GLS-IV for Time Series Regressions with Application to the "New Keynesian Phillips Curve"*

Marco Olivari[†] Pierre Perron[‡]

Boston University Boston University

August 13, 2024

Abstract

Consider a linear model $y = X\beta + u$ with $u = (u_1, ..., u_T)$ and u_t a serially correlated linear process given by $u_t = \sum_{j=-h}^{\infty} c_j e_{t-j}$ for a sequence of innovations e_t. Given a set of instruments Z, the "optimal GMM" estimator based on the moment condition $\mathbb{E}(Zu) = 0$ is by far the most commonly used method to estimate such models. It can, however, be inconsistent unless the instruments are exogenous with respect to past innovations e_{t-j} for $j > 0$, when $c_j \neq 0$ for $j > 0$. We propose a GLS-IV estimator valid in the general case with instruments exogenous or not, as long as they are predetermined. It is shown to be much more efficient than GMM whether the moment condition $\mathbb{E}(Zu) = 0$ is satisfied or not. We discuss issues of consistency by casting the estimators in a GMM framework with different moment conditions and instruments. To analyze the relative merits of the estimators when all are consistent, we cast them as some GLS estimator using different instruments and first-stage regression. It then becomes clear that GLS-IV involves "stronger instruments", while GMM is more likely to be affected by issues of weak instruments. Other motivating elements and extensive simulations are presented to argue that our proposed GLS-IV estimator has better properties. As an empirical application, we revisit the extensive study of Mavroeidis et al. (2014) about the empirical relevance of the forward looking New Keynesian Phillips curve. Using the GLS-IV procedure on the same dataset, our estimates are all in the right quadrant, consistent with theoretical expectations.

Keywords: Feasible Generalized Least-Squares, GMM, Phillips Curve, Autoregressive Filtering, Instrumental Variables.

JEL Codes: C22, C31, E27, E31, E47, E52.

*We are grateful to Emilio González-Coya and Zhongjun Qu for useful comments.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (olivari@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

"*The literature has reached a limit on how much can be learned about the New Keynesian Phillips curve from aggregate macroeconomic time series;*" Mavroeidis et al. (2014).

·Well, maybe not. But we need to revisit good old-fashioned efficiency arguments.

1 Introduction

Consider the linear model $y = X\beta + u$, where $y = (y_1, ..., y_T)$, $X = (x_1, ..., x_T)$, $x_t = (x_{1t}, ..., x_{kT})$ and $u = (u_1, ..., u_T)$. The errors are serially correlated and, for simplicity assumed to be a linear process of the form $u_t = C(L)e_t = \sum_{-h}^{\infty} c_j e_{t-j}$ for some sequence of independent random variables e_t with mean 0 and variance σ_e^2. The regressors are said to be endogenous if $\mathbb{E}(x_t e_t) \neq 0$, i.e., if at least one element of x_t is correlated with the current innovation. Suppose there is a set of instruments Z available. In this context, the generalized method of moments (GMM), introduced by Hansen (1982), is the most widely used estimator and is based on the moment condition $\mathbb{E}(Zu) = 0$. In the majority of cases, the optimal weighting matrix is used and the resulting estimator labelled "optimal GMM".

Distinguishing two cases allows to grasp the usefulness and the limitations of GMM. The "rational expectation (RE) case" occurs when $c_j = 0$ for all $j \geq 0$. Models with forward looking regressors when the rational expectation hypothesis is assumed is an example. This implies that $e_{t+1}, ..., e_{t+h}$ are expectation errors and, given the rational expectations hypothesis, they are uncorrelated with the contemporaneous instruments, hence $\mathbb{E}(Zu) = 0$ is always satisfied. The "general case" instead occurs when $c_j \neq 0$ for some $j \geq 0$. The moment condition $\mathbb{E}(Zu) = 0$ here requires, in general, instruments that are exogenous with respect to the past innovations e_{t-j} for $j > 0$ as well as the condition that $\mathbb{E}(z_t e_t) = 0$ to be valid. This requirement contrasts with that of pre-determined instruments, which are uncorrelated with current and future values of e_t. Unfortunately, it is common practice to simply assume $\mathbb{E}(Zu) = 0$ without investigating the implications about the exogeneity requirement it imposes on the instruments. The belief that no other estimators are available or the practice to think solely in terms of contemporaneous correlation between the errors e_t and the instruments, without properly investigating the temporal exogeneity issue are two possible reasons for this practice. One of the few alternative estimator proposed is the Forward Filter (FF) method of Hayashi and Sims (1983). It aims to improve upon GMM by allowing non-exogenous instruments but is valid only for the "RE case".

We propose a GLS-IV estimator as an alternative method of estimation that is valid in the general case with instruments exogenous or not, as long as they are pre-determined. It is shown to be much more efficient than GMM or FF in both the "RE case" and especially the "general case". We discuss issues of consistency by casting all estimators in a GMM framework with different moment conditions and instruments. We analyze the relative merits of the estimators when all are consistent, by casting all of them as some GLS estimator using different instruments and first-stage regression. It then becomes clear that GLS-IV involves "stronger instruments", while GMM and FF are more likely to be affected by issues of weak instruments, thereby decreasing their efficiency. We also make several other arguments suggesting that GLS-IV is likely to provide better estimates.

Our method is based on a traditional autoregressive approximation of order, say, k_T to approximate the unknown serial correlation in the errors, and we select the order using the Bayesian Information Criterion, see Schwarz (1978). We follow the method proposed by Perron and González-Coya (2023), based on a generalization of the Durbin (1970) regression, to account for the non-endogenous regressors. The method involves a few steps but is easy to implement using standard least-squares procedures. We show that, in most cases, the feasible version of our GLS-IV procedure is as good as the infeasible version with known structure and parameter values of the error process. The gains in efficiency are substantial even with exogenous instruments. This disparity becomes even more pronounced with non-exogenous instruments, with the optimal GMM exhibiting significant bias and confidence intervals that are meaningless. Our method relies on $C(L)$ being invertible. We discuss what happens when it is not. In a nutshell, as expected, all methods remain consistent when the instruments are exogenous. When they are not, GLS-IV is inconsistent, while GMM and FF are and their feasible versions provide tests with the correct nominal level.

As an empirical application, we revisit the study of Mavroeidis et al. (2014) about the empirical relevance of the forward looking New Keynesian Phillips curve (NKPC). For an early influential treatment, see Galí and Gertler (1999). Note that as they argue a feature of importance is the possibility of serially correlated cost-push shocks, in which case issues of exogenous instruments (and non-endogenous regressors) become central. The theory suggests that the expected inflation parameter, β, should be close to but smaller than one and that the marginal cost parameter, λ, also referred to as the slope of the Phillips curve, should be small and positive. Mavroeidis et al. (2014) analyzed a myriad of possibilities for various measures of inflation, the output gap or the marginal cost of labor (or labor share). They mostly used "optimal GMM" based on the moment condition $\mathbb{E}(Zu) = 0$. Their point estimates are all over the map and completely uninformative. They conclude that *"the literature has reached a limit on how much can be learned about the New Keynesian Phillips curve from aggregate* *macroeconomic time series"*. Our findings refute this claim. By using the GLS-IV procedure on the same dataset, our estimates are all in the right quadrant, values of β below 1 and of λ above 0. Our median estimates are $\lambda = 0.03$ and $\beta = 0.95$, consistent with theoretical expectations and also in line with estimates obtained with micro data. Our results suggest that, by applying an estimator that is consistent and more efficient in a wider range of cases, there is still much to be learned from aggregate macroeconomic time series data.

