WORKING PAPERS

16 | 2011

MOMENT CONDITIONS MODEL AVERAGING WITH AN APPLICATION TO A FORWARD-LOOKING MONETARY POLICY REACTION FUNCTION

Luis F. Martins

June 2011

Please address correspondence to
Banco de Portugal
Av. Almirante Reis 71, 1150-012 Lisboa, Portugal;
estudos@bportugal.pt

Banco de Portugal
EUROSYSTEM

BANCO DE PORTUGAL

Av. Almirante Reis, 71

1150-012 Lisboa

www.bportugal.pt

Edition

Economics and Research Department

Pre-press and Distribution

Administrative Services Department

Documentation, Editing and Museum Division

Editing and Publishing Unit

Printing

Administrative Services Department

Logistics Division

Lisbon, June 2011

Number of copies

150

ISBN 978-989-678-085-2

ISSN 0870-0117 (print)

ISSN 2182-0422 (online)

Legal Deposit no. 3664/83

Moment Conditions Model Averaging with an Application to a Forward-Looking Monetary Policy Reaction Function*

Luis F. Martins

Department of Quantitative Methods and UNIDE, ISCTE-IUL, Portugal (luis.martins@iscte.pt)

This version: October 3^{rd}, 2009

Abstract

In this paper, we examine the empirical validity of the baseline version of the 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 and Generalized Empirical Likelihood setups. We derive some of their asymptotic properties under correctly specified and misspecified models. Although the model averaging estimates and the standard procedures point to a stabilizing policy rule during the Paul Volcker and Alan Greenspan tenures but not so during the pre-Volker period, our results cast serious doubts on the significance of the cyclical output variable as a forcing variable in the FED funds dynamics during the Volcker-Greenspan period.

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

JEL Classification: C22; C52; E43; E52

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

1 Introduction

The forward-looking monetary policy reaction function proposed by Clarida, Gali, and Gertler (2000, henceforth CGG), based on Clarida, Gali, and Gertler (1998), has become a fundamental macroeconomic specification in the context of the United States monetary policy in the postwar. In this model, the central bank forms beliefs about the future state of the economy based on the available information so far. The target rate 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 adjusts it smoothly over time.

They employ the Generalized Method of Moments (GMM) methodology to estimate the monetary policy using the Federal Funds rate as the instrument of policy making[1]. In particular, they suggest that the FED 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 policy appeared to be much more sensitive to changes in the 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, which we call moment conditions model averaging estimator, in the GMM and Generalized Empirical Likelihood (GEL) setups. For completeness, we also employ existent moment and model selection criteria methods and the Empirical Likelihood (EL) estimation approach. 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. Another drawback is that the results hinge on the weighting matrix used in the estimation[2].

Given these disappointing properties of GMM, it has recently been proposed the GEL class

[1] There has been a considerable interest in the properties of the GMM estimator to the analysis of time series dependent data. In fact, since the seminal contribution by Hansen (1982), one has witnessed the remarkable growth of the theoretical and empirical research on this issue over the recent decades. One of the main reasons why it became so popular in moment conditions models is its computational accessibility.

[2] In CGG paper, it is not clear what weighting matrix is used. If by "optimal" they mean the HAC matrix then still it is to find what kernel and bandwidth selection method are used.

of estimators. Newey and Smith (2004) have shown that while GMM and GEL estimators have identical first-order asymptotic properties, the latter are higher order efficient, in the sense that these estimators are able to eliminate some sources of GMM biases. For example, they show that the bias of the EL estimator does not grow with the number of moment conditions, unlike GMM. Similar properties have been established by Anatolyev (2005), in a time series setting.

The second main reason for re-evaluating the empirical results by CGG concerns the econometric analysis of model selection in the context of moment condition models. In the standard GMM and GEL estimation approaches (as in CGG, for the case of GMM), parameter estimates are obtained with a given/fixed list of instruments (moment conditions, in general). On the other hand, in model selection, one ranks the available moment conditions models (combinations of instruments, in the linear IV case) according to the particular goal undertaken. Instead of just one, in model selection there are "many" competing models.

Model selection based on information criteria, hypothesis testing or shrinkage-type estimators has already been studied in the GMM and GEL framework. See, for example, Smith (1992) and Smith and Ramalho (2002) for model testing and Caner (2009) for a LASSO-type GMM estimator. Out of these three lines of model selection research, we only apply the information criteria to the CGG model. 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 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, imposing zero restrictions on the parameters. Hong, Preston and Shum (2003) extend these two previous papers to the GEL framework. Hall, Inoue, Jana and Shin (2007) take a different approach: The information criteria used for moment selection is based on the entropy of the limiting distribution of the GMM estimator.

We pursue an alternative methodological direction in model selection – Model Averaging (MA). In MA the estimation procedure is based on a weighted combination of all candidate models/estimators. The weights are chosen according to some relevant criterion. To smooth estimators across several models is a neat strategy to improve the bias and variance balance. Studies on least squares (Hansen, 2005), likelihood-based (see, for example, Hjort and Claeskens, 2003) and Bayesian (see, for example, Hoeting, Madigan, Raftery and Volinsky, 1999) MA have already been developed. At the notable study of Bruce Hansen, in the regression setup, it is through the well-known Mallows criterion that the weights are estimated. To our best knowledge, there is no published work yet in GMM/GEL MA estimators.

Thus, we propose GMM and GEL model averaging estimators and discuss some of their asymptotic properties under correctly specified and misspecified models. 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 at Andrews (1999), Andrews and Lu (2001), Hong, Preston and Shum (2003) and Hall et all (2007).

Although the MA estimates and the standard procedures point to the same conclusion, which is the evidence for a stabilizing policy rule during the Paul Volcker and Alan Greenspan tenures but not so during the pre-Volker period with respect to inflation, our results raise serious doubts on the significance of the cyclical output variable as a forcing variable in the FED funds dynamics during the Volcker-Greenspan period. Contrary to our results, CGG found that the parameter associated with output gap was statistical significant for most of the policy rule specifications. Before us, Jondeau, Le-Bihan and Gallès (2004) also questioned this result at CGG using standard GMM, CU and MLE methods.

In the next two sections, we briefly review the forward-looking monetary policy rule proposed by CGG and the econometrics of moment conditions models. In Section 4, we discuss the existence of a linear combination of instruments that give rise to a GMM/GEL estimator that attains the Chamberlain efficiency bound relative to the set of all available instruments. The approach on moment conditions model averaging is presented in Section 5. The empirical application of the existing methods and the MA procedure to the baseline CGG model is in Section 6 and a conclusion finalizes this paper.

2 Forward-Looking Monetary Policy Reaction Function

In order to provide a strong empirical evidence that the FED 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, Clarida, Gali, and Gertler (2000), based on Clarida, Gali, and Gertler (1998), estimates a policy rule for which the central bank has forward-looking expectations. Due to this type of specification, they used the GMM methodology to estimate the monetary policy with the Federal Funds rate as the instrument of policy making during the aforementioned periods. In this paper, we illustrate the merits of our approach on GMM model averaging using the CGG forward-looking specification of the monetary policy. To better understand the main aspects of model, we now discuss the monetary policy dynamics following their work in a very close manner[3].

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) at period t, i_t^*, is given by

$$i_t^* = i^* + \beta\left(E_t \pi_{t,k} - \pi^*\right) + \gamma E_t x_{t,q}, \tag{1}$$

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, π^* 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, Ω_t. Hence, $E_t \pi_{t,k}$ should be read as $E\left(\pi_{t,k} | \Omega_t\right)$.

In this model, the central bank forms beliefs about the future state of the economy based on the available information so far. The target rate at 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 (1) is equivalent to

$$r_t^* = r^* + (\beta - 1)\left(E_t \pi_{t,k} - \pi^*\right) + \gamma E_t x_{t,q}, \tag{2}$$

where $r^* = i^* - \pi^*$ is the equilibrium real interest rate and $r_t^* = i_t^* - E_t \pi_{t,k}$ is the (ex-ante) real interest rate target. Here, stability occurs as a result of low real interest rates which stimulate economic activity and inflation.

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 deviate randomly from the target rate due to monetary shocks e_t, such that $E_{t-1} e_t = 0$, and that the adjustment goes smooth over time according to

$$\rho(L)\, i_t = (1 - \rho)\, i_t^* + e_t, \tag{3}$$

with the $p^{th}-$order autoregressive lag polynomial $\rho(L) = 1 - \rho_1 L - ... - \rho_p L^p$ and

$$\rho \equiv 1 - \rho(1) = \rho_1 + ... + \rho_p. \tag{4}$$

[3] For more details, read Clarida, Gali, and Gertler (2000) and Clarida, Gali, and Gertler (1998), among others.

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

$$i_t = \rho_1 i_{t-1} + ... + \rho_p i_{t-p} + (1-\rho)\, i_t^* + e_t, \tag{5}$$

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 parameter ρ is 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 (1) and the partial adjustment model that adjusts i_t gradually towards i_t^*, (5). Substituting terms, yields

$$i_t = \rho_1 i_{t-1} + ... + \rho_p i_{t-p} + (1-\rho)\left[\alpha + \beta E_t \pi_{t,k} + \gamma E_t x_{t,q}\right] + e_t, \tag{6}$$

where

$$\alpha = i^* - \beta \pi^* = r^* + (1-\beta)\,\pi^*. \tag{7}$$

By the law of iterated expectations, equation (6) can be written as

$$i_t = \rho_1 i_{t-1} + ... + \rho_p i_{t-p} + (1-\rho)\left[\alpha + \beta \pi_{t+k} + \gamma x_{t+q}\right] + \varepsilon_t, \tag{8}$$

where the innovation ε_t follows the process

$$\varepsilon_t = e_t - (1-\rho)\left[\beta\left(\pi_{t+k} - E_t \pi_{t+k}\right) + \gamma\left(x_{t+q} - E_t x_{t+q}\right)\right]. \tag{9}$$

The most appropriate estimation method to the unknown quantities α, ρ, β and γ is GMM (or GEL, for that matter)[4]. Indeed, the forecast errors $\pi_{t+k} - E_t \pi_{t+k}$ and $x_{t+q} - E_t x_{t+q}$ are, by construction, orthogonal to any variable at the information set Ω_t and, most likely, correlated with π_{t+k} and x_{t+q}. The instrumental variables z_t that belong to Ω_t are, most probably, correlated with past i_t, π_{t+k} and x_{t+q}, as well.

In this paper, we build upon theoretical results on averaging GMM (and GEL) estimators. In this sense, we take the scenario of studying a macroeconomic model but, for which, there is possibly not a 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

$$E\left(\left(i_t - \rho_1 i_{t-1} - ... - \rho_p i_{t-p} - (1-\rho)\left[\alpha + \beta \pi_{t+k} + \gamma x_{t+q}\right]\right) z_t^{(i)}\right) = 0, \tag{10}$$

which most likely provide distinct GMM estimates for different i.

[4] In this model, α is identifiable but not i^* and π^*, jointly (notice that β is identified through π_{t+k}). Thus, with the argument that π^* is of some interest in the characterization of the monetary policy, following CGG, the parameter of interest subject to estimation is π^* and i^* is measured as the observed sample average.

3 Econometric Framework

In this section, we review the most important and well-established results regarding the GMM and GEL estimation procedures as well as the moment and model selection criteria in moment condition models.

3.1 Moment Conditions Model

Let $\mathcal{M}$ be the collection of candidate moment conditions models. Here, $\mathcal{M}$ is a countable/finite or an uncountable set and a model M_i belongs to the family of models $\mathcal{M} : M_i \in \mathcal{M}$. The "true" model may or may not be a member of $\mathcal{M}$. Take any particular moment conditions model, M_i, which, in our application, is characterized by a particular set of instruments (for example, model (10) in CGG setup). 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. Next, we distinguish a correctly specified model from a misspecified one, as in Hall and Inoue (2003).

Correctly Specified Model Consider the estimation of a p-dimensional parameter vector $\theta_0 = (\theta_{0,1}, ..., \theta_{0,p}) \in \Theta \subset \Re^p$ based on $m \geq p$ moment conditions of the form

$$E[g(y_t, \theta_0)] \equiv E[g_t(\theta_0)] = 0, \tag{11}$$

for all t, where, usually, $g(y_t, \theta_0) \equiv g_t(\theta_0) = \varepsilon(x_t, \theta_0) \otimes z_t$ for some set of variables x_t and instruments z_t such that $y_t = (x_t', z_t')'$. When ε_t is univariate, $g_t(\theta_0) = z_t \varepsilon(x_t, \theta_0)$ and z_t is $m \times 1$, which, for the linear regression model,[5]

$$g_t(\theta_0) = z_t \left(y_t - x_t'\theta_0\right) \text{ and } E[z_t \left(y_t - x_t'\theta_0\right)] = 0. \tag{12}$$

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

The $m \times p$ Jacobian matrix is defined as

$$G(\theta_0) \equiv G = E\left(\left. \frac{\partial g(y_t; \theta)}{\partial \theta'} \right|_{\theta = \theta_0} \right) \tag{13}$$

[5] There 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. Similarly, x_t at the monetary model is the economic variable "output gap" and not a predefined set of covariates.

and, under some regularity conditions, a CLT can be invoked:

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} g(y_t,\theta_0)\right) \xrightarrow{d} N_m\left(0, S\left(\theta_0\right)\right), \tag{14}$$

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

$$S\left(\theta_0\right) \equiv S = \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} g(y_t,\theta_0)\right]. \tag{15}$$

In the linear case,

$$G = E\left(z_t x_t^{'}\right) \text{ and } S = \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} z_t\left(y_t - x_t'\theta_0\right)\right], \tag{16}$$

which, under no-dependence (martingale difference sequence),

$$S = \Gamma_0 = E\left(g_t(\theta_0)g_t(\theta_0)'\right) = E\left(z_t z_t' \varepsilon_t^2\right). \tag{17}$$

Definition 1 (Correctly specified model): The model is said to be correctly specified if there exists a unique value θ_0 in $\Theta \subset \Re^p$ such that

$$E[g(y_t,\theta_0)] = 0. \tag{18}$$

In this definition, the orthogonality condition is $E[g(y_t,\theta_0)] = 0$ and the identification condition (uniqueness) results from G full-column ranked[6].

Misspecified Model *Definition 2 (Misspecified model): A model is said to be misspecified if there is no value of θ which satisfies the orthogonality condition, that is,*

$$E[g(y_t,\theta)] = \mu\left(\theta\right) \tag{19}$$

where $\mu : \Theta \to \Re^m$ such that $\|\mu\left(\theta\right)\| > 0$ for all $\theta \in \Theta$.

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). Also, $m > p$ because if $m = p$ then there must exist some value of θ such that $E[g(y_t,\theta)] = 0$. This is a non-local misspecification as we are not considering local misspecification where $E[g(y_t,\theta_0)] = T^{-1/2}\mu, \mu \neq 0$, say.

According to Schennach (2007), in a misspecified model $\inf_{\theta} \|E[g(y_t,\theta)]\| > 0$, whereas for linear models, Maasoumi and Phillips (1982) define misspecification by $E\left(z_t\varepsilon_t\right) = \alpha_\varepsilon$. Chen,

[6] In this paper,we are not considering the many and weak instruments issues in the linear IV model nor weak identification in the general GMM estimation procedure.

Hong and Shum (2007), among others, define a misspecified model differently. For each $\theta \in \Theta$, let $\mathcal{P}_\theta = \left\{ P | \int g(y,\theta)\, dP = 0 \right\}$ be a nonparametric family of measures for y which are consistent with the moment conditions. Then, we can define $\mathcal{P} = \cup_{\theta \in \Theta} \mathcal{P}_\theta$ as the family of measures that are compatible with the moment conditions model. The model $\mathcal{P}$ is misspecified if the true population distribution $P_0 \notin \mathcal{P}$ (in fact, $\mathcal{P}_\theta$ and $\mathcal{P}$ are induced by g and should be read as $\mathcal{P}_\theta^g$ and $\mathcal{P}^g$, instead).

