ASSESSING THE FED’S REACTION FUNCTION WITH A
MOMENT CONDITIONS MODEL AVERAGING ESTIMATOR*

LUI S F. MARTINS
Department of Quantitative Methods, ISCTE-LUI, Portugal
Centre for International Macroeconomic Studies (CIMS), UK
(luis.martins@iscte.pt)

VASCO J. GABRIEL
CIMS, University of Surrey, UK and NIPE-UM
(v.gabriel@surrey.ac.uk)

This version: December 2010 (Very preliminary)

Abstract

In this paper, we examine the empirical validity of the baseline forward-looking monetary
policy reaction function proposed by Clarida, Gali, and Gertler (2000). For that purpose,
we propose a moment conditions model averaging estimator in the Generalized Method of
Moments setup. The weights for averaging are chosen to minimize the moment selection
criteria of Andrews (1999). We derive some asymptotic properties under correctly speciﬁed
and misspeciﬁed models. Monte Carlo experiments show that our procedure outperforms
the existing alternatives in the most relevant setups. Both model averaging estimates and
standard procedures point to a stabilizing policy rule during the Paul Volcker and Alan
Greenspan tenures, but not so during the pre-Volcker period. However, our results cast
doubts on the signiﬁcance of the cyclical output variable as a forcing variable in the Fed
funds rate dynamics during the Volcker-Greenspan period.

Keywords: Forward-Looking Monetary Policy Rule; Stabilizing Policy; Generalized Method
of Moments; Model Selection; Model Averaging; Misspeciﬁcation

JEL Classiﬁcation: C22; C52; E43; E52

*This study was initiated while the ﬁrst author was visiting the Economics and Research Department (ERD)
of the Bank of Portugal (BP) for a short-term scholarship. The author acknowledges ﬁnancial support from
the ERD, BP, and thanks its members for useful comments. The analysis, opinions and ﬁndings of this paper
represent the views of the authors, they are not necessarily those of the BP or the Eurosystem.
1 Introduction

The forward-looking monetary policy reaction function proposed by Clarida, Gali, and Gertler (2000, henceforth CGG, see also Clarida, Gali, and Gertler, 1998), has become a fundamental macroeconomic specification in the context of monetary policy analysis. In this model, the central bank forms beliefs about the future state of the economy based on currently available information. The target rate then depends on the expected inflation and output gaps with respect to their equilibrium values. Moreover, the monetary authorities do not immediately set the actual interest rate to its targeted counterpart, but rather adjust it smoothly over time.

In order to estimate the monetary policy using the Federal Funds rate as the instrument of policy making, CGG employ the Generalized Method of Moments (GMM) methodology. Indeed, the presence of forward-looking variables leads naturally to orthogonality conditions that can be explored using this method. As a result, these authors suggest that the Federal Reserve monetary policy during the Paul Volcker and Alan Greenspan period was more stable than during the fifteen or so years prior to Volcker’s appointment. The reasoning for this claim is that the Volcker-Greenspan stance appeared to be much more aggressive in response to changes in expected inflation.

In this study, we re-evaluate the empirical validity of the baseline model discussed by CGG. For that purpose, we propose a new estimation method in the GMM setup, based on a moment conditions model averaging estimator. For completeness, we also employ existing moment and model selection criteria methods.

We do so for several reasons. First, the CGG papers rely on a standard two-step GMM estimator, which may deviate substantially from its small sample distribution - as discussed in Hansen, Heaton and Yaron (1996), for example, and in the two special issues of the *Journal of Business and Economic Statistics* (1996, vol. 14(3) and 2002, vol. 20(4)) dedicated to GMM. Furthermore, the GMM estimation is not invariant to the specification of the moment conditions, which means that the results depend on the normalization adopted for the estimation. An additional drawback is that the results hinge on the weighting matrix used in the estimation. Some of these issues are explored in Jondeau, Le-Bihan and Gallès (2004) in the context of an estimated forward-looking Taylor rule, highlighting the problems with by GMM estimation.

Another reason for re-evaluating the empirical results obtained by CGG concerns the econometric analysis of model selection in the context of moment condition models. In the standard GMM estimation approach, parameter estimates are obtained with a given/fixed list of instruments (in general, a list of moment conditions). On the other hand, for model selection purposes,
one ranks the available moment conditions models according to the particular goal undertaken. Instead of just one, in model selection there are "many" competing models and the data is used to select one of the models under consideration.

Model selection based on information criteria, hypothesis testing or shrinkage-type estimators have been studied in the GMM framework. See, for example, Smith (1992) and Smith and Ramalho (2002) for model testing and Caner (2009) for a LASSO-type GMM estimator. In this paper, we focus on the application of information criteria-based model selection to the CGG model.

Given that the rejection of the $J$-statistic is an indicator that some moment conditions are invalid, Andrews (1999) conceived a GMM information criteria procedure for consistently selecting the correct moment conditions. Andrews and Lu (2001) extend Andrews’ paper to the case of jointly picking the moments and the parameter (model) vector, that is, by imposing zero restrictions on the parameters. As an alternative, Hall, Inoue, Jana and Shin (2007) suggest selecting a model according to the relevant moment selection criterion. The information criteria used for moment selection are based on the entropy of the limiting distribution of the GMM estimator. Also, Donald, Imbens and Newey (2009) suggest a criterion that balances bias and efficiency for conditional moment restriction models.

On the other hand, Hall and Peixe (2003) and Donald and Newey (2001) propose measures to optimally choose instruments. Hall and Peixe (2003) consider the problem of instrument selection based on a combination of the efficiency and non-redundancy conditions and avoiding, therefore, poor approximations of the asymptotic distribution to finite samples by the inclusion of redundant moment restrictions. In Donald and Newey (2001), the selected instruments are such that an approximate mean-square error is minimized over all existing instruments deemed to be valid. Nevertheless, both procedures are intended for model selection in the linear IV context and not for the general GMM framework.

We pursue the alternative methodology in model selection of Model Averaging (MA). In MA, the estimation procedure is based on a number of possible models and forming a weighted average of the resulting estimates. The weights are chosen according to some relevant criterion. Smoothing estimates across several models is a neat strategy to improve the bias and variance balance. Studies on least squares MA (Hansen, 2007), likelihood-based MA (see, for example, Hjort and Claeskens, 2003) and Bayesian MA (see, for example, Hoeting, Madigan, Raftery and Volinsky, 1999) have already been developed.

In Hansen (2007), the weights are estimated through the well-known Mallows criterion. In
an IV framework, Kuersteiner and Okui (2010) extend Donald and Newey (2001) approach to construct optimal instruments with averaging schemes. This method is suited to the 2SLS, LIML and Fuller’s estimators where the averaging approach of Hansen (2007) is applied in the first-step least squares estimation. The MA weights are chosen to minimize the approximate MSE (as in Donald and Newey, 2001). Okui (2008) proposes a shrinkage-type 2SLS estimator for addressing the “many instruments” problem in the context of instrumental variable estimation. According to Kuersteiner and Okui (2010), Okui’s estimator is a special case of their own 2SLS-MA estimator and it can be interpreted in terms of kernel weighted 2SLS. Canay (2010) also proposes a procedure to estimate linear models when the number of instruments is large. The estimator uses a trapezoidal kernel to select and weight the number of moments by shrinking the first-stage coefficient estimators toward zero. Kuersteiner and Okui (2010) claim that kernel methods make the weighting scheme somewhat inflexible once a particular kernel is chosen.

Although MA in the linear IV context has been recently developed, MA in general GMM estimation is still an open issue. We intend to fill this gap in the literature. Thus, we propose GMM model averaging estimators and discuss some of their asymptotic properties under correctly specified and misspecified models. Contrary to the above-mentioned methods, we do not confine the analysis to linear IV models, but consider a general GMM setup; our approach is not two-stepped; and, in our case, the list of candidate models does not depend on ordered instruments from the full-instrument matrix\(^1\). In conclusion, we are able to combine the estimation of general moment conditions models with one-step and information criteria-based MA estimation, as in Hansen’s spirit.

We show that the MA GMM asymptotic theory under misspecification is not standard in the sense that the consistency and distributional results depend on the weighting matrices and the pseudo-true values. The optimal MA weights are found by means of particular moment and model selection criteria as defined in Andrews (1999), Hall and Peixe (2003) and Hall et al. (2007). Monte Carlo experiments show that, in the most relevant setups, our MA estimation procedure outperforms the optimal instrument selection method of Donald and Newey (2001) and the IV model averaging estimation approach of Kuersteiner and Okui (2010), in terms of median bias and absolute deviation, namely in models with possibly weak instruments.

When applied to the US forward-looking Taylor rule, both MA estimates and standard procedures point to a stabilizing policy rule during the Paul Volcker and Alan Greenspan tenures.

\(^1\)In our method, with \(m\) instruments we can consider \(m\) models (each model including an extra instrument), but also any possible combination of these. This makes our approach more general and not conditioned on how to order the instruments.
but not so during the pre-Volker period. However, our results raise serious doubts on the significance of the cyclical output variable as a forcing variable in the Fed funds rate dynamics during the Volcker-Greenspan period. Indeed, CGG found that the parameter associated with output gap was statistically significant for most of the policy rule specifications, which contrasts with the evidence we provide.

In the next section, we briefly review estimation and model selection in moment conditions models, and we present our moment conditions model averaging approach. In Section 3, we discuss the properties of model averaging in misspecified models. The Monte Carlo simulation study is developed in Section 4. The empirical application of the existing methods and the MA procedure to the baseline forward-looking monetary policy rule proposed by CGG can be found in Section 5 and a conclusion finalizes this paper.

2 Moment Conditions Model Averaging

In this section, we review the main results regarding the GMM estimation procedures, including the moment and model selection criteria of Andrews (1999). Then, we present a new MA estimator for moment conditions models whose weights are chosen according to Andrews’ measure. Finally, we compare alternative strategies for selecting weights based on alternative selection criteria.

2.1 GMM Estimation and Model Selection Procedures: A Brief Review

Let $\mathcal{M}$ be the collection of candidate moment conditions models. Here, $\mathcal{M}$ is a countable/finite or an uncountable set, such that model $M_i$ belongs to the family of models $\mathcal{M} : M_i \in \mathcal{M}$. The "true" model may or may not be an element of $\mathcal{M}$. In our model averaging procedure, we specify a subset of $\mathcal{M}$ from which we define the (MA) estimator. For now, take any particular moment conditions model, $M_i$, which is characterized by a particular set of instruments. The data $\{y_t\}$ is assumed to be an infinite sequence of stationary and ergodic (or weakly dependent) variables such that the standard assumptions in the GMM context hold.

In a correctly specified model, the estimation of the unique $p$-dimensional parameter vector $\theta_0 = (\theta_{0,1}, ..., \theta_{0,p}) \in \Theta \subset \mathbb{R}^p$ is based on $m \geq p$ moment conditions of the form

$$E[g(y_t, \theta_0)] = E[g_t(\theta_0)] = 0,$$

for all $t$, where, in standard IV/2SLS/LIML estimation, $g_t(\theta_0) = \varepsilon(x_t, \theta_0) \otimes z_t$ for some set of variables $x_t$ and instruments $z_t$, such that $y_t = (x'_t, z'_t)'$. When $\varepsilon_t$ is univariate, $z_t$ is a vector
and the model is a linear regression, 2

\[ g_t(\theta_0) = z_t (y_t - x_t^t \theta_0) \quad \text{and} \quad E[z_t (y_t - x_t^t \theta_0)] = 0. \tag{2} \]

Due to linearity, \( g_t(\theta) \) is most certainly an unbounded function in the data: \( \sup_y g(y, \theta) = \infty \) for any \( \theta \) and any unit vector \( \iota \).

The \( m \times p \), full-column ranked, Jacobian matrix is defined as

\[ G(\theta_0) \equiv G = E \left( \frac{\partial g(y_t; \theta)}{\partial \theta^t} \bigg|_{\theta=\theta_0} \right) \tag{3} \]

and, under some regularity conditions, a CLT can be invoked: \( \sqrt{T} \left( \frac{1}{T} \sum_{t=1}^{T} g(y_t, \theta_0) \right) \xrightarrow{d} N_m (0, S(\theta_0)) \), as \( T \to \infty \), where the long-run variance of the process \{\( g(y_t, \theta_0) \)\} is some \( m \times m \) positive definite matrix

\[ S(\theta_0) \equiv S = \lim_{T \to \infty} \text{Var} \left[ T^{-1/2} \sum_{t=1}^{T} g(y_t, \theta_0) \right]. \tag{4} \]

In the linear case,

\[ G = E (z_t x_t^t) \quad \text{and} \quad S = \Gamma_0 = E (g_t(\theta_0) g_t(\theta_0)^t) = E (z_t z_t^t \epsilon_t^2), \tag{5} \]

under no-dependence, for example a martingale difference sequence.

**Estimation Procedure** In order to estimate consistently and efficiently the unknown quantity \( \theta_0 \), we discuss the GMM estimator, the typical procedure in moment condition models.