The estimator we propose has a long history dating back to the early 60s but which, for some reasons, the essential ingredients have been incorrectly dismissed as inappropriate, following the same arguments advanced by Perron and González-Coya (2023) in the context of using OLS versus GLS when the errors are serially correlated. Here, we briefly review such antecedents. Estimators similar to the GLS-IV have been proposed in the context of the estimation of simultaneous equations with non-spherical errors. A few prominent examples are Sargan (1961), Theil (1961), Amemiya (1966), Fair (1970), and Fair (1984). Similar to the GLS-IV, they apply a quasi-difference filter to the model to obtain spherical errors. The main difference resides in the transformation of the instruments. Our GLS-IV uses the same quasi-difference filter proposed in Theil (1961) ("Theil's method") and applies it to all instruments, while Fair (1970) applies the quasi-difference filter to the subset of instruments that appear in only one equation of a simultaneous equations system. Theil's method, which coincides with the infeasible version of GLS-IV in the case of a single equation model, has been neglected in the subsequent literature either because of hurdles in estimating the variance-covariance matrix or for incorrect reasoning about exogeneity issues. However, studies by Moazzami and Buse (1990) and Buse and Moazzami (1991) highlight its strong performance in finite samples, suggesting that it should be *"placed on equal footing with the procedure currently in use"* (i.e., GMM). The Fair method instead has found widespread application, including in the estimation of rational expectations and error-in-variable models. Nevertheless, complications arise when dealing with expectation errors, as emphasized by Flood and Garber (1980) and Cumby et al. (1983). They highlight that neither Fair's nor Theil's methods are applicable in such cases. Subsequently, researchers have been using the GMM estimator. Our approach can be viewed as an extension of Theil's method to a general stationary stochastic structure for the error process. For that, we use an autoregressive approximation, which is a standard tool nowadays. Also, we pay more attention to issues related to exogeneity of the instruments and regressors with respect to past innovations.

The rest of the paper is organized as follows. Section 2 motivates the analysis by introducing the NKPC and discusses the framework used for our analysis, carefully defining issues of exogeneity and pre-determinedness of the instruments (and regressors) with respect to the basic innovations. Section 3 reviews the GMM and FF estimators, and introduces our GLS-IV estimator. It also highlights how one can cast them in a GMM framework with different moment conditions. Section 4 provides a framework to interpret all estimators as some GLS procedure, thereby allowing to extract what are the instruments used and the implicit first-stage regression, which is useful to assess their relative efficiency. Section 6 presents a step-by-step algorithm to obtain the feasible GLS-IV estimator. Section 5 present additional advantages of GLS-IV over GMM and FF. Section 7 presents various simulation results that substantiate the various claims made throughout the technical discussion. Section 8 presents the empirical results related to the New Keynesian Phillips curve highlighting the differences between the GLS-IV and GMM estimates. Section 9 presents brief concluding remark. An online appendix discusses additional issues and provides more extensive simulations and empirical estimates to illustrate the robustness of the results.

2 Motivation and Framework

Let us consider an example that will be the focus of our empirical study in Section 8, namely the forward looking New Keynesian Phillips Curve (NKPC)

$$y_t = \beta \mathbb{E}_t[y_{t+h}] + \lambda x_t + \eta_t, \tag{1}$$

where y_t is the current level of inflation. According to New Keynesian theory, inflation is determined by three factors: future expected inflation $\mathbb{E}_t[y_{t+h}]$, the deviation of a real variable from its steady state level x_t; e.g., some measure of the output gap or the marginal cost of labor and a cost-push shock η_t. The cost-push shock can cover many things, a recent example being increases in inflation due to supply chain bottlenecks, which are in general serially correlated. Another is an oil price shock, which can also be viewed as serially correlated, i.e., having an effect that persists but gradually decreases over time. The value h can be any positive integer to represent future expected inflation; the commonly used value with quarterly data is $h = 4$ so that $\mathbb{E}_t[y_{t+4}]$ is the expected inflation one year ahead.

As $\mathbb{E}_t[y_{t+h}]$ is unobservable, we cannot directly estimate the original NKPC equation (1). Therefore, we need to consider an alternative specification using the actual value of y_{t+h}. The forecast error, under rational expectations, is $\nu_t = \mathbb{E}_t[y_{t+h}] - y_{t+h}$. It can be shown, that ν_t follows an $MA(h-1)$ process with innovations $\epsilon_{t+j} = y_{t+j} - \mathbb{E}_{t-j-1}[y_{t+j}]$ given by:

$$\nu_t = \epsilon_{t+h} + \phi_1 \epsilon_{t+(h-1)} + ... + \phi_{h-1} \epsilon_{t+1}, \tag{2}$$

Plugging into the original NKPC equation (1), we obtain the following modified NKPC

$$y_t = \beta y_{t+h} + \lambda x_t + u_t. \tag{3}$$

where $u_t = \nu_t + \eta_t$, which can be some ARMA model or a general linear process. Note first that, since y_{t+h} is observed, we can potentially estimate (3). However, y_{t+h} is endogenous: it appears both as a regressor and in the error ν_t. Standard ordinary least squares estimation is not consistent, and we need to rely on an instrumental variable estimator. The second is that, assuming some regularity conditions to ensure that u_t is stationary, by the Wold decomposition Theorem, we can model u_t as a linear process of the form $u_t = \sum_{i=-h}^{\infty} c_h e_{t+h}$. Our paper discuss various methods to estimate equation (3), which features endogenous regressors, making efficient use of the information about the non-spherical errors u_t.

2.1 Setup

The general model for y_t a scalar and $x_t' = (x_{1t}, ..., x_{kt})$ a vector of regressors is

$$y_t = x_t'\beta + u_t, \qquad t = 1, ..., T, \tag{4}$$

where $\beta = (\beta_{1t}, ..., \beta_{kt})'$ is a vector of unknown coefficients. Assume that the errors u_t follows the linear process $u_t = C(L)e_t$, with $e_t \overset{i.i.d.}{\sim} (0, \sigma_e^2)$,where $C(L) = \sum_{j=-h}^{\infty} c_j L^j$, $e_t = 0$ for $t \leq 0$, and the roots of $C(L)$ are outside the unit circle. Also, with $u = (u_1, ..., u_T)$, $\mathbb{E}(uu') \equiv \mathbb{V}(u) = \sigma^2\Omega$. We define a regressors $x_{i,t}$ as being endogenous if $\mathbb{E}(x_{i,t}e_{t+j}) \neq 0$ for some $j = 0, ..., h$, and we assume that only the first m elements of x_t are endogenous, i.e.,

$$\begin{aligned} \mathbb{E}(x_{i,t}e_{t+j}) &\neq 0 \quad \forall i = 1, ..., m \text{ and some } j = 0, ..., h \text{ for which } c_j \neq 0 \\ \mathbb{E}(x_{i,t}e_{t+j}) &= 0 \quad \forall i = m+1, ..., k \quad \text{ and } \forall j \geq 0. \end{aligned} \tag{5}$$

To deal with endogeneity, consider a vector of instruments $z_t = (z_{1,t}, ..., z_{l,t})'$ with $T > l \geq m$. Note that the non-endogenous regressors should be included in z_t, i.e. $z_{1,t} = x_{i,t}$ for $i = m+1, ..., k$, so henceforth all statements about the conditions on the instruments apply equally well to the non-endogenous regressors. Since the issues related to the conditions between the regressors and errors are extensively analyzed in Perron and González-Coya (2023), we shall focus solely on the issues related to the instruments.

There are two leading cases of interest. The first is with $c_j \neq 0$ for some $j > 0$. We shall label this as the "general case". Then, the regressors are said to be non-endogenous if $\mathbb{E}(z_{i,t}e_t) = 0$. The instrument are exogenous if $\mathbb{E}(z_{i,t}e_{t-j}) = 0$ for all $j > 0$,and they are said to be pre-determined if $\mathbb{E}(z_{i,t}e_{t+j}) = 0$, for all $j > 0$. This applies as well to the non-endogenous regressors since they act as their own instruments. Accordingly, this follows the classification made in Perron and González-Coya (2023). As an example, suppose that the output gap next period is forecasted to be positive, i.e., $g_{t+1} > 0$ for a given level of current output g_t. Inflation y_t increases immediately as firms that can reset prices will take this into account in their price-setting decision. Given that inflation four periods ahead remains unchanged, this news is reflected in the regression as a positive error realization e_t. As a result, we will observe a positive correlation between the errors and the future values of the output gap. This means that the output gap is not exogenous. Moreover, since the output gap acts as its own instrument, the full set of instruments will not be exogenous either. A less stringent requirement that we can hope for is that the output gap be predetermined.