So that an extremum estimator has a well defined probability limit in a misspecified model, we need to impose the following identification condition.

Assumption 1 (Identification for a misspecified model): There exists a pseudo-true value $\theta_* \in \Theta$ such that $Q_0(\theta_*) < Q_0(\theta), \forall \theta \in \Theta \backslash \{\theta_*\}$, where $Q_0(\theta)$ is the population objective function, that is,

$$\theta_* = \arg \min_\theta Q_0(\theta). \tag{20}$$

Note that two different estimators may converge to different pseudo-values (due to two different well-defined objective functions). Given the existence of θ_*, we define the following quantities:

$$\begin{aligned} \mu_* &\equiv \mu(\theta_*) = E[g(y_t, \theta_*)], \\ G_* &= E\left(\left. \frac{\partial g(y_t;\theta)}{\partial \theta'} \right|_{\theta=\theta_*} \right), \text{ and} \\ S_* &= \lim_{T\to\infty} Var\left[T^{-1/2} \sum_{t=1}^{T} (g(y_t,\theta_*) - \mu_*) \right] \text{ (positive definite).} \end{aligned} \tag{21}$$

In the linear model, $G_* = G$. Under some conditions given at Hall and Inoue (2003),

$$\sqrt{T}\left(\frac{1}{T} \sum_{t=1}^{T} (g(y_t,\theta_*) - \mu_*) \right) \xrightarrow{d} N_m(0, S_*), \text{ as } T \to \infty. \tag{22}$$

When the number of moment conditions is large, it is possible that no value of θ 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. Although the imperfections of the model can be, in some cases, avoided through the use of specification tests, the consequences on estimation may have little impact on the results. Also, misspecified (and parsimonious) models may have reasonable predictive properties. Just like in MLE, the object of interest is the pseudo-true value of the parameter vector, which may not be unique since distinct objective functions may be specified.

3.2 Estimation Procedures

In order to estimate (consistently and efficiently) the unknown quantities θ_0 or θ_*, we discuss the typical estimation procedures in moment condition models: GMM (IV for linear models), CUE and GEL. For a sample of size T, define the sample counterparts of the population moments as

$$\begin{aligned}\widehat{g}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} g(y_t,\theta), \widehat{G}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(y_t,\theta)}{\partial\theta'} \text{ and} \\ \widehat{S}_T(\theta) &= \text{(HAC formula. See Den Haan and Levin, 1996, for example).}\end{aligned} \tag{23}$$

For the linear model,

$$\widehat{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} z_t\left(y_t - x_t'\theta\right) \text{ and } \widehat{G}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} z_t x_t' = \frac{Z'X}{T}. \tag{24}$$

The GMM estimator is defined as

$$\widehat{\theta}_{GMM,T}(W) = \arg\min_{\theta\in\Theta} \widehat{g}_T(\theta)' W_T \widehat{g}_T(\theta), \tag{25}$$

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 $\sqrt{T}\left(\widehat{\theta}_{GMM,T} - \theta_0\right)$ depends on the $p\lim W_T = W$. For the two-step efficient estimator, $W_T = \widehat{S}_T\left(\widehat{\theta}_{FS}\right)^{-1} \equiv \widehat{S}_T^{-1}$, where $\widehat{\theta}_{FS}$ is a first-step consistent GMM estimator (take $W_T = I_m$, for example):

$$\widehat{\theta}_{EGMM,T} = \arg\min_{\theta\in\Theta} \widehat{g}_T(\theta)' \widehat{S}_T^{-1} \widehat{g}_T(\theta), \tag{26}$$

where the random matrix $\widehat{S}_T \xrightarrow{p} S$, as $T \to \infty$. The GMM estimator depends on the weighting matrix $\widehat{S}_T^{-1}$ which is influenced by the choice made for a first step consistent estimator. To overcome this issue, the continuous-updating (CU-GMM) objective function contains the weighting matrix as a function itself of the unknown θ :

$$\widehat{\theta}_{CUE,T} = \arg\min_{\theta\in\Theta} \widehat{g}_T(\theta)' \widehat{S}_T^{-}(\theta)\, \widehat{g}_T(\theta), \tag{27}$$

where $\widehat{S}_T^{-}$ is the generalized inverse of $\widehat{S}_T$.

Solving for θ at FOC of the GMM objective function for the linear model,

$$\left(\frac{1}{T}\sum_{t=1}^{T} z_t x_t'\right)' W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t\left(y_t - x_t'\theta\right)\right) = 0, \tag{28}$$

we obtain the IV estimator

$$\widehat{\theta}_{IV,T} = \left(\left(X'Z\right) W_T \left(Z'X\right)\right)^{-1} \left(X'Z\right) W_T \left(Z'y\right). \tag{29}$$

For the time-series version, $W_T = \widehat{S}_T^{-1}$ evaluated at $\widehat{\theta}_{FS} = \left((X'Z)(Z'X)\right)^{-1}(X'Z)(Z'y)$, whereas for the homoskecedastic with an error variance of one and no-dependence cross-section version, $W_T = \left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right)^{-1} = \left(\frac{Z'Z}{T}\right)^{-1}$ (first step estimator calculated using the inverse of an instrument cross product matrix as the weighting matrix).

Given the often disappointing small sample properties of the GMM estimator, alternative methods have been proposed recently such as those belonging to the GEL class of estimators. Newey and Smith (2004) demonstrate that, while GMM and GEL estimators have identical first-order asymptotic properties, the latter are able to eliminate some sources of GMM biases. In particular, the bias of the EL estimator does not grow with the number of moment conditions, unlike GMM. Similar properties have been recently established by Anatolyev (2005), in a time series setting. This author demonstrates that, in the presence of correlation in $g(y_t, \theta)$, the smoothed GEL estimator of Kitamura and Stutzer (1997) is efficient, obtained by smoothing the moment function with the truncated kernel, so that it solves the saddle point problem

$$\widehat{\theta}_{SGEL,T} = \arg\min_{\theta\in\Theta}\sup_{\lambda\in\Lambda_T(\theta)}\frac{1}{T}\sum_{t=K_T+1}^{T-K_T}\rho[\lambda' g_{tT}(\theta)] \tag{30}$$

with

$$g_{tT}(\theta) = \frac{1}{2K_T+1}\sum_{k=-K_T}^{K_T} g\left(y_{t-k}, \theta\right). \tag{31}$$

Here,

$$\Lambda_T(\theta) = \{\lambda : \lambda' g_{tT}(\theta) \in O, t = K_T+1, ..., T-K_T\}, \tag{32}$$

where the open set O includes the zero number and λ is the $m \times 1$ vector of lagrange multipliers each associated with the j^{th} moment condition, $j = 1, ..., m$. Moreover, the real function $\rho : \Re \rightarrow \Re$ is twice differentiable and concave on O and defines the specific GEL estimator. When $\rho(v) = -(1+v)^2/2$, the GEL estimator coincides with the CUE of Hansen et al (1996). If $\rho(v) = \ln(1-v)$ we have the EL estimator of Kitamura (1997), whereas $\rho(v) = -\exp(v)$ leads to the ET case presented by Kitamura and Stutzer (1997). In the *i.i.d.* case, the Newey and Smith (2004) GEL typology sets $K_T = 0$. The SEL variant, in particular, removes important sources of bias associated with the GMM, namely the correlation between the moment function and its derivative[7], as well as third-order biases. Furthermore, Anatolyev (2005) shows that even when there is no serial correlation, using smoothing and an appropriate HAC weight matrix, as in Andrews (1991) or Newey and West (1994), leads to a reduction in estimation biases.

[7] This correlation leads to an increasing bias deterioration as the number of moment conditions increase.

It has been shown that, under some regularity conditions, the GMM and the GEL estimators have some probability limit and converge in distribution to some random variable under a correct or misspecified model. These results are summarized in the following lines.

Correctly Specified Model Under correct model specification, the estimators are consistent: $\widehat{\theta}_{GMM,T} \xrightarrow{p} \theta_0$ and $\widehat{\theta}_{GEL,T} \xrightarrow{p} \theta_0$, as $T \to \infty$. Also, the EGMM and the GEL are (first-order) equivalent:

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\lambda} \end{pmatrix} \xrightarrow{d} N_{p+m}\left(0, diag\left(V, P\right)\right), \tag{33}$$

where

$$V = \left(G^{'}S^{-1}G\right)^{-1} \text{ and } P = S^{-1} - S^{-1}GVG^{'}S^{-1}. \tag{34}$$

Recall that for linear IV, $G = E\left(z_t x_t'\right)$ and S depends on the properties of $\{z_t\varepsilon_t\} \equiv \{g_t\}$. For the GMM estimator with an arbitrary weighting matrix W_T,

$$\sqrt{T}\left(\widehat{\theta}_{GMM,T} - \theta_0\right) \xrightarrow{d} N\left(0, \left(G^{'}WG\right)^{-1}\left(G^{'}WSWG\right)\left(G^{'}WG\right)^{-1}\right). \tag{35}$$

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,

$$\left(G^{'}S^{-1}G\right)^{-1} = \left(E\left(x_t z_t'\right)S^{-1}E\left(z_t x_t'\right)\right)^{-1} \tag{36}$$

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. The GEL estimator also attain this efficiency bound. It can also be shown that adding moment conditions improves asymptotic efficiency but it increases the small sample bias (and it can increase the small sample variance).

Misspecified Model The limiting distribution theory of the GMM estimator under misspecification is derived by Hall and Inoue (2003). Its importance can be justified through the rejection of the model using the J-statistic and the need to keep the whole set of moment conditions. The combination of overidentification and misspecification leads to a GMM estimator whose $p\lim$ depends on the limit of the weigthing matrix and whose limiting distribution depends on the limiting distribution of the elements of the weigthing matrix (its rate of convergence included). As a consequence, there is no one single limiting distribution theory for the

GMM estimator[8]. This fact has also been proved by Maasoumi and Phillips (1982) for the IV estimator in linear models and a particular weighting matrix - the matrix $\left(\frac{Z'Z}{T}\right)^{-1}$.

The identification condition states that there exists $\theta_*(W) \in \Theta$ such that $Q_0(\theta_*(W)) < Q_0(\theta), \forall \theta \in \Theta \backslash \{\theta_*(W)\}$, where $Q_0(\theta) = E[g(y_t, \theta)]' W E[g(y_t, \theta)]$. That is, the GMM estimator $\widehat{\theta}_T$ is consistent for the pseudo-true value

$$\theta_*(W) = \arg\min_{\theta} E[g(y_t, \theta)]' W E[g(y_t, \theta)], \tag{37}$$

where

$$\widehat{Q}(\theta) = \frac{1}{T} J_T(\theta) = \widehat{g}_T(\theta)' W_T \widehat{g}_T(\theta) \xrightarrow{p} E[g(y_t, \theta)]' W E[g(y_t, \theta)] = Q_0(\theta) \tag{38}$$

uniformly in θ, if $W_T \xrightarrow{p} W$.

Hall and Inuoe (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, $\sqrt{T}\left(\widehat{\theta}_T - \theta_*\right)$ converges in distribution to a normal process with zero expectation and a variance that depends on several quantities and distinct from the correctly specified model. The first case includes the FS estimation, $W_T = I_m$, and the second case includes another FS estimator, $W_T = \left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right)^{-1} \xrightarrow{p} E(z_t z_t') = W$, and a second step estimator based on the assumption that $\{z_t(y_t - x_t'\theta_*) - \mu_*\}$ is a martingale difference sequence,

$$W_T = \left(\frac{1}{T}\sum_{t=1}^{T}\left[z_t\left(y_t - x_t'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]\left[z_t\left(y_t - x_t'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]'\right)^{-1} \tag{39}$$

where $\widehat{\theta}_T(1)$ denotes the GMM estimator on the first-step such that $p\lim \widehat{\theta}_T(1) = \theta_*(1)$ and $\widehat{\theta}_T(1) - \theta_*(1) = O_p\left(T^{-1/2}\right)$ and $\widehat{\mu}_{*1} = \widehat{g}_T\left(\widehat{\theta}_T(1)\right)$. In the third case, W_T is the inverse of a centred HAC estimator, $W_T = \widehat{S}_T\left(\widehat{\theta}_T(1)\right)^{-1}$, where

$$\widehat{S}_T\left(\widehat{\theta}_T(1)\right) = \sum_{i=-T+1}^{T-1} \omega(i/b_T)\, \widetilde{\Gamma}_i, \tag{40}$$

with

$$\widetilde{\Gamma}_i = \begin{cases} \frac{1}{T}\sum_{t=i+1}^{T}\left[z_t\left(y_t - x_t'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]\left[z_{t-i}\left(y_{t-i} - x_{t-i}'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]', & \text{if } i \geq 0 \\ \frac{1}{T}\sum_{t=-i+1}^{T}\left[z_{t+i}\left(y_{t+i} - x_{t+i}'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]\left[z_t\left(y_t - x_t'\widehat{\theta}_T(1)\right) - \widehat{\mu}_{*1}\right]', & \text{if } i < 0 \end{cases} \tag{41}$$

Let $\theta_*(2) = \theta_*\left(S_*^{-1}\right)$. If the bandwidth *does not increase too quickly,* $\sqrt{\frac{T}{b_T}}\left(\widehat{\theta}_T(2) - \theta_*(2)\right)$ converges in distribution to a normal process with zero or non-zero expectation and a certain

[8] The iterated GMM changes its distribution at each iteration! Also, inference on the pseudo-true values is troubling. Finally, we no longer have first-order equivalence.

variance; Otherwise, $b_T^k\left(\widehat{\theta}_T - \theta_*\right) \xrightarrow{p}$constant (degenerates), where $k > 0$ is the characteristic exponent of the kernel. Finally, the case where W_T is the inverse of a uncentred HAC estimator: $b_T\left(\widehat{\theta}_T - \theta_*\right)$ converges in probability to a constant, in most of the cases.

Contrary to the GMM procedure, the limiting properties of the GEL class of estimators under misspecification has not been fully derived yet. Recently, Schennach (2007) provides some results for the EL and the ET estimators under some specific conditions. She proves that the EL estimator may cease to be $\sqrt{T}$-consistent in the *i.i.d.* setting under model misspecification and unbounded moment conditions (relevant for the linear IV estimator) even when their expectations are bounded. The objective function $Q_0(\theta)$ is a KLIC discrepancy measure for which the EL pseudo-value λ_* does not exist because $\log\left(1 + \lambda' g(y_t, \theta)\right)$ is ill-defined for unbounded g. Without a non-zero λ_*, one cannot define a EL pseudo-true θ_* which satisfies the model moment conditions. The existence and definition of a EL pseudo-true value for which the EL estimator converges is still to be discussed. For bounded conditions and misspecification, $\sqrt{T}$-consistent of the EL estimator is possible.

In contrast, the ET estimator avoids this problem because, even with unbounded moments, its objective does not restrict the values for λ_*. Therefore, the ET is more robust than EL under misspecification since their pseudo-true values are well-defined[9]. In her paper, Schennach proposes a hybrid estimator, the so-called ETEL (Exponentially Tilted Empirical Likelihood), that combines the EL and the ET estimators to exhibit the advantages of both. Under misspecification, she shows that the ETEL avoids EL's pitfalls maintaining root $\sqrt{T}$ convergence with pseudo-true values (θ_*, λ_*) that are generically well defined.