For a sample of size \( T \), define the sample counterparts of the population moments as

\[ \hat{g}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} g(y_t, \theta), \quad \hat{G}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(y_t, \theta)}{\partial \theta^t} \quad \text{and} \]

\[ \hat{S}_T(\theta) = (\text{HAC formula. See Den Haan and Levin, 1996, for example}). \]

The GMM estimator is defined as

\[ \hat{\theta}_{GMM,T}(W) = \arg\min_{\theta \in \Theta} \hat{g}_T(\theta)^t W_T \hat{g}_T(\theta), \tag{7} \]

where \( W_T \) is a weighting matrix such that \( W_T \xrightarrow{p} W \), a positive definite matrix. When \( m > p \), the asymptotic variance of \( \sqrt{T} \left( \hat{\theta}_{GMM,T} - \theta_0 \right) \) depends on the \( p \lim W_T = W \). For the two-step efficient estimator, \( W_T = \hat{S}_T \left( \hat{\theta}_{FS} \right)^{-1} \equiv \hat{S}_T^{-1} \), where \( \hat{\theta}_{FS} \) is a first-step consistent GMM estimator (take \( W_T = I_m \), for example):

\[ \hat{\theta}_{EGMM,T} = \arg\min_{\theta \in \Theta} \hat{g}_T(\theta)^t \hat{S}_T^{-1} \hat{g}_T(\theta), \tag{8} \]

2There should be no confusion in terms of notation: in the general case, \( y_t \) is the set of all variables in the model; in the linear case, \( y_t \) is the dependent variable and \( x_t \) the covariates.
where \( \hat{S}_T \xrightarrow{p} S \), as \( T \to \infty \). The GMM estimator depends on the weighting matrix \( \hat{S}_T^{-1} \) which is influenced by the choice made for a first step consistent estimator.

Solving for \( \theta \) at the FOC of the GMM objective function for the linear model,

\[
\left( \frac{1}{T} \sum_{t=1}^{T} z_t x_t' \right) W_T \left( \frac{1}{T} \sum_{t=1}^{T} z_t (y_t - x_t' \theta) \right) = 0,
\]

we obtain the IV/2SLS estimator

\[
\hat{\theta}_{IV,T} = \left( (X'Z) W_T (Z'X) \right)^{-1} (X'Z) W_T (Z'y).
\]

For the dependent time-series version, \( W_T = \hat{S}_T^{-1} \) evaluated at \( \hat{\theta}_{FS} = \left( (X'Z) (Z'X) \right)^{-1} (X'Z) (Z'y) \), whereas for the homoskedastic version, with an error variance of one and no-dependence cross-section, \( W_T = \left( \frac{1}{T} \sum_{t=1}^{T} z_t x_t' \right)^{-1} = \left( \frac{Z'Z}{T} \right)^{-1} \).

It has been shown that, under some regularity conditions, \( \hat{\theta}_{GMM,T} \xrightarrow{p} \theta_0 \) and

\[
\sqrt{T} \left( \hat{\theta}_{GMM,T} - \theta_0 \right) \xrightarrow{d} N \left( 0, \left( G'WG \right)^{-1} \left( G'WSWG \right) \left( G'WG \right)^{-1} \right),
\]

as \( T \to \infty \). The EGMM estimator is efficient in the sense that it attains the smallest asymptotic variance over the class of GMM estimators with alternative weighting matrices \( W_T \) for a given set of moment conditions. Chamberlain (1987) shows that the EGMM estimator is semiparametrically efficient, that is,

\[
\left( G'WG \right)^{-1} \left( G'WSWG \right) \left( G'WG \right)^{-1} \equiv \left( G' S^{-1} G \right)^{-1}
\]

is the lower bound for the variance of any estimation procedure based solely on the information \( E[g(y_t, \theta_0)] = 0 \) and with unknown distribution.

**Moments and Model Selection Criteria** Due to the well-known bias/variance trade-off in any estimation method, Donald and Newey (2001) discuss how to choose among a list of instruments in a system of linear simultaneous equations, using the 2SLS and LIML instrumental variables estimators. In this setup, one chooses the (optimal) instruments, with the corresponding estimator, such that the estimated mean square error is minimized. In the general, not necessarily linear, GMM literature, the choice of moments is achieved according to some general information criteria instead.

Based upon the evidence that the rejection of the \( J \)-statistic is an indicator that some moment conditions are invalid, Andrews (1999) conceived a GMM procedure for consistently selecting the correct moment conditions. Following his notation, let \( m \) denote the total number
of available moment conditions and let the GMM moment selection criteria for a given model be defined as

\[ MSC_T(c) = J_T(c) - \kappa_T(|c| - p), \]  

where \( c \in \mathbb{R}^m \) is a moment selection vector that represents a list of "selected" moment conditions (a subset of \( g \)), \(|c|\) denotes the cardinality (number) of the "selected" moments \( c \) (here, \(|c| \leq m\)), \( J_T(c) \) is the \( J \)-statistic computed using the "selected" moments \( c, |c| - p \) is the number of overidentifying restrictions and \( \kappa_T = o(T) \) is a sequence that defines the selection criterion (\( \kappa_T = 2 \) for the AIC; \( \kappa_T = \log T \) for the BIC; and \( \kappa_T = Q \log \log T \) for some \( Q > 2 \) for the HQ-type criterion). Defining the unit-simplex set

\[ C = \{ c \in \mathbb{R}^m \setminus \{ 0 \} : c_j = 0 \text{ or } 1, \forall 1 \leq j \leq m, \text{ where } c = (c_1, ..., c_m) \} \]  

\( c \) is a vector of zeros (excluded conditions) and ones (included conditions) and \(|c| = \sum_j c_j \) for \( c \in C \). Accordingly, for a GMM estimator based on the moment conditions \( c, \hat{\theta}_T(c) \),

\[ J_T(c) = \inf_{\theta \in \Theta} \hat{g}_{Tc}(\theta) W_T(c) \hat{g}_{Tc}(\theta) = T \hat{g}_{Tc} \left( \hat{\theta}_T(c) \right) W_T(c) \hat{g}_{Tc} \left( \hat{\theta}_T(c) \right) \]  

where \( W_T(c) \) is the \(|c| \times |c| \) weight matrix employed with the moment conditions \( \hat{g}_{Tc}(\theta) \).

The moment selection criteria estimator is defined as

\[ \hat{c}_{msc} = \arg \min_{c \in \mathcal{C}} MSC_T(c) = \arg \min_{c \in \mathcal{C}} (J_T(c) - \kappa_T(|c| - p)), \]  

where \( \mathcal{C} \subset C \), with \( \{ 0 \} \in \mathcal{C} \), is some parameter space for the moment selection vector. The estimator \( \hat{c}_{msc} \) picks the moment conditions \( c \) over \( \mathcal{C} \) such that the increase in \( J_T(c) \) that typically occurs when moment conditions are added (even if correct) is offset by the "bonus term" \( \kappa_T(|c| - p) \) that rewards selection vectors that utilize more moment conditions. Under some technical conditions, Andrews (1999) shows that \( \hat{c} \) is a consistent estimator of \( c_0 \), assumed to be the single "correct" selection vector. If, additionally, one assumes that \( E(g_{c_0}(\theta)) = 0 \) has a unique solution \( \theta_0 \in \Theta \) (the "true" value of \( \theta \), set at \( c_0 \)), then \( \hat{c} \) consistently estimates both \( c_0 \) and \( \theta_0 \). In Andrews (1999), the selection of correct moments is conditional on correct modeling. Andrews and Lu (2001) extend Andrews’ paper to the case of jointly picking the moments and the parameter (model) vector, that is, imposing zero restrictions on the parameters.

### 2.2 Model Averaging Estimators

In model "averaging" instruments, we are not averaging model estimators, but only over instruments. In that approach, we estimate the model once after obtaining an optimal set of

\[ ^3 \text{The GMM-AIC is not consistent and it has positive probability (even asymptotically) of selecting too few moments.} \]
instruments. The quantity of interest is this best linear combination of instruments. In this section, we present a methodology whereby we average a list of candidate models to obtain a truly averaged one. The optimal weights associated with each estimator are to be chosen according to information criteria. We adopt the measure $MSC_T$ of Andrews (1999) to find the weights, but acknowledge the fact that the criteria need not be unique. In moment conditions models, averaged estimators can be discussed for general $g$ functions, which have the linear IV/2SLS as a special case. Next, we define a model averaging estimator and then we discuss its statistical properties relating to alternative criteria for choosing the weights.

2.2.1 Moment Conditions Model Averaging Estimator

Consider $m$ and $c$ as defined in (14). Quantities such as $\hat{y}_{Tc}(\theta), W_{Tc}$ and $\hat{\theta}_{Tc}$ are obtained after deleting the moments $j$ corresponding to $c_j = 0$. Here, $\hat{y}_{Tc}(\theta)$ is a $|c| \times 1$ vector. Now, let $\omega = (\omega_1, ..., \omega_{|C|})'$ be a weight vector in the unit-simplex in $\Re^{|C|}$, where $|C| = 2^m - \sum_{j=0}^{p-1} \left( \begin{array}{c} m \\ j \end{array} \right) = \sum_{j=p}^{m} \left( \begin{array}{c} m \\ j \end{array} \right)$, with the binomial coefficients $\left( \begin{array}{c} m \\ j \end{array} \right) = \frac{m!}{j!(m-j)!}$ - which reads $m$ choose $j$ is the number of different elements in $C$:

$$H_m = \left\{ \omega \in [0,1]^{|C|}: \sum_{c \in C} \omega_c = 1 \right\}. \quad (17)$$

A model averaging estimator of the unknown $p \times 1$ vector $\theta_0$ is

$$\hat{\theta}_T(\omega) = \sum_{c \in C} \omega_c \hat{\theta}_{Tc}. \quad (18)$$

Clearly, the standard GMM estimation is a special case for which no model averaging occurs: $\omega_{c^*} = 1$ for some $c^*$ and $\omega_{c'} = 0$ for $c' \neq c^*$ and $\hat{\theta}_T(\omega) = \hat{\theta}_{Tc^*}$.

For an arbitrarily given $\tilde{\omega} \in H_m$, we have a MA estimator $\hat{\theta}_T(\tilde{\omega})$. However, in general, $\omega$ is assumed to be unknown and, therefore, needs to be estimated according to some criterion. In this paper, we suggest alternative data-based criteria to optimally find estimated weights $\tilde{\omega}$ with corresponding averaged estimate

$$\hat{\theta}_T(\tilde{\omega}) = \sum_{c \in C} \tilde{\omega}_c \hat{\theta}_{Tc}. \quad (19)$$

We propose the estimation of the weight vector $\omega$ based on existing moment selection criteria, namely the $MSC_T(c)$, evaluated at $\hat{\theta}_T(\omega)$. The empirical $MSC$ selected weight vector is defined

$^4$We need to exclude $\sum_{j=0}^{p-1} \left( \begin{array}{c} m \\ j \end{array} \right)$ from the total of combinations $2^m$, those for which $m < p$. 

9
as

$$\hat{\omega} = \arg \min_{\omega \in H_m} \text{MSC}_{T\omega} (\omega) = \arg \min_{\omega \in H_m} (J_{T\omega} (\omega) - \kappa_T (|\bar{\tau}| - p)),$$  

(20)

where, for a chosen set of moment conditions $\bar{\tau}$ and a given $W_{T\bar{\tau}}$ (usually, $W_{T\bar{\tau}} = \tilde{S}_{T\bar{\tau}}^{-1}$),

$$J_{T\bar{\tau}} (\omega) = T \tilde{g}_{T\bar{\tau}} \left( \hat{\theta}_T (\omega) \right)' W_{T\bar{\tau}} \tilde{g}_{T\bar{\tau}} \left( \hat{\theta}_T (\omega) \right).$$  

(21)

The least squares MA estimator of Hansen (2007) and the two-step MA instruments estimators of Kuersteiner and Okui (2010) have distinct number of parameters to estimate for each individual model. On the contrary, in our case, $p_c = p$ for all $c$ and therefore $\min_{\omega \in H_m} \text{MSC}_{T\omega} (\omega) = \min_{\omega \in H_m} J_{T\bar{\tau}} (\omega)$ for any penalty term $\kappa_T$. That is to say, our MA estimator for unconditional moment conditions models is based on the well-known $J_T (\omega)$-statistic. To achieve maximum efficiency, one can pick $\bar{\tau} = \iota_m$, a vector of ones, which implies using the whole set of moment conditions (in this case, $|\bar{\tau}| = m$ and, in terms of notation, “c” is dropped):

$$J_T (\omega) = T \tilde{g}_T \left( \hat{\theta}_T (\omega) \right)' W_T \tilde{g}_T \left( \hat{\theta}_T (\omega) \right).$$  

(22)

For the linear IV/2SLS case,

$$J_{T\bar{\tau}} (\omega) = T \left( \frac{1}{T} \sum_{t=1}^{T} z_{\tau,t} \left( y_t - x_t' \sum_{c \in C} \omega_c \hat{\theta}_T \right) \right)' W_{T\bar{\tau}} \left( \frac{1}{T} \sum_{t=1}^{T} z_{\tau,t} \left( y_t - x_t' \sum_{c \in C} \omega_c \hat{\theta}_T \right) \right).$$  

(23)

It is worth mentioning that averaging the MSC criterion itself is not a valid method to estimate $\omega$. In this case, the empirical selected weight vector is defined as

$$\hat{\omega} = \arg \min_{\omega \in H_m} \text{AMSC}_T (\omega) = \arg \min_{\omega \in H_m} \sum_{c \in C} \omega_c \text{MSC}_{Tc},$$  

(24)

where $\text{MSC}_{Tc} = J_{Tc} - \kappa_T (|c| - p)$ and $J_{Tc} = T \tilde{g}_{Tc} \left( \hat{\theta}_{Tc} \right)' W_{Tc} \tilde{g}_{Tc} \left( \hat{\theta}_{Tc} \right)$. Notice that, in general, $J_T (\omega) \neq \sum_{c \in C} \omega_c J_{Tc}$. The AMSC criterion is linear in $\omega$ and, therefore, no weight is given other than to the selected (smallest) MSC model. Hence, the MA GMM estimator would coincide with the estimator for the model selected by means of the MSC: $\omega_{\hat{c}} = 1$ and $\omega_c = 0$, otherwise (for all $c \neq \hat{c}$).