The second case of interest is when $C(L) = \sum_{j=-h}^{-1} c_j L^j$ or $u_t = e_{t+h} + c_1 e_{t+(h-1)} + ... + c_{h-1}e_{t+1}$. Here, $\mathbb{E}(y_{t+h}e_{t+j}) \neq 0$ for all $j = 1, ..., h$, hence y_{t+h} is endogenous. The instruments are strictly exogenous if $\mathbb{E}(z_{i,t}e_{t-j}) = 0$ for all $j \geq 0$. They are pre-determined if $\mathbb{E}(z_{i,t}e_{t+j}) = 0$ for all $j > 0$. This occurs often in pure rational expectations models. For instance, in the NKPC example discussed above without a cost push shock, when u_t is only a function of future forecast errors, then, $\mathbb{E}(y_{t+h}e_{t+j}) \neq 0$ for all $j = 1, ..., h$. It is inconsequential if any of the instruments is not strictly exogenous since e_{t-j} does not occur in the model for any $j \geq 0$. The instruments are also guaranteed to be pre-determined given the rational expectations hypothesis, which states that any variable dated at time t is uncorrelated with future forecast errors from an optimal predictor that uses information efficiently. We shall label this as the "RE case" for Rational Expectations.

3 The Estimators

In the following we adopt the general regression framework (4), understanding the subtleties involved in the various cases. In matrix notation, we have $y = X\beta + u$, where $y = (y_1, ..., y_T)$, $X = (x_1, ..., x_m, x_{m+1}, ...x_k)$ with $x_t' = (x_{1t}, ..., x_{kt})$ and u as defined above. We assume that at least one of the regressor is endogenous as defined in the previous section, i.e., $m > 0$. We suppose there is a set of pre-determined instruments $Z = (z_1, ..., z_l)$ with $z_t' = (z_{1t}, ..., z_{lt})$ ($l \geq k$) such that $\mathbb{E}(z_{i,t}e_{t+j}) = 0$ for all $i = 1, ..., l$ and all $j = 0, ..., h$, for which $c_j \neq 0$. Note that z_t contains the non-endogenous regressors $\{x_{i,t}; i = m+1, ..., k\}$. The "RE case" simply involves non-endogenous regressors dated h periods before t. As a matter of notation, we write an infeasible estimate without a "^" and a feasible estimate with one.

It is useful to start with a GMM framework. Consider first the moment condition

$\mathbb{E}(Zu) = 0$. A first option would be to use sub-optimal GMM using the identity as the weighting matrix. This leads to the standard Instrumental Variable (IV) estimator $\hat{\beta}_{IV} = (X'P_Z X)^{-1}X'P_Z y$, such that $T^{1/2}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, \sigma^2\mathbb{V})$, with

$$\mathbb{V} = plim_{T\to\infty} \left[T^{-1}X'P_Z X\right]^{-1} (\sigma^2 T^{-1}X'P_Z\Omega P_Z X) \left[T^{-1}X'P_Z X\right]^{-1},$$

The latter can be consistently estimated by $\mathbb{V} = [T^{-1}X'P_Z X]^{-1}\,\hat{\Omega}_Z\,[T^{-1}X'P_Z X]^{-1}$, where $\hat{\Omega}_{P_Z X}$ is a consistent estimate of $\Omega_{P_Z X} = plim_{T\to\infty}\sigma^2 T^{-1}X'P_Z\Omega P_Z X$. Note that, in the context of time series data, $\hat{\beta}_{IV}$ is consistent under the condition of exogenous instruments, i.e., $\mathbb{E}[z_t u_t] = 0$. This estimator is not often used in practice, since one can do better by applying the so-called "optimal GMM" estimator for the moment condition stated above.

3.1 Optimal GMM based on $\mathbb{E}(Zu) = 0$.

Using the same moment condition $\mathbb{E}(Zu) = 0$ but with the optimal weighting matrix leads to the estimator that solves the following minimization problem

$$\min_\beta (y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta),$$

and the closed form solution in the linear model

$$\beta_{GMM} = (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y. \tag{6}$$

Since $plim_{T\to\infty}\sigma^2 T^{-1}Z'\Omega Z = \Omega_Z$ an asymptotically equivalent feasible GLS estimate is

$$\hat{\beta}_{GMM} = (X'Z\hat{\Omega}_Z^{-1}Z'X)^{-1}X'Z\hat{\Omega}_Z^{-1}Z'y.$$

where $\hat{\Omega}_Z$ is a consistent estimate of Ω_Z. There are a variety of estimates that were proposed. A commonly used one is $\hat{S}_{w,T} = \hat{R}_v(0) + 2\sum_{j=1}^{T-1} w(j, m_T)\hat{R}_v(j)$, where $w(j, m_T)$ is some weighting function and m_T a bandwidth, with $\hat{R}_v(j) = T^{-1}\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}$, and $\hat{v}_t = z_t\hat{u}_t$ with $\hat{u}_t = y_t - X_t\hat{\beta}_{IV}$, where $\hat{\beta}_{IV} = (X'P_Z X)^{-1}X'P_Z y$ is the standard IV estimate constructed assuming $i.i.d.$ errors, which acts as a preliminary consistent estimate (with exogenous instruments). Such an estimate is asymptotically equivalent to the infeasible GMM estimate and the variance can be estimated using

$$\hat{\mathbb{V}}(\hat{\beta}_{GMM}) = (T^{-1}X'Z\hat{\Omega}_Z^{-1}Z'X)^{-1}. \tag{7}$$

This is the most commonly used method. Note, however, that for this method to be consistent in the general case, strictly exogenous instruments are needed (except for some knife-edge cases), otherwise the moment condition $\mathbb{E}(Zu) = 0$ is not satisfied. In the "RE case", the estimate is consistent with non-exogenous instruments, since such issues are irrelevant. These results follow directly from those discussed in Perron and González-Coya (2023).

3.2 The Forward Filter Estimator

Hayashi and Sims (1983) proposed a so-called forward-filter estimator. In terms of a GMM setting it amounts to applying optimal GMM for the moment condition $\mathbb{E}(Z'Fu) = 0$, where F is some upper triangular matrix such that $F\Omega F' = I$ (e.g., the matrix obtained from a Choleski decomposition). As discussed in more details below, the idea is to apply a quasi-GLS transformation to obtain residuals that are serially uncorrelated. This leads to the estimate, labelled as the forward filtered estimate,

$$\beta_{FF} = (\widetilde{X}'P_Z\widetilde{X})^{-1}\widetilde{X}'P_Z\widetilde{y}, \tag{8}$$

where $\widetilde{X} = FX$ and $\widetilde{y} = Fy$. Let $Fu = \widetilde{e}$, where $\widetilde{e} = (\widetilde{e}_1, .., \widetilde{e}_T)$ is a sequence of *i.i.d.* random variables. Note that we write $\widetilde{e}$ instead of e since in general the sequence $\widetilde{e} = (\widetilde{e}_1, .., \widetilde{e}_T)$ does not correspond to $e = (e_1, .., e_T)$, but is rather a linear combination of the elements of the latter. The estimator will be consistent provided $plim_{T\to\infty} T^{-1}\sum_{t=1}^{T} z_t\widetilde{e}_t = 0$. This condition is satisfied in the "RE case" since z_t is uncorrelated with any elements of the sequence of errors $(e_{t+1}, ..., e_{t+h})$ given that these are rational expectations errors. It will also be satisfied in the general case provided the instruments are exogenous, i.e., uncorrelated with any elements of the sequence $(e_1, ..., e_t)$. For the general case, the estimator is, in general, not consistent if the instruments are not exogenous with respect to $(e_1, ..., e_t)$ since $\widetilde{e}_t$ will be some linear combination of all their elements. To give some perspective, consider u_t being an $AR(1)$ process of the form $u_t = \rho u_{t-1} + e_t$. Ignoring the last observation, the forward filer is given by $(1 - \rho L^{-1})$ so that $\widetilde{e}_t = (1 - \rho L^{-1})u_t$. Then,

$$\begin{aligned} \widetilde{e}_t &= u_t - \rho u_{t+1} = u_t - \rho^2 u_t - \rho e_{t+1} = (1-\rho^2)u_t - \rho e_{t+1} \\ &= (1-\rho^2)\textstyle\sum_{j=0}^{t-1} \rho^j e_{t-j} - \rho e_{t+1} \end{aligned} \tag{9}$$