3.3 Moments and Model Selection Criteria

Due to the well-known bias/variance trade-off in any estimation method, in this section we present the main existing results on how to choose among a finite number of instruments/moment conditions in the GMM and GEL setup. Donald and Newey (2001) discuss on 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 GMM and GEL literature, the choice of moments is achieved according to some general

[9] Schennach (2007) refers to a 2000 paper by Yuichi Kitamura to justify the existence of a finite asymptotic variance at the $\sqrt{T}$-rate limiting distribution of the ET under misspecification.

information criteria instead[10]. The procedures described below assume that the selected moment conditions (the correct model itself) is not misspecified. On the other hand, the papers by Hall and Inoue (2003) and Schennach (2007) on misspecified models do not treat the issue of model selection.

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

$$MSC_T(c) = J_T(c) - \kappa_T(|c| - p), \tag{42}$$

where $c \in \Re^m$ is a moment selection vector that represents a list of "selected" moment conditions (subset of g), $|c|$ denotes the cardinality (number) of the "selected" moments c ($|c| \leq m$), $J_T(c)$ 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

$$C = \left\{c \in \Re^m \backslash \{0\} : c_j = 0 \text{ or } 1, \forall 1 \leq j \leq m, \text{ where } c = (c_1, ..., c_m)'\right\}, \tag{43}$$

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

$$J_T(c) = T \inf_{\theta \in \Theta} \widehat{g}_{Tc}(\theta)' W_T(c) \widehat{g}_{Tc}(\theta) = T \widehat{g}_{Tc}\left(\widehat{\theta}_T(c)\right)' W_T(c) \widehat{g}_{Tc}\left(\widehat{\theta}_T(c)\right), \tag{44}$$

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

The moment selection criteria estimator is defined as

$$\widehat{c}_{msc} = \arg\min_{c \in \mathcal{C}} MSC_T(c) = \arg\min_{c \in \mathcal{C}} \left(J_T(c) - \kappa_T(|c| - p)\right), \tag{45}$$

where $\mathcal{C} \subset C$, with $\{0\} \in \mathcal{C}$, is some parameter space for the moment selection vector. The estimator $\widehat{c}_{msc}$ 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

[10] Quite often, empirical information criteria are used when proved to be consistent in choosing the correct model and or when shown to equal the (unknown) MSE up to a constant.

some technical conditions, Andrews (1999) shows that $\widehat{c}$ is a consistent[11] estimator of c_0, assumed to be the single "correct" selection vector. If, additionally, one assumes that $E\left(g_{c_0}(\theta)\right)=0$ has a unique solution $\theta_0 \in \Theta$ (the "true" value of θ, set at c_0), then $\widehat{c}$ consistently estimates both c_0 and θ_0.

At Andrews (1999), the selection of corrects 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. Now, let (b,c) denote a pair of model and moment selection vectors and $|b|$ and $|c|$ denote the number of parameter from the vector θ (not necessarily all of them) and moments, respectively, selected by (b,c). The MMSC selects the pair (b,c) that minimizes $J_T(b,c)-\kappa_T(|c|-|b|)$, where

$$J_T(b,c)=T\inf_{\theta_{[b]}\in\Theta_{[b]}}\widehat{g}_{Tc}\left(\theta_{[b]}\right)'W_T(b,c)\,\widehat{g}_{Tc}\left(\theta_{[b]}\right)=T\widehat{g}_{Tc}\left(\widehat{\theta}_T(b,c)\right)'W_T(b,c)\,\widehat{g}_{Tc}\left(\widehat{\theta}_T(b,c)\right). \tag{46}$$

Here, $\widehat{\theta}_T(b,c)\in\Theta_{[b]}\subseteq\Theta$ is the GMM estimator based on the model and moments selection (b,c). It can be shown that the pair $\left(\widehat{b}_{mmsc},\widehat{c}_{mmsc}\right)$ is a consistent estimator. Hong, Preston and Shum (2003) extend these two previous papers to the GEL framework: At the definition of MMSC, replace $J_T(b,c)$ by

$$GEL_T(b,c)=2T\min_{\theta_{[b]}}\sup_{\lambda_c}\frac{1}{T}\sum_{t=K_T+1}^{T-K_T}\rho\left(\lambda_c' g_{tTc}(\theta_{[b]})\right), \tag{47}$$

where $g_{tTc}\left(\theta_{[b]}\right)=\frac{1}{2K_T+1}\sum_{k=-K_T}^{K_T}g_c\left(y_{t-k},\theta_{[b]}\right)$.

4 Model Averaging Instruments

So far, the literature on moment conditions models has essentially focused on estimation methods for *a given model* and on optimally selecting *a model* among a list of candidate alternatives. In this section, we build upon the principle that gains can be obtained once we consider all the available moment conditions in hand and average them out to obtain an alternative estimator. Although the setup could be defined for general g functions, we take the special case of linear IV moment conditions because averaging instruments makes the study more interesting and appealing. After defining some key concepts and optimality criteria we present instrument averaging under correct and misspecified models which, in this case, imply valid and invalid instruments.

[11] The GMM-AIC is not consistent and it has positive probability (even asymptotically) of selecting too few moments.

If one is not willing to do instrument averaging without dropping invalid instruments, then previous to the analysis of model averaging under correct specification one can do instruments selection as proposed by Andrews (1999), Andrews and Lu (2001) and Hong, Preston and Shum (2003).

4.1 Definitions

For a m-dimensional vector $z_t, t = 1, ..., T$ of available (valid or invalid) instruments, with $m \geq p$, define a $p \times m$ matrix $\pi_z \equiv \pi$ of p linear combinations of the instruments such that $z_t^e = \pi z_t$ is p-dimensional. Define the set of all possible instrument averages (up to a constant)

$$\mathcal{Z}_t = \{z_t^e : z_t^e = \pi z_t, \text{ for some } p \times m \text{ matrix } \pi \text{ such that } \pi \neq 0 \text{ and } \pi_{11} = 1\}. \tag{48}$$

For a well-defined criteria, the goal is to find an optimal weight $\widehat{\pi}$ that give rise to a selected vector $\widehat{z}_t^e \in \mathcal{Z}_t$ in a way that the estimation of an overidentified system is reduced to one that is exactly identified. We build optimal instruments instead of selecting instruments (as in Andrews, 1999 and Donald and Newey, 2001, among others). In our optimality criteria, the resulting estimator ought to be consistent and, whenever possible, attain the Chamberlain efficiency bound relative to the set z_t. When averaging instruments we do not consider the standard information criteria since all the instruments are assumed to be used: $c = \iota_m$ (vector of ones) and $|c| = m$.

4.2 Correct Specification

The identification condition under a correctly specified model does not change with a linear transformation of the instruments. In fact, if θ_0 is unique in model (12) then the same parameter of interest solves

$$E\left[z_t^e\left(y_t - x_t'\theta_0\right)\right] = \pi E\left[z_t\left(y_t - x_t'\theta_0\right)\right] = 0, \tag{49}$$

for any given π. Moreover,

$$G(\pi) \equiv G_\pi = E\left(z_t^e x_t'\right) = \pi E\left(z_t x_t'\right) = \pi G \tag{50}$$

and

$$S(\pi) \equiv S_\pi = \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} z_t^e\left(y_t - x_t'\theta_0\right)\right] = \pi S \pi', \tag{51}$$

both $p \times p$ matrices. The IV estimator for this exactly identified model (the weight matrix W_T does not play any role) is known to equal

$$\widehat{\theta}_{\pi,T} = \left(Z^{e\prime}X\right)^{-1} Z^{e\prime}y = \left(\pi Z'X\right)^{-1}\pi Z'y = \left(\pi\sum_{t=1}^{T} z_t x_t'\right)^{-1}\left(\pi\sum_{t=1}^{T} z_t y_t\right) \tag{52}$$

with an asymptotic variance given by

$$V_\pi = G_\pi^{-1} S_\pi G_\pi'^{-1} = (\pi G)^{-1} \pi S \pi' \left(G' \pi'\right)^{-1}. \tag{53}$$

To find the optimal matrix π, we need the estimator $\widehat{\theta}_{\pi,T}$ to have an asymptotic variance that equal the Chamberlain efficiency bound relative to the set z_t, $\left(E\left(x_t z_t'\right) S^{-1} E\left(z_t x_t'\right)\right)^{-1} = \left(G' S^{-1} G\right)^{-1}$. The result is presented in the following Theorem.

Theorem 1 (MA instruments in correctly specified models): Assume that model (12) is correctly specified. The efficient general IV estimator (29) coincides to the IV estimator with MA instruments

$$\widehat{\theta}_T = \left(\widehat{Z}' X\right)^{-1} \widehat{Z}' y = \left(\sum_{t=1}^T \widehat{z}_t x_t'\right)^{-1} \left(\sum_{t=1}^T \widehat{z}_t y_t\right), \tag{54}$$

where the optimal instruments are given by $\widehat{z}_t = \widehat{\pi}^o z_t$ *with weights*

$$\widehat{\pi}^o = \left(\frac{1}{T} \sum_{t=1}^T x_t z_t'\right) \widehat{S}_T^{-1} \tag{55}$$

which attains the efficiency bound $\left(E\left(x_t z_t'\right) S^{-1} E\left(z_t x_t'\right)\right)^{-1}$.

Proof: It is straightforward to show that

$$(\pi G)^{-1} \pi S \pi' \left(G' \pi'\right)^{-1} = \left(G' S^{-1} G\right)^{-1} \tag{56}$$

is solved for $\pi^o = E\left(x_t z_t'\right) S^{-1} = G' S^{-1}$, up to a constant. See also Anatoliev (2005) for more details. QED.

The J-statistic with MA instruments is the same as the original one:

$$\begin{aligned} J_T\left(\pi^o\right) &= T \inf_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^T \pi^o z_t \left(y_t - x_t'\theta\right)\right)' S^{-1}\left(\pi^o\right) \left(\frac{1}{T} \sum_{t=1}^T \pi^o z_t \left(y_t - x_t'\theta\right)\right) \\ &= T \inf_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^T z_t \left(y_t - x_t'\theta\right)\right)' S^{-1} \left(\frac{1}{T} \sum_{t=1}^T z_t \left(y_t - x_t'\theta\right)\right) = J_T \end{aligned} \tag{57}$$

because

$$\pi^{o\prime} \left(\pi^o S \pi^{o\prime}\right)^{-1} \pi^o = S^{-1} G \left(G' S^{-1} S S^{-1} G\right)^{-1} G' S^{-1} = S^{-1} G \left(G' S^{-1} G\right)^{-1} G' S^{-1} \tag{58}$$

equals S^{-1} when we post(pre)-multiply both sides by G' (G).

The matrix π^o is not necessarily an "averaging" matrix. So that π^o is some sort of (instrument) weighting matrix it would have to be true that $\pi_{ij}^o \geq 0$ and $\sum_j \pi_{ij}^o = 1$ for all $i = 1, ..., p$ combinations. That is not usually the case for $\pi^o = E\left(x_t z_t'\right) S^{-1}$, which depends on $E\left(x_t z_t'\right)$ and S^{-1}. Recall that the GEL and the GMM estimators are first-order equivalent. For these reasons, we do not discuss here the GEL procedure.

4.3 Misspecification

Building MA instruments under the assumption that not all of them are valid but none to be discarded is more challenging than under correct model specification. The task remains the same: Reduce the m-dimensional misspecified model to a p-dimensional model according to some optimally criteria. Regardless of the estimation method, by averaging the instruments we no longer have a misspecified model because when $m = p$ there must exist some value of θ such that $E[g(y_t, \theta)] = 0$. That is, the original model is misspecified

$$E[z_t\left(y_t - x_t'\theta\right)] = \mu\left(\theta\right), \tag{59}$$

where $\mu : \Theta \to \Re^m$ such that $\|\mu(\theta)\| > 0$ for all $\theta \in \Theta$ but the average model is correctly specified

$$E\left[z_t^e\left(y_t - x_t'\theta_{0\pi}\right)\right] = \pi E\left[z_t\left(y_t - x_t'\theta_{0\pi}\right)\right] = 0, \tag{60}$$

for any given π. Consequently, for any given π, there must exist some θ (call it, $\theta_0(\pi) \equiv \theta_{0\pi}$) such that[12]

$$\pi\mu\left(\theta_{0\pi}\right) \equiv \mu_\pi\left(\theta_{0\pi}\right) = 0_{p\times 1}, \|\mu\left(\theta_{0\pi}\right)\| > 0 \text{ for } \theta_{0\pi} \in \Theta. \tag{61}$$

Consider the GMM estimation procedure. Contrary to the well-specified model case, it is not guaranteed that the pseudo-true value

$$\theta_*\left(W\right) = \arg\min_\theta E\left[z_t\left(y_t - x_t'\theta\right)\right]' W E\left[z_t\left(y_t - x_t'\theta\right)\right], \tag{62}$$

for the larger misspecified model, coincide with the true value

$$\theta_0\left(\pi\right) = \arg\min_\theta E\left[z_t^e\left(y_t - x_t'\theta\right)\right]' E\left[z_t^e\left(y_t - x_t'\theta\right)\right] = E\left(\pi z_t x_t'\right)^{-1} E\left(\pi z_t y_t\right), \tag{63}$$

for the just-identified averaged model. For a given π,

$$E\left[z_t^e\left(y_t - x_t'\theta_{0\pi}\right)\right]' E\left[z_t^e\left(y_t - x_t'\theta_{0\pi}\right)\right] = 0, \tag{64}$$

whereas, for a given W,

$$E\left[z_t\left(y_t - x_t'\theta_{*W}\right)\right]' W E\left[z_t\left(y_t - x_t'\theta_{*W}\right)\right] = \mu\left(\theta_{*W}\right)' W \mu\left(\theta_{*W}\right) > 0, \tag{65}$$

because W is assumed to be positive definite and $\|\mu\left(\theta_{*W}\right)\| > 0$. The assumption of $rank\left(W\right) = m$ is important because it rules out the case where both true values coincide for $W = \pi'\pi$, which

[12] We maintain the assumption of identification in any misspecified or correctly specified model. Moreover, we assume that the $m-p$ free variables of the homogeneous system $\pi\mu\left(\theta_{0\pi}\right) = 0_{p\times 1}$ are not zero (the $m \times 1$ solution $\mu\left(\theta_{0\pi}\right)$ have no zero component at $\theta_{0\pi}$) so that the condition $\|\mu\left(\theta\right)\| > 0$ for all $\theta \in \Theta$ is not violated.

is of reduced rank, $rank\left(\pi'\pi\right) = p < m$. Despite different values at the objective function, both true values coincide for the following mapping between π and W :

Theorem 2 (True-values in correctly and misspecified models): For

$$\pi = E\left(x_t z_t'\right) W, \tag{66}$$

the true value $\theta_0\left(\pi\right)$ and the pseudo-true value $\theta_*\left(W\right)$ coincide.

Proof: Solving the FOC $E\left(x_t z_t'\right) W E\left[z_t\left(y_t - x_t'\theta\right)\right] = 0$ with respect to θ, we have

$$\theta_*\left(W\right) = \left[E\left(x_t z_t'\right) W E\left(z_t x_t'\right)\right]^{-1} E\left(x_t z_t'\right) W E\left(z_t y_t\right), \tag{67}$$

which equals $\theta_0\left(\pi\right) = E\left(\pi z_t x_t'\right)^{-1} \pi E\left(z_t y_t\right)$ when $\pi = E\left(x_t z_t'\right) W$ noting that $E\left(E\left(x_t z_t'\right) W z_t x_t'\right) = E\left(x_t z_t'\right) W E\left(z_t x_t'\right)$. QED.