The $\text{MSC}_T (\omega)$ criterion is a (normalized) weighted squared loss for correctly specified models, up to the constant $\kappa_T (m - p)$. The loss function is $L_T (\omega) = \frac{1}{T} J_T (\omega)$ - see (22)- which can be decomposed as $\left( \tilde{g}_T \left( \hat{\theta}_T (\omega) \right) - 0 \right)' W_T \left( \tilde{g}_T \left( \hat{\theta}_T (\omega) \right) - 0 \right)$, where the vector $\tilde{g}_T \left( \hat{\theta}_T (\omega) \right)$ is an estimator of $E [g_t (\theta_0)] = 0$ and for which the distance is weighted by $W_T$. The risk function is

$$R_T (\omega) = E (L_T (\omega)) = E \left( \tilde{g}_T \left( \hat{\theta}_T (\omega) \right)' W_T \left( \tilde{g}_T \left( \hat{\theta}_T (\omega) \right) \right) \right).$$  

(25)
Therefore, for a fixed \( \omega \) and \( W_T \), the quantity \( \frac{1}{T} MSC_T (\omega) \) is an unbiased estimator of the risk function up to a \( o(1) \) constant, that is,

\[
E \left( \frac{1}{T} MSC_T (\omega) \right) = E \left( \frac{1}{T} J_T (\omega) \right) + \frac{1}{T} \kappa_T (m - p) = R_T (\omega) + o(1).
\]  

The solution \( \hat{\omega} \) is found by numerical algorithms. It solves a constraint optimization problem in which the constraints are nonnegativity \( (\omega_c \in [0, 1], \text{for all } c) \) and a summation that equals one \( (\sum_{c \in C} \omega_c = 1) \). The (asymptotic) distribution of \( \hat{\omega} \) is beyond the scope of this paper. This is not an easy task, despite its relevance for specific inferences such as a null of \( \omega_c = 0, c \in C \) (note that this is not even known for the case of least squares MA estimators, see Hansen, 2007, for details).

### 2.2.2 Properties of the MA Estimator and Alternative Measures

For a given \( \omega \), the limit statistical properties of \( \hat{\theta}_T (\omega) = \sum_{c \in C} \omega_c \hat{\theta}_{Tc} \) follows a linear combination of the random processes \( \hat{\theta}_{Tc}, c \in C \). Under correct model specification, \( \text{plim} \hat{\theta}_{Tc} = \theta_0 \) for all \( c \in C \) and \( \hat{\theta}_{Tc} \) is \( \sqrt{T} \)-gaussian with asymptotic variance

\[
V_c = \left( G'_c W_c G_c \right)^{-1} \left( G'_c W_c S_c W_c G_c \right) \left( G'_c W_c G_c \right)^{-1}.
\]

The asymptotic variance of the efficient GMM estimator is given by

\[
V_c = \left( G'_c S^{-1}_c G_c \right)^{-1}.
\]

Hence, for a fixed \( \omega, \hat{\theta}_T (\omega) \) is also consistent and also \( \sqrt{T} \)-normal.

**Theorem 1** (Distribution of the MA estimator under correct specification): Assume that the model is correctly specified. As \( T \to \infty \), for any \( \omega \in H_m \),

\[
\hat{\theta}_T (\omega) = \sum_{c \in C} \omega_c \hat{\theta}_{Tc} \xrightarrow{p} \theta_0,
\]

where \( \hat{\theta}_{Tc}, c \in C \), is the GMM estimator. Moreover,

\[
\sqrt{T} \left( \hat{\theta}_T (\omega) - \theta_0 \right) \xrightarrow{d} \eta = \sum_{c \in C} \omega_c \eta_c,
\]

where the \( k \times 1 \) random variable \( \eta_c \sim N(0, V_c) \), with \( V_c = (27) \) or \( V_c = (28) \). The limit process \( \eta \) is gaussian with zero expectation.

**Proof:** Consistency follows from

\[
\hat{\theta}_T (\omega) = \sum_{c \in C} \omega_c \hat{\theta}_{Tc} \xrightarrow{p} \sum_{c \in C} \omega_c \theta_0 = \theta_0.
\]
The asymptotic distribution follows from the limiting law for \( \sqrt{T} \left( \hat{\theta}_{T_c} - \theta_0 \right) \) noting that \( \sqrt{T} \left( \hat{\theta}_{T} (\omega) - \theta_0 \right) \) equals
\[
\sqrt{T} \left( \sum_{c \in C} \omega_c \hat{\theta}_{T_c} - \sum_{c \in C} \omega_c \theta_0 \right) = \sum_{c \in C} \omega_c \sqrt{T} \left( \hat{\theta}_{T_c} - \theta_0 \right)
\]
(32)
because \( \sum_{c \in C} \omega_c = 1 \). Alternatively, the result can be shown by taking the FOC for a given \( c \in C, \hat{G}_{T_c} \left( \hat{\theta}_{T_c} \right)^T W_{T_c} \hat{g}_{T_c} \left( \hat{\theta}_{T_c} \right) = 0 \), and expand \( \hat{g}_{T_c} \left( \hat{\theta}_{T_c} \right) \) around \( \hat{g}_{T_c} (\theta_0) \) using the Mean Value Theorem:
\[
\hat{G}_{T_c} \left( \hat{\theta}_{T_c} \right)^T W_{T_c} \hat{g}_{T_c} (\theta_0) + \hat{G}_{T_c} \left( \hat{\theta}_{T_c} \right)^T W_{T_c} \hat{G}_{T_c} \left( \bar{\theta}_{T_c} \right) \left( \bar{\theta}_{T_c} - \theta_0 \right) = 0,
\]
(33)
where \( \bar{\theta}_{T_c} \) is some value "between" \( \hat{\theta}_{T_c} \) and \( \theta_0 \). Rearranging terms,
\[
\sqrt{T} \left( \hat{\theta}_{T_c} - \theta_0 \right) = - \left[ \hat{G}_{T_c} \left( \hat{\theta}_{T_c} \right)^T W_{T_c} \hat{G}_{T_c} \left( \bar{\theta}_{T_c} \right) \right]^{-1} \hat{G}_{T_c} \left( \hat{\theta}_{T_c} \right)^T W_{T_c} \sqrt{T} \hat{g}_{T_c} (\theta_0).
\]
(34)
Finally, \( \eta \) is gaussian because it is a linear combination of normal variables and it has zero expectation because \( E(\eta_c) = 0, c \in C \). QED.

In general, the variance of \( \eta \) does not equal the (squared) weighted sum of variances \( V_c \), that is,
\[
V(\eta) = V \left( \sum_{c \in C} \omega_c \eta_c \right) = \sum_{c \in C} \omega_c^2 V(\eta_c) + \sum_{c_1, c_2 \in C} \omega_{c_1} \omega_{c_2} E(\eta_{c_1} \eta_{c_2}') \neq \sum_{c \in C} \omega_c^2 V(\eta_c),
\]
(35)
since there are pairs \( c_1, c_2 \in C, c_1 \neq c_2 \), such that \( \eta_{c_1} \) and \( \eta_{c_2} \) are not independent, in particular when \( c_1 \) and \( c_2 \) have common moment conditions. Consequently, in terms of the smallest \( V(\eta) \), it is not clear that this is attained for \( \omega_c^* = 1 \), where \( c^* = \mu_m \), and \( \omega_{c^*} = 0 \), for all \( c' \neq c^* \) (the MA estimator coincides with the estimator obtained using the whole set of moment conditions), due to the existence of the covariance components \( E(\eta_{c_1} \eta_{c_2}') \). In practice, finding the argument \( \hat{\omega} \) that minimizes the trace of \( V(\eta) \),
\[
tr \left( V(\eta) \right) = \sum_{c \in C} \omega_c^2 tr \left( V(\eta_c) \right) + \sum_{c_1, c_2 \in C} \omega_{c_1} \omega_{c_2} tr \left( E(\eta_{c_1} \eta_{c_2}') \right),
\]
(36)
can be hard to accomplish because a consistent estimator for \( E(\eta_{c_1} \eta_{c_2}') \) is not trivial to find. In fact, larger weights given to models \( c \) with more moment conditions (for efficiency matters) can be offset by the covariances \( E(\eta_{c_1} \eta_{c_2}') \).

In summary, despite knowing the asymptotic law of the MA estimator, uniformly in \( \omega \), the variance criterion to choose the optimal vector \( \hat{\omega} \) might not be that helpful in practice. Fortunately, there are alternative moment and model selection criteria that are related to \( V(\eta_c) = V_c \)
which can be considered as valid alternatives to the measure $MSC_T(\omega)$ for determining $\hat{\omega}$, namely the relevant moment selection criterion $RMSC_T(c)$ of Hall et al. (2007) and the canonical correlations information criterion $CCIC_T(c)$ of Hall and Peixe (2003). Eryuruk et al. (2009a) show the close relationship between $RMSC_T(c)$ and $CCIC_T(c)$ in the normal linear simultaneous equations model.

Contrary to Andrews (1999), Hall et al. (2007) suggest selecting a model according to the relevant moment selection criterion

$$RMSC_T(c) = \ln \left( \left[ \hat{V}_c \right] \right) + \kappa_T (|c| - p),$$

where the efficient GMM variance-covariance matrix $\hat{V}_c$ is evaluated at $\hat{\theta}_{Tc}$, whereas Hall and Peixe (2003) consider the problem of instrument selection based on a combination of the efficiency and non-redundancy conditions

$$CCIC_T(c) = T \sum_{i=1}^{p} \ln \left( 1 - r^2_{i,T}(c) \right) + \kappa_T (|c| - p),$$

where $r_{i,T}(c)$ is the $i^{th}$ sample canonical correlation between $d_t(\tilde{\theta}_T)$ and $z_t(c)$, with $d_t(\theta) = \frac{\partial m(\theta)}{\partial \theta}$ and $\tilde{\theta}_T$ is a $\sqrt{T}$-consistent preliminary estimator. Here, $u_t(\theta)$ is scalar and, if the model is linear, $d_t(\theta) = -x_t$. Accordingly, weight estimation in unconditional moment conditions models could also be pursued through

$$\hat{\omega} = \mathop{\arg\min}_{\omega \in H_m} RMSC_T(\omega) = \mathop{\arg\min}_{\omega \in H_m} \left[ \ln \left( \omega^\prime \text{diag} \left( \left[ \hat{V}_1 \right], ... , \left[ \hat{V}_{|c|} \right] \right) \omega \right) \right]$$

$$= \mathop{\arg\min}_{\omega \in H_m} \left[ \omega^\prime \text{diag} \left( \left[ \hat{V}_1 \right], ... , \left[ \hat{V}_{|c|} \right] \right) \omega \right]$$

and/or

$$\hat{\omega} = \mathop{\arg\min}_{\omega \in H_m} CCIC_T(c)(\omega) = \mathop{\arg\min}_{\omega \in H_m} \left[ \omega^\prime \text{diag} \left( \sum_{i=1}^{p} \ln \left( 1 - r^2_{i,T,1} \right), ... , \sum_{i=1}^{p} \ln \left( 1 - r^2_{i,T,|c|} \right) \right) \omega \right],$$

where $\text{diag} (\cdot)$ refers to a $|C| \times |C|$ diagonal matrix. Nevertheless, this two alternative measures have an important caveat. In theory, $\hat{\theta}_T(\hat{\omega}) = \hat{\theta}_{TC1}$, where $c_1 = t_m$ (model with the largest number of instruments/conditions), since $\hat{\omega}_{c_1} = 1$ and $\hat{\omega}_{c_j} = 0, j \neq 1$. This result follows from the fact that the objective functions are of the type $\omega^\prime A \omega$, for $A$ diagonal, positive definite and $A \neq 0$. In practice, numerical algorithms may give a solution $\hat{\omega}$ different from the theoretical one whenever parameter estimation is almost indistinguishable across models with different moment conditions (and, consequently, with very similar $\left[ \hat{V}_j \right]$ and $\sum_{i=1}^{p} \ln \left( 1 - r^2_{i,T,j} \right)$ for $j = 1, ... , |C|$).
3 Model Averaging and Misspecification

The procedures described so far assume that the selected moment conditions, and the correct model itself, is correctly specified. However, when the number of instruments is large, it is possible that no value of the parameter vector simultaneously satisfies all the moment restrictions exactly in the population, resulting in a misspecified model. Another reason for considering misspecification stems from the fact that most models are only approximations to the underlying phenomena. In this section, we discuss model averaging estimation under moment restrictions model misspecification.\footnote{Building optimal instruments in misspecified models can be found at Martins (2009). In their optimality criteria, the resulting estimator ought to be consistent and, whenever possible, attain the Chamberlain efficiency bound relative to the set $z_t$.} We begin by reviewing some established results for GMM under misspecification.