Note that $\widetilde{e}_t$ is a sequence of *i.i.d.* errors since

$$\begin{aligned} \mathbb{E}[\widetilde{e}_t\widetilde{e}_{t+j}] &= \mathbb{E}[(u_t - \rho u_{t+1})(u_{t+j} - \rho u_{t+j+1})] \\ &= \mathbb{E}[u_t u_{t+j}] - \rho\mathbb{E}[u_t u_{t+j+1}] - \rho\mathbb{E}[u_{t+1}u_{t+j}] + \rho^2\mathbb{E}[u_{t+1}u_{t+j+1}] \\ &= \rho^j\mathbb{E}[u_t u_t] - \rho\rho^{j+1}\mathbb{E}[u_t u_t] - \rho\rho^{j-1}\mathbb{E}[u_{t+1}u_{t+1}] + \rho^2\rho^j\mathbb{E}[u_{t+1}u_{t+1}] \\ &= (\rho^j - \rho^{j+2} - \rho^j + \rho^{j+2})\mathbb{V}(u_t) = 0. \end{aligned}$$

However, $\widetilde{e}_t$ from (9) depends on e_{t+1}, e_t, e_{t-1},...,e_1. Hence, both pre-determined and exogenous instruments are required.

3.3 The GLS-IV Estimator

The estimate proposed in this paper can also be cast in a GMM framework using the moment condition $\mathbb{E}((DZ)'Du) = 0$, where D is any matrix that satisfies $D\Omega D' = I$. In practice, we shall use D as a lower triangular matrix generated by an autoregressive approximation. However, in principle, any D can be used provided there exists an invertible representation for the errors u_t in terms of a (possibly) infinite autoregression. The arguments are similar to those in **?**. Applying optimal GMM,

$$\beta_{GLS-IV} = \arg\min_{\beta}(y - X\beta)\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}(y - X\beta),$$

and the closed form expression for the estimator is given by

$$\begin{aligned}\beta_{GLS-IV} &= (X^{*\prime}P_{Z^*}X^*)^{-1}X^{*\prime}P_{Z^*}y^* \\ &= [X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y, \qquad (10)\end{aligned}$$

where $y^* = Dy$, $X^* = DX$ and $Z^* = DZ$, with P_{Z^*}, the usual projection matrix $P_{Z^*} = Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}$. Its limit distribution is given by

$$\sqrt{T}(\beta_{GLS-IV} - \beta) \xrightarrow{d} N(0, plim_{T\to\infty}\sigma^2 T(X^{*\prime}P_{Z^*}X^*)^{-1}). \qquad (11)$$

Let $Du = e^*$. The condition for the consistency of the estimate is $plim_{T\to\infty}(DZ)'e^* = 0$. Following the arguments in Perron and González-Coya (2023), this holds requiring only pre-determined instruments whatever the form of D. In a nutshell, one can always choose D lower triangular such that a) $Du = e$, the original sequence of shocks, b) DZ only involves past values of the instruments z_t. Since $D'D = \Omega^{-1}$ by construction, the estimator is invariant to the choice of D. Hence, only pre-determined instruments are needed whatever the choice of D. To give an example, consider again the $AR(1)$ model with the forward filter, i.e., D chosen to be upper triangular. Then, the condition for consistency is

$$\mathbb{E}[(z_t - \rho z_{t+1})(u_t - \rho u_{t+1})] = \mathbb{E}[(z_t - \rho z_{t+1})((1 - \rho^2)u_t - \rho e_{t+1})] = 0$$

which requires a) $\mathbb{E}[z_{t+1}e_{t+1}] = 0$, which holds by assumption; b) $\mathbb{E}[z_t e_{t+1}] = 0$, satisfied with predetermined instruments; and c) $\mathbb{E}[(z_t - \rho z_{t+1})u_t] = 0$, which also holds with instruments that are not strictly exogenous. To see this, assume $\mathbb{E}[z_t e_{t-j}] \neq 0$ (i.e., errors correlated with future instruments), then

$$\begin{aligned}\mathbb{E}[(z_t - \rho z_{t+1})u_t] &= \mathbb{E}[(z_t - \rho z_{t+1})\textstyle\sum_{j=0}^{t}\rho^j e_{t-j}] = \mathbb{E}[z_t\rho^j e_{t-j}] - \rho\mathbb{E}[z_{t+1}\rho^{j-1}e_{t-j+1}] \\ &= \rho^j\mathbb{E}[z_t e_{t-j}] - \rho^j\mathbb{E}[z_{t+1}e_{t-j+1}] = 0\end{aligned}$$

Hence, only predetermined instruments are needed even if Du depends on the past e_t's. It also shows the importance of transforming both the regressors X and the instruments Z when using the upper triangular transformation F, unlike what is done with the Forward Filter estimator suggested by Hayashi and Sims (1983), which only transforms X not Z, in which case it is inconsistent from the arguments above. Transforming only X but not Z when using the backward filer D would result in a consistent but inefficient estimator.

3.4 Summary

The analysis so far suggested three estimators: GMM, Forward-Filter and our GLS-IV. In the "RE case" or with exogenous instruments, all are consistent. In the "general case" with non-exogenous instruments, only the GLS-IV is consistent. This is by itself a strong argument in favor of adopting GLS-IV. However, when all estimates are consistent, we have no clear guidance about which estimator might be more efficient. In the traditional literature on GMM estimation, the so-called "optimal GMM" estimate are optimal for the particular moment condition selected. The theory of optimal instruments is a way to guide us on how to modify the moment condition to obtain an estimate that is actually optimal for all possible moment conditions and achieves the semi-parametric efficiency bound. This theory has been used to great benefits in cases dealing with random samples of data. Unfortunately, in time series contexts, especially under the level of generality adopted here, no useful results are available. But still there are ways to make useful comparisons to guide us about the relative efficiency of the estimates. We do so by first casting the estimators into a GLS framework. This allows comparing the strength of the instruments implicitly used. We then use other arguments that show that GLS-IV is likely to be superior using a variety of scenarios.

4 The Estimators as GLS Procedures

In this section, we assume that $\mathbb{E}(Zu) = 0$ so that all estimators are consistent. Casting them as GLS procedures is useful as it will permit gauging the nature and strength of the instruments from the implied first-stage regression. In matrix notation, the model is still

$$y = X\beta + u \tag{12}$$

with the same conditions as stated previously. Recall that the IV estimator is of the form $\hat{\beta}_{IV} = (\breve{Z}'X)^{-1}\breve{Z}'y$, where $\breve{Z} = ZR$ is some linear combination of the instrument, with R a $\ell \times k$ of rank k. We seek the matrix R that yields the best IV estimator, in the sense that it is consistent and has minimal variance in the class of estimators using some linear combinations of the instruments Z. We start by analyzing how GMM falls into this framework and our treatment follows McFadden (1999). First, multiply (12) by Z' so that

$$Z'y = Z'X\beta + Z'u = Z'X\beta + v, \tag{13}$$

where $v = Z'u$ and $\mathbb{V}(v) = \sigma^2 Z'\Omega Z$. Note that $\mathbb{E}(X'Zv) = \mathbb{E}(X'ZZ'u) = 0$, and GLS applied to (13) yields

$$\beta_{GMM} = (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y,$$

with $\sqrt{T}(\beta_{GMM} - \beta) \stackrel{d}{\to} N(0, \sigma^2[Q'_{ZX}\Omega_Z^{-1}Q_{ZX}]^{-1})$, where $Q_{ZX} = plim_{T\to\infty}T^{-1}Z'X$ and $\Omega_Z = plim_{T\to\infty}\sigma^2 T^{-1}Z'\Omega Z$. Hence, GMM first applies some IV estimation and then the GLS transformation. This corresponds to setting $R = (Z'\Omega Z)^{-1}Z'X$. To see that this is the optimal choice, consider the estimate

$$\beta_R = (R'Z'X)^{-1}R'Z'y = \beta + (R'Z'X)^{-1}R'Z'u,$$

so that $T^{1/2}(\beta_R - \beta) \stackrel{d}{\to} N(0, (R'Q_{ZX})^{-1}(R'\Omega_Z R)(Q'_{ZX}R)^{-1})$. One can show that

