE models, i.e., with maximum likelihood and Bayesian approaches. Finally, a regularized estimation considering the choice of optimal regularization parameter as well as an optimal number of horizon can be studied.

Appendices

A Technical Proofs

A.1 Proof of Lemma 1

Proof. Let us write:

$$W^{\alpha} f = (\alpha I + W^{*}W)^{-1} W^{*} f$$

$$\hat{W}^{\alpha} f = (\alpha I + \hat{W}^{*}\hat{W})^{-1} \hat{W}^{*} f$$

Then

$$\left\| \hat{W}^{\alpha} - W^{+} \right\|^{2} \leq \underbrace{\left\| \hat{W}^{\alpha} - W^{\alpha} \right\|^{2}}_{I} + \underbrace{\left\| W^{\alpha} - W^{+} \right\|^{2}}_{II}$$

To prove the Lemma 1, we need to show that the first and the second term on the right hand side converges to zero as. Let us begin by the first term:

$$\left\| \hat{W}^{\alpha} - W^{\alpha} \right\|^{2} = \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1}\hat{W}^{*} - (\alpha I + W^{*}W)^{-1}W^{*} \right\|^{2}$$

$$= \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1}(\hat{W} - W) + (\alpha I + \hat{W}^{*}\hat{W})^{-1}W^{*} - (\alpha I + W^{*}W)^{-1}W^{*} \right\|^{2}$$

$$\leq \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1} \right\|^{2} \left\| (\hat{W} - W) \right\|^{2} + \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1} - (\alpha I + W^{*}W)^{-1} \right\|^{2} \left\| W^{*} \right\|^{2}$$

The first term in (I) is $O(\frac{1}{\alpha^2})$ by Darolles, Fan, Florens, and Renault (2010), while the second term is $O(\frac{1}{T})$. The third term need a further investigation:

$$\left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1} - (\alpha I + W^{*}W)^{-1} \right\|^{2} \leq \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1}(\hat{W} - W)(\alpha I + \hat{W}^{*}\hat{W})^{-1}W \right\|^{2}$$

$$\leq \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1} \right\|^{2} \left\| (\hat{W} - W) \right\|^{2} \left\| (\alpha I + \hat{W}^{*}\hat{W})^{-1}W \right\|^{2}$$

The first term is $O(\frac{1}{\alpha^2})$ by Darolles, Fan, Florens, and Renault (2010), the second term is $O(\frac{1}{T})$ and finally the third term is $O(\frac{1}{\alpha})$ by Carrasco and Florens (2000). So (I) converges to zero as $T \to \infty$, $\alpha^3 T \to \infty$ and $\alpha \to 0$.

To show that (II) converges to 0, if we use Fourier decompositions:

$$W^{+} f = \sum_{j=1}^{r} \frac{1}{\lambda_j} \langle f, \phi_j \rangle \phi_j$$

$$W^{\alpha}f = \sum_{j=1}^{r} \frac{\lambda_j}{\alpha + \lambda_j^2} \langle f, \phi_j \rangle \phi_j$$

where λ_j and ϕ_j are the eigen values and the eigen functions of the operator W, respectively and r is number of moments of the process {X}. Then:

$$\left\| W^{\alpha} - W^{+} \right\|^2 = \sum_{j=1}^{r} \left(\frac{\lambda_j}{\alpha + \lambda_j^2} - \frac{1}{\lambda_j} \right)^2 \langle f, \phi_j \rangle^2$$

$$= \sum_{j=1}^{r} \left(\frac{-\alpha}{(\alpha + \lambda_j^2)\lambda_j} \right)^2 \langle f, \phi_j \rangle^2$$

$$= \sum_{j=1}^{r} \frac{\alpha^2}{\left((\alpha + \lambda_j^2)\lambda_j\right)^2} \langle f, \phi_j \rangle^2$$

$$= \alpha^2 \sum_{j=1}^{r} \frac{1}{\left((\alpha + \lambda_j^2)\lambda_j\right)^2} \langle f, \phi_j \rangle^2 \leq \alpha^2 \sum_{j=1}^{r} \frac{1}{(\lambda_j^3)^2} \langle f, \phi_j \rangle^2$$

So (II) is $O(\frac{1}{\alpha^2})$. Then we can conclude that:

$$\left\| \hat{W}^{\alpha} - W^{+} \right\| \to 0 \quad as \quad \alpha \to 0, T \to \infty and \alpha^3 T \to \infty$$

■

A.2 Proof of Theorem 2

Proof.