According to Hall and Inoue (2003) and Schennach (2007), a model is said to be misspecified if there is no value of $\theta$ which satisfies the orthogonality condition, that is,

$$E[g(y_t, \theta)] = \mu(\theta)$$ \hfill (41)

where $\mu : \Theta \to \mathbb{R}^m$ such that $\|\mu(\theta)\| > 0$ for all $\theta \in \Theta$ (alternatively, $\inf_{\theta} \|E[g(y_t, \theta)]\| > 0$).\footnote{These authors do not treat the issue of model selection under model misspecification. Similar definitions of misspecification can be found at Maasoumi and Phillips (1982) and Chen, Hong and Shum (2007), among others. In local misspecification, $E[g(y_t, \theta_0)] = T^{-1/2} \mu, \mu \neq 0$, say.} In the previous definition, $E[g(y_t, \theta)]$ is assumed to be constant for all $t$ (it rules out misspecification due to structural instability) and $m > p$ because, otherwise, if $m = p$ then there must exist some value of $\theta$ such that $E[g(y_t, \theta)] = 0$. In order for an extremum estimator to have well defined probability limit in a misspecified model, we need to impose the identification condition: there exists a pseudo-true value $\theta_\star \in \Theta$ such that $Q_0(\theta_\star) < Q_0(\theta), \forall \theta \in \Theta \setminus \{\theta_\star\}$, where $Q_0(\theta)$ is the population objective function, that is,

$$\theta_\star = \arg \min_{\theta} Q_0(\theta).$$ \hfill (42)

Notice that two different estimators may converge to different pseudo-values (due to two different well-defined objective functions). Following Hall and Inoue (2003), given the existence of $\theta_\star$, we
define the following quantities:

\[ \mu_* \equiv \mu(\theta_*) = E[g(y_t, \theta_*)], \]

\[ G_* = E \left( \frac{\partial g(y_t; \theta)}{\partial \theta'} \right) \theta = \theta_* \) (in the linear model, \( G_* = G \)), and

\[ S_* = \lim_{T \to \infty} \text{Var} \left[ T^{-1/2} \sum_{t=1}^{T} (g(y_t, \theta_* - \mu_*) \right] \) (positive definite).

Under some conditions given in Hall and Inoue (2003),

\[ \sqrt{T} \left( \frac{1}{T} \sum_{t=1}^{T} (g(y_t, \theta_* - \mu_*)) \right) \xrightarrow{d} N_m(0, S_*), \text{ as } T \to \infty. \] (44)

The limiting distribution theory of the GMM estimator under misspecification was derived by Hall and Inoue (2003). The combination of overidentification and misspecification leads to a GMM estimator whose \( \rho \lim \) depends on the limit of the weighting matrix and whose limiting distribution depends on the limiting distribution of the elements of the weighting matrix (its rate of convergence included). As a consequence, there is no one single limiting distribution theory for the GMM estimator.

The GMM estimator \( \hat{\theta}_T \) is consistent for the pseudo-true value

\[ \theta_* (W) = \arg \min_{\theta} E [g(y_t, \theta)]' W E [g(y_t, \theta)], \] (45)

where

\[ \hat{Q}(\theta) = \frac{1}{T} J_T(\theta) = \hat{g}_T(\theta)' W_T \hat{g}_T(\theta) \xrightarrow{p} E [g(y_t, \theta)]' W E [g(y_t, \theta)] = Q_0(\theta) \] (46)

uniformly in \( \theta \), if \( W_T \xrightarrow{p} W \). Hall and Inoue (2003) consider four cases, each with its own specific limiting distribution. Whenever \( W_T = W \) for all \( T \) or \( \sqrt{T} (W_T - W) \) is asymptotically normal, \( \sqrt{T} (\hat{\theta}_T - \theta_*) \) converges in distribution to a normal process with zero expectation and a variance that depends on several quantities and is distinct from the case of a correctly specified model.

The first case includes the FS estimation, \( W_T = I_m \), and the second case includes another FS estimator, \( W_T = (\frac{1}{T} \sum_{t=1}^{T} z_t z_t')^{-1} \xrightarrow{p} E (z_t z_t') = W \), and a second step estimator based on the assumption that \( \{z_t (y_t - x_t' \theta_* - \mu_*) \) is a martingale difference sequence. When \( W_T \) is the inverse of a centred or uncentred HAC estimator, the distribution can be degenerated, under some specific conditions. See Hall and Inoue (2003) for details.

Now, we derive the asymptotic properties of the MA GMM estimator \( \hat{\theta}_T (\omega) = \sum_{c \in C} \omega_c \hat{\theta}_{Tc} \), where, given the previous lines, each individual estimator \( \hat{\theta}_{Tc} \) has properties that depend on each specific model \( c \). For model \( c^* = \iota_m \) (full set of restrictions), we assume misspecification \( E[g_{c^*}(y_t, \theta)] = \mu_{c^*}(\theta), \) although there might exist some other model \( \tilde{c} \neq c^* \) such that
\[ E[g_c(y_t, \theta_0)] = 0 \] for some \( \theta_0 \in \Theta \subset \mathbb{R}^p \). That is, adding conditions may increase (asymptotic) efficiency, but will potentially create model misspecification and bias. Moreover, \( \hat{\theta}_{Tc} \) is consistent for the pseudo-true value \( \theta_{sc}(W_c) = \arg \min_{\theta} E[g_c(y_t, \theta)]W_cE[g_c(y_t, \theta)]. \)

Clearly, for a fixed \( \omega \) and as \( T \to \infty \),

\[
p \lim \hat{\theta}_T(\omega) = \sum_{c \in C} \omega_c p \lim \hat{\theta}_{Tc} = \sum_{c \in C} \omega_c \theta_{sc}(W_c) \equiv \sum_{c \in C} \omega_c \theta_{sc}.
\]

The limit in probability of \( \hat{\theta}_T(\omega) \) is a linear combination of the pseudo-values \( \theta_{sc} \), each depending on \( W_c \), meaning that the \( p \lim \hat{\theta}_T(\omega) \) depends on the choice for \( W_{Tc}, c \in C \).

The rate of convergence at which \( \hat{\theta}_T(\omega) \) converges to \( \sum_{c \in C} \omega_c \theta_{sc} \) is influenced by the way \( W_{Tc} \) converges to \( W_c \), for each given model \( c \in C \). In some cases, \( \hat{\theta}_T(\omega) \) is \( \sqrt{T} \)-gaussian but it might happen that the distribution collapses or diverges in the limit (recall the four cases of convergence of the GMM estimator under model misspecification.) Due to the equalities

\[
a_T \left( \hat{\theta}_T(\omega) - \sum_{c \in C} \omega_c \theta_{sc} \right) = \sum_{c \in C} \omega_c a_T \left( \hat{\theta}_{Tc} - \theta_{sc} \right) = \sum_{c \in C} \omega_c \frac{a_T}{a_{Tc}} a_{Tc} \left( \hat{\theta}_{Tc} - \theta_{sc} \right),
\]

where the rate of convergence of the MA GMM estimator is \( a_T \to \infty \), as \( T \to \infty \), the MA estimator \( \hat{\theta}_T(\omega) \) is \( \sqrt{T} \)-gaussian whenever \( W_{Tc} \) is chosen in a way that \( a_T = a_{Tc} = \sqrt{T} \) and \( \sqrt{T} \left( \hat{\theta}_{Tc} - \theta_{sc} \right) \) is gaussian for at least one \( c \in C \) and when \( a_{Tc} \left( \hat{\theta}_{Tc} - \theta_{sc} \right) = O_p(1) \) and \( \frac{\sqrt{T}}{a_{Tc}} = o(1) \) for the remaining models \( c \). This covers the case of all models \( c \in C \) having \( \sqrt{T} \left( \hat{\theta}_{Tc} - \theta_{sc} \right) \) asymptotic normality. Assuming that \( a_{Tc} \left( \hat{\theta}_{Tc} - \theta_{sc} \right) = O_p(1) \) for some \( a_{Tc} \), for all \( c \), the distribution of the MA GMM estimator collapses or diverges in the limit depending on the orders of magnitude \( \frac{a_T}{a_{Tc}}, c \in C \).

To simplify the analysis we consider a "local" specification, in the spirit of White (1982) for the MLE. Suppose that the pseudo-true value is indexed by the sample size through \( W_{cT} \):

\[
\theta_{sc,T}(W_{cT}) \equiv \theta_{sc,T} = \arg \min_{\theta} E[g_c(y_t, \theta)]W_{cT}E[g_c(y_t, \theta)].
\]

With \( W_{Tc} \overset{p}{\to} W_c \), we have \( \theta_{sc,T} = \theta_{sc} + o_p(1) \). The "local" characteristic of this setup is a result of the following assumption.

**Assumption 1 (Local model misspecification):** Assume that, for all \( c \in C \), the function \( \mu_c : \Theta \to \mathbb{R}^m \) is such that

\[
\sqrt{T} \mu_c(\theta_{sc,T}) \to 0, \text{ as } T \to \infty.
\]

The previous assumption\(^1\) keeps the key properties of \( \mu_c \) of time-invariance and \( \| \mu_c(\theta) \| > 0 \) for all \( \theta \in \Theta \). Now, we add the "local" condition that the sequence \( \theta_{sc,T} \) is such that \( \mu_c(\theta_{sc,T}) =\)

\(^1\)Local misspecification is usually defined as \( S^{-1/2} E_T[g(y_t, \theta_0)] = T^{-1/2} \mu \), where \( \mu \) is a vector of finite constants, see Newey (1985) or Hall (2005) for details.
which means that there exists a sequence $\theta_{sc,T}$ responsible for a model misspecification that disappears at a rate that is faster than $\sqrt{T}$. As expected, by imposing "local" misspecification, the MA GMM estimator is now gaussian and $\sqrt{T}$-consistent, regardless of the rate of convergence of $W_Tc$ to $W_c$, as it is the case of correct specification.

**Theorem 2** (Distribution of the MA GMM estimator under misspecification): Assume that the model is misspecified according to Assumption 1. As $T \to \infty$, for any $\omega \in H_m$,

$$
\hat{\theta}_T(\omega) \xrightarrow{p} \sum_{c \in C} \omega_c \theta_{sc}(W_c),
$$

where $\hat{\theta}_{Tc}, c \in C$, is the GMM estimator and

$$
\theta_{sc}(W_c) = \arg \min_{\theta} E[g_c(y_t, \theta)]' W_c E[g_c(y_t, \theta)].
$$

Moreover, for the GMM estimator,

$$
\sqrt{T} \left( \hat{\theta}_T(\omega) - \sum_{c \in C} \omega_c \theta_{sc,T} \right) \xrightarrow{d} \eta_s = \sum_{c \in C} \omega_c \eta_{sc},
$$

where the $k \times 1$ random variable $\eta_{sc} \sim N(0, V_{sc})$, with

$$
V_{sc} = \left( G_{sc}' W_c G_{sc} \right)^{-1} \left( G_{sc}' W_c S_{sc} W_c G_{sc} \right) \left( G_{sc}' W_c G_{sc} \right)^{-1}.
$$

Here,

$$
\mu_{sc} = \mu_{c}(\theta_{sc}) = E[g_c(y_t, \theta_{sc})], G_{sc} = E \left( \frac{\partial g_c(y_t, \theta)}{\partial \theta} \bigg| \theta = \theta_{sc} \right),
$$

and

$$
S_{sc} = \lim_{T \to \infty} \text{Var} \left[ T^{-1/2} \sum_{t=1}^{T} (g_c(y_t, \theta_{sc}) - \mu_{sc}) \right].
$$

For $W_c = S_{sc}^{-1}, c \in C$, we have $V_{sc} = \left( G_{sc}' S_{sc}^{-1} G_{sc} \right)^{-1}$. The limit process $\eta_s$ is gaussian with zero expectation.