$$(R'Q_{ZX})^{-1}(R'\Omega_Z R)(Q'_{ZX}R)^{-1} - (Q'_{ZX}\Omega_Z^{-1}Q_{ZX})^{-1} \geq 0, \tag{14}$$

in the sense that it is positive semi-definite when $R = (Z'\Omega Z)^{-1}Z'X$. The proof is omitted. The derivations above apply IV first then GLS. One can, of course, do the reverse. We first transform the model pre-multiplying by the matrix D:

$$Dy = DX\beta + Du = DX\beta + e^*,$$

where $\mathbb{V}(e^*) = \sigma^2$. Now the errors are *i.i.d.* but still correlated with the regressors. Let $\check{Z} = (D')^{-1}Z$ and apply IV with the instruments $\check{Z}$. Let $D\hat{X} = P_{\check{Z}}DX$ be the projection of DX on $\check{Z}$, i.e., $P_{\check{Z}} = \check{Z}(\check{Z}'\check{Z})^{-1}\check{Z}'$. Since $\mathbb{V}(e^*) = \sigma^2$, the optimal IV is

$$\begin{aligned}
\beta_{IV} &= ((P_{\check{Z}}DX)'DX)^{-1}(P_{\check{Z}}DX)'Dy \\
&= (X'Z(Z'(D'D)^{-1}Z)^{-1}Z'X)^{-1}X'Z(Z'(D'D)^{-1}Z)^{-1}Z'X)y \\
&= (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y = \beta_{GMM}.
\end{aligned}$$

Recall that for any square matrix C, $(C')^{-1} = (C^{-1})'$ and $(D'D)^{-1} = (\Omega^{-1})^{-1} = \Omega$.

We now address the Forward Filter estimator. An alternative way to derive and interpret this estimator is the following. We start by applying the GLS transformation to the model

$\widetilde{y} = \widetilde{X}\beta + \widetilde{e}$, where $\widetilde{y} = Fy$ and $\widetilde{X} = FX$ with F an upper triangular matrix such that $F\Omega F' = I$. Now, since the instruments are Z, the optimal IV is

$$\begin{aligned}\beta_{IV} &= (\widetilde{X}'Z(Z'Z)^{-1}Z'\widetilde{X})^{-1}\widetilde{X}'Z(Z'Z)^{-1}Z'\widetilde{y} \equiv \beta_{FF} \\ &= \beta + (\widetilde{X}'Z(Z'Z)^{-1}Z'\widetilde{X})^{-1}\widetilde{X}'Z(Z'Z)^{-1}Z'\widetilde{e},\end{aligned}$$

so that $T^{1/2}(\beta_{FF} - \beta) \xrightarrow{d} N(0, \sigma^2_{\widetilde{e}}[Q'_{Z\widetilde{X}}\Omega_Z^{-1}Q_{Z\widetilde{X}}]^{-1})$, where $Q_{Z\widetilde{X}} = plim_{T\to\infty}T^{-1}Z'FX$, $\Omega_Z = plim_{T\to\infty}T^{-1}Z'Z$ and $\sigma^2_{\widetilde{e}} = \mathbb{V}(\widetilde{e}_t)$.

We now turn to the GLS-IV estimator. The approach is similar but uses the instruments $Z^* = DZ$. The idea is that if Z are good instruments for X than DZ should be better instruments for DX than Z. This leads to the infeasible estimate, labelled β_{GLS-IV} given that it first applies a GLS transformation and then an IV estimation:

$$\begin{aligned}\beta_{GLS-IV} &= (X^{*\prime}P_{Z^*}X^*)^{-1}X^{*\prime}P_{Z^*}y^* \\ &= [X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y\end{aligned}$$

with limit distribution $T^{1/2}(\beta_{GLS-IV} - \beta) \xrightarrow{d} N(0, plim_{T\to\infty}\sigma^2 T(X^{*\prime}P_{Z^*}X^*)^{-1})$. An alternative way to derive and interpret this estimator is to start by applying the GLS transformation to the model $y^* = X^*\beta + e^*$, where $y^* = Dy$ and $X^* = DX$ with D a matrix such that $D'D = \Omega^{-1}$ and $D\Omega D' = I$. The optimal linear instruments are then $P_{Z^*}X^*$, which implies

$$\beta_{GLS-IV} = (X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*)^{-1}X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^*$$

so that

$$T^{1/2}(\beta_{GLS-IV} - \beta) \xrightarrow{d} N(0, \sigma^2 plim_{T\to\infty}[X^{*\prime}P_{Z^*}X^*]^{-1}). \tag{15}$$

4.1 Summary

As before, when framing the estimates according to a GMM framework, we cannot draw any conclusion about either estimator having higher precision in general. They differ according to the GLS transformation applied before implementing the IV step. Here we argue that using GLS-IV imply stronger instruments, i.e., with a higher F-statistic in the first stage. The intuition is that GLS-IV preserves the relationship between the regressors and the instruments that one specifies while the other methods do not. In practice, this property is particularly important if it allows us to move away from the weak instrument zone to avoid the use of weak identification techniques. To better understand the difference between the estimators, we start by analyzing the first stage of GMM, FF, and GLS-IV obtained from their interpretation as GLS procedures. The first stages are:

$$\begin{aligned} \text{GLS-IV} &: \quad DX = \Pi_1 DZ + \nu_1, \\ \text{GMM} &: \quad DX = \Pi_2 (D')^{-1} Z + \nu_2, \\ \text{FF} &: \quad FX = \Pi_3 Z + \nu_3. \end{aligned}$$

We first highlight that on the left-hand side of the equations, we have DX or FX as the endogenous regressors in the GLS transformed model. The instruments, on the right-hand side instead depend on the estimator used, namely DZ for GLS-IV, $(D')^{-1}Z$ for GMM, and Z for FF. By applying the same transformation on both sides of the equation, GLS-IV preserves the relationship between the two variables, while the other estimators do not. This means that in principle the correlation between the filtered endogenous variable and the instruments is likely, in general, to be maximized by GLS-IV. The intuition is simple. If Z is deemed a good instrument for X, then DZ should be the obvious candidate as the corresponding instrument for DX. Simulations reported below confirm that this is the case in all our experiments. Consider a simple example with errors following an $AR(1)$ process with autoregressive coefficient ρ. The first stage of the FF estimator is

$$(x_t - \rho x_{t+1}) = \Pi_3 z_t + \nu_3.$$

Then if ρ is close to 1, we are trying to fit the increments in x_t with the level of z_t. On the opposite, for GLS-IV the first stage is

$$(x_t - \rho x_{t-1}) = \Pi_1 (z_t - \rho z_{t-1}) + \nu_1,$$

where we fit the increments in x_t with the increments in z_t. It seems from common sense that the latter would yields better instruments. For the case of GMM, things are more complex as the relationship between the instruments and DX is rather ad hoc.

5 Other Advantages of the GLS-IV Estimator

As highlighted by Hayashi and Sims (1983) the performance of the GMM and the FF estimators depends on the entire distribution of the variables. This also translates to the comparison with the GLS-IV estimator. However, we can still present other arguments suggesting that the GLS-IV estimator has some advantages compared to FF and GMM in the case with exogenous instruments (and regressors) or the "RE case" so that all estimates are consistent. Issues related to non-exogenous instruments (regressors) were addressed previously.