Now, we can continue with the proof of the Theorem 2. We will begin by showing that the optimal weighting matrix is equal to the inverse of covariance matrix or IRFs. Then we will show that the regularized estimator, which uses the regularized inverse of the covariance matrix, is consistent.

The estimator of θ_0, $\hat{\theta}$ is given by:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left\langle \hat{K}\hat{f}, \hat{K}\hat{f} \right\rangle$$

where $\hat{f} = \varphi(\hat{m}) - \psi(\theta)$. So, by definition:

$$\left\langle \hat{K}\frac{\partial \hat{f}}{\partial \theta'}(\hat{\theta}), \hat{K}\hat{f}(\hat{\theta}) \right\rangle = 0 \tag{5}$$

A mean value expansion of $\hat{f}(\hat{\theta})$ about θ_0 gives:

$$\hat{f}(\hat{\theta}) = \hat{f}(\theta_0) + \frac{\partial \hat{f}}{\partial \theta'}(\bar{\theta})(\hat{\theta} - \theta_0) \tag{6}$$

where $\bar{\theta}$ is on the line segment joining $\hat{\theta}$ and θ_0. If we replace Equation (6) in Equation (5):

$$\left\langle \hat{K}\frac{\partial \hat{f}(\hat{\theta})}{\partial \theta'}, \hat{K}\hat{f}(\theta_0) + \frac{\partial \hat{f}}{\partial \theta'}(\bar{\theta})(\hat{\theta} - \theta_0) \right\rangle = 0$$

Then we get:

$$(\hat{\theta} - \theta_0) = -\left\langle \hat{K}\frac{\partial \hat{f}(\bar{\theta})}{\partial \theta'}, \hat{K}\frac{\partial \hat{f}(\hat{\theta})}{\partial \theta'} \right\rangle^{-1} \left\langle \hat{K}\frac{\partial \hat{f}(\hat{\theta})}{\partial \theta'}, \hat{K}\hat{f}(\theta_0) \right\rangle \tag{7}$$

Assumption 8 implies the invertibility of the first matrix for T large. Since $\hat{\theta} \xrightarrow{p} \theta_0$ and $\bar{\theta} \xrightarrow{p} \theta_0$, by Slutsky's Theorem and by Assumption 1:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\left\langle \hat{K}\frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \hat{K}\frac{\partial \hat{f}(\theta_0)}{\partial \theta'} \right\rangle^{-1} \left\langle \hat{K}\frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \sqrt{T}\hat{K}\hat{f}(\theta_0) \right\rangle + o_p(1) \tag{8}$$

Let us now analyze the term $\sqrt{T}\hat{K}\hat{f}(\theta_0)$. We know that $\psi(\theta_0) = \varphi(m(X))$. So, in fact, we need to check the distribution of $\sqrt{T}(\varphi(\hat{m}) - \varphi(m(X))$.

$$\sqrt{T}(\varphi(\hat{m}) - \varphi(m(X))) \sim \mathcal{N}\left(0, \frac{\partial \varphi}{\partial m'} V(m) \left(\frac{\partial \varphi}{\partial m'}\right)^*\right)$$

. Let W be the operator such that:

$$W : \mathcal{E} \mapsto \mathcal{E}, W(z) = \left[\frac{\partial a}{\partial m'} V(m) \left(\frac{\partial a}{\partial m'}\right)^*\right](z)$$

Finally we get the asymptotic distribution of $\hat{\theta}$:

$$\sqrt{T}(\hat{\theta} - \theta_0)$$

$$\sim \mathcal{N}\left(0, \left\langle K^*K\frac{\partial f}{\partial \theta'}(\theta_0), \frac{\partial f}{\partial \theta'}(\theta_0)\right\rangle^{-1} \left\langle K^*K\frac{\partial f}{\partial \theta'}(\theta_0), WK^*K\frac{\partial f}{\partial \theta'}(\theta_0)\right\rangle \left\langle K^*K\frac{\partial f}{\partial \theta'}(\theta_0), \frac{\partial f}{\partial \theta'}(\theta_0)\right\rangle^{-1}\right)$$

We get the optimal estimator if $WK^*K = I$ so, $K^*K = W^{-1}$. We already mentioned that W has zero eigenvalues and its inverse does not exist, instead we can use the generalized inverse, which will give a consistent estimator.