**Proof**: The first result was shown above. To derive the asymptotic distribution consider the Mean Value Theorem

$$
\hat{G}_{Tc} \left( \hat{\theta}_{Tc} \right)' W_T c \hat{g}_{Tc}(\theta_{sc,T}) + \hat{G}_{Tc} \left( \hat{\theta}_{Tc} \right)' W_T c \hat{G}_{Tc}(\hat{\theta}_{Tc}) \left( \hat{\theta}_{Tc} - \theta_{sc,T} \right) = 0,
$$

where $\hat{\theta}_{Tc}$ is some value "between" $\hat{\theta}_{Tc}$ and $\theta_{sc,T}$. Rearranging terms,

$$
\sqrt{T} \left( \hat{\theta}_{Tc} - \theta_{sc,T} \right) = - \left[ \hat{G}_{Tc} \left( \hat{\theta}_{Tc} \right)' W_T c \hat{G}_{Tc}(\hat{\theta}_{Tc}) \right]^{-1} \hat{G}_{Tc} \left( \hat{\theta}_{Tc} \right)' W_T c \sqrt{T} \hat{g}_{Tc}(\theta_{sc,T}).
$$
Then, for a fixed $\omega$,

$$
\sqrt{T} \left( \hat{\theta}_T (\omega) - \sum_{c \in C} \omega_c \theta_{sc,T} \right) = - \sum_{c \in C} \omega_c \left[ \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \tilde{G}_{Tc} (\tilde{\theta}_{Tc}) \right]^{-1} \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \sqrt{T} \tilde{g}_{Tc} (\theta_{sc,T})
$$

$$
= - \sum_{c \in C} \omega_c \left[ \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \tilde{G}_{Tc} (\tilde{\theta}_{Tc}) \right]^{-1} \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \sqrt{T} \left( \tilde{g}_{Tc} (\theta_{sc,T}) - \mu_c (\theta_{sc,T}) \right)
$$

$$
= - \sum_{c \in C} \omega_c \left[ \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \tilde{G}_{Tc} (\tilde{\theta}_{Tc}) \right]^{-1} \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \sqrt{T} \mu_c (\theta_{sc,T})
$$

$$
= - \sum_{c \in C} \omega_c \left[ \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \tilde{G}_{Tc} (\tilde{\theta}_{Tc}) \right]^{-1} \tilde{G}_{Tc} (\tilde{\theta}_{Tc})' W_{Tc} \sqrt{T} \left( \tilde{g}_{Tc} (\theta_{sc,T}) - \mu_c (\theta_{sc,T}) \right) + o_p (1),
$$

by Assumption 1, which converges in distribution to $\sum_{c \in C} \omega_c \eta_{sc}$, where $\eta_{sc} \sim N (0, V_{sc})$, with

$$
V_{sc} = \left( G'_{sc} W_{sc} G_{sc} \right)^{-1} \left( G'_{sc} W_{sc} S_{sc} W_{sc} G_{sc} \right) \left( G'_{sc} W_{sc} G_{sc} \right)^{-1}. \tag{60}
$$

QED.

Assumption 1 is important to guarantee a zero expectation of $\eta_s$. If, instead, the sequence $\theta_{sc,T}$ is such that $\mu_c (\theta_{sc,T}) = O \left(T^{-1/2}\right)$, then $\sqrt{T} \mu_c (\theta_{sc,T}) \xrightarrow{P} \mu_{sc} \neq 0$, for all $c \in C$, then the expectation of $\eta_s$ is

$$
- \sum_{c \in C} \omega_c \left( G'_{sc} W_{sc} G_{sc} \right)^{-1} G'_{sc} W_{sc} \mu_{sc} \neq 0. \tag{61}
$$

Another important result is that the $\lim \hat{\theta}_T (\omega)$ can be regarded as a linear combination of a true value and pseudo-true values when it is assumed that there is a correct specification, for some models $c \neq \iota_m$. This follows from the decomposition

$$
\lim \hat{\theta}_T (\omega) = \sum_{c \in C} \omega_c \theta_{sc} = \sum_{c \in C_0} \omega_c \theta_0 + \sum_{c \in C_\ast} \omega_c \theta_{sc} (W_c)
$$

$$
= \omega_0 \theta_0 + (1 - \omega_0) \sum_{c \in C_\ast} \left( \frac{\omega_c}{1 - \omega_0} \right) \theta_{sc} (W_c), \tag{62}
$$

where $\omega_0 = \sum_{c \in C_0} \omega_c$ and $C_0$ is the set of correctly specified models. The larger $C_0$ is, the "closer" to $\theta_0$ the $\lim \hat{\theta}_T (\omega)$ gets.

Searching for the optimal weight $\hat{\omega}$ is even harder than the correctly specified case: The $\lim \hat{\theta}_T (\omega)$ and the variance of $\eta_s$ depend on the arbitrarily choice for $W_c$ for all models $c \in C$. Hopefully, the criteria based on $MSC_T (c)$ lead to $\hat{\omega}$, found by numerical algorithms. Also, the relationship between $\frac{1}{T} MSC_T (\omega)$ and $RT (\omega)$ is hard to establish, for misspecified models.

\[8\] Notice that this is still a "local" misspecification result as $\mu_c (\theta_{sc,T}) \xrightarrow{P} 0$ for all $c$. 

18
4 Simulations

In this section, we address the properties of the proposed model averaging estimators for unconditional moment conditional models in finite samples. We use the experimental design suggested in Donald and Newey (2001), and subsequently used in Donald, Imbens and Newey (2009), Eryuruk et al. (2009b) and Kuersteiner and Okui (2010), to facilitate comparability of our results with theirs. The generated data is

\[ y_i = \beta_0 Y_i + \varepsilon_i, \quad Y_i = \pi' Z_i + u_i, \quad i = 1, \ldots, N, \tag{63} \]

where the true parameter of interest is the scalar \( \beta_0 = 0.1 \); \( Y_i \) is a scalar;

\[ (\varepsilon_i, u_i, Z_i)' \sim i.i.d. N(0, \Sigma), \quad \text{where} \quad \Sigma = \begin{pmatrix} 1 & 0.5 & 0_{1 \times M} \\ 0.5 & 1 & 0_{1 \times M} \\ 0_{M \times 1} & 0_{M \times 1} & I_M \end{pmatrix}, \tag{64} \]

the number of observations is \( N \in \{100, 1000\} \) and the number of instruments is

\[ M = \begin{cases} 20, \text{ if } N = 100 \\ 30, \text{ if } N = 1000 \end{cases}. \]

In terms of specifications for \( \pi \), we have, for \( m = 1, \ldots, M \),

Model A (equal coefficients) : \( \pi_m = \frac{R_f^2}{M (1 - R_f^2)} \) and

Model B (declining coefficients) : \( \pi_m = c(M) \left( \frac{1}{M + 1} - \frac{m}{M + 1} \right)^4 \),

where \( c(M) \) is set so that \( \pi \) satisfies \( \pi' \pi = R_f^2 / \left(1 - R_f^2\right) \), where \( R_f^2 \in \{0.1, 0.01\} \). The value of \( R_f^2 = 0.01 \) can be interpreted as the "weak" instrument case, quite common in empirical applications. The number of replications is 5000.

The estimation is conducted for the unconditional moments \( E[Z_{c,i} (y_i - \beta Y_i)] = 0 \), where \( Z_{c,i} \) is \( c \times 1 \) for all models \( c = m = 1, \ldots, M \). For comparison with Kuersteiner and Okui (2010), we compute the MA estimators as a linear combination of \( M \) weights, where each \( Z_m \) is the matrix of the first \( m \) elements of \( Z \). By not taking all possible combinations of instruments, we avoid estimating a huge number of weights (not very informative) and assume implicitly that \( \beta_0 \) is identified by all the subsets of the candidate set considered. We consider the following MA estimators: \( MSC_{T\tau} \) with \( W_{T\tau} = \hat{S}_{T\tau}^{-1} \) (HAC formula) or \( W_{T\tau} = \left(Z'Z\right)^{-1} \) and with \( \tau = \tau_m \) (all instruments) or \( \tau = (1, 0, m-1)' \) (only the first instrument); \( RMSC_T \); and \( CCIC_T \). In terms of notation, \( MA-MSC^C_{1} \) corresponds to the case of \( W_{T\tau} = \left(Z'Z\right)^{-1} \) and \( \tau = (1, 0, m-1)' \) whereas \( MA-MSC^{hac}_m \) stands for the case of \( W_{T\tau} = \hat{S}_{T\tau}^{-1} \) and \( \tau = \tau_m \).
Table 1: Relative Median Bias

<table>
<thead>
<tr>
<th>Model</th>
<th>RMB</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 100</td>
<td>$R^2_f = 0.01$</td>
<td>$R^2_f = 0.1$</td>
</tr>
<tr>
<td>MA-MSC$^z_1$</td>
<td>1.0133</td>
<td>0.9218</td>
<td>0.9680</td>
</tr>
<tr>
<td>MA-MSC$^z_m$</td>
<td>0.9768</td>
<td>0.8622</td>
<td>0.8895</td>
</tr>
<tr>
<td>MA-MSC$^{hac}_1$</td>
<td>1.0133</td>
<td>0.9218</td>
<td>0.9680</td>
</tr>
<tr>
<td>MA-MSC$^{hac}_m$</td>
<td>0.9809</td>
<td>0.8678</td>
<td>0.8903</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>0.9864</td>
<td>0.9120</td>
<td>0.9623</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>0.9815</td>
<td>0.9131</td>
<td>0.9490</td>
</tr>
<tr>
<td></td>
<td>N = 1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-MSC$^z_1$</td>
<td>0.9409</td>
<td>0.8310</td>
<td>0.8368</td>
</tr>
<tr>
<td>MA-MSC$^z_m$</td>
<td>0.9194</td>
<td>0.6602</td>
<td>0.6631</td>
</tr>
<tr>
<td>MA-MSC$^{hac}_1$</td>
<td>0.9409</td>
<td>0.8310</td>
<td>0.8368</td>
</tr>
<tr>
<td>MA-MSC$^{hac}_m$</td>
<td>0.9214</td>
<td>0.6625</td>
<td>0.6634</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>0.9223</td>
<td>0.8064</td>
<td>0.8575</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>0.9454</td>
<td>0.8388</td>
<td>0.8031</td>
</tr>
</tbody>
</table>

The results can be found in Tables 1-3. As in Kuersteiner and Okui (2010), we compute, for each estimator, the median absolute deviation relative to that of 2SLS-DN of Donald and Newey (2001) and the measure $KW = \sum_{m=1}^{M} m \omega_m$. We extend the analysis to the median bias relative to that of 2SLS-DN as well.

In general, the MA estimators perform well in these experiments, namely in terms of relative median bias. In fact, our MA estimates compare favorably to the 2SLS-DN estimator in all cases, with the exception of Model B, $R^2_f = 0.1$ and $N = 100$. The MA and 2SLS-DN estimators have similar results for Model A, $R^2_f = 0.01$ and $N = 100$. Comparing to the MA estimators of Kuersteiner and Okui (2010), we conclude that our method outperforms theirs when we have Model B, $R^2_f = 0.1$ and $N = 100$ and for Model B, $R^2_f = 0.01$ and $N = 1000$ (see our Table 2 and their Table III, positive weights). In summary, our MA estimation approach produces gains in estimating unconditional moment conditions models, relatively to the existing methods, in models with "weak" identification and declining coefficients for large data sets, at least.