Argument 1: GMM versus GLS-IV objective functions. An important feature of the difference between GMM and GLS-IV is the difference in how the weighting is done. It can be shown that β_{GMM} solves the following problem

$$\min_\beta (y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta).$$

Hence, it weight the moments $Z'u$ by the square-root of their long-run variance. Note, however, that β_{GLS-IV} solves the minimization problem

$$\begin{aligned} & \min_\beta (y - X\beta)'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}(y - X\beta) \\ = \ & \min_\beta (DY - DX\beta)'DZ(Z'\Omega^{-1}Z)^{-1}Z'D'(DY - DX\beta). \end{aligned}$$

Hence, it weights the transformed observations by some matrix $DZ(Z'\Omega^{-1}Z)^{-1}Z'D'$, in a similar way as what is done by GLS. An argument can be made that weighting all observations can lead to more efficient estimates than weighting aggregate moments.
Argument 2: The case with no endogeneity. As mentioned before there is nothing in the GMM framework that guides us in the choice of the moment restriction. This is a disadvantage of the GMM framework as it may lead to inefficient estimators. To illustrate this issue, we review two cases, one in which GMM is optimal and one in which it is not. Consider a standard linear regression model without serial correlation $y = \beta X + u$. The standard GMM theory suggests using the $\mathbb{E}[X'u] = 0$ as the moment conditions. Consider homoskedastic errors: $\mathbb{V}(u) = \sigma^2 I$. Then, the GMM estimator is OLS: $\beta_{GMM} = (X'X)^{-1} X'y \equiv \hat{\beta}_{OLS}$. With non-spherical errors, i.e., $\mathbb{V}(u) = \sigma^2\Omega \neq \sigma^2 I$, optimal GMM is no longer optimal. It does not reduce to GLS, which is optimal in this case, i.e.,

$$\beta_{GMM} = \left(X'X(X'\Omega X)^{-1}X'X\right)^{-1} X'X(X'\Omega X)^{-1}X'y \neq \beta_{GLS}.$$

Note that in both cases we assess the optimality of the GMM estimator by comparing it to other estimators which we know are optimal in that situation. Clearly, in the second case, in principle we can modify the GMM moments as follows to recover GLS so that GMM becomes optimal: $\mathbb{E}[X'D'Du] = \mathbb{E}[X'D'e] = 0$. Then, $\beta_{GMM} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y \equiv \beta_{GLS}$. But in this case β_{GMM} is β_{GLS-IV}. It can be shown that GLS-IV is the only method that reduces to the optimal GLS estimator in this case. Note that we achieved this optimality result not because of the GMM theory but by modifying the moments to exactly yield the GLS, which we knew ex-ante to be optimal. Unfortunately, for the general setup where, on top of non-spherical errors, we also have endogeneity of the regressors, a clear comparison between the estimators is not possible as the optimal estimator would depend on the correlation structure between the regressors, the dependent variable, and the instruments.

Argument 3: The case with only heteroskedasticity. Probably the most compelling argument in favor of GLS-IV when all estimators are consistent is that when only heteroskedasticity is present, it is the only that reduces to GLS, which is optimal, i.e., the GMM estimator with optimal instruments, which can be obtained here in the context of a random sample of data. Consider the standard linear regression model and assume both endogeneity and heteroskedastic errors, $\mathbb{V}(u_t|z_t) = \sigma_t^2$. We know from the optimal instruments theory of GMM that in this case the optimal instrument is

$$z_t^* = c\mathbb{E}\left[\partial(y_t - x_t'\beta)/\partial\beta|z_t\right]\sigma_t^{-2} = -c\sigma_t^{-2}\mathbb{E}\left[x_t'|z_t\right] = \sigma_t^{-2}\mathbb{E}\left[x_t'|z_t\right],$$

where c is a free constant that we set to -1 in the last step. Using this optimal instrument for GMM is equivalent to applying IV to the following model with z_t/σ_t as the instrument

$$y_t/\sigma_t = (x_t'/\sigma_t)\beta + u_t/\sigma_t. \tag{16}$$

To obtain the estimator, express the model in matrix notation where D is diagonal with entries $1/\sigma_t$, so that $DY = DX\beta + Du$. Applying IV using DZ as an instrument,

$$\begin{aligned}\beta_{IV} &= \left[X'D'DZ(Z'D'Z)^{-1}Z'D'DX\right]^{-1}X'D'DZ(Z'D'Z)^{-1}Z'D'DY \\ &= \left[X'^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\right]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'^{-1}Y \equiv \beta_{GLS-IV}\end{aligned}$$

which is the GLS-IV estimator. This shows that the GLS-IV framework includes the case of optimal GMM instruments when only heteroskedasticity is present as a special case. On the other hand, FF reduces to applying IV to the regression (16) with z_t as the instrument instead of z_t/σ_t, which is not optimal. By definition, the GMM estimate is also sub-optimal. It reduces to an estimator originally proposed by White (1982).

6 Feasible GLS-IV

The GLS-IV requires the knowledge of the variance-covariance matrix $\sigma^2\Omega = \mathbb{V}(u)$, which is unknown in the majority of the applications. Here we discuss a feasible version which has the same limiting distribution as the infeasible GLS-IV under the general class of processes discussed in Section 2.1. We start by explaining the intuition of the procedure and then we state the formal algorithm. Let us first consider a simple case. Assume that the errors follow a known $AR(1)$ process. Substituting the errors $u_t = \rho_1 u_{t-1} + e_t$ in the model and subtracting $\rho_1 y_{t-1}$ from both sides we obtain

$$y_t = \rho_1 y_{t-1} + \beta(x_t - \rho_1 x_{t-1}) + e_t. \tag{17}$$

While in the initial regression model the errors are $AR(1)$, the errors in (17) are *i.i.d.*. Assuming that the regressors are predetermined so that $\mathbb{E}[y_{t-1}e_t] = 0$, $\mathbb{E}[x_t e_t] = 0$, and $\mathbb{E}[x_{t-1}e_t] = 0$, we can estimate ρ_1 consistently by IV. As discussed in Perron and González-Coya (2023) for the case with no endogeneity, the Cochrane-Orcutt procedure will not work.

We now consider the case with an error component u_t that follows a linear process of the form $u_t = C(L)e_t$ with the roots of $C(L)$ outside the unit circle, we have $A(L) = C(L)^{-1}u_t = e_t$. Here, $A(L)$ is, in general, an infinite order polynomial. However we can use an approximation that involves a finite number of lags, say k_T, so that $A(L)^* = (1 - \sum_{j=1}^{k_T} \rho_j)$ is close to $C(L)^{-1}$. This leads to applying OLS to the following regression model

$$y_t = \sum_{j=1}^{k_T} \rho_j y_{t-j} + x_t'\beta - \sum_{j=1}^{k_T} x_{t-j}'\delta_j + e_{t,k}, \tag{18}$$

where the sequence $e_{t,k}$ is nearly *i.i.d.*. Since some of the regressors are correlated with the errors e_t, one needs to estimate the parameters ρ_j using an IV procedure. Provided the regressors are pre-determined, following the work of Berk (1974) and others, the estimates will be consistent provided $k_T \to \infty$ and $k_T^3/T \to 0$ as $T \to \infty$. One can then construct the matrix D as a lower triangular matrix with the proper powers of the estimates of ρ_j as the entries. This amounts to using the transformed regression

$$y_t^* = x_t^{*\prime}\beta + e_{t,k}, \tag{19}$$

where $y_t^* = y_t - \sum_{j=1}^{k_T} \hat{\rho}_j y_{t-j}$ and $x_t^* = x_t - \sum_{j=1}^{k_T} \hat{\rho}_j x_{t-j}$ are the quasi-differenced data and $\hat{\rho}_j$ $(j = 1, ..., k_T)$ are the estimates obtained applying IV to (18). We then simply use IV applied to (19) to estimate β using the instruments $z_t^* = z_t - \sum_{j=1}^{k_T} \hat{\rho}_j z_{t-j}$. This is a completely standard approach in the time series literature with a long history of useful applications. In order to select the lag length k_T in such a way that the technical conditions are satisfied, we use the Bayesian Information Criterion (BIC), see Schwarz (1978). For this we need to specify a sufficiently large value k_{max} and repeat the IV estimation described above for all $k \in [0, k_{max}]$. Then, we select the model with the lowest BIC and use the corresponding estimates of ρ_j to construct the quasi-differenced data. Of particular interest is the fact that when k_T is properly chosen, the residuals $e_{t,k}$ from the regression (19) will be nearly the same as the original time t shock e_t and more so as T increase when the order of the autoregressive process is infinite. Intuitively, not many lags are needed as a short-memory stationary process implies a geometric rate of decay on the parameters. Our simulations will show that it performs well even in the case of finite order moving-average processes.