Let θ^α denote the regularized estimator and θ^+ denote the estimator obtained with generalized inverse. To show the consistency of the regularized estimator we need to show that $\theta^\alpha \xrightarrow{p} \theta^+$ since:

$$lim_{T\to\infty}\sqrt{T}(\theta^\alpha - \theta_0) = \sqrt{T}(\theta^\alpha - \theta^+) + \sqrt{T}(\theta^+ - \theta_0)$$

Let us write:

$$\sqrt{T}(\theta^\alpha - \theta^+) = \sqrt{T}(\theta^\alpha - \theta_0) - \sqrt{T}(\theta^+ - \theta_0)$$

Moreover, using Equation (8):

$$\sqrt{T}(\theta^\alpha - \theta^+) = \left\langle \hat{W}^\alpha \frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \frac{\partial \hat{f}(\theta_0)}{\partial \theta'}\right\rangle^{-1} \left\langle \hat{W}^\alpha \frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \sqrt{T}\hat{f}(\theta_0)\right\rangle$$

$$- \left\langle (W^+)\frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \frac{\partial \hat{f}(\theta_0)}{\partial \theta'}\right\rangle^{-1} \left\langle (W^+)\frac{\partial \hat{f}(\theta_0)}{\partial \theta'}, \sqrt{T}\hat{f}(\theta_0)\right\rangle$$

where $\hat{W}^\alpha = (\alpha I + \hat{W}^*\hat{W})^{-1}\hat{W}^*$, is the regularized inverse and W^+ is the generalized inverse of the operator W. To simplify exposition, let us denote $\frac{\partial \hat{f}(\theta_0)}{\partial \theta'}$ by f' and $\hat{f}(\theta_0)$ by f.

$$\sqrt{T}(\theta^\alpha - \theta^+) = \left(\left\langle \hat{W}^\alpha f', f'\right\rangle^{-1} \left\langle \hat{W}^\alpha f', \sqrt{T}f\right\rangle - \langle W^+ f', f'\rangle^{-1} \left\langle W^+ f', \sqrt{T}f\right\rangle\right)$$

$$= \left\langle \hat{W}^\alpha f', f' \right\rangle^{-1} \left\langle \hat{W}^\alpha f', \sqrt{T} f \right\rangle - \langle W^+ f', f' \rangle^{-1} \left\langle W_\alpha^{-1} f', \sqrt{T} f \right\rangle$$
$$+ \langle W^+ f', f' \rangle^{-1} \left\langle \hat{W}^\alpha f', \sqrt{T} f \right\rangle - \langle W^+ f', f' \rangle^{-1} \left\langle W^+ f', \sqrt{T} f \right\rangle$$
$$= \left(\underbrace{\left(\left\langle \hat{W}^\alpha f', f' \right\rangle^{-1} - \langle W^+ f', f' \rangle^{-1} \right) \left\langle \hat{W}^\alpha f', \sqrt{T} f \right\rangle}_{I} + \underbrace{\langle W^+ f', f' \rangle^{-1} \left\langle (\hat{W}^\alpha - W^+) f', \sqrt{T} f \right\rangle}_{II} \right)$$

Let us check the first term:

$$I = \left\langle \hat{W}^\alpha f', f' \right\rangle^{-1} \left[\left\langle \hat{W}^\alpha f', f' \right\rangle - \langle W^+ f', f' \rangle \right] \langle W^+ f', f' \rangle^{-1} \left\langle \hat{W}^\alpha f', f \right\rangle$$

$\left\langle \hat{W}^\alpha f', f' \right\rangle^{-1}$ and $\langle W^+ f', f' \rangle^{-1}$ are bounded.

$$\|I\|^2 \leq \left\| (\hat{W}^\alpha - W^+)^{1/2} f' \right\| \left\| \hat{W}^\alpha f' \right\| \left\| \sqrt{T} f \right\|$$

The first term on the RHS converges to zero by Lemma 1 as $\alpha \to 0$, $T \to \infty$ and $\alpha^3 T \to \infty$. The second and the third term together converges to a normal distribution as $T \to \infty$ and $T\alpha \to \infty$. So, we conclude that the I converges to 0. For the second term:

$$\|II\|^2 \leq \left\| (W^+) f' \right\| \left\| (W^+ - \hat{W}^\alpha) f' \right\| \|f\|$$

The first term on the RHS is bounded and the second term is $o_p(1)$ by Lemma 1 as $\alpha \to 0$, $T \to \infty$ and $\alpha^3 T \to \infty$. And, this completes the proof. ■

B Monte Carlo Simulation for Generalized Inverse

Generalized inverse would be an optimal and easy solution to the problem of invertibility of optimal weighting matrix in IRFME. Nonetheless, none of the aforementioned paper in DSGE literature uses generalized inverse. In this section, we present a Monte Carlo Simulation analysis to examine the performance of the estimation with generalized inverse (now on generalized estimator).

We use the same simulation design as in *Section 5.1* and simulate a sample of 500. For each sample, we estimate the model by two methods. The first method, uses the generalized inverse of the covariance matrix of IRF whilst the second one uses the regularized inverse. We estimate the model with $h = 2$, 5 and 7. Results are presented in Table (4.1).

Except the estimation where $h = 2$, regularized estimator performs better than the generalized one. For $h = 2$ the MSE of the two estimators are equal. Moreover, the variance of the generalized estimator increases as the number of included horizons increases. In other words, it gets less stable as the number of zero eigenvalues increases. On the other hand, regularized estimator can deal well with the problem of increasing zero eigenvalues as a result of the penalty, α it impose to the inverse.

Table 1: Simulation Results for Generalized Estimator

	Estimation Method	BIAS	VAR	MSE
h=2	Generalized	0.0026	0.0021	0.0021
	Regularized	0.0021	0.0021	0.0021
h=5	Generalized	0.0029	0.0103	0.0103
	Regularized	0.0019	0.0020	0.0020
h=7	Generalized	0.0130	0.0347	0.0349
	Regularized	0.0036	0.0020	0.0020

C Tables and Figures

Table 2: Simulation Results with $h = 2$

Estimation Method	BIAS	VAR	MSE
$n = 100$			
Regularized	0.0067	0.0093	0.0093
Diagonal	0.0117	0.0097	0.0097
Identity	0.0123	0.0098	0.0098
$n = 200$			
Regularized	0.0030	0.0056	0.0056
Diagonal	0.0030	0.0057	0.0057
Identity	0.0034	0.0057	0.0057
$n = 500$			
Regularized	0.0015	0.0020	0.0020
Diagonal	0.0020	0.0021	0.0021
Identity	0.0021	0.0021	0.0021
$n = 1000$			
Regularized	0.0011	0.0010	0.0010
Diagonal	0.0018	0.0010	0.0010
Identity	0.0019	0.0010	0.0010

Table 3: Simulation Results with $h = 5$

Estimation Method	BIAS	VAR	MSE
$n = 100$			
Regularized	0.0136	0.0098	0.0099
Diagonal	0.0524	0.1898	0.1923
Identity	0.0169	0.0179	0.0180
$n = 200$			
Regularized	0.0057	0.0048	0.0048
Diagonal	0.0483	0.1096	0.1119
Identity	0.0087	0.0092	0.0093
$n = 500$			
Regularized	0.0020	0.0020	0.0020
Diagonal	0.0207	0.0482	0.0486
Identity	0.0048	0.0040	0.0040
$n = 1000$			
Regularized	0.0015	0.0010	0.0010
Diagonal	0.0098	0.0263	0.0264
Identity	0.0027	0.0020	0.0020

Table 4: Simulation Results with $h = 7$

Estimation Method	BIAS	VAR	MSE
$n = 100$			
Regularized	0.0145	0.0108	0.0108
Diagonal	0.0604	0.2223	0.2258
Identity	0.0147	0.0241	0.0242
$n = 200$			
Regularized	0.0098	0.0048	0.0047
Diagonal	0.0432	0.1197	0.1215
Identity	0.0094	0.0110	0.0110
$n = 500$			
Regularized	0.0042	0.0020	0.0020
Diagonal	0.0196	0.0547	0.0550
Identity	0.0042	0.0046	0.0046
$n = 1000$			
Regularized	0.0013	0.0010	0.0010
Diagonal	0.0070	0.0257	0.0257
Identity	0.0020	0.0021	0.0021

Table 5: Optimal α's for different simulations

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
$h = 2$	4.71×10^{-6}	1.09×10^{-5}	1.05×10^{-7}	6.87×10^{-8}
$h = 5$	1.33×10^{-6}	1.60×10^{-7}	1.93×10^{-8}	1.05×10^{-7}
$h = 7$	3.08×10^{-6}	1.33×10^{-6}	1.05×10^{-7}	1.26×10^{-8}

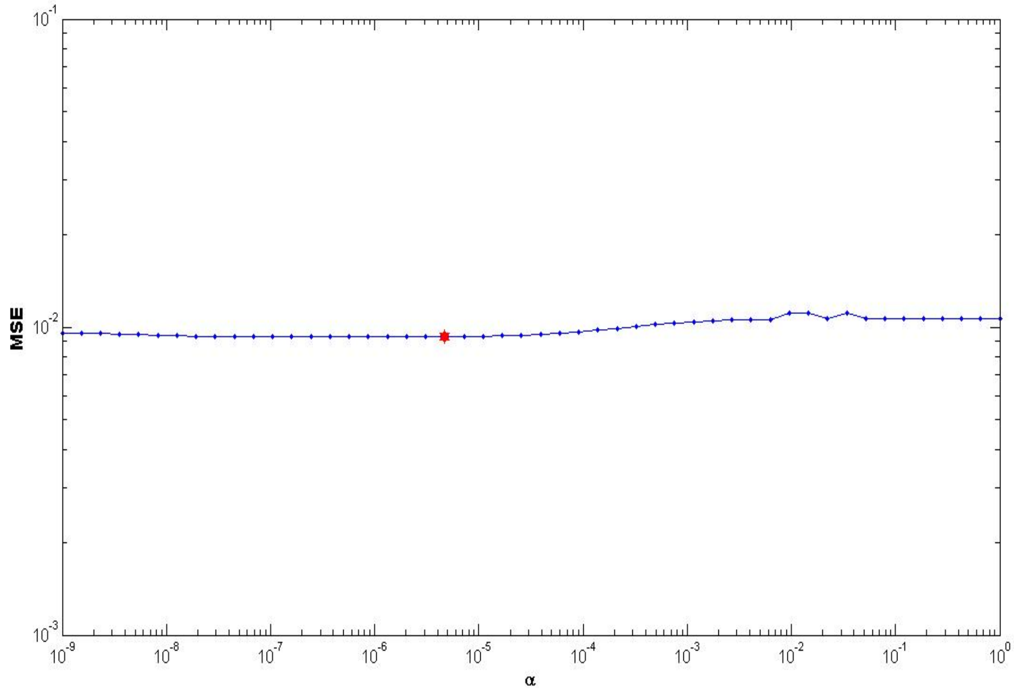

Figure 1: *selection of α for $n = 100$ and $h = 2$*

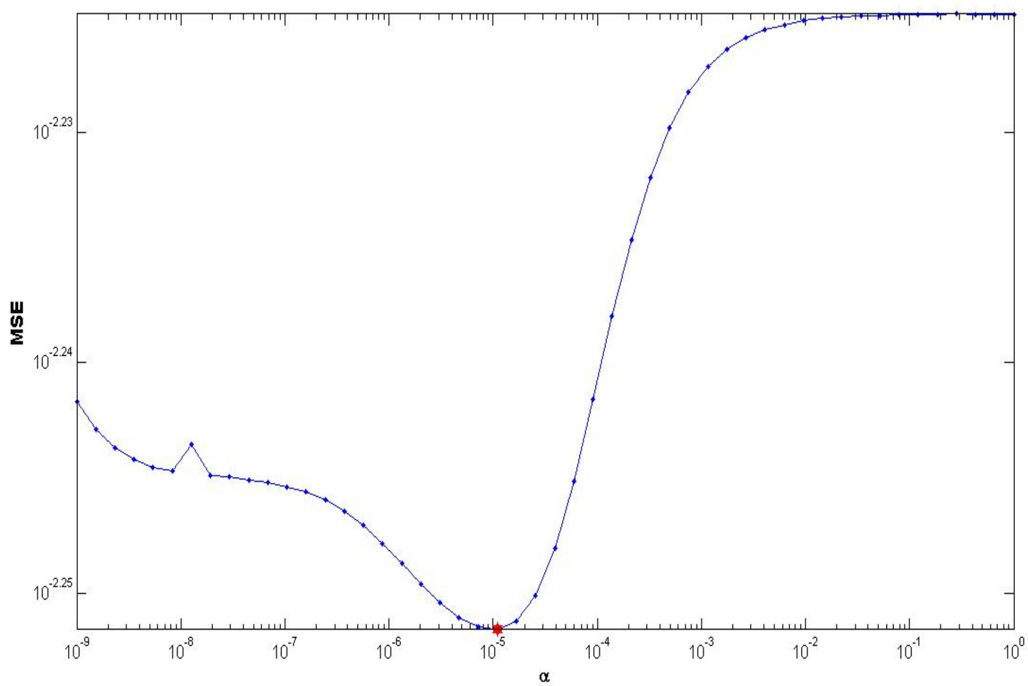

Figure 2: *selection of* α *for* $n = 200$ *and* $h = 2$

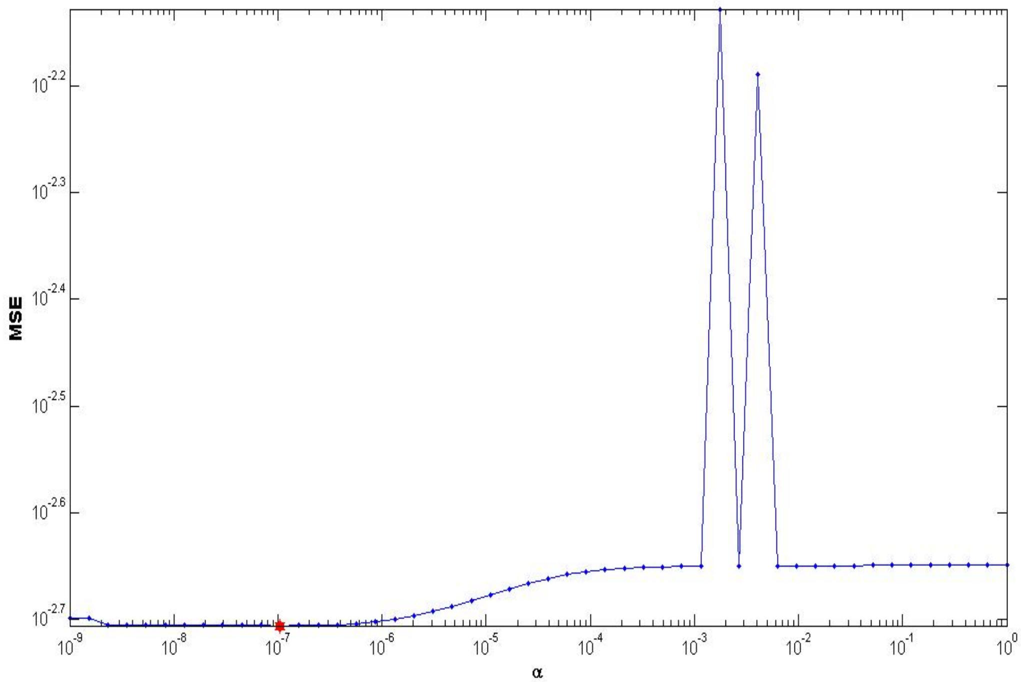

Figure 3: *selection of α for $n = 500$ and $h = 2$*

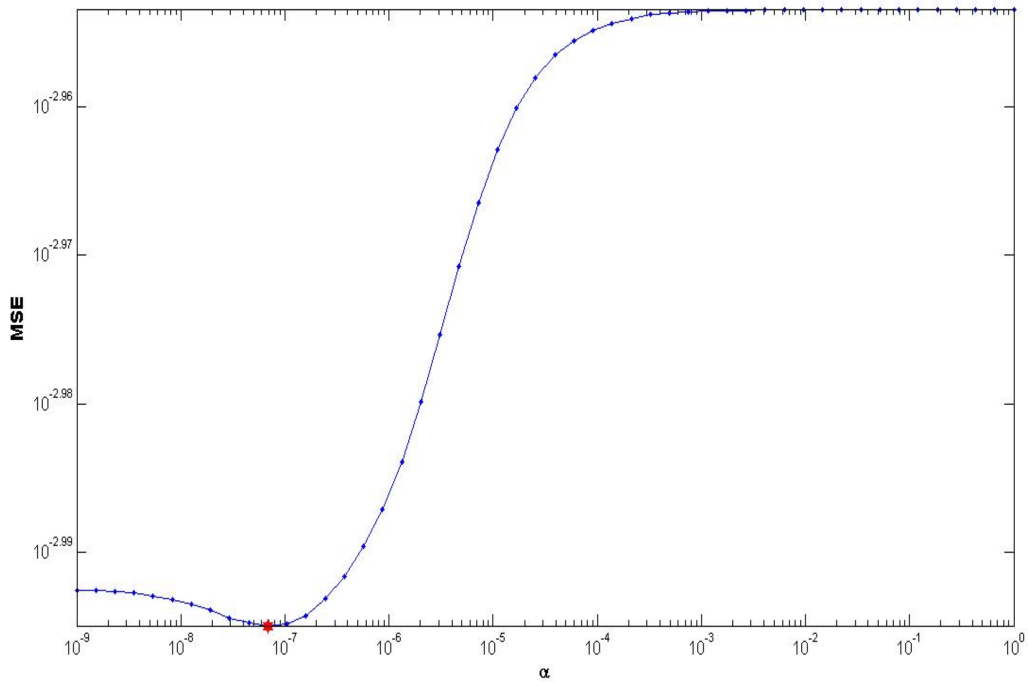

Figure 4: *selection of* α *for* $n = 1000$ *and* $h = 2$

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2004): "Firm-specific capital, nominal rigidities and the business cycle," Discussion paper.

Carrasco, M. (2006): "A Regularization Approach to the Many Instruments Problem," Discussion paper, Universite de Montreal.

Carrasco, M., and J.-P. Florens (2000): "Generalization Of Gmm To A Continuum Of Moment Conditions," *Econometric Theory*, 16(06), 797–834.

——— (2003): "On the Asymptotic Efficiency of GMM," Discussion paper.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 77. Elsevier.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2010): "Nonparametric Instrumental Regression," Discussion paper, Toulouse School of Economics.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117(1), 55–93.

Donald, S. G., and W. K. Newey (2001): "Choosing the number of instruments," *Econometrica*, 69(5), 1161–1191.

Engl, H. W., M. Hanke, and A. Neubauer (1996): *Regularization of Inverse Problems*. Kluwer Academic Publications, Dordrecht.

Feve, P., J. Matheron, and J.-G. Sahuc (2009): "Minimum Distance Estimation and Testing of DSGE Models From Structural VARs," Discussion paper, Toulouse School of Economics.

Florens, J.-P. (2007): "Regularized Non-parametric estimation of game theoretic models," Discussion paper, IDEI.

Giannoni, M., and M. Woodford (2004): “Optimal Inflation-Targeting Rules,” in *The Inflation-Targeting Debate*, NBER Chapters, pp. 93–172. National Bureau of Economic Research, Inc.

Golub, G. H., M. Heath, and G. Wahba (1979): “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” *Technometrics*, 21(2), 215–223.

Gourieroux, C., A. Monfort, and Q. Vuong (1995): *Statistics and Econometric Models*, no. 9780521477451 in Cambridge Books. Cambridge University Press.

Hall, A., A. Inoue, M. J. Nason, and B. Rossi (2007): “Information Criteria for Impulse Response Function Matching Estimation of DSGE Models,” Discussion paper.

Hansen, C., J. Hausman, and W. Newey (2006): “Estimation with many instrumental variables,” CeMMAP working papers CWP19/06, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Rotemberg, J., and M. Woodford (1997): “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy,” in *NBER Macroeconomics Annual 1997, Volume 12*, NBER Chapters, pp. 297–361. National Bureau of Economic Research, Inc.

Ruud, P. A. (2000): *An Introduction to Classical Econometric Theory*. Oxford University Press.