With respect to the measure $KW$, the results are very similar to Kuersteiner and Okui (2010) when $N = 100$ and slightly larger for $N = 1000$, meaning that our procedure tends to add extra
## Table 2: Relative Median Absolute Deviation

<table>
<thead>
<tr>
<th>RMAD</th>
<th>Model A</th>
<th>Model A</th>
<th>Model B</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_f = 0.01$</td>
<td>$R^2_f = 0.1$</td>
<td>$R^2_f = 0.01$</td>
<td>$R^2_f = 0.1$</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-$MSC^z_1$</td>
<td>1.0047</td>
<td>0.9202</td>
<td>0.9678</td>
<td>0.8190</td>
</tr>
<tr>
<td>MA-$MSC^z_m$</td>
<td>1.0933</td>
<td>1.0542</td>
<td>1.0238</td>
<td>0.8358</td>
</tr>
<tr>
<td>MA-$MSC^{hac}_1$</td>
<td>1.0047</td>
<td>0.9202</td>
<td>0.9678</td>
<td>0.8190</td>
</tr>
<tr>
<td>MA-$MSC^{hac}_m$</td>
<td>1.1049</td>
<td>1.0579</td>
<td>1.0256</td>
<td>0.8372</td>
</tr>
<tr>
<td>MA-$RMSC$</td>
<td>0.9719</td>
<td>0.8888</td>
<td>0.9511</td>
<td>0.8229</td>
</tr>
<tr>
<td>MA-$CCIC$</td>
<td>0.9661</td>
<td>0.8910</td>
<td>0.9372</td>
<td>0.7649</td>
</tr>
<tr>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-$MSC^z_1$</td>
<td>0.9379</td>
<td>0.8405</td>
<td>0.8209</td>
<td>1.0228</td>
</tr>
<tr>
<td>MA-$MSC^z_m$</td>
<td>1.0004</td>
<td>0.9773</td>
<td>0.7816</td>
<td>1.1318</td>
</tr>
<tr>
<td>MA-$MSC^{hac}_1$</td>
<td>0.9379</td>
<td>0.8405</td>
<td>0.8209</td>
<td>1.0228</td>
</tr>
<tr>
<td>MA-$MSC^{hac}_m$</td>
<td>1.0006</td>
<td>0.9781</td>
<td>0.7817</td>
<td>1.1235</td>
</tr>
<tr>
<td>MA-$RMSC$</td>
<td>0.9112</td>
<td>0.7705</td>
<td>0.8374</td>
<td>1.0430</td>
</tr>
<tr>
<td>MA-$CCIC$</td>
<td>0.9343</td>
<td>0.8015</td>
<td>0.7841</td>
<td>1.0111</td>
</tr>
</tbody>
</table>
Table 3: KW Measure

<table>
<thead>
<tr>
<th>KW</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N = 100$</td>
<td>$R^2_f = 0.01$</td>
</tr>
<tr>
<td>MA-$MSC_{1}^{z}$</td>
<td>11.1253</td>
<td>11.1926</td>
</tr>
<tr>
<td>MA-$MSC_{m}^{z}$</td>
<td>9.7932</td>
<td>9.7177</td>
</tr>
<tr>
<td>MA-$MSC_{1}^{hac}$</td>
<td>11.1253</td>
<td>11.1926</td>
</tr>
<tr>
<td>MA-$MSC_{m}^{hac}$</td>
<td>9.7938</td>
<td>9.7241</td>
</tr>
<tr>
<td>MA-$RMSC$</td>
<td>13.9373</td>
<td>13.7680</td>
</tr>
<tr>
<td>MA-$CCIC$</td>
<td>11.3449</td>
<td>11.3020</td>
</tr>
<tr>
<td>$N = 1000$</td>
<td>$R^2_f = 0.01$</td>
<td>$R^2_f = 0.1$</td>
</tr>
<tr>
<td>MA-$MSC_{1}^{z}$</td>
<td>16.1841</td>
<td>16.9432</td>
</tr>
<tr>
<td>MA-$MSC_{m}^{z}$</td>
<td>14.6297</td>
<td>13.9622</td>
</tr>
<tr>
<td>MA-$MSC_{1}^{hac}$</td>
<td>16.1841</td>
<td>16.9432</td>
</tr>
<tr>
<td>MA-$MSC_{m}^{hac}$</td>
<td>14.6324</td>
<td>13.9820</td>
</tr>
<tr>
<td>MA-$RMSC$</td>
<td>20.5868</td>
<td>19.3898</td>
</tr>
<tr>
<td>MA-$CCIC$</td>
<td>16.4316</td>
<td>16.5388</td>
</tr>
</tbody>
</table>
instruments in large samples. Comparing our six MA estimation criteria, we observe that the MA-\textit{MSC} estimator is the best in terms of bias and that no estimator clearly dominates in terms of absolute deviation (in this case, the MA-\textit{RMSC} and MA-\textit{CCIC} estimators can do at least as well as the MA-\textit{MSC} procedures).

5 Application to a Monetary Policy Rule

In this section, we estimate the Fed’s forward-looking monetary policy reaction function for the period 1960:1-1996:4. as in CGG, employing GMM and our model averaging approaches. The relative merits of each method are evaluated by building error measures of the differences between the actual and the targeted Fed funds rate during the last four decades of the 20th century.

5.1 A Forward-Looking Monetary Policy Reaction Function

Making use of the orthogonality conditions arising from the rational expectations framework, CGG used the GMM methodology to estimate a forward-looking policy rule, using with the Federal Funds rate as the instrument of policy making during the aforementioned periods. They find that monetary policy during the Paul Volcker and Alan Greenspan period was more stable than during the fifteen or so years prior to Volcker’s appointment.

CGG derived the forward-looking monetary policy reaction function without specifying a central bank’s objective function that would lead to an optimal monetary instrument rule. The baseline policy rule for the target nominal interest rate (nominal Federal Funds rate) in period $t$, $i^*_t$, is given by

$$i^*_t = i^* + \beta (E_t \pi_{t+k} - \pi^*) + \gamma E_t x_{t,q},$$

(67)

where $\pi_{t+k}$ is the percent change in the price level between periods $t$ and $t+k$, expressed in annual rates, and $x_{t,q}$ is a measure of the average output gap between $t$ and $t + q$. The output gap is defined as the percent deviation between actual GDP and the corresponding target. Moreover, $\pi^*$ denotes the target for inflation and, by model construction, $i^*$ is the desired nominal interest rate when both the inflation rate and output are expected to be at their target levels. $E_t$ is the expectation operator conditional on the information set available at time $t$, $\Omega_t$. Hence, $E_t \pi_{t+k}$ should be read as $E (\pi_{t+k} | \Omega_t)$.

In this model, the central bank forms beliefs about the future state of the economy based on available information in period $t$. The target rate in period $t$, $i^*_t$, is a linear function of the
expected inflation and output gaps with respect to their target levels. The interest rate policy rules tend to be stabilizing for $\beta > 1$ and for $\gamma > 0$ (the monetary rules are more likely to be destabilizing for $\beta \leq 1$ and $\gamma < 0$), as model (67) is equivalent to

$$r_t^* = r^* + (\beta - 1) (E_t \pi_{t,k}^* - \pi^*) + \gamma E_t x_{t,q},$$

(68)

where $r^* = i^* - \pi^*$ is the equilibrium real interest rate and $r_t^* - E_t \pi_{t,k}^*$ is the (ex-ante) real interest rate target.

Another key feature of the model is that the monetary authorities do not immediately set the actual interest rate to its targeted counterpart. To be in line with the literature, let us assume that the actual interest rate deviates randomly from the target rate due to monetary shocks $e_t$, such that $E_t e_t = 0$, and that the adjustment goes smoothly over time according to $\rho (L) i_t = (1 - \rho) i_t^* + e_t$, with the $p$th-order autoregressive lag polynomial $\rho (L) = 1 - \rho_1 L - \ldots - \rho_p L^p$ and

$$\rho \equiv 1 - \rho (1) = \rho_1 + \ldots + \rho_p.$$  

(69)

The partial adjustment of the actual rate to the target value is observed through the equation

$$i_t = \rho_1 i_{t-1} + \ldots + \rho_p i_{t-p} + (1 - \rho) i_t^* + e_t,$$

(70)

where $i_t$ depends on a linear combination of its past values and on the current target rate (plus a zero mean exogenous interest rate shock). The parameters $\rho_i$ may be interpreted as the degree of smoothing of interest rate changes.

The CGG policy reaction rule for $i_t$ results from combining the target nominal policy (67) and the partial adjustment model that adjusts $i_t$ gradually towards $i_t^*$, (70). Substituting terms, yields

$$i_t = \rho_1 i_{t-1} + \ldots + \rho_p i_{t-p} + (1 - \rho) [\alpha + \beta E_t \pi_{t,k} + \gamma E_t x_{t,q}] + e_t,$$

(71)

where $\alpha = i^* - \beta \pi^* = r^* + (1 - \beta) \pi^*$. By the law of iterated expectations, equation (71) can be written as

$$i_t = \rho_1 i_{t-1} + \ldots + \rho_p i_{t-p} + (1 - \rho) [\alpha + \beta \pi_{t+k} + \gamma x_{t+q}] + \varepsilon_t,$$

(72)

where the innovation $\varepsilon_t$ follows the process

$$\varepsilon_t = e_t - (1 - \rho) [\beta (\pi_{t+k} - E_t \pi_{t+k}) + \gamma (x_{t+q} - E_t x_{t+q})].$$

(73)

A natural and appropriate estimation methods to the unknown quantities $\alpha, \rho, \beta$ and $\gamma$ that exploits the rational expectations structure is GMM. Indeed, the forecast errors $\pi_{t+k} - E_t \pi_{t+k}$

\footnote{In this model, $\alpha$ is identifiable but not $i^*$ and $\pi^*$, jointly (notice that $\beta$ is identified through $\pi_{t+k}$). Thus, with the argument that $\pi^*$ is of some interest in the characterization of the monetary policy, following CGG, the parameter of interest subject to estimation is $\pi^*$ and $i^*$ is measured as the observed sample average.}
and \(x_{t+q} - E_t x_{t+q}\) are, by construction, orthogonal to any variable in the information set \(\Omega_t\) and, most likely, correlated with \(\pi_{t+k}\) and \(x_{t+q}\). The instrumental variables \(z_t\) that belong to \(\Omega_t\) are, most probably, correlated with past \(i_t, \pi_{t+k}\) and \(x_{t+q}\), as well.

Using the two-step GMM estimation method with quarterly US data (1960:1-1996:4), CGG found that during the Volcker-Greenspan policy period the Fed funds rate was more sensitive to changes in expected inflation than in the pre-Volcker period (prior to 1979:3). Though with a smaller magnitude, the same conclusion was obtained with respect to the output gap variable. They show that during the Volcker-Greenspan period the monetary rule was stabilizing.

Nonetheless, the impact of output on the interest rate policy was sensitive to the particular choice of the output gap measure. In fact, \(\gamma\) is not statistically significant for the Volcker-Greenspan period when \(x\) is obtained either using detrended output or the unemployment rate. Using the series constructed by the CBO, \(\hat{\gamma}\) doubles its previous values and \(\gamma\) is statistically significant in both periods. Jondeau et al. (2004), using the CBO series with GMM and Continuous-Updating GMM estimation methods, obtained a wide range of estimates for \(\gamma\) depending on the choice of the weighting matrix (from 0.3 to 3.4, statistically significant in some cases, but not in all). The estimates of the inflation target \(\pi^*\) seemed plausible: 4.25% for the pre-Volcker period and about 3.5% post-Volcker. Finally, the estimate of the smoothing parameter \(\rho\) is high (about 0.7 pre-Volcker and 0.8 post-Volcker) reflecting the inertia in the interest rate dynamics and that the Fed smooths adjustments in its monetary instrument.

In this paper, we build upon the theoretical results on averaging GMM estimators. In this sense, we take the scenario of studying a macroeconomic model, but for which there is possibly a non-unique set of instruments to estimate the unknown parameters. To fix ideas, let \(p, q\) and \(k\) be any given values. For a particular set of instruments \(i\) (collected in a \(m_i \times 1\) vector \(z_t^{(i)}\)) one can define a specific model \(M_i\) with orthogonality conditions

\[
E \left( \left( i_t - \rho_1 i_{t-1} - ... - \rho_p i_{t-p} - (1 - \rho) [\alpha + \beta \pi_{t+k} + \gamma x_{t+q}] \right) z_t^{(i)} \right) = 0, \quad (74)
\]

which most likely provide distinct GMM estimates for different \(i\).

We adopt their baseline specification for which \(k = q = 1\) (one period forward)\(^{10}\) and where the monetary authorities set an expected interest rate that is a linear combination of the target rate and the observed rate in the two previous periods, \(p = 2\):

\[
i_t = \rho_1 i_{t-1} + \rho_2 i_{t-2} + (1 - \rho) [\alpha + \beta E_{t} \pi_{t+1} + \gamma E_{t} x_{t+1}] + e_t. \quad (75)
\]

\(^{10}\)CGG also consider the more realistic cases of \(k = 4\) and \(q = 1\) and of \(k = 4\) and \(q = 2\), but concluded that the results are qualitatively very similar to the baseline specification.
Moreover, we take their baseline inflation and "output gap" measures. By fixing $p, k, q$ and the definitions of inflation and output gap, we focus on the differences that occur when one uses different estimation procedures and how the choice of the instruments may affect the empirical conclusions. On the other hand, combining different set of instruments allows us to derive estimates for the MA GMM procedures.

5.2 Data and List of Available Instruments

The data is the same as in CGG\textsuperscript{11}, that is, US quarterly data for the period 1960:1-1996:4. This way, we can compare the standard GMM results with the alternative estimation procedures such as our model averaging techniques. The sample period is divided into two subsamples: One spanning from 1960:1 to 1979:2 (pre-Volker) and the second from 1979:3 to 1996:4 corresponding to the Paul Volcker and Alan Greenspan as the Fed's chairmen. It is argued that these two periods correspond to the unstable and stable eras of recent history, respectively.