Note that this amounts to some approximate GLS since the matrix D is truncated to eliminate the first k_T rows so that the initial errors are set to their unconditional mean 0. Other approaches might be possible; e.g., using the method of Prais and Winsten (1954). This would complicate matters greatly without efficiency gains for commonly sample sizes.

6.1 Estimation Algorithm

We now outline the detailed steps of the algorithm to obtain a feasible GLS-IV estimate. Step 1) Select $k_{\max}$, usually a number like 12 should be sufficient. Of course, other values are possible depending on the sample size. Initialize $k_T = 0$. Step 2) Consider the regression

$$y_t = \sum_{j=1}^{k_T} \rho_j y_{t-j} + x_t'\beta - \sum_{j=1}^{k_T} x_{t-j}'\delta_j + e_{t,k}, \quad t = k_{\max}+1, ..., T, \tag{20}$$

where $\delta_j = \rho_j\beta$. Note that we use the same number of observations for all values of k, namely $T - k_{\max}$. This is needed to ensure a proper comparison across models as suggested by Ng and Perron (2005). Given the exogeneity restriction $\mathbb{E}[z_t e_{t,k}] = 0$ we can use IV to estimate the parameters. Note that the estimate of β is consistent but inefficient. Step 3) Compute the sample variance $\hat{\sigma}_{e,k}^2 = N^{-1}\sum_{t=k_{\max}+1}^{T} \hat{e}_{t,k}^2$, where $N = T - k_{\max}$ and

$$\hat{e}_{t,k} = y_t - \sum_{j=1}^{k_T} \hat{\rho}_j y_{t-j} - x_t'\hat{\beta} + \sum_{j=1}^{k_T} x_{t-j}'\hat{\delta}_j.$$

Step 4) With $N = T - k_{\max}$, compute the BIC for the model associated with k_T, $BIC_{k_T} = \ln(\hat{\sigma}_{ek}^2) + (\ln N/N)k_T$ with $N = T - k_{\max}$. Step 5) Increase k_T by 1 and, provided that k_T is in the range $[0, k_{\max}]$, repeat Steps 2-4. If $k_T > k_{\max}$ move to the next step. Step 6) Choose the autoregressive order that minimizes the BIC criterion, i.e., $k_T^* = \arg\min_{k_T} BIC_{k_T}$ and store the $\hat{\rho}_j$ associated with k_T^*. Step 7) Compute the autoregressive-based GLS-IV estimator, $\hat{\beta}_{GLS-IV}$. This is amounts to applying IV to the transformed regression $y_t^* = x_t^{*\prime}\beta + e_{t,k}$, where $y_t^* = y_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j y_{t-j}$ and $x_t^* = x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j x_{t-j}$ using $z_t^* = z_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j z_{t-j}$ as the instruments. The following optional steps can improve the accuracy of the estimates in finite samples. Step 8) Given the estimate of $\hat{\beta}_{GLS-IV}$, which is consistent, compute the residuals $\hat{u} = y - X\hat{\beta}_{GLS-IV}$. Step 9) Select $k_{\max}$, something like 12 should be sufficient. For $k = 1, ..., k_{\max}$, re-estimate the regression $\hat{u}_t = \sum_{i=1}^{k} \rho_i \hat{u}_{t-i} + \nu_t$, and store the values $\hat{\rho}_i$. Compute the BIC for each model and choose the value k_T^* associated with the model that minimizes the BIC criterion. Step 10) Repeat Step 8. Of course, further iterations are possible, but we failed to find any benefits in doing so, though we did find some benefit of applying the additional Steps 8-10.

The limit distribution is given by (15), whose covariance matrix can be consistently estimated by $\hat{\sigma}^2[(T^{-1}Z^{*\prime}X^*)'(T^{-1}Z^{*\prime}Z^*)^{-1}(T^{-1}Z^{*\prime}X^*)]^{-1}$, where $\hat{\sigma}^2 = T^{-1}\sum_{t=k_T^*+1}^{T}\hat{e}_t^2$, with $\hat{e}_t$ the residuals from Step (7), or Step (10) if further iterations are applied.

The estimation algorithm for the FF estimator is similar, except that one does not apply a quasi-difference operation on Z and one transform the data using $F = D'$, which amounts to a forward filtering of the form $x_t^* = x_t - \sum_{j=1}^{k_T^*}\hat{\rho}_j x_{t+j}$. This ignores the last observations but without loss in large samples, so that the estimate of the covariance matrix is $\hat{\sigma}^2[(T^{-1}Z'X^*)'(T^{-1}Z'Z)^{-1}(T^{-1}Z'X^*)$.

6.1.1 Dealing with Heteroskedastic Errors

If heteroskedasticity in the errors is a concern, two avenues are possible. The first is to correct the standard errors of the estimate using a heteroskedasticity-robust covariance matrix as suggested by, e.g., White (1980) or variations suggested afterwards. In obvious matrix notation, the limit covariance matrix of the estimate $\hat{\beta}_{GLS-IV}$ is

$$\mathbb{V} = plim_{T\to\infty}\left[T^{-1}X^{*\prime}P_{Z^*}X^*\right]^{-1}(T^{-1}Z^{*\prime}\Omega Z)\left[T^{-1}X^{*\prime}P_{Z^*}X^*\right]^{-1},$$

which can be consistently estimated by $\mathbb{V} = [T^{-1}X^{*\prime}P_{Z^*}X^*]^{-1}\hat{\Omega}_W[T^{-1}X^{*\prime}P_{Z^*}X^*]^{-1}$, where $\hat{\Omega}_W = T^{-1}\sum_{t=1}^{T}\hat{v}_t\hat{v}_t'$, with $\hat{v}_t = z_t^*\hat{e}_{t,k}$ and $\hat{e}_{t,k} = y_t^* - x_t^*\hat{\beta}_{GLS-IV}$. Of course, various finite sample refinements are possible. Alternatively, one can apply a further feasible GLS correction as suggested by González-Coya and Perron (2024). It is based on an Adaptive Lasso procedure to fit the skedastic function. Overall, further reduction in the MSE of the estimates are possible even using incorrect covariates to estimate the skedastic function as long as there is some correlation between the covariates used in the Lasso specification and those in the true skedastic function. The coverage rate of the confidence intervals have an exact size close to the nominal level and the lengths are smaller than obtained when applying OLS or correcting only for serial correlation. With homoskedastic errors, the results are equivalent to those obtained correcting only for serial correlation. Hence, correcting for heteroskedasticity when it is not present has no detrimental effect on the precision of the estimate. They showed that it provides improvements in the precision of the estimates, as well as good coverage rates of the confidence intervals in the linear model. Perron and González-Coya (2023) show the same for GLS regression in models with errors that are both serially correlated and heteroskedastic. They note, as many others do (e.g., Romano and

Wolf (2017)) that it is difficult to obtain a feasible version that achieves what could be done using the true structure. Still, additional gains in efficiency are possible. Since these features have been documented at length elsewhere, we shall not consider them further and focus on the relative performance of the various estimates when correcting only for serial correlation.

7 Simulations

As mentioned above the comparison between the properties of the various estimators requires specifying the entire joint distribution of x_t, y_t, z_t. We instead run some simulations with reasonable data-generating processes and parameter specifications to assess the performance of our GLS-IV estimator relative to GMM and FF. We first describe the data-generating processes, then present results assuming exogenous instruments, followed by results with predetermined but not exogenous instruments. In all cases, we present the case with known values of the parameters and the structure of the model and the feasible versions. This is done in order to a) verify the theoretical properties discussed above when using the true structure; b) assess how close the feasible procedures are to the best that can be done.

7.1 Data-generating Process and Estimation

For the simulations, we specify a simple setting where we have only one endogenous regressor x_t and one instrument z_t. The data-generating process is the following with $x_t' = (1, x_{1t})$

$$y_t = x_t\beta + u_t \tag{21}$$

$$\begin{aligned} x_{1t} &= \mu + \omega_{1,t} + v_t \\ z_t &= \sqrt{\gamma}\omega_{1,t} + \sqrt{2-\gamma}\omega_{2,t}, \end{aligned} \tag{22}$$

and the errors u_t can either follow an $AR(1)$ or $MA(1)$ process

$$u_t = \rho_u u_{t-1} + \epsilon_t, \text{ or } u_t = \varphi_u \epsilon_{t-1} + \epsilon_t. \tag{23}$$

where ϵ_t, v_t, $\omega_{1,t}$, and $\omega_{2,t}$ have mean 0 and variance σ_e^2, σ_v^2, σ_{z1}^2 and σ_{z2}^2, respectively. They are all uncorrelated except that $\mathbb{E}(\epsilon_t v_t) = \phi$, which measure the strength of the endogeneity, i.e., the correlation between x_{1t} and ϵ_t. The coefficient γ lies between 0 and 2 and modifies the strength of the instrument z_t, a higher value implies stronger correlation between the regressor and the instrument. The term $\omega_{2,t}$ in (22) is used to keep the R^2 constant as we vary γ. Table 1 reports the baseline values of the parameters in our simulations which consists of $N = 1000$ replications with a sample size of $T = 200$. For both AR or MA processes, the coefficients ρ_u and φ_u assume values $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

We consider the following regression model

$$y_t = \beta_1 + \beta_2 x_{1t} + u_t.$$

When Ω is assumed known we compute the estimators by plugging it into (6), (8), (10). For Ω_Z, we use the empirical analog $T^{-1}Z'\Omega Z$ and obtain D and F we use the Cholesky decomposition. For the case in which Ω is unknown, we use the procedure described in Section 6 to estimate D. For F, we simply take the transpose of D and for GMM we compute the variance of the moment condition using the methods suggested by Andrews (1991) with a quadratic spectral kernel and a bandwidth selected using the plug-in method based on an $AR(1)$ approximation. For every estimator (GMM, FF, GLS-IV), we plot the value of the bias, mean square error (MSE), the average length of the confidence interval (Length), the coverage rate of the confidence interval (Coverage) of β_2 and the F-statistic (F-Mean) from the first-stage regression applicable to each case as explained in Section 4.1.

7.2 Simulation Results with Exogenous Instruments

We start with cases involving exogenous instruments. Figure 1 presents the results when Ω is known and u_t follows $AR(1)$ process, while Figure 2 shows the results using the feasible version of the estimators. The results shows that they are virtually the same. They confirm the theoretical expectations concerning bias: all estimators provide consistent estimates (GMM is slightly biased when $\rho_u = \pm\, 0.9$, i.e., near the boundary of the stationary region). Moreover, the coverage rate of the confidence interval is near the nominal 90% level for all estimators, except for GMM with $\rho_u = 0.9$, a well documented feature. Hence, in most cases, they all provide valid statistical inference. Importantly, the MSE, and the length of confidence intervals for GLS-IV are consistently lower than those for all other estimators. Notably, when the AR coefficient is ± 0.9, GLS-IV exhibits an MSE that is about 10 times smaller than that of GMM. This indicates that GLS-IV yields more precise and efficient estimates across the entire range of values considered while maintaining the correct coverage for the confidence intervals. This can partly be explained by looking at the F-Test from the first-stage regression. When the AR parameter is ± 0.9, GLS-IV has by far the highest value. Compared with GMM, it can be as large as 15 times higher; e.g., below 10 for GMM, which is a sign of weak instruments and larger than 40 for GLS-IV, which indicates very strong instruments. The FF estimator has a performance intermediate between the two other estimators. The results seem not to depend on the value of the AR coefficient. In none of the cases considered does FF perform better than GLS-IV. Overall, these findings highlight the advantages of using GLS-IV in this scenario.

The results for the $MA(1)$ case are presented in Figures 3-4. They are qualitatively similar. The bias is very small for all estimators, especially GLS-IV. The MSE is smallest for GLS-IV and considerably so as the MA parameter φ_u is near the boundary of the non-invertible region, i.e., $\varphi_u = \pm\, 0.9$. All estimators have a coverage rate near the 90% nominal level. However, the lengths of the confidence intervals are considerably shorter for GLS-IV and the largest for GMM. Again, this can be explained by looking at the F-Test from the first-stage regression, which is highest for GLS-IV and by a wide margin when $\varphi_u = \pm\, 0.9$.

All results points in the direction of favoring GLS-IV. A noteworthy comment is the following. It is often the case that the correlation structure is such that the process is near the unit root region for the AR case, e.g., $\rho_u = 0.9$, and near the noninvertible region in the MA case, e.g., $\varphi_u = -0.9$. Much work has been done to document the fact that this leads to problems of weak instruments, thereby leading to estimates with poor properties. As can be seen from the results presented, this issue disappears when using GLS-IV. Hence, one need not consider more complex refined alternative methods of inference.

7.3 Simulations with non-Exogenous Instruments

We now assess the performance of the estimators when the instruments are predetermined but not exogenous. Of course, the estimation methods are the same for all estimators. All that is being changed is the data-generating process. It is identical to the one described in Section (7.1), except for the equation that defines the instrument, which is replaced by

$$z_t = \sqrt{\gamma}\omega_{1,t} + \sqrt{2-\gamma}\omega_{2,t} + \alpha\epsilon_{t-1}.$$

Adding the component ϵ_{t-1} makes the instruments non-exogenous with respect to past innovations, though they remain pre-determined. For the simulations we set $\alpha = 1$ and we keep the remaining parameters at the baseline values defined in Table 1.

Figure 5 considers the case when Ω is known and u_t follows an $AR(1)$ process. The results are drastically different from the case with exogenous instruments. GLS-IV is nearly unbiased (consistent with the fact that it is consistent), while the bias of FF and GMM is substantial. Additional simulations show that it does not decrease as T increases, in line with the fact that they are inconsistent. GLS-IV has by far the smallest MSE, the smallest length of the confidence intervals with coverage rates that are near the 90% nominal level.

The coverage rates of the confidence intervals for FF and GMM are very poor and liberal (under-coverage). Again, the F-test in the first stage regression is highest with GLS-IV, but now this is only part of the explanation since FF and GMM are inconsistent. The results for the case with Ω is unknown are presented in Figure 6. They are virtually the same. Hence, the feasible GLS-IV procedure performs as well as if the true structure and parameter values were known. All these results are consistent with the theoretical arguments discussed and show that the feasible GLS-IV method performs as well as the infeasible one. Of importance here is that FF and GMM break down while GLS-IV remains adequate.

The results for the MA case are presented in Figure 7 for the case with Ω known and in Figure 8 for the feasible procedures with Ω unknown. When Ω is known, the results are qualitatively the same as in the $AR(1)$ case, except that the differences in performance between the estimators are even larger. Again, the bias of GLS-IV is near zero while those of FF and GMM are now very large. The same applies to the MSE. The coverage rate of the confidence intervals for GLS-IV is very near the nominal 90% level, while for FF and GMM they are very low. The lengths of the confidence intervals are again smaller with GLS-IV. The differences in the fist-stage F-tests are even larger. When Ω is unknown the same conclusions apply to the relative ranking of the estimators: GLS-IV has smaller bias and MSE, shorter length and better coverage rates for the confidence intervals. The only caveat is some bias when the MA parameter φ_u is near ± 0.9 and an associated coverage rate that is below the 90% nominal level, even though better than with FF and GMM. It is to be noted that the design adopted implies a high degree of correlation between the instruments and the past innovations when setting $\alpha = 1$. As shown in Section A.2 in the appendix, when $\alpha = 0.25$, arguably a more realistic value, the coverage rate of GLS-IV is very near the nominal 90% level, and according to all other measures (bias, MSE, length of the confidence intervals) still outperforms GMM and FF by a wide margin. Still, it remains an open question about how to obtain reliable coverage rates for the confidence interval for GLS-IV when the strength of the correlation between the instruments and past errors is high.

7.4 Comments about Non-invertible MA Processes

We now discuss the properties of the estimators when the MA process generating the errors u_t is non-invertible, i.e., $u_t = C(L)e_t$ with some roots of $C(L)$ inside the unit circle. For simplicity, we focus on the $MA(1)$ case. This should be enough to provide a clear overview of the issues and results, without much loss of generality and considerably less technical derivations. All details are provided in Appendices A.5 and A.6.

Consider first the general model given by $y_t = x_t'\beta + u_t$, where u_t is serially correlated. If the instruments are exogenous