So that the two GMM estimators (overidentified and misspecified model and the just-identified and correctly specified model) converge in probability to the same value, $\pi = E\left(x_t z_t'\right) W$. Hence, in this case, for a given W, we have

$$E\left(x_t z_t'\right) W \mu\left(\theta_{*W}\right) = 0_{p\times 1} \text{ where } \theta_{*W} = (67) \text{ with } \left\|\mu\left(\theta_{*W}\right)\right\| > 0. \tag{68}$$

For the efficient case we saw previously that $W = S_*^{-1}$ implying $\pi^o = E\left(x_t z_t'\right) S_*^{-1}$. Naturally, for $\widehat{z}_t = \widehat{\pi} z_t$, where

$$\widehat{\pi} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t z_t'\right) W_T, \tag{69}$$

the two GMM estimators coincide (see (29) and (54) with $\widehat{z}_t$ defined in the previous line). For $W_T = \widehat{S}_T\left(\widehat{\theta}_T(1)\right)^{-1}$, defined by (40), this is a two-step estimator and it attains the efficiency bound $\left(E\left(x_t z_t'\right) S_*^{-1} E\left(z_t x_t'\right)\right)^{-1}$.

According to Hall and Inoue (2003), the distribution and its rate of convergence depends on W (limiting distribution of the elements of W_T including its rate of convergence). We first consider the cases where the general IV estimator is $\sqrt{T}$-consistent. In case (i), $W_T = W$ for all T and $\sqrt{T}\left(\widehat{\theta}_T - \theta_*\right) \xrightarrow{d} N(0, \Sigma_1)$, where

$$\Sigma_1 = \left(E\left(x_t z_t'\right) W E\left(z_t x_t'\right)\right)^{-1} \begin{pmatrix} E\left(x_t z_t'\right) W \Omega_{11} W E\left(z_t x_t'\right) + \\ E\left(x_t z_t'\right) W \Omega_{12} + \Omega_{21} W E\left(z_t x_t'\right) \\ +\Omega_{22} \end{pmatrix} \left(E\left(x_t z_t'\right) W E\left(z_t x_t'\right)\right)^{-1} \tag{70}$$

and Ω_{ij} are the asymptotic variances-covariances of the processes $T^{-1/2}\sum_{t=1}^{T}\left(z_t\left(y_t - x_t'\theta_*\right) - \mu_*\right)$ and $\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} z_t x_t' - E\left(z_t x_t'\right)\right)' W \mu_*$ (in terms of notation, $\Omega_{11} \equiv S_*$). To have the same asymptotic efficiency as the estimator without misspecification,

$$\begin{aligned} V_W &= (\pi G)^{-1} \pi S \pi' \left(G'\pi'\right)^{-1} \\ &= \left(E\left(x_t z_t'\right) W E\left(z_t x_t'\right)\right)^{-1} E\left(x_t z_t'\right) W S W E\left(z_t x_t'\right) \left(E\left(x_t z_t'\right) W E\left(z_t x_t'\right)\right)^{-1}, \end{aligned} \tag{71}$$

it must hold

$$E\left(x_t z_t'\right) W S W E\left(z_t x_t'\right) + E\left(x_t z_t'\right) W \Omega_{12} + \Omega_{21} W E\left(z_t x_t'\right) + \Omega_{22} = E\left(x_t z_t'\right) W S W E\left(z_t x_t'\right), \tag{72}$$

that is,

$$D_1 \equiv E\left(x_t z_t'\right) W \Omega_{12} + \Omega_{21} W E\left(z_t x_t'\right) + \Omega_{22} = 0, \tag{73}$$

a condition that is hardly met. The efficiency "difference" D_1 will let us conclude how misspecification contaminates the efficiency of the GMM estimator. If D_1 is positive definite then efficiency drops; otherwise, if D_1 is negative definite misspecification leads to efficiency gains. For a given W, if $E\left(x_t z_t'\right) W \Omega_{12}$ is PD, then so it is D_1 because Ω_{22} is PD. In case (ii), $\sqrt{T}\left(W_T - W\right)$ converges to a normal distribution. Then, $\sqrt{T}\left(\widehat{\theta}_T - \theta_*\right) \xrightarrow{d} N(0, \Sigma_2)$, where Σ_2 is similar to Σ_1 but includes extra terms (asymptotic variances-covariances of $\sqrt{T}\left(W_T - W\right)\mu_*$ and the two processes in case (i)) in the inside brackets. In this case, the efficiency "difference" is

$$\begin{aligned} D_2 \equiv\; & D_1 + E\left(x_t z_t'\right) \Omega_{33} E\left(z_t x_t'\right) + E\left(x_t z_t'\right) W \Omega_{13} E\left(z_t x_t'\right) \\ & + E\left(x_t z_t'\right) \Omega_{31} W E\left(z_t x_t'\right) + \Omega_{23} E\left(z_t x_t'\right) + E\left(x_t z_t'\right) \Omega_{32}. \end{aligned} \tag{74}$$

With respect to the last two cases, the rate of convergence of the general IV estimator is smaller than $\sqrt{T}$ or it is even degenerated. In case (iii), where W_T is the inverse of a centred HAC estimator, the limiting law depends on the rate of increase of the bandwidth b_T, converging to a normal (can even have a nonzero mean) or being degenerated. In case (iv), where W_T is the inverse of a uncentred HAC estimator, $b_T\left(\widehat{\theta}_T - \theta_*\right)$ converges in probability to a constant in most of the cases. For more details, see Section 2.3 of this paper or Hall and Inuoe (2003).

Now, let $\pi \neq E\left(x_t z_t'\right) W$ for all possible W, up to a constant. In this case, the two estimators do not coincide nor converge in probability to the same value, which makes any comparison unreasonable and purely mathematical. The general IV estimator (29) converges in probability to (67) and has the above distribution, which depends on W, whereas the IV estimator for the just-identified model with (restricted) MA instruments has the following properties.

Definition 3 (IV estimator with restricted MA instruments): For a given $p \times m$ weight matrix π such that $\pi \neq E\left(x_t z_t'\right) W$ for all possible W, up to a constant, where $W = p\lim W_T$ of the *overidentified and misspecified model, define the IV estimator with MA instruments as*

$$\widehat{\theta}_{T\pi} = \left(Z^{e\prime}X\right)^{-1} Z^{e\prime}y = \left(\sum_{t=1}^{T} \pi z_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} \pi z_t y_t\right), \tag{75}$$

where $z_t^e = \pi z_t$ *and*

$$\sqrt{T}\left(\widehat{\theta}_{T\pi} - \theta_{0\pi}\right) \xrightarrow{d} N\left(0, V_\pi\right), \tag{76}$$

where $\theta_{0\pi} = (63)$ *and*

$$V_\pi = G_\pi^{-1} S_\pi G_\pi^{\prime -1} = \left(\pi E\left(z_t x_t'\right)\right)^{-1} \pi S \pi' \left(E\left(x_t z_t'\right)\pi'\right)^{-1} \neq V_W. \tag{77}$$

Very often, $\widehat{\theta}_{T\pi}$ is unfeasible (not arbitrarily given nor observable) and a consistent estimator for π is required. Ruling out the efficient case where $\widehat{\pi} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t z_t'\right) \widehat{S}_T \left(\widehat{\theta}_T(1)\right)^{-1}$, or any other $\widehat{\pi}$ such that $p\lim \widehat{\pi} = \pi = E\left(x_t z_t'\right) W$, the asymptotic distribution is derived through

$$\sqrt{T}\left(\widehat{\theta}_{T\widehat{\pi}} - \theta_{0\pi}\right) = -\left[\left(\frac{1}{T}\sum_{t=1}^{T} \widehat{\pi} z_t x_t'\right)'\right]^{-1} \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \widehat{\pi} z_t \left(y_t - x_t'\theta_{0\pi}\right)\right), \tag{78}$$

under the assumption that a CLT can be invoked for the process $\{\pi z_t \left(y_t - x_t'\theta_{0\pi}\right)\}$ for some $\pi = p\lim \widehat{\pi}$.

The theory behind GEL estimators using averaged instruments is not pursued due to the lack of established results under misspecified results. As explained in the third chapter, Schennach (2007) proves that the EL estimator may cease to be $\sqrt{T}$-consistent but a pseudo-true value for which the EL estimator converges is not well-defined besides its limit distribution being unknown. With respect to the ET we only know that it is $\sqrt{T}$-consistent. This is a topic that deserve further developments as, contrary to the GMM case, the GEL pseudo-true value does not necessarily depend on a matrix such as W.

5 Model Averaging Estimators

In model averaging instruments we are not averaging estimators but only instruments. In that approach, we estimate the model once after obtaining an optimal set of instruments. The quantity of interest is this best linear combination of instruments. In this section, we present a methodology where we average a list of candidate estimators to obtain a truly averaged one. The optimal weights associated with each estimator are to be chosen according to some criteria. By the fact that the criteria need not be unique, it is important to notice that more than one weighted estimator can be defined. In moment conditions models, averaged estimators can be discussed for general g functions, which have the linear IV as a special case. Next, we define a model averaging estimator and then we define the criteria to find the weights and discuss the statistical properties of the averaged estimator under correct and misspecified models.

5.1 Definitions

Following some previous notation, let m denote the full set of available moment conditions and $c \in \Re^m$ a moment selection vector that belongs to the unit-simplex

$$C = \left\{ c \in \Re^m \backslash \{0\} : c_j = 0 \text{ or } 1, \forall 1 \leq j \leq m, \text{ where } c = (c_1, ..., c_m)' \right\}. \tag{79}$$

Quantities such as $\widehat{g}_{Tc}(\theta), W_{Tc}$ and $\widehat{\theta}_{Tc}$ are obtained after deleting the moments corresponding to $c_j = 0$. For example, $\widehat{g}_{Tc}(\theta)$ is a $|c| \times 1$ vector.

Let $\omega = \left(\omega_1, ..., \omega_{|C|}\right)'$ be a weight vector in the unit-simplex in $\Re^{|C|}$, where

$$|C| = 2^m - \sum_{j=0}^{p-1} \begin{pmatrix} m \\ j \end{pmatrix} = \sum_{j=p}^{m} \begin{pmatrix} m \\ j \end{pmatrix}, \tag{80}$$

with the binomial coefficients $\begin{pmatrix} m \\ j \end{pmatrix} = \frac{m!}{j!(m-j)!}$ which reads m *choose* j[13], is the number of different elements at C :

$$H_m = \left\{ \omega \in [0,1]^{|C|} : \sum_{c \in C} \omega_c = 1 \right\}. \tag{81}$$

A model average estimator of the unknown $p \times 1$ vector θ is

$$\widehat{\theta}_T(\omega) = \sum_{c \in C} \omega_c \widehat{\theta}_{Tc}. \tag{82}$$

Clearly, no model average occurs when $\omega_{c^*} = 1$ for some c^* and $\omega_{c'} = 0$ for $c' \neq c^*$ in which $\widehat{\theta}_T(\omega) = \widehat{\theta}_{Tc^*}$.

For an arbitrarily given ω we have an estimator $\widehat{\theta}_T(\omega)$. But ω is assumed to be unknown and, therefore, needs to be estimated according to some criterion. In this paper, we suggest two alternative data-based criteria to optimally find estimated weights $\widehat{\omega}$ with corresponding averaged estimate

$$\widehat{\theta}_T(\widehat{\omega}) = \sum_{c \in C} \widehat{\omega}_c \widehat{\theta}_{Tc}. \tag{83}$$

[13] To the total of combinations 2^m we need to exclude $\sum_{j=0}^{p-1} \begin{pmatrix} m \\ j \end{pmatrix}$, those for which $m < p$.

The first criteria is based on the asymptotic distribution of $\widehat{\theta}_T(\omega)$, which depends on ω. Whenever $\widehat{\theta}_T(\omega)$ is $\sqrt{T}$-consistent, ω can be chosen such that it minimizes the asymptotic variance (and the MSE as well). In this scenario, we need to distinguish the cases of correct model specification and of model misspecification. This is dealt bellow in the text. The analysis of the distributional properties of $\widehat{\theta}_T(\omega)$ is also useful for the understanding of the asymptotic distribution of the model averaging estimators that arise from any other criterion. The minimization of a weighted asymptotic variance-covariance matrix can be related to one of the moment selection criteria suggested by Hall et al (2007). Contrary to Andrews (1999), they suggest selecting a model according to the relevant moment selection criterion

$$RMSC_T(c) = \ln\left(\left|\widehat{V}_c\right|\right) + \kappa_T(|c| - p)/T, \tag{84}$$

where the efficient GMM variance-covariance matrix $\widehat{V}_c$ is evaluated at $\widehat{\theta}_{Tc}$. This note can also be relevant to what follows next.

In the other criterion, the selection of the weight vector ω is based on the existing moment selection criteria, namely, the $MSC_T(c)$ (see (42)) and the $GEL_T(c)$ (see (47)), evaluated at the estimator $\widehat{\theta}_T(\omega)$. The empirical MSC selected weight vector is defined as

$$\widehat{\omega} = \arg\min_{\omega\in H_m} MSC_{T\overline{c}}(\omega) = \arg\min_{\omega\in H_m}\left(J_{T\overline{c}}(\omega) - \kappa_T(|\overline{c}| - p)\right), \tag{85}$$

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

$$J_{T\overline{c}}(\omega) = T\widehat{g}_{T\overline{c}}\left(\widehat{\theta}_T(\omega)\right)' W_{T\overline{c}}\widehat{g}_{T\overline{c}}\left(\widehat{\theta}_T(\omega)\right). \tag{86}$$

To achieve maximum efficiency, one can pick $\overline{c} = \iota_m$, a vector of ones that implies using the whole set of moment conditions (in this case, $|\overline{c}| = m$ and, in terms of notation, "c" is dropped):

$$J_T(\omega) = T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)' W_T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right). \tag{87}$$

For the linear IV case,

$$J_{T\overline{c}}(\omega) = T\left(\frac{1}{T}\sum_{t=1}^{T} z_{\overline{c},t}\left(y_t - x_t'\sum_{c\in C}\omega_c\widehat{\theta}_{Tc}\right)\right)' W_{T\overline{c}}\left(\frac{1}{T}\sum_{t=1}^{T} z_{\overline{c},t}\left(y_t - x_t'\sum_{c\in C}\omega_c\widehat{\theta}_{Tc}\right)\right). \tag{88}$$

The corresponding GEL selected weight vector (with $\overline{c} = \iota_m$ and $K_T = 0$, for sake of simplicity of exposition) is

$$\widehat{\omega} = \arg\min_{\omega\in H_m}\left(GEL_T(\omega) - \kappa_T(m - p)\right), \tag{89}$$

where

$$GEL_T(\omega) = 2T\sup_{\lambda(\omega)}\frac{1}{T}\sum_{t=1}^{T}\rho\left(\lambda(\omega)'\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)\right). \tag{90}$$

One may think of an alternative averaged MSC criterion. In this case, the empirical selected weight vector is defined as

$$\widehat{\omega} = \arg\min_{\omega \in H_m} AMSC_T(\omega) = \arg\min_{\omega \in H_m} \sum_{c \in C} \omega_c MSC_{Tc}, \tag{91}$$

where $MSC_{Tc} = J_{Tc} - \kappa_T(|c| - p)$ and

$$J_{Tc} = T\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc}\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right). \tag{92}$$

Note that, in general,

$$J_T(\omega) = T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)' W_T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right) \neq \sum_{c \in C} \omega_c T\widehat{g}_{c,T}\left(\widehat{\theta}_{Tc}\right)' W_{cT}\widehat{g}_{c,T}\left(\widehat{\theta}_{Tc}\right) = \sum_{c \in C} \omega_c J_{Tc}. \tag{93}$$

The GEL counterpart is

$$\widehat{\omega} = \arg\min_{\omega \in H_m} AGEL_T(\omega) = \arg\min_{\omega \in H_m} \sum_{c \in C} \omega_c\left(GEL_{Tc} - \kappa_T(|c| - p)\right), \tag{94}$$

where (for $K_T = 0$)

$$GEL_{Tc} = 2T \min_{\theta \in \Theta} \sup_{\lambda_c} \frac{1}{T} \sum_{t=1}^{T} \rho\left(\lambda_c' g_{Tc}(\theta)\right). \tag{95}$$

Nonetheless, this criteria is not interesting in practice because, obviously, the optimal weights are given by $\omega_{\widehat{c}} = 1$ and $\omega_{\widehat{c}} = 0$, otherwise. This is because the AMSC criteria is linear in ω and, therefore, no weight is given other than to the selected (smallest) MSC model. Hence, the MA GMM estimator coincides with the estimator for the selected model by means of the MSC.

The $MSC_T(\omega)$ criteria 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 (87)) that can be decomposed as

$$\left(\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right) - 0\right)' W_T\left(\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right) - 0\right) \tag{96}$$

where the vector $\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)$ is an estimator of $E[g_t(\theta_0)] = 0$ and for which the distance is weighted by W_T. The risk function is

$$R_T(\omega) = E(L_T(\omega)) = E\left(\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)' W_T\left(\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)\right)\right). \tag{97}$$

Therefore, for a fixed ω 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,

$$E\left(\frac{1}{T}MSC_T(\omega)\right) = E\left(\frac{1}{T}J_T(\omega)\right) + \frac{1}{T}\kappa_T(m-p) = R_T(\omega) + o(1). \tag{98}$$

For misspecified models, the relationship between $\frac{1}{T}MSC_T(\omega)$ and $R_T(\omega)$ is hard to establish[14].

With respect to the $AMSC_T(\omega)$ criteria, it is also the $J-$statistic that dominates asymptotically:

$$\frac{1}{T}AMSC_T(\omega)=\sum_{c\in C}\omega_c\frac{1}{T}J_{Tc}-\frac{1}{T}\kappa_T\sum_{c\in C}\omega_c(|c|-p)=\sum_{c\in C}\omega_c\frac{1}{T}J_{Tc}+o(1), \tag{99}$$

where $\frac{1}{T}J_{Tc}=\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)$. As long as $W_{Tc}\xrightarrow{p}W_c$,

$$\frac{1}{T}J_{Tc}(\theta)=\widehat{g}_{Tc}(\theta)'W_{Tc}\widehat{g}_{Tc}(\theta)\xrightarrow{p}E\left[g_c(y_t,\theta)\right]'W_cE\left[g_c(y_t,\theta)\right], \tag{100}$$

as $T\to\infty$, a result that holds for correct or misspecified models. Hence, $\frac{1}{T}AMSC_T(\omega)$ converges in probability to a linear combination of the $p\lim\left(\frac{1}{T}J_{Tc}\left(\widehat{\theta}_{Tc}\right)\right)'s$ and where each weight is given by ω_c. Once again, deriving $p\lim\left(\frac{1}{T}J_{Tc}\left(\widehat{\theta}_{Tc}\right)\right), c\in C$, in misspecified models is not straightforward. On the other hand, in correctly specified models, if $p\lim\left(\frac{1}{T}J_{Tc}\left(\widehat{\theta}_{Tc}\right)\right)=0$ for all $c\in C$ then ω cannot be identified asymptotically because, in this case, $\sum_{c\in C}\omega_c\frac{1}{T}J_{Tc}\to 0$ for all ω.

Any of the solutions $\widehat{\omega}$ are found by numerical algorithms. The solution solves a constraint optimization problem in which the constraints are nonnegativity ($\omega_c\in[0,1]$, for all c) and a summation that equals one ($\sum_{c\in C}\omega_c=1$). The solution $\widehat{\omega}$ can put zero weights on some of the candidate models, specially for large $|C|$ (large m). The (asymptotic) distribution of $\widehat{\omega}$ is beyond the scope of this paper. This is not an easy exercise (we do not even know for the case of least squares MA estimators - see Hansen, 2007, for details) despite its relevance for inference such as a null of $\omega_c=0, c\in C$.

Still, some open questions deserve special attention. First, does the choice of $\bar{c}$ at the criteria $MSC_{T\bar{c}}(\omega)$ condition the solution $\widehat{\omega}$? Our guess is that $\widehat{\omega}$ is not neutral to $\bar{c}$ when using $MSC_{T\bar{c}}(\omega)$. It might be reasonable to admit that $MSC_{T\bar{c}}(\omega)$ draws a solution $\widehat{\omega}$ where $\widehat{\omega}_{\bar{c}}=1:\sum_{c\in C}\widehat{\omega}_c\widehat{\theta}_{Tc}=\widehat{\theta}_{T\bar{c}}$ and $J_{T\bar{c}}(\widehat{\omega})=J_{T\bar{c}}$. Second, and maybe the most interesting question, will models with invalid instruments (misspecified moment conditions) get zero weight? If so, the criteria suggested in this paper using MA estimators can be interpreted as equivalent procedures to those of Andrews (1999), Andrews and Lu (2001), Hong, Preston and Shum (2003).

[14] Basically, the MA estimator $\widehat{\theta}_T(\omega)$ converges to a linear combination of pseudo-true values, each depending on W, and $E\left[g_t\left(p\lim\widehat{\theta}_T(\omega)\right)\right]=\mu\neq 0$ by definition.

5.2 Correct Specification

For a given ω, the limit statistical properties of $\widehat{\theta}_T(\omega) = \sum_{c\in C}\omega_c\widehat{\theta}_{Tc}$ follow a linear combination of the random processes $\widehat{\theta}_{Tc}, c \in C$. Under correct model specification, the $p\lim\widehat{\theta}_{Tc} = \theta_0$ for all $c \in C$ and $\widehat{\theta}_{Tc}$ is $\sqrt{T}$-gaussian (this is true for both GMM and GEL) and the asymptotic variance of the GMM $\widehat{\theta}_{Tc}$ is

$$V_c = \left(G_c^{'}W_cG_c\right)^{-1}\left(G_c^{'}W_cS_cW_cG_c\right)\left(G_c^{'}W_cG_c\right)^{-1}. \tag{101}$$

Recall that $G_c = E\left(z_{c,t}x_t'\right)$, for the linear IV case. The asymptotic variance of the GEL estimator and the efficient GMM $\left(W_c = S_c^{-1}\right)$ is given by

$$V_c = \left(G_c^{'}S_c^{-1}G_c\right)^{-1}. \tag{102}$$

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

Theorem 3 (Distribution of MA estimator under correct specification): Assume that the model is correctly specified. As $T \to \infty$, for any $\omega \in H_m$,

$$\widehat{\theta}_T(\omega) \xrightarrow{p} \theta_0, \tag{103}$$

where $\widehat{\theta}_{Tc}, c \in C$, is the GMM or the GEL estimator. Moreover, for the GMM estimator,

$$\sqrt{T}\left(\widehat{\theta}_T(\omega) - \theta_0\right) \xrightarrow{d} \eta = \sum_{c\in C}\omega_c\eta_c, \tag{104}$$

where the $k \times 1$ random variable $\eta_c \sim N(0, V_c)$, with $V_c = (101)$. For the efficient GMM or the GEL estimator,

$$V_c = \left(G_c^{'}S_c^{-1}G_c\right)^{-1}. \tag{105}$$

The limit process η is gaussian with zero expectation.

Proof: Consistency follows from

$$\widehat{\theta}_T(\omega) = \sum_{c\in C}\omega_c\widehat{\theta}_{Tc} \xrightarrow{p} \sum_{c\in C}\omega_c\theta_0 = \theta_0. \tag{106}$$

The asymptotic distribution follows from the limiting law for $\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_0\right)$ noting that $\sqrt{T}\left(\widehat{\theta}_T(\omega) - \theta_0\right)$ equals

$$\sqrt{T}\left(\sum_{c\in C}\omega_c\widehat{\theta}_{Tc} - \sum_{c\in C}\omega_c\theta_0\right) = \sum_{c\in C}\omega_c\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_0\right) \tag{107}$$

because $\sum_{c\in C}\omega_c = 1$. Alternatively, the result can be shown by taking the FOC for a given $c \in C$,

$$\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right) = 0, \tag{108}$$

and expand $\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)$ around $\widehat{g}_{Tc}\left(\theta_0\right)$ using the Mean Value Theorem:

$$\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc}\widehat{g}_{Tc}\left(\theta_0\right) + \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc}\widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\left(\widehat{\theta}_{Tc} - \theta_0\right) = 0, \tag{109}$$

where $\overline{\theta}_{Tc}$ is some value "between" $\widehat{\theta}_{Tc}$ and θ_0. Rearranging terms,

$$\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_0\right) = -\left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc}\widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc}\sqrt{T}\widehat{g}_{Tc}\left(\theta_0\right), \tag{110}$$

where, for the linear IV case,

$$\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right) = \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right) = -\frac{1}{T}\sum_{t=1}^{T} z_t x_t'. \tag{111}$$

Finally, η is gaussian because it is a linear combination of normal variables and it has zero expectation because $E\left(\eta_c\right) = 0, c \in C$. QED.

In general, the variance of η does not equal the (squared) weighted sum of variances V_c, that is,

$$V\left(\eta\right) = V\left(\sum_{c\in C}\omega_c\eta_c\right) \neq \sum_{c\in C}\omega_c^2 V\left(\eta_c\right). \tag{112}$$

This is because there are pairs $c_1, c_2 \in C, c_1 \neq c_2$, such that η_{c_1} and η_{c_2} are not independent, in particular when c_1 and c_2 have common moment conditions. It is also not unusual to have distinct moment conditions with positive correlation.

In terms of the smallest $V\left(\eta\right)$ it is not clear that it is attained for $\omega_{c^*}^* = 1$, where $c^* = \iota_m$, and $\omega_{c'}^* = 0$, for all $c' \neq c^*$ (the MA estimator coincides with the estimator obtained using the whole set of moment conditions). For sake of simplicity, consider the efficient case (102). We conjecture that it is not always true that, for all p-dimensional vector $\xi \neq 0$,

$$\xi' V\left(\eta^*\right)\xi \leq \xi' V\left(\eta\right)\xi, \tag{113}$$

where

$$V\left(\eta^*\right) = V\left(\eta_{\iota_m}\right) = E\left(\eta_{\iota_m}\eta_{\iota_m}'\right) = \left(G' S^{-1} G\right)^{-1} \tag{114}$$

and

$$\begin{aligned} V\left(\eta\right) &= E\left(\left(\sum_{c\in C}\omega_c\eta_c\right)\left(\sum_{c\in C}\omega_c\eta_c\right)'\right) \\ &= \sum_{c\in C}\omega_c^2\left(G_c' S_c^{-1} G_c\right)^{-1} + \sum_{\substack{c_1,c_2\in C \\ c_1\neq c_2}}\omega_{c_1}\omega_{c_2}E\left(\eta_{c_1}\eta_{c_2}'\right). \end{aligned} \tag{115}$$

In practice, finding the argument $\widehat{\omega}$ that minimizes the asymptotic variance of $\sqrt{T}\left(\widehat{\theta}_T(\omega)-\theta_0\right)$ can be achieved by means of the minimization of the trace of $V(\eta)$:

$$tr\left(V\left(\eta\right)\right)=\sum_{c\in C}\omega_c^2.tr\left(\left(G_c'S_c^{-1}G_c\right)^{-1}\right)+\sum_{\substack{c_1,c_2\in C\\ c_1\neq c_2}}\omega_{c_1}\omega_{c_2}tr\left(E\left(\eta_{c_1}\eta_{c_2}'\right)\right). \tag{116}$$

Replacing $V(\eta_c)=\left(G_c'S_c^{-1}G_c\right)^{-1}$ by a consistent estimator, larger weights given to models c with more moment conditions (for efficiency matters[15]) can be offset by the covariances $E\left(\eta_{c_1}\eta_{c_2}'\right)$. Also, the optimal weight vector $\widehat{\omega}$ in terms of asymptotic variance is not necessarily the same as the best weight in small samples.

In summary, despite knowing the asymptotic law of the MA estimator, uniformly in ω, the variance criterion to choose the optimal vector $\widehat{\omega}$ is not that handy in practice. Fortunately, there are alternative criteria based on the existing moment selection criteria $MSC_T(c)$ and $GEL_T(c)$ although $\widehat{\omega}$ needs to be found by numerical algorithms and whose (asymptotic) properties are yet to be known.

5.3 Misspecification

In this section, we assume model misspecification in the sense that, for $c^*=\iota_m$,

$$E[g_{c^*}(y_t,\theta)]=\mu_{c^*}(\theta), \tag{117}$$

where $\mu_{c^*}:\Theta\to\Re^m$ such that $\|\mu_{c^*}(\theta)\|>0$ for all $\theta\in\Theta$, although there might exist some other $\widetilde{c}\neq c^*$ such that

$$E[g_{\widetilde{c}}(y_t,\theta_0)]=0. \tag{118}$$

Here, we assume that not all the moment conditions are necessarily introducing model misspecification (this is more likely to happen in the case of instruments). Adding conditions may increase (asymptotic) efficiency but will create model misspecification and bias. In terms of notation, we may use either g_{ι_m} or g (similar for μ_{c^*} among others). We restrict attention to the MA GMM estimator, following the results by Hall and Inoue (2003). As explained before, Schennach (2007) does not provide an asymptotic theory for the GEL estimator under misspecification and the ETEL estimator is beyond the scope of this paper[16].

[15] Recall that adding moment conditions improves asymptotic efficiency.

[16] Schennach (2007) derives the asymptotic distribution of the ETEL estimator under misspecification. It is $\sqrt{T}$-consistent and gaussian and therefore it would be interesting to study the properties of a MA ETEL estimator under model misspecification.

Under model misspecification, the GMM estimator $\widehat{\theta}_{Tc}$ is consistent for the pseudo-true value

$$\theta_{*c}(W_c) = \arg\min_{\theta} E\left[g_c(y_t,\theta)\right]' W_c E\left[g_c(y_t,\theta)\right], \tag{119}$$

where

$$\frac{1}{T} J_{Tc}(\theta) = \widehat{g}_{Tc}(\theta)' W_{Tc} \widehat{g}_{Tc}(\theta) \xrightarrow{p} E\left[g_c(y_t,\theta)\right]' W_c E\left[g_c(y_t,\theta)\right] \tag{120}$$

if $W_{Tc} \xrightarrow{p} W_c$. The way we assumed model misspecification implies that necessarily $\theta_{*\iota_m} \equiv \theta_*$ is a pseudo-true value that depends on $W_{\iota_m} \equiv W$ but there may exist a true unknown $\theta_{*c}(W_c) = \theta_0$ for some $c \neq \iota_m$ and for all W_c. Consequently, $\|\mu(\theta_*)\| > 0$ but $\mu_c(\theta_{*c}(W_c)) = \mu_c(\theta_0) = 0$, for some $c \neq \iota_m$.

Next, we derive the asymptotic properties of the MA GMM estimator $\widehat{\theta}_T(\omega) = \sum_{c\in C} \omega_c \widehat{\theta}_{Tc}$. Clearly, for a fixed ω and as $T \to \infty$,

$$p\lim \widehat{\theta}_T(\omega) = \sum_{c\in C} \omega_c . p\lim \widehat{\theta}_{Tc} = \sum_{c\in C} \omega_c \theta_{*c}(W_c) \equiv \sum_{c\in C} \omega_c \theta_{*c}. \tag{121}$$

Note that the $p\lim \widehat{\theta}_T(\omega)$ is a linear combination of the pseudo-values θ_{*c}, each depending on W_c. This means that the $p\lim \widehat{\theta}_T(\omega)$ depends on the choice for $W_{Tc}, c \in C$. That is, for each given model c one can specify different weight matrices W_{Tc} that converge in probability to distinct matrices W_c, which, consequently, will imply different probability limits for the MA estimator $\widehat{\theta}_T(\omega)$.

The rate of convergence at which $\widehat{\theta}_T(\omega)$ converge to $\sum_{c\in C} \omega_c \theta_{*c}$ is influenced by the way W_{Tc} converges to W_c, for each given model $c \in C$. In some cases, $\widehat{\theta}_T(\omega)$ 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 equality

$$a_T\left(\sum_{c\in C} \omega_c \widehat{\theta}_{Tc} - \sum_{c\in C} \omega_c \theta_{*c}\right) = \sum_{c\in C} \omega_c a_T\left(\widehat{\theta}_{Tc} - \theta_{*c}\right) = \sum_{c\in C} \omega_c \frac{a_T}{a_{Tc}} a_{Tc}\left(\widehat{\theta}_{Tc} - \theta_{*c}\right), \tag{122}$$

where the rate of convergence of the MA GMM estimator is $a_T \to \infty$, as $T \to \infty$, the MA estimator $\widehat{\theta}_T(\omega)$ is $\sqrt{T}$-gaussian whenever W_{Tc} is chosen in a way that $a_T = a_{Tc} = \sqrt{T}$ and $\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_{*c}\right)$ is gaussian for at least one $c \in C$ and when $a_{Tc}\left(\widehat{\theta}_{Tc} - \theta_{*c}\right) = O_p(1)$ and $\frac{\sqrt{T}}{a_{Tc}} = o(1)$ for the remaining models c. This covers the case of all models $c \in C$ having $\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_{*c}\right)$ –asymptotic normality. Assuming that $a_{Tc}\left(\widehat{\theta}_{Tc} - \theta_{*c}\right) = O_p(1)$ 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} :

$$\theta_{*c,T}\left(W_{cT}\right) \equiv \theta_{*c,T} = \arg\min_{\theta} E\left[g_c\left(y_t,\theta\right)\right]' W_{cT} E\left[g_c\left(y_t,\theta\right)\right]. \tag{123}$$

With $W_{Tc} \xrightarrow{p} W_c$, we have something like $\theta_{*c,T} = \theta_{*c} + o_p(1)$. 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 $\mu_c : \Theta \to \Re^m$ is such that

$$\sqrt{T}\mu_c\left(\theta_{*c,T}\right) \to 0, \text{ as } T \to \infty. \tag{124}$$

The previous assumption[17] keeps the key properties of μ_c of time-invariance and $\|\mu_c(\theta)\| > 0$ for all $\theta \in \Theta$. Now, we add the "local" condition that the sequence $\theta_{*c,T}$ is such that $\mu_c\left(\theta_{*c,T}\right) = o\left(T^{-1/2}\right)$, which means that there exists a sequence $\theta_{*c,T}$ 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, as the correctly specified case, regardless of the rate of convergence of W_{Tc} to W_c.

Theorem 4 (Distribution of MA GMM estimator under misspecification): Assume that the model is misspecified according to Assumption 1. As $T \to \infty$, for any $\omega \in H_m$,

$$\widehat{\theta}_T(\omega) \xrightarrow{p} \sum_{c \in C} \omega_c \theta_{*c}\left(W_c\right), \tag{125}$$

where $\widehat{\theta}_{Tc}, c \in C$, is the GMM estimator and $\theta_{*c} = (119)$. Moreover, for the GMM estimator,

$$\sqrt{T}\left(\widehat{\theta}_T(\omega) - \sum_{c \in C} \omega_c \theta_{*c,T}\right) \xrightarrow{d} \eta_* = \sum_{c \in C} \omega_c \eta_{*c}, \tag{126}$$

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

$$V_{*c} = \left(G_{*c}' W_c G_{*c}\right)^{-1}\left(G_{*c}' W_c S_{*c} W_c G_{*c}\right)\left(G_{*c}' W_c G_{*c}\right)^{-1}. \tag{127}$$

Here,

$$\mu_{*c} \equiv \mu_c\left(\theta_{*c}\right) = E[g_c(y_t, \theta_{*c})], \tag{128}$$

$$G_{*c} = E\left(\left.\frac{\partial g_c\left(y_t;\theta\right)}{\partial \theta'}\right|_{\theta=\theta_{*c}}\right), \text{ and} \tag{129}$$

$$S_{*c} = \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T}\left(g_c(y_t,\theta_{*c}) - \mu_{*c}\right)\right]. \tag{130}$$

[17] Local misspecification is usually defined as $S^{-1/2}E_T[g(y_t,\theta_0)] = T^{-1/2}\mu$, where μ is a vector of finite constants. See Newey (1985) or Hall (2005) for details.

For $W_c = S_{*c}^{-1}, c \in C$,

$$V_{*c} = \left(G_{*c}' S_{*c}^{-1} G_{*c} \right)^{-1}. \tag{131}$$

The limit process η_* is gaussian with zero expectation.

Proof: Consistency was shown above. To derive the asymptotic distribution we start with the MVT (see the proof of Theorem 3)

$$\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{g}_{Tc}\left(\theta_{*c,T}\right) + \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\left(\widehat{\theta}_{Tc} - \theta_{*c,T}\right) = 0, \tag{132}$$

where $\overline{\theta}_{Tc}$ is some value "between" $\widehat{\theta}_{Tc}$ and $\theta_{*c,T}$. Rearranging terms,

$$\sqrt{T}\left(\widehat{\theta}_{Tc} - \theta_{*c,T}\right) = -\left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \sqrt{T} \widehat{g}_{Tc}\left(\theta_{*c,T}\right). \tag{133}$$

Then, for a fixed ω,

$$\begin{aligned}
& \sqrt{T}\left(\widehat{\theta}_T(\omega) - \sum_{c \in C} \omega_c \theta_{*c,T}\right) \qquad (134) \\
= & -\sum_{c \in C} \omega_c \left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \sqrt{T} \widehat{g}_{Tc}\left(\theta_{*c,T}\right) \\
= & -\sum_{c \in C} \omega_c \left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \sqrt{T}\left(\widehat{g}_{Tc}\left(\theta_{*c,T}\right) - \mu_c\left(\theta_{*c,T}\right)\right) \\
& -\sum_{c \in C} \omega_c \left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \sqrt{T} \mu_c\left(\theta_{*c,T}\right) \\
= & -\sum_{c \in C} \omega_c \left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1} \widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)' W_{Tc} \sqrt{T}\left(\widehat{g}_{Tc}\left(\theta_{*c,T}\right) - \mu_c\left(\theta_{*c,T}\right)\right) + o_p(1),
\end{aligned}$$

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

$$V_{*c} = \left(G_{*c}' W_c G_{*c} \right)^{-1} \left(G_{*c}' W_c S_{*c} W_c G_{*c} \right) \left(G_{*c}' W_c G_{*c} \right)^{-1}. \tag{135}$$

QED.

A few comments are worth mentioning. The Assumption 1 is important to guarantee a zero expectation of η_*. If, instead, the sequence $\theta_{*c,T}$ is such that $\mu_c\left(\theta_{*c,T}\right) = O\left(T^{-1/2}\right)$, that is, $\sqrt{T}\mu_c\left(\theta_{*c,T}\right) \xrightarrow{p} \widetilde{\mu}_{*c} \neq 0$, for all $c \in C$, then the expectation of η_* is

$$-\sum_{c \in C} \omega_c \left(G_{*c}' W_c G_{*c} \right)^{-1} G_{*c}' W_c \widetilde{\mu}_{*c} \neq 0^{18}. \tag{136}$$

Another key feature is that the $p\lim \widehat{\theta}_T(\omega)$ can be regarded as a linear combination of a true value and pseudo-true values when it is accepted that, for some models $c \neq \iota_m$, there is correct

[18] Notice that this is still a "local" misspecification result as $\mu_c\left(\theta_{*c,T}\right) \xrightarrow{p} 0$ for all c.

specification. This follows from the decomposition

$$
\begin{aligned}
p\lim \widehat{\theta}_T(\omega) &= \sum_{c\in C} \omega_c \theta_{*c} = \sum_{c\in C_0} \omega_c \theta_0 + \sum_{c\in C_*} \omega_c \theta_{*c}(W_c) \\
&= \omega_0 \theta_0 + (1-\omega_0) \sum_{c\in C_*} \left(\frac{\omega_c}{1-\omega_0} \right) \theta_{*c}(W_c), \qquad (137)
\end{aligned}
$$

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 θ_0 the $p\lim \widehat{\theta}_T(\omega)$ gets. Finally, picking the optimal weight $\widehat{\omega}$ is even harder than the correctly specified case: The $p\lim \widehat{\theta}_T(\omega)$ and the variance of η_* depend on the arbitrarily choice for W_c for all models $c \in C$. Hopefully, the alternative criteria based on $MSC_T(c)$ and $GEL_T(c)$ lead to $\widehat{\omega}$, found by numerical algorithms.

6 Empirical Results

In this section, we estimate the CGG forward-looking monetary policy reaction function for the same period as theirs by GMM and extend the analysis to the GEL and model averaging approaches. The relative merits of each approach 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 twentieth century.

6.1 A Monetary Policy Rule

We apply the MA GMM approach to the CGG benchmark model. Recall that for any AR lag p and inflation and output delays k and q, the CGG model is given by

$$
i_t = \rho_1 i_{t-1} + ... + \rho_p i_{t-p} + (1-\rho)\left[\alpha + \beta E_t \pi_{t,k} + \gamma E_t x_{t,q}\right] + e_t, \qquad (138)
$$

where $\alpha = i^* - \beta\pi^*$. We adopt their baseline specification for which $k = q = 1$ (one period forward) and where the monetary authorities set an expected interest rate that is a linear combination of the target rate and the observed rate at the two previous periods, $p = 2$[19] :

$$
i_t = \rho_1 i_{t-1} + \rho_2 i_{t-2} + (1-\rho)\left[\alpha + \beta E_t \pi_{t+1} + \gamma E_t x_{t+1}\right] + e_t. \qquad (139)
$$

Moreover, we take their baseline inflation and "output gap" measures. These are the (annualized) rate of change of the GDP deflator between two subsequent quarters and the series constructed

[19] 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.

by the Congressional Budget Office (CBO), respectively. By fixing p, k, q and the definitions of inflation and output gap we focus, in this paper, on the differences that occur when one uses different estimation procedures (GMM versus GEL) 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 and GEL procedures. The choice of the weight matrix for the GMM methodology is also taken into account.

Using the two-step GMM estimation method with quarterly US data (1960:1-1996:4), CGG found that during the Volcker-Greenspan policy period there were more sensitive changes in expected inflation than in the pre-Volcker period (prior to 1979:3). In a smaller parameter scale, 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, γ is not statistically significant for the Volcker-Greenspan period when x is obtained over either the detrended output or the unemployment rate. Using the series constructed by the CBO, $\widehat{\gamma}$ is twice its value and γ is statistically significant in both periods. Jondeau et al (2004), using the CBO series and the GMM and Continous-Updating GMM estimation methods, obtained a wide range of estimates for γ 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 π^* seemed plausible: 4.25% for the pre-Volcker period and about 3.5% post-Volcker. Finally, the estimate of the smoothing parameter ρ is high (about 0.7 pre-Volcker and 0.8 post-Volcker) reflecting the inertia at the interest rate dynamics and that the FED smooths adjustments in its monetary instrument.

6.2 Data and List of Available Instruments

The data is the same as in CGG[20]. This way, we can compare the standard GMM results with the alternative estimation procedures such as the GEL and the model averaging techniques. We have US quarterly data for the period 1960:1-1996:4. This 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 FED's chairmen. It is argued that these two periods correspond to the unstable and stable eras of recent history. Due to lagged/leaded variables in the model, the sample period is in fact 1961:1-1996:3.

[20] We thank the authors for providing us the data used in their paper. See CGG paper for more details about the data.

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 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

$$z_{t-1} = (1, i_{t-1}, ..., i_{t-4,} \pi_{t-1}, ..., \pi_{t-4,} x_{t-1}, ..., x_{t-4,} dc_{t-1}, ..., dc_{t-4,} spr_{t-1}, ..., spr_{t-4})' . \quad (140)$$

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 has to do with the theoretical approach presented at the previous sections. For a particular model M_i, the orthogonality conditions for the baseline specification are

$$E\left(\left(i_t - \rho_1 i_{t-1} - \rho_2 i_{t-2} - (1-\rho)\left[\alpha + \beta\pi_{t+1} + \gamma x_{t+1} \right] \right) z_{t-1}^{(i)} \right) = 0, \quad (141)$$

where $z_{t-1}^{(i)}$ is a subset of z_{t-1}.

6.3 Estimation and Model Selection Criteria

We begin by discussing the results based on the GMM/GEL estimation of (141). We conduct estimations for the period 1960:1 to 1979:2 (PV stands for pre-Volcker), the more recent vintage of the data, which spans from 1979:3 to 1996:4 (VG stands for Volcker-Greenspan) and for the whole period (W stands for "whole" period from 1960:1 to 1996:4). The GMM estimation refers to the two-step efficient procedure and the Empirical Likelihood (EL) is the GEL-type method under consideration. For now, we let the number of lagged instruments to be either four (4l) or two (2l). The results are presented in Table 1. The standard errors are reported in parentheses and the p-value of the J-statistic is denoted by "J". No values are reported for the EL, 4l, VG case due to convergence problems at the estimation algorithm.

Some issues are worth mentioning. First, we observe that, overall, the estimators produce consistent and comparable results for each given period. Secondly, estimates of β and γ tend

Table 1: GMM and EL Estimates with a Fixed Number of Instruments

		π^*	ρ	β	γ	J
GMM	PV	4.681 (0.119)	0.509 (0.063)	0.970 (0.049)	0.173 (0.064)	0.969
4l	VG	3.916 (0.141)	0.357 (0.114)	1.750 (0.185)	-0.155 (0.146)	0.907
	W	4.065 (0.208)	0.711 (0.038)	1.394 (0.144)	0.361 (0.177)	0.799
	PV	4.556 (0.158)	0.554 (0.092)	0.884 (0.058)	0.139 (0.085)	0.406
2l	VG	3.839 (0.152)	0.567 (0.096)	2.396 (0.331)	0.045 (0.271)	0.855
	W	4.221 (0.290)	0.722 (0.040)	1.243 (0.233)	0.445 (0.225)	0.514
EL	PV	4.908 (0.236)	0.876 (0.029)	1.497 (0.249)	0.794 (0.229)	0.829
4l	VG	n/a	n/a	n/a	n/a	n/a
	W	4.066 (0.298)	0.874 (0.020)	1.544 (0.212)	0.610 (0.252)	0.841
	PV	5.174 (0.465)	0.687 (0.076)	0.593 (0.148)	0.321 (0.172)	0.177
2l	VG	3.815 (0.179)	0.679 (0.065)	2.874 (0.381)	0.206 (0.251)	0.208
	W	3.984 (0.338)	0.892 (0.041)	2.456 (0.646)	0.929 (0.819)	0.122

to suggested that the pre-Volcker policy was unstabilizing through inflation and the Volcker-Greenspan rule was stabilizing despite neutral with respect to the product. Third, estimates of β, γ and ρ for the whole period suggest a stabilizing FED monetary policy rule over the second half of the century (γ is marginally significant).

With the exception of the EL with 4 lags, the point estimates of β 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 and even for 2 lags these are about 2.5. This makes evidence for stabilizing rules during VG tenure and not so much for the PV period with respect to inflation. Once we consider the whole sample, point estimates of β are barely above one.

In terms of γ, we conclude that it is statistical significant for the PV period (with small but positive point estimates) but 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. For the whole sample, γ is statistically significant with slightly larger point estimates than those for the PV period.

There is evidence that the FED smooths adjustments in its monetary instrument. The inertia at the rate dynamics is similar for the two subperiods (with the exception of the GMM with 4 lags for the VG period in which $\widehat{\rho} = 0.357$, a relatively small value) despite important differences across estimation methods. The GMM estimate is about 0.55, the EL with 4 lags is 0.87 and it is 0.68 considering 2 lags. The estimates are larger for the whole sample than for any of the two subperiods. The point estimates of the inflation target π^* are relatively close to the expected ones ($4.5\% - 5\%$ during the PV period and $3.5\% - 4\%$ for the VG tenures). The J-statistic does not allow us to rejected the model for any of the estimation procedures.

Next, we performed the moment/instrument selection exercise as described by Andrews (1999). Rather than fixing the list of instruments to either 2 or 4 lags of the variables, this method allows the data to determine the "best" model out of the possible combinations of instruments. Accordingly, the instrument selection criteria estimator is defined as

$$\widehat{c}_{msc} = \arg\min_{c \in \mathcal{C}} MSC_T(c) = \arg\min_{c \in \mathcal{C}} \left(J_T(c) - \kappa_T(|c| - p) \right), \tag{142}$$

where $\kappa_T = 2$ for the AIC; $\kappa_T = \log T$ for the BIC; and $\kappa_T = 2 \log \log T$ for the HQ-type criterion. Recall that the AIC is not consistent and it has positive probability (even asymptotically) of selecting too few moments. We considered two weight matrices: The efficient one $W_T^e = \widehat{S}_T^{-1}$, evaluated at $\widehat{\theta}_{FS} = \left((X'Z)(Z'X)\right)^{-1}(X'Z)(Z'y)$, and also $W_T^z = \left(\frac{1}{T}\sum_{t=1}^T z_t z_t'\right)^{-1} = \left(\frac{Z'Z}{T}\right)^{-1}$.

Table 2: GMM Estimates with Moment Selection Criteria

		$\widehat{c}$	π^*	ρ	β	γ	J
W_T^e	PV	4l	see Table 1				
	VG	4l	see Table 1				
	W	4l	see Table 1				
	W	4l-π_{t-3}	4.026 (0.230)	0.725 (0.035)	1.278 (0.143)	0.371 (0.182)	0.896
W_T^z	PV	4l-$dc_{t-1}, dc_{t-2}, dc_{t-4}$	4.899 (0.186)	0.721 (0.071)	0.862 (0.059)	0.426 (0.043)	0.970
	VG	n/a	n/a				
	W	n/a	n/a				

We do not have results for the EL model selection criteria because there were convergence problems at the estimation procedure for most of the models.

The total number of instrument combinations is

$$|C| = 2^m - \sum_{j=0}^{p-1} \begin{pmatrix} m \\ j \end{pmatrix} = \sum_{j=p}^{m} \begin{pmatrix} m \\ j \end{pmatrix} = \sum_{j=p}^{m} \frac{m!}{j!\,(m-j)!}, \tag{143}$$

which equals 2089605, in our case where $m = 21$ 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,

$$z_{F,t-1} = (1, i_{t-1}, i_{t-2,}\pi_{t-1}, \pi_{t-2,}x_{t-1}, x_{t-2})' \tag{144}$$

(the two lags of the variables of the model belong to any list of instruments) and the 14 remaining are combined to construct the z matrix. This way, it is considered $\sum_{j=0}^{14} \frac{14!}{j!(14-j)!} = 2^{14} = 16384$ different models. The same model was chosen for the three criteria, except for the case of the whole sample and the weight matrix W_T^e. In this case, the model with 4 lags (full set of instruments) was chosen for the BIC and HQ criteria but for the AIC criterion the best model was the 4 lags without π_{t-3} as an instrument (4 lags was the third best model). The results are in Table 2. The 4 lags model was the preferred specification when the weight matrix is W_T^e (see Table 1 for the economic implications of the results). For the case of W_T^z and VG and W periods, all models have a zero pvalue for the J-statistic and due to the non-acceptance of the models we do not report the results. For the PV period, (most of) the lags of the commodity price inflation were taken out in order to obtain the selected model. Relatively to the model with four lags, the former specification gave rise to larger point estimates of ρ and γ.

6.4 MA Estimators

Now, we apply the model averaging method described earlier in the paper to the baseline forward-looking monetary policy rule. The MA estimator is given by

$$\widehat{\theta}_T\left(\widehat{\omega}\right) = \sum_{c\in C} \widehat{\omega}_c \widehat{\theta}_{Tc}, \tag{145}$$

where the optimal weights $\widehat{\omega}_c$ are estimated by some criteria and $\widehat{\theta}_{Tc}$ is a GMM or GEL estimator for model c[21]. We basically proposed two criteria for choosing $\widehat{\omega}$. One results from the minimization of a consistent estimator of

$$tr\left(V\left(\eta\right)\right) = \sum_{c\in C} \omega_c^2 .tr\left(\left(G_c' S_c^{-1} G_c\right)^{-1}\right) + \sum_{\substack{c_1,c_2\in C \\ c_1\neq c_2}} \omega_{c_1}\omega_{c_2} tr\left(E\left(\eta_{c_1}\eta_{c_2}'\right)\right). \tag{146}$$

Because we are not able to derive the second quantity, we only minimize

$$\sum_{c\in C} \omega_c^2 .tr\left(\left(G_c' S_c^{-1} G_c\right)^{-1}\right). \tag{147}$$

It is clearly assumed that some bias exists in the estimation of ω without the covariance terms. The other criteria is

$$\widehat{\omega} = \arg\min_{\omega\in H_m} MSC_{T\overline{c}}\left(\omega\right) = \arg\min_{\omega\in H_m} \left(J_{T\overline{c}}\left(\omega\right) - \kappa_T\left(\left|\overline{c}\right| - p\right)\right), \tag{148}$$

for a fixed model $\overline{c}$. We considered $\overline{c} = \iota_m$ (4 lags) and $\overline{c} =$ "F" (only the fixed instruments, $Z_{\overline{c}} = Z_F$).

For computational reasons and so that the estimated weights $\widehat{\omega}_c$ were not excessively small and, for that reason, meaningless for interpretation, we only considered the 100 models with smallest $MSC_T\left(c\right) - HQ$ (note that $1/16384 = 6.103 \times 10^{-5}$). All of this best models have p-values for the J-statistic (way) larger than 10% and, therefore, one can invoke the mathematical and statistical properties of the MA GMM estimator under correct specification that we have shown earlier in this paper. The MA parameter and weight estimation results are displayed in Tables 3 and 4, respectively. In the first place, we observe extremely similar results for $\overline{c} = \iota_m$ and $Z_{\overline{c}} = Z_F$ and for the Trace criteria, in some extension. A possible explanation is that the MA GMM estimator is averaging out the 100 different point estimators, which are common for distinct criteria.

[21] Due to the convergence problems with the EL, we only report the results for the MA GMM estimator. Also, we only considered W_T^e as the GMM weight matrix due to rejection of the models with W_T^z, as explained at the previous subsection.

Table 3: MA GMM Estimates

		π^*	ρ	β	γ
	PV	4.703	0.569	0.962	0.241
Trace	VG	3.876	0.448	1.895	-0.066
	W	4.114	0.722	1.298	0.375
MSC	PV	4.699	0.557	0.966	0.231
$\bar{c} = \iota_m$	VG	3.879	0.408	1.853	-0.081
	W	4.109	0.713	1.293	0.368
MSC	PV	4.699	0.557	0.966	0.231
$Z_{\bar{c}} = Z_F$	VG	3.879	0.408	1.854	-0.082
	W	4.109	0.713	1.293	0.369

The point estimates confirm the main conclusions drawn from standard methods. There is evidence support for 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 β are (slightly) bellow one and (way) above one for the PV and VG periods, respectively, and the estimates for γ are positive and negative (probably not statistical significant since the sign is not the expected one) for the PV and VG periods, respectively. The value for $\widehat{\gamma}$ for the PV period is larger than the one obtained from standard estimation procedures. For the whole sample, there is evidence of stabilizing rules. The point estimates for π^* and ρ are similar to those from standard methods, noting that $\widehat{\rho}$ by MA GMM is a value between the standard estimates with 4 and 2 lags.

In terms of the estimated weights, a number of interesting results stand out. First, the model that gets the largest estimated weight is never the selected one by means of the standard MSC. The model with the largest estimated weight is consistently the one with 4 lags without two or three instruments. The model selected by MSC can be ranked from number 12 (with $\widehat{\omega}_{\widehat{c}} = 0.1008$ - clearly above 0.01) to number 93 (with $\widehat{\omega}_{\widehat{c}} = 0.007$ - just bellow 0.01). Note secondly that the range of estimated weights is relatively wide for the trace criterion but not so much for the MSC criteria and that it does not differ for different sample periods. Finally, we conclude that although the trace and the MSC criteria provide very similar point estimates, the corresponding estimated weights do not seem to be exactly the same across models.

Table 4: MA GMM Weights

		ω	$\omega_{\widehat{c}}$ (rank)	$\widehat{c} : \max \omega$
Trace	PV	$[0.007, 0.029]$	0.007 (93)	4l-dc_{t-1}, dc_{t-4}
	VG	$[0.006, 0.030]$	0.010 (20)	4l-$dc_{t-2}, spr_{t-3}, x_{t-3}$ 4l-i_{t-3}, dc_{t-2}
	W	$[0.008, 0.023]$	0.010 (34)	4l-dc_{t-1} 4l-π_{t-4}, dc_{t-1}
MSC	PV	$[0.00979, 0.01014]$	0.00993 (76)	4l-$\pi_{t-3}, spr_{t-4}, x_{t-4}$
$\overline{c} = \iota_m$	VG	$[0.00950, 0.01111]$	0.00981 (79)	4l-$i_{t-3}, \pi_{t-4}, dc_{t-2}$
	W	$[0.00979, 0.01043]$	0.01008 (14)	4l-π_{t-4}, dc_{t-1}
MSC	PV	$[0.00978, 0.01016]$	0.00992 (78)	4l-$\pi_{t-3}, spr_{t-4}, x_{t-4}$
$Z_{\overline{c}} = Z_F$	VG	$[0.00944, 0.01109]$	0.00979 (81)	4l-$i_{t-3}, \pi_{t-4}, dc_{t-2}$
	W	$[0.00973, 0.01058]$	0.01008 (12)	4l-π_{t-4}, dc_{t-1}

6.5 Targeting the Interest Rate

Following CGG, we now measure how well the estimated target rules

$$i_t^* = i^* + \widehat{\beta}\left(E_t \pi_{t+1} - \widehat{\pi}^*\right) + \widehat{\gamma} E_t x_{t+1}, \tag{149}$$

where the point estimates are found by standard methods and by the MA GMM approach, characterize the behavior of the actual Funds rate[22]. For matter of comparison, we also consider the Taylor-type rule (not forward-looking and without unknown coefficients) as in Woodford (2001):

$$i_t^* = i^* + 1.5\left(\pi_t - \pi^*\right) + 0.5 x_t. \tag{150}$$

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

$$i_t^* = r^* + \pi_t + 0.5\left(\pi_t - \pi^*\right) + 0.5 x_t, \tag{151}$$

where $r^* = \pi^* = 0.02$. At the 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 effectively to inflation giving top priority

[22] 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 5: Error Measures when Targeting the Interest Rate: Standard Approaches

		MSE	RMSE	MAD
TR	PV	6.716	2.591	2.214
	VG	7.663	2.768	2.097
	W	11.445	3.383	2.643
GMM 4l	PV	2.351	1.533	1.292
	VG	7.100	2.665	2.005
	W	11.614	3.408	2.616
GMM 2l	PV	2.103	1.450	1.199
	VG	11.784	3.433	2.443
	W	11.440	3.382	2.558
$W_T^z, \widehat{c}$ =4l-$dc_{t-1}, dc_{t-2}, dc_{t-4}$	PV	2.837	1.684	1.410
$W_T^e, \widehat{c}$ =4l-π_{t-3}	W	11.071	3.327	2.557
EL 4l	PV	9.733	3.119	2.651
	VG	n/a	n/a	n/a
	W	15.209	3.900	3.008
EL 2l	PV	2.459	1.568	1.198
	VG	17.415	4.173	2.871
	W	32.222	5.676	4.472

to price-presures 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 at the real economy) and $\gamma > 0$. Otherwise, the monetary policy may generate an inflation spiral.

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 5 and 6.

Overall, the results are relatively poor once we observe a RMSE of around three and a MAD of approximately 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 a bit off the observed value.

Table 6: Error Measures when Targeting the Interest Rate: MA GMM Approach

		MSE	RMSE	MAD
	PV	2.460	1.568	1.327
Trace	VG	7.934	2.817	2.101
	W	11.127	3.336	2.550
MSC	PV	2.449	1.565	1.323
$\overline{c} = \iota_m$	VG	7.698	2.774	2.074
	W	11.051	3.324	2.542
MSC	PV	2.449	1.565	1.323
$Z_{\overline{c}} = Z_F$	VG	7.700	2.775	2.075
	W	11.055	3.324	2.542

Moreover, the target rules consistently do better during the PV period than during the VG tenures. The worst performances occur for the entire sample. 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 rule does better during the VG than the PV period but this may be because Taylor studied the 1987-1992 period originally.

As expected, the performance of the MA GMM approach is essentially the same for different criteria because the point estimates were very similar. For the whole sample, the MA GMM beats all other methods. For the PV and VG periods, the best performance is not achieved by the MA GMM method but it gets very close to best one and, in fact, it outperforms the EL approach. See Table 7 for details on the ranking performances.

In general, the best results are obtained by GMM and the models selected by the standard MSC (usually, 4 lags of instruments) do as well as the GMM with 2 lags. The EL and the TR (even for the VG period) are not as good as the GMM methodologies, averaged or standard. In conclusion, the MA GMM procedure is as valid as the standard GMM approaches to estimate the FED's target rule. To illustrate, we present in Figures 1 and 2 the MA GMM 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[23].

[23] CGG present the same plots for GMM in pages 158 and 159. The truth is that for the VG period our MA GMM captures the swings better than theirs.

Table 7: Error Measures when Targeting the Interest Rate (Ranked)

		RMSE		MAD
PV	GMM2l	1.450	EL2l	1.198
	GMM4l	1.533	GMM2l	1.199
	MAGMM	1.565	GMM4l	1.292
	EL2l	1.568	MAGMM	1.323
	TR	2.591	TR	2.214
VG	GMM4l	2.665	GMM4l	2.005
	TR	2.768	MAGMM	2.074
	MAGMM	2.774	TR	2.097
	GMM2l	3.433	GMM2l	2.443
	EL2l	4.173	EL2l	2.871
W	MAGMM	3.324	MAGMM	2.542
	GMM 2l	3.382	GMM2l	2.558
	TR	3.383	GMM4l	2.616
	GMM4l	3.408	TR	2.643
	EL4l	3.900	EL4l	3.008

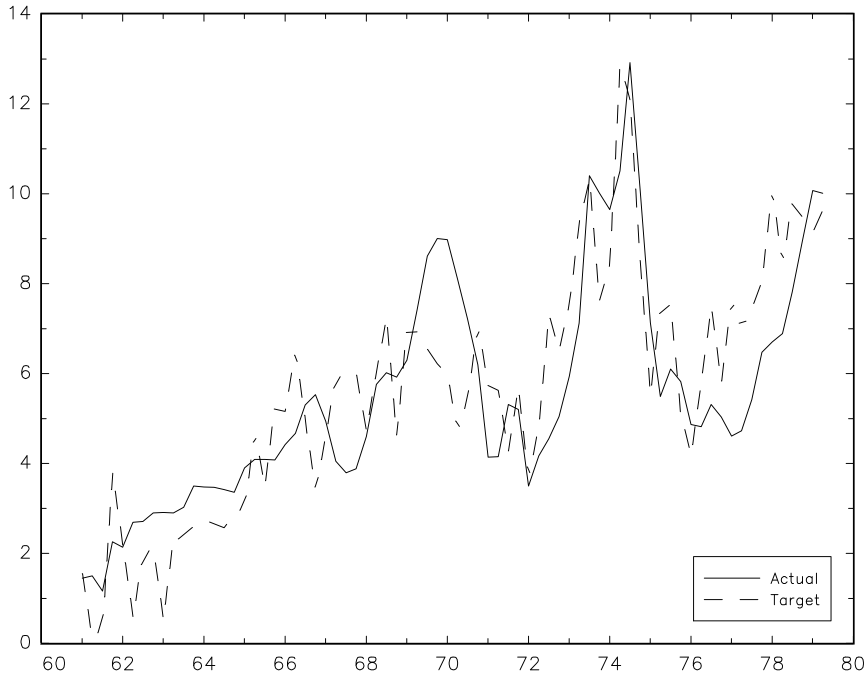

Figure 1: Actual versus Target Rates: Pre-Volcker Era and MA GMM Estimates

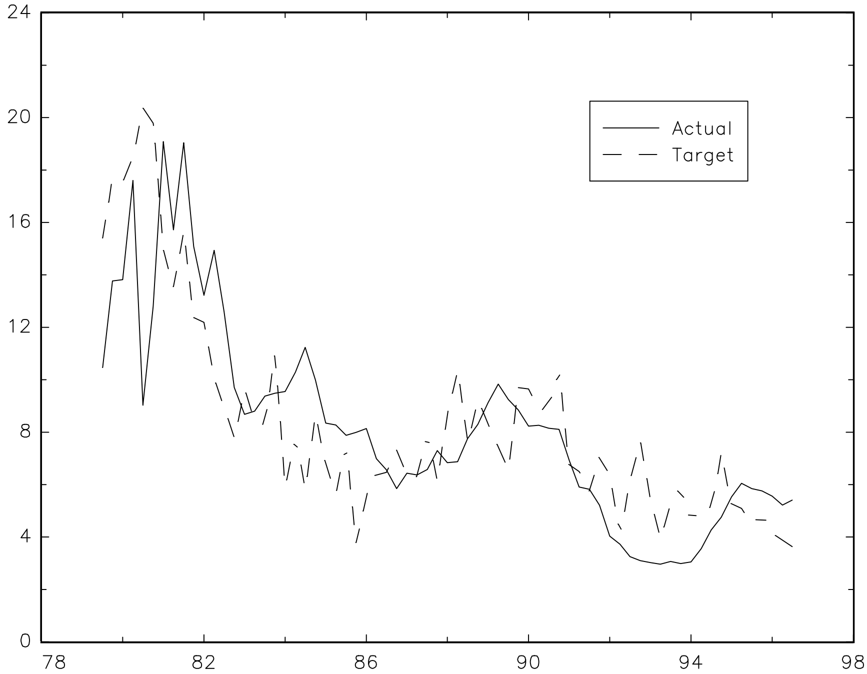

Figure 2: Actual versus Target Rates: Volcker-Greenspan Era and MA GMM Estimates

7 Conclusion

The United States monetary policy in the postwar era has become of particular interest in macroeconomics. In this paper, we revisit the model and the results presented in Clarida, Gali, and Gertler (2000) by employing a new estimation procedure, which we call Model Averaging estimator. Their baseline forward-looking monetary policy reaction function is estimated by moment condition procedures but we resort to procedures that averages instruments and estimators. This way, we focus on the potential gains in averaging estimators rather than assuming a fixed moment specification as implicitly assumed by standard techniques.

Thus, we define GMM and GEL model averaging estimators and discuss some of their asymptotic properties under correctly specified and misspecified models. 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. The MA estimators are a weighted average of a list of standard GMM or GEL estimators for each possible model specification. The optimal weights are found by means of particular moment and model selection criteria that share some good statistical properties. We also discuss the existence of a linear combination (average) of instruments that results in a GMM/GEL estimator that attains the

Chamberlain efficiency bound relative to the set of all 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 by standard approaches: Efficient GMM with fixed instruments and with instruments selected by information criteria and by Empirical Likelihood. The methods point to similar conclusions where, according to the point estimates, there is evidence for a stabilizing policy rule during the Paul Volcker and Alan Greenspan tenures, albeit neutral to the cyclical output variable, but not so during the pre-Volker period with respect to inflation.

That is, before Volcker came to office, the FED raised the nominal interest rates by a smaller proportion than the increase of expected inflation. This would led to a decline of short-run real interest rates. In short, it seems that the FED's primary objective was growth as the funds rate responded to output fluctuations and not so much the control for prices. On the contrary, Volcker and Greenspan tenures were characterized by an anti-inflationary policy. By increasing the nominal rates by more than the expected inflation, the real interest rates tent to raise as well. During this period, there is quantitative evidence for a monetary policy that did not react to product fluctuations.

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 method do as well as the standard approaches and the Taylor rule for the Volcker-Greenspan period and that it does even better than the EL. The target rules consistently do better during the pre-Volcker period than during the stabilizing Volcker-Greenspan tenures. 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 this 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, such as the New Keynesian model. Also, it would be important to investigate further on the statistical properties of the MA estimator, and to compare them to the standard approaches, by means of Monte Carlo experiments. We leave this for future research.

References

[1] Anatolyev, S. (2005), "GMM, GEL, Serial Correlation and Asymptotic Bias," *Econometrica*, 73, 983-1002.

[2] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica,* 59, 817-858.

[3] Andrews, D. W. K. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica,* 67, 543-564.

[4] Andrews, D. W. K. and Lu, B. (2001), "Consistent Model and Moment Selection Criteria for GMM Estimation with Applications to Dynamic Panel Data Models," *Journal of Econometrics,* 101, 123-164.

[5] Caner, M. (2009), "LASSO Type GMM Estimator" *Econometric Theory*, 25, 1-23.

[6] Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics,* 34, 305-334.

[7] Chen, X., Hong, H. and Shum, M. (2007), "Nonparametric Likelihood Ratio Model Selection Tests Between Parametric Likelihood and Moment Condition Models," *Journal of Econometrics*, 141, 109-140.

[8] Clarida, R., Gali, J. and Gertler, M. (1998), "Monetary Policy Rules in Practice: Some International Evidence," *European Economic Review*, 42, 1033-1067.

[9] Clarida, R., Gali, J. and Gertler, M. (2000), "Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory," *The Quarterly Journal of Economics*, 115, 147-180.

[10] Den Haan, W. and Levin, A. (1996), "A Practitioner's Guide to Robust Covariance Matrix Estimation," Technical Working Paper Series, 197, NBER.

[11] Donald, S. G., Newey, W. K. (2001), "Choosing the Number of Instruments," *Econometrica,* 69, 1161-1192.

[12] Hall, A. R. (2005), "Generalized Method of Moments," Oxford, UK: Oxford University Press.

[13] Hall, A. R. and Inoue, A. (2003), "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361-394.

[14] Hall, A. R., Inoue, A., Jana, K., Shin, C. (2007), "Information in Generalized Method of Moments Estimation and Entropy Based Moment Selection," *Journal of Econometrics,* 138, 488-512.

[15] Hansen, B. E. (2005), "Least Squares Model Averaging," *Econometrica*, 75, 1175-1189.

[16] Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[17] Hansen, L. P., Heaton, J. and Yaron, A. (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-80.

[18] Hjort, N. L. and Claeskens, G. (2003), "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 98, 879-899.

[19] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999), "Bayesian Model Averaging: A Tutorial", *Statistical Science*, 14, 382-417.

[20] Hong, H., Preston, B. and Shum, M. (2003), "Generalized Empirical Likelihood-Based Model Selection Criteria for Moment Condition Models," *Econometric Theory*, 19, 923-943.

[21] Jondeau, E., Le-Bihan, H. and Gallès, C. (2004), "Assessing Generalized Method-of-Moments Estimates of the Federal Reserve Reaction Function," *Journal of Business and Economic Statistics*, 22, 225-239.

[22] Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes," *The Annals of Statistics*, 25, 2084-2102.

[23] Kitamura, Y. and Stutzer, M. (1997), "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861-874.

[24] Maasoumi, E. and Phillips, P.C.B. (1982), "On the Behavior of Inconsistent Instrumental Variable Estimators," *Journal of Econometrics,* 19, 183-201.

[25] Newey, W.K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics,* 29, 229-256.

[26] Newey, W. K. and Smith, R. J. (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[27] Newey, W. K. and West, K. D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.

[28] Schennach, S. M. (2007), "Point Estimation with Exponentially Tilted Empirical Likelihood," *The Annals of Statistics*, 35, 634-672.

[29] Smith, R. J. (1992), "Non-nested.Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60, 973-980.

[30] Smith, R. J. and Ramalho, J. S. (2002), "Generalized Empirical Likelihood Non-Nested Tests," *Journal of Econometrics*, 107, 99-125.

[31] Taylor, J. B. (1993), "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

[32] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica,* 50, 1-25.

[33] Woodford, M. (2001), "The Taylor Rule and Optimal Monetary Policy," *The American Economic Review*, 91, 232-237.

WORKING PAPERS

2010

1/10 MEASURING COMOVEMENT IN THE TIME-FREQUENCY SPACE
— *António Rua*

2/10 EXPORTS, IMPORTS AND WAGES: EVIDENCE FROM MATCHED FIRM-WORKER-PRODUCT PANELS
— *Pedro S. Martins, Luca David Opromolla*

3/10 NONSTATIONARY EXTREMES AND THE US BUSINESS CYCLE
— *Miguel de Carvalho, K. Feridun Turkman, António Rua*

4/10 EXPECTATIONS-DRIVEN CYCLES IN THE HOUSING MARKET
— *Luisa Lambertini, Caterina Mendicino, Maria Teresa Punzi*

5/10 COUNTERFACTUAL ANALYSIS OF BANK MERGERS
— *Pedro P. Barros, Diana Bonfim, Moshe Kim, Nuno C. Martins*

6/10 THE EAGLE. A MODEL FOR POLICY ANALYSIS OF MACROECONOMIC INTERDEPENDENCE IN THE EURO AREA
— *S. Gomes, P. Jacquinot, M. Pisani*

7/10 A WAVELET APPROACH FOR FACTOR-AUGMENTED FORECASTING
— *António Rua*

8/10 EXTREMAL DEPENDENCE IN INTERNATIONAL OUTPUT GROWTH: TALES FROM THE TAILS
— *Miguel de Carvalho, António Rua*

9/10 TRACKING THE US BUSINESS CYCLE WITH A SINGULAR SPECTRUM ANALYSIS
— *Miguel de Carvalho, Paulo C. Rodrigues, António Rua*

10/10 A MULTIPLE CRITERIA FRAMEWORK TO EVALUATE BANK BRANCH POTENTIAL ATTRACTIVENESS
— *Fernando A. F. Ferreira, Ronald W. Spahr, Sérgio P. Santos, Paulo M. M. Rodrigues*

11/10 THE EFFECTS OF ADDITIVE OUTLIERS AND MEASUREMENT ERRORS WHEN TESTING FOR STRUCTURAL BREAKS IN VARIANCE
— *Paulo M. M. Rodrigues, Antonio Rubia*

12/10 CALENDAR EFFECTS IN DAILY ATM WITHDRAWALS
— *Paulo Soares Esteves, Paulo M. M. Rodrigues*

13/10 MARGINAL DISTRIBUTIONS OF RANDOM VECTORS GENERATED BY AFFINE TRANSFORMATIONS OF INDEPENDENT TWO-PIECE NORMAL VARIABLES
— *Maximiano Pinheiro*

14/10 MONETARY POLICY EFFECTS: EVIDENCE FROM THE PORTUGUESE FLOW OF FUNDS
— *Isabel Marques Gameiro, João Sousa*

15/10 SHORT AND LONG INTEREST RATE TARGETS
— *Bernardino Adão, Isabel Correia, Pedro Teles*

16/10 FISCAL STIMULUS IN A SMALL EURO AREA ECONOMY
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix, José Francisco Maria*

17/10 FISCAL INSTITUTIONS AND PUBLIC SPENDING VOLATILITY IN EUROPE
— *Bruno Albuquerque*

18/10 GLOBAL POLICY AT THE ZERO LOWER BOUND IN A LARGE-SCALE DSGE MODEL
— *S. Gomes, P. Jacquinot, R. Mestre, J. Sousa*

19/10 LABOR IMMOBILITY AND THE TRANSMISSION MECHANISM OF MONETARY POLICY IN A MONETARY UNION
— *Bernardino Adão, Isabel Correia*

20/10 TAXATION AND GLOBALIZATION
— *Isabel Correia*

21/10 TIME-VARYING FISCAL POLICY IN THE U.S.
— *Manuel Coutinho Pereira, Artur Silva Lopes*

22/10 DETERMINANTS OF SOVEREIGN BOND YIELD SPREADS IN THE EURO AREA IN THE CONTEXT OF THE ECONOMIC AND FINANCIAL CRISIS
— *Luciana Barbosa, Sónia Costa*

23/10 FISCAL STIMULUS AND EXIT STRATEGIES IN A SMALL EURO AREA ECONOMY
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix, José Francisco Maria*

24/10 FORECASTING INFLATION (AND THE BUSINESS CYCLE?) WITH MONETARY AGGREGATES
— *João Valle e Azevedo, Ana Pereira*

25/10 THE SOURCES OF WAGE VARIATION: AN ANALYSIS USING MATCHED EMPLOYER-EMPLOYEE DATA
— *Sónia Torres,Pedro Portugal, John T.Addison, Paulo Guimarães*

26/10 THE RESERVATION WAGE UNEMPLOYMENT DURATION NEXUS
— *John T. Addison, José A. F. Machado, Pedro Portugal*

27/10 BORROWING PATTERNS, BANKRUPTCY AND VOLUNTARY LIQUIDATION
— *José Mata, António Antunes, Pedro Portugal*

28/10 THE INSTABILITY OF JOINT VENTURES: LEARNING FROM OTHERS OR LEARNING TO WORK WITH OTHERS
— *José Mata, Pedro Portugal*

29/10 THE HIDDEN SIDE OF TEMPORARY EMPLOYMENT: FIXED-TERM CONTRACTS AS A SCREENING DEVICE
— *Pedro Portugal, José Varejão*

30/10 TESTING FOR PERSISTENCE CHANGE IN FRACTIONALLY INTEGRATED MODELS: AN APPLICATION TO WORLD INFLATION RATES
— *Luis F. Martins, Paulo M. M. Rodrigues*

31/10 EMPLOYMENT AND WAGES OF IMMIGRANTS IN PORTUGAL
— *Sónia Cabral, Cláudia Duarte*

32/10 EVALUATING THE STRENGTH OF IDENTIFICATION IN DSGE MODELS. AN A PRIORI APPROACH
— *Nikolay Iskrev*

33/10 JOBLESSNESS
— *José A. F. Machado, Pedro Portugal, Pedro S. Raposo*

2011

1/11 WHAT HAPPENS AFTER DEFAULT? STYLIZED FACTS ON ACCESS TO CREDIT
— *Diana Bonfim, Daniel A. Dias, Christine Richmond*

2/11 IS THE WORLD SPINNING FASTER? ASSESSING THE DYNAMICS OF EXPORT SPECIALIZATION
— *João Amador*