Following CGG, inflation is measured as the (annualized) rate of change of the GDP deflator between two subsequent quarters and the output gap is the series constructed by the Congressional Budget Office (CBO). Moreover, the interest rate corresponds to the average Federal Funds rate in the first-month of each quarter, expressed in annual rates. Lagged variables are used as instruments, as well as the lags of commodity price inflation and the "spread" between the long-term bond rate and the three-month Treasury Bill rate. We have four lags of each variable available in the data. CGG used these four lags to estimate the baseline model by GMM (see CGG, Table II, page 157).

In terms of notation, the list of available instruments is

$$Z_{t-1} = (1, i_{t-1}, \ldots, i_{t-4}, \pi_{t-1}, \ldots, \pi_{t-4}, x_{t-1}, \ldots, x_{t-4}, dc_{t-1}, \ldots, dc_{t-4}, spr_{t-1}, \ldots, spr_{t-4})' . \quad (76)$$

Nevertheless, we also considered the estimation with only two fixed lags and, for the moments and model selection criteria and the model averaging procedure, we estimated the model for (almost) all possible combinations of instruments out of the available set. The two lags are chosen with the purpose of minimizing potential biases due to the large number of identifying restrictions. The array of instrument combinations is related to the theoretical approach presented in the previous sections. For a particular model $M_i$, the orthogonality conditions for the baseline specification are

$$E \left( \left( i_t - \rho_1 i_{t-1} - \rho_2 i_{t-2} - (1 - \rho) [\alpha + \beta \pi_{t+1} + \gamma x_{t+1}] \right) z_{t-1}^{(i)} \right) = 0, \quad (77)$$

\textsuperscript{11}We thank the authors for providing us the data used in their paper. See CGG paper for more details about the data.
where $z_{t-1}^{(i)}$ is a subset of $z_{t-1}$.

5.3 Estimation and Model Selection Criteria

We begin by discussing the results based on the GMM estimation of model (77). We conduct estimations for the period 1960:1 to 1979:2 (PV stands for pre-Volcker) and the more recent vintage of the data, which spans from 1979:3 to 1996:4 (VG stands for Volcker-Greenspan). The GMM estimations refer to the two-step efficient procedure. For now, we let the number of lagged instruments to be either four (4l) or two (2l). The results are presented in Table 4. The standard errors are reported in parentheses and the $p$-value of the $J$-statistic is denoted by "$J$".

We observe that, overall, the estimators produce consistent and comparable results for each given period. In particular, estimates of $\beta$ and $\gamma$ tend to suggest that the pre-Volcker policy was passive with regards to inflation, while the Volcker-Greenspan rule was aggressive, although neutral with respect to the product. The point estimates of $\beta$ are smaller than one for the PV period (a 95% confidence interval for the GMM with 4 lags includes values for $\beta \geq 1$). On the other hand, for the VG period all estimates are larger than one. This makes evidence for stabilizing rules during VG tenure and less so for the PV period with respect to inflation. In terms of $\gamma$, we conclude that it is statistically significant for the PV period (with small but positive point estimates), though not significant for the VG tenures. Hence, it reflects stabilizing rules for the PV period and a neutral policy for the VG period with respect to the output gap.

There is evidence that the Fed smooths adjustments in its monetary instrument. The inertia in the interest rate dynamics is similar for the two subperiods, with the exception of the GMM with 4 lags for the VG period in which $\hat{\beta} = 0.357$, a relatively small value. The GMM estimate is about 0.55. The point estimates of the inflation target $\pi^*$ are relatively close to the expected ones (4.5% – 4.7% during the PV period and 3.8% – 4% for the VG tenures). The $J$-statistic

Table 4: GMM Estimates with a Fixed Number of Instruments

<table>
<thead>
<tr>
<th></th>
<th>$\pi^*$</th>
<th>$\rho$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>4.681</td>
<td>0.509</td>
<td>0.970</td>
<td>0.173</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.063)</td>
<td>(0.049)</td>
<td>(0.064)</td>
<td></td>
</tr>
<tr>
<td>4l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>3.916</td>
<td>0.357</td>
<td>1.750</td>
<td>-0.155#</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.114)</td>
<td>(0.185)</td>
<td>(0.146)</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>4.556</td>
<td>0.554</td>
<td>0.884</td>
<td>0.139</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td>(0.158)</td>
<td>(0.092)</td>
<td>(0.058)</td>
<td>(0.085)</td>
<td></td>
</tr>
<tr>
<td>2l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VG</td>
<td>3.839</td>
<td>0.567</td>
<td>2.396</td>
<td>0.045#</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.096)</td>
<td>(0.331)</td>
<td>(0.271)</td>
<td></td>
</tr>
</tbody>
</table>
does not allow us to reject the model for any of the estimation procedures.

Next, we perform some of the existing moment/instrument selection procedures in the literature. These include the moment selection criterion MSC of Andrews (1999), the relevant moment selection criterion RMSC of Hall et al. (2007), the canonical correlations information criterion CCIC of Hall and Peixe (2003) and the AMSE instrument selection method of Donald and Newey (2001), DN. Rather than fixing the list of instruments to either 2 or 4 lags of the variables, these methods allow the data to determine the "best" model out of the possible combinations of instruments.

The total number of instrument combinations is $|C| = \sum_{j=p}^{m} \binom{m}{j}$ which equals 2089605, in our case, with $m = 21$ and $p = 5$. To make the procedure more tractable, we split the list of available instruments in two groups. There are seven instruments that are kept fixed,

$$z_{F,t-1} = (1, \pi_{t-1}, \pi_{t-2}, x_{t-1}, x_{t-2})'$$

(78)

(the two lags of the variables of the model belong to any list of instruments) and the 14 remaining are combined to construct the matrix of instruments. This way, we considered 16384 different models. The chosen models are for the BIC criterion (same with AIC and HQ for almost all cases). The results are displayed in Table 5.

First, it confirms the practice of stabilizing rules for the PV period, contrasting with a neutral policy for the VG period with respect to the output gap. Second, point estimates for $\beta$ give different conclusions. There is evidence for stabilizing rules during VG tenure, but with passive policy for the PV period with respect to inflation only for the MSC procedure. For both periods, with the RMSC method we observe non-stabilizing rules whereas for the CCIC and DN approaches there is evidence of stabilizing rules. Point estimates of $\rho$ and $\pi^*$ are similar to those obtained with a fixed number of lags. Finally, we acknowledge the fact that the selected model by the RMSC criterion contains the smallest number of instruments and the opposite occurs with the MSC and DN methods, despite far from the 4-lags full set of restrictions.

5.4 MA Estimators

Now, we apply the model averaging methods described earlier in the paper to the baseline forward-looking monetary policy rule. The MA estimator is given by $\hat{\theta}_T(\hat{\omega}) = \sum_{c \in C} \hat{\omega}_c \hat{\theta}_{Tc}$, where the optimal weights $\hat{\omega}_c$ are estimated by the proposed criteria and $\hat{\theta}_{Tc}$ is the efficient GMM estimator for model $c$. For the $MSC_{T\tau}$ method, we considered $\tau = \iota_m$ (4 lags) and $\tau = "F"$ (only the fixed instruments, $Z_{\tau} = Z_F$).
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MSC</td>
<td>PV</td>
<td>$\hat{c}$</td>
<td>$\pi^*$</td>
<td>$\rho$</td>
<td>$\beta$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.867</td>
<td>0.571</td>
<td>0.897</td>
<td>0.287</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.186)</td>
<td>(0.079)</td>
<td>(0.071)</td>
<td>(0.106)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.800</td>
<td>0.588</td>
<td>2.434</td>
<td>0.028$^#$</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.146)</td>
<td>(0.088)</td>
<td>(0.329)</td>
<td>(0.240)</td>
<td></td>
</tr>
<tr>
<td>RMSC</td>
<td>PV</td>
<td>4.867</td>
<td>0.571</td>
<td>0.897</td>
<td>0.287</td>
<td>0.394</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.211)</td>
<td>(0.104)</td>
<td>(0.096)</td>
<td>(0.042)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>4.000</td>
<td>0.600</td>
<td>0.800</td>
<td>0.300$^#$</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.567)</td>
<td>(0.110)</td>
<td>(0.439)</td>
<td>(0.375)</td>
<td></td>
</tr>
<tr>
<td>CCIC</td>
<td>PV</td>
<td>4.630</td>
<td>0.554</td>
<td>1.018</td>
<td>0.209</td>
<td>0.781</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.152)</td>
<td>(0.069)</td>
<td>(0.072)</td>
<td>(0.082)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.887</td>
<td>0.728</td>
<td>2.614</td>
<td>0.163$^#$</td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.186)</td>
<td>(0.079)</td>
<td>(0.455)</td>
<td>(0.395)</td>
<td></td>
</tr>
<tr>
<td>DN</td>
<td>PV</td>
<td>4.682</td>
<td>0.604</td>
<td>1.038</td>
<td>0.266</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.181)</td>
<td>(0.063)</td>
<td>(0.083)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.831</td>
<td>0.483</td>
<td>1.868</td>
<td>$-0.030^#$</td>
<td>0.770</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.153)</td>
<td>(0.088)</td>
<td>(0.244)</td>
<td>(0.196)</td>
<td></td>
</tr>
</tbody>
</table>
For computational reasons and so that the estimated weights $\hat{w}_c$ were not excessively small and, for that reason, meaningless for interpretation, we only considered the 100 models with smallest $MSC_T(c) - BIC$. All of this "best" models have p-values for the $J$-statistic much larger than 10% and, therefore, we assume a correct specification setup, as discussed earlier in the paper. We also considered an alternative procedure of defining the class of available models for the MA estimation. Instead of the "best" 100, we performed a sequential procedure where we add one instrument at a time, starting from $Z_F$ up to $Z$, in which each instrument is added according to the minimization of $MSC_T(c) - BIC$ over the list of instruments available in each sequence. This approach is closer in spirit to the methods of Hansen (2007) and Kuersteiner and Okui (2010) and follows the strategy in the Simulations section, resulting in 15 models.\textsuperscript{12} The MA parameter and weight estimation results are displayed in Tables 6 and 7, respectively.

As expected, we observe very similar results for $\tau = \tau_m$ and $Z_\tau = Z_F$. For the "Best-100" approach the conclusions are essentially the same for all methods but for the "Sequential" strategy we observe important differences in the estimation of $\gamma$ for the VG period.

The point estimates confirm the main conclusions drawn from standard methods. There is evidence of a stabilizing policy rule during the VG tenures, albeit neutral to the cyclical output variable, but not stabilizing during the PV period with respect to inflation. In fact, the estimates for $\beta$ are (slightly) below one and above one for the PV and VG periods, respectively, and the estimates for $\gamma$ are positive and negative (probably not statistically significant, since the sign is not the expected one) for the PV and VG periods, respectively. Still, $\gamma$ is positive for the VG period using the $MSC$ and $CCIC$, but only for the "Sequential" strategy. The point estimates for $\pi^*$ and $\rho$ are similar to those from standard methods.

In terms of the estimated weights, a number of interesting results stands out. First, the model that gets the largest estimated weight is never the selected one by means of the standard criterion. The model with the largest estimated weight is consistently one with 4 lags without a few instruments (more so for the "Sequential" strategy and for the $RMSC$ and $CCIC$ cases). Models selected by standard criteria are ranked below 48. Finally, the range of estimated weights is relatively wide for the $RMSC$ criterion, but less so for the other.

\textsuperscript{12} The first model is $Z_F$; the second one is $Z_F, \pi_{t-3}$; the third is $Z_F, \pi_{t-3}, x_{t-4}$; then we sequently added $x_{t-3}, dc_{t-2}, dc_{t-3}, i_{t-3}, i_{t-4}, spr_{t-1}, \pi_{t-4}, spr_{t-2}, spr_{t-4}, spr_{t-3}, dc_{t-4}$ and $dc_{t-1}$.
<table>
<thead>
<tr>
<th>Best-100</th>
<th>$\pi^*$</th>
<th>$\rho$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>4.820</td>
<td>0.616</td>
<td>0.885</td>
</tr>
<tr>
<td>$Z_{\tau} = Z_F$</td>
<td>VG</td>
<td>3.799</td>
<td>0.617</td>
<td>2.370</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>4.820</td>
<td>0.616</td>
<td>0.885</td>
</tr>
<tr>
<td>$\bar{\tau} = \bar{t}_m$</td>
<td>VG</td>
<td>3.799</td>
<td>0.617</td>
<td>2.370</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV</td>
<td>4.658</td>
<td>0.526</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.866</td>
<td>0.425</td>
<td>1.851</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV</td>
<td>4.691</td>
<td>0.577</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.885</td>
<td>0.410</td>
<td>1.193</td>
</tr>
<tr>
<td>Sequential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>4.799</td>
<td>0.657</td>
<td>0.836</td>
</tr>
<tr>
<td>$Z_{\tau} = Z_F$</td>
<td>VG</td>
<td>3.978</td>
<td>0.733</td>
<td>2.365</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>4.799</td>
<td>0.657</td>
<td>0.836</td>
</tr>
<tr>
<td>$\bar{\tau} = \bar{t}_m$</td>
<td>VG</td>
<td>3.978</td>
<td>0.733</td>
<td>2.365</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV</td>
<td>4.751</td>
<td>0.596</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.918</td>
<td>0.366</td>
<td>1.761</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV</td>
<td>4.795</td>
<td>0.656</td>
<td>0.838</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>3.979</td>
<td>0.728</td>
<td>2.348</td>
</tr>
</tbody>
</table>
Table 7: MA GMM Weights

<table>
<thead>
<tr>
<th>Best-100</th>
<th>( \omega )</th>
<th>( \omega_{c} ) (rank)</th>
<th>( \hat{c} : \max \omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA-MSC</td>
<td>PV [0.00995, 0.01004]</td>
<td>0.00997 (91)</td>
<td>( z_{F}, i_{t-3}, x_{t-3}, dc_{t-1}, dc_{t-3}, dc_{t-4}, spr_{t-2}, spr_{t-4} )</td>
</tr>
<tr>
<td></td>
<td>VG [0.00927, 0.01032]</td>
<td>0.01001 (48)</td>
<td>( z_{F}, dc_{t-2}, dc_{t-3}, dc_{t-4} )</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV [0.00995, 0.01003]</td>
<td>0.00998 (93)</td>
<td>( z_{F}, i_{t-3}, x_{t-3}, dc_{t-1}, dc_{t-3}, dc_{t-4}, spr_{t-2}, spr_{t-4} )</td>
</tr>
<tr>
<td>( \tau = t_{m} )</td>
<td>VG [0.00924, 0.01035]</td>
<td>0.00999 (49)</td>
<td>( z_{F}, dc_{t-2}, dc_{t-3}, dc_{t-4} )</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV [0.00657, 0.01680]</td>
<td>0 (&gt;100)</td>
<td>4lags</td>
</tr>
<tr>
<td></td>
<td>VG [0.00251, 0.10367]</td>
<td>0 (&gt;100)</td>
<td>( z_{F}, x_{t-3}, i_{t-3}, \pi_{t-3}, \pi_{t-4}, dc_{t-1}, dc_{t-2} )</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV [0.00994, 0.01012]</td>
<td>0 (&gt;100)</td>
<td>4lags</td>
</tr>
<tr>
<td></td>
<td>VG [0.00990, 0.01016]</td>
<td>0 (&gt;100)</td>
<td>4lags</td>
</tr>
<tr>
<td>Sequential</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV [0.06590, 0.06810]</td>
<td>-</td>
<td>4lags except ( dc_{t-1} )</td>
</tr>
<tr>
<td>( Z_{F} = Z_{F} )</td>
<td>VG [0.05878, 0.07230]</td>
<td>-</td>
<td>4lags except ( dc_{t-1} )</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV [0.06608, 0.06834]</td>
<td>-</td>
<td>4lags except ( dc_{t-1} )</td>
</tr>
<tr>
<td>( \tau = t_{m} )</td>
<td>VG [0.05946, 0.06983]</td>
<td>-</td>
<td>4lags except ( dc_{t-1} )</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV [0.01414, 0.41526]</td>
<td>-</td>
<td>4lags</td>
</tr>
<tr>
<td></td>
<td>VG [0.00016, 0.97632]</td>
<td>-</td>
<td>4lags</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV [0.05739, 0.07746]</td>
<td>-</td>
<td>4lags</td>
</tr>
<tr>
<td></td>
<td>VG [0.05973, 0.07364]</td>
<td>-</td>
<td>4lags</td>
</tr>
</tbody>
</table>
5.5 Targeting the Interest Rate

Following CGG, we now measure how well the estimated target rules characterize the behavior of the actual Funds rate, where the point estimates are found by standard methods and by the MA GMM approaches. For comparison, we also consider the simple Taylor-type rule as in Woodford (2001):

\[ i_t^* = i^* + 1.5 (\pi_t - \pi^*) + 0.5 x_t. \]  

(80)

Taylor rules go back to Taylor (1993), who claimed that the following original rule was appropriate for the Fed during the period 1987-1992,

\[ i_t^* = r^* + \pi_t + 0.5 (\pi_t - \pi^*) + 0.5 x_t, \]  

(81)

where \( r^* = \pi^* = 0.02 \). In equilibrium \( (\pi_t = \pi^*, x_t = 0) \) we have \( i_t^* = i^* \). The terms \( 1.5 \) and \( 0.5 \) represent the Fed responses to inflation and output deviations from equilibrium. Basically, the Fed respond positively to both variables, but more aggressively to inflation, giving top priority to price-pressures instead of growth. If \( \pi_t < \pi^* \), then the Fed would have the opposite reaction.

For stabilization, we need \( \beta > 1 \) (the proportional reaction does not suffice so that the real interest rate has effects in the real economy) and \( \gamma > 0 \). Otherwise, the monetary policy may generate an inflation spiral (indeterminacy).

We measure the quality of the policy rules by the mean squared error (MSE) and its root (RMSE) and the mean absolute deviation (MAD). The error is given by \( i_t - i_t^* \) and the average is over the period under discussion. The results are presented in Tables 8 and 9.

Overall, the results are relatively poor once we observe RMSE’s of more than three and MAD’s of at least two percentage points. Despite the fact that we are not comparing the actual rate with the fitted rate, but with the target rate, this result may indicate that the CGG forward-looking model does not fully capture the interest rate dynamics and that it targets a Fed rate that is somewhat off the observed value.

Moreover, the target rules consistently do better during the PV period than during the VG tenures. If it is possible to extrapolate on this result then one may conclude that the Taylor and the CGG rules are worse suited for stabilizing periods. In fact, in terms of MAD, the Taylor

---

\[ \text{As CGG point out, we do not compare the actual rate with the fitted model that allows for partial adjustment. This way, the estimated target rate does not perform as well as the fitted model.} \]
Table 8: Error Measures when Targeting the Interest Rate: Standard Approaches

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR</td>
<td>6.716</td>
<td>2.591</td>
<td>2.214</td>
</tr>
<tr>
<td></td>
<td>7.663</td>
<td>2.768</td>
<td>2.097</td>
</tr>
<tr>
<td>GMM 4l</td>
<td>2.351</td>
<td>1.533</td>
<td>1.292</td>
</tr>
<tr>
<td></td>
<td>7.100</td>
<td>2.665</td>
<td>2.005</td>
</tr>
<tr>
<td>GMM 2l</td>
<td>2.103</td>
<td>1.450</td>
<td>1.199</td>
</tr>
<tr>
<td></td>
<td>11.784</td>
<td>3.433</td>
<td>2.443</td>
</tr>
<tr>
<td>c: MSC</td>
<td>2.375</td>
<td>1.541</td>
<td>1.300</td>
</tr>
<tr>
<td></td>
<td>12.188</td>
<td>3.491</td>
<td>2.450</td>
</tr>
<tr>
<td>c: RMSC</td>
<td>2.375</td>
<td>1.541</td>
<td>1.300</td>
</tr>
<tr>
<td></td>
<td>10.244</td>
<td>3.201</td>
<td>2.389</td>
</tr>
<tr>
<td>c: CCIC</td>
<td>2.616</td>
<td>1.617</td>
<td>1.359</td>
</tr>
<tr>
<td></td>
<td>14.154</td>
<td>3.762</td>
<td>2.669</td>
</tr>
<tr>
<td>c: DN</td>
<td>2.826</td>
<td>1.681</td>
<td>1.429</td>
</tr>
<tr>
<td></td>
<td>7.897</td>
<td>2.810</td>
<td>2.097</td>
</tr>
</tbody>
</table>
Table 9: Error Measures when Targeting the Interest Rate: MA GMM Approach

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best-100</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>2.535</td>
<td>1.592</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>11.538</td>
<td>3.396</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>2.535</td>
<td>1.592</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>11.538</td>
<td>3.396</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV</td>
<td>2.378</td>
<td>1.542</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>6.141</td>
<td>2.478</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV</td>
<td>2.475</td>
<td>1.573</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>7.740</td>
<td>2.782</td>
</tr>
<tr>
<td><strong>Sequential</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>2.240</td>
<td>1.496</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>13.144</td>
<td>3.625</td>
</tr>
<tr>
<td>MA-MSC</td>
<td>PV</td>
<td>2.240</td>
<td>1.496</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>13.144</td>
<td>3.625</td>
</tr>
<tr>
<td>MA-RMSC</td>
<td>PV</td>
<td>2.230</td>
<td>1.493</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>7.165</td>
<td>2.676</td>
</tr>
<tr>
<td>MA-CCIC</td>
<td>PV</td>
<td>2.233</td>
<td>1.494</td>
</tr>
<tr>
<td></td>
<td>VG</td>
<td>13.008</td>
<td>3.606</td>
</tr>
</tbody>
</table>
rule does better during the VG than the PV period, but this may be explained by the fact that Taylor himself studied the 1987-1992 period originally.

As expected, the performance of the MA GMM approach is essentially the same for different criteria, since the point estimates are very similar. More importantly, it outperforms the standard approaches for the VG period with the MA-RMSC method and it is only beaten by the GMM with fixed 2 lags for the PV period. The MA-MSC method does very well for the PV period. See Table 10 for details on the top-5 ranking performances.

In conclusion, the MA GMM procedure appears to be a valid approach to estimate the Fed’s target rule. To illustrate, we present in Figures 1 and 2 the MA-MSC target estimates and the actual rates for the two subperiods. Despite significant gaps, the upward and downward swings are reasonably captured the estimated policy rule.

6 Conclusion

In this paper, we revisit the model and the results presented in Clarida, Gali, and Gertler (2000) by employing a new estimation procedure, an averaging estimator for Unconditional Moment Conditions Models. Their baseline forward-looking monetary policy reaction function is estimated by moment condition procedures but we resort to procedures that averages estimators. This way, we focus on the potential gains in averaging estimators rather than assuming a fixed
Figure 1: Actual versus Target Rates: Pre-Volcker Era and MA GMM Estimates

Figure 2: Actual versus Target Rates: Volcker-Greenspan Era and MA GMM Estimates
moment specification as imposed by standard techniques.

Thus, we define GMM model averaging estimators and discuss some of their asymptotic properties under correctly specified and misspecified models. The MA estimators are a weighted average of a list of standard GMM estimators for each model specification under consideration. The optimal weights are found by means of particular moment and model selection criteria that share some good statistical properties. We show that the asymptotic theory under misspecification is not standard in the sense that the consistency and distributional results depend on the weight matrices and the pseudo-true values. Some Monte Carlo experiments show that our MA estimation procedure outperforms the optimal instrument selection method of Donald and Newey (2001) and the two-step IV model averaging estimation approach of Kuersteiner and Okui (2010) in some setups, including models with weak instruments.

We apply our MA GMM procedure to the same data and baseline policy model as in Clarida, Gali, and Gertler (2000) and compare the results with those obtained with standard approaches, namely the efficient GMM with fixed instruments and with instruments selected by information criteria. The methods point to similar conclusions. There is evidence for a stabilizing policy rule during the Paul Volcker and Alan Greenspan tenures, albeit neutral to the cyclical output variable, but less so during the pre-Volcker period with respect to inflation. On the contrary, CGG found that the monetary policy was not neutral to the output variable during the Paul Volcker and Alan Greenspan periods.

To evaluate the merits of our approach, we measure the quality of the policy target rules estimated by the different methods relative to the actual rates by the mean squared error and the mean absolute deviation. We conclude that the MA GMM methods do better than the standard approaches and the Taylor rule for the Volcker-Greenspan period and that it is the second best for the Pre-Volcker period. The upward and downward swings of the data are reasonably captured the estimated policy rules, despite a disappointing mean absolute deviation of approximately two percentage points. Along these lines, it seems that there is room for additional research on how to improve the CGG forward-looking rule in terms of explaining the data.

Although the empirical analysis of the paper resorts on the CGG model, our theoretical results on the moment conditions model averaging estimator suggest that alternative applications to economic models should be considered. Also, it would be important to further investigate the statistical properties of the MA estimators for models with many moments, weak identification and with conditional moment restrictions. In particular, linear IV model averaging estimators
for models with many and weak instruments deserve additional attention.

The estimated weights of our MA estimator are based on the moment selection criterion of Andrews (1999) for the GMM estimation. Valid alternatives would include the objective function of the Continuous-Updated GMM estimator of Hansen et al. (1996) and the selection criteria for estimators that belong to the class of Generalized Empirical Likelihood (GEL) estimators, such as the empirical likelihood of Kitamura (1997) and the exponential tilting of Kitamura and Stutzer (1997) (see Newey and Smith, 2004 and Anatolyev, 2005 for a general framework). With respect to GEL estimation method, it would be interesting to develop MA estimators with weights based on the GEL model selection criteria of Hong, Preston and Shum (2003). Contrary to our GMM approach, this alternative does not depend on any weight matrix in the $J$-statistic, but has the disadvantage of needing to estimate a Lagrange multiplier which, in model averaging, is likely complicate the analysis. Still, this is an important topic in the literature on MA for moment conditions models that is left for future research.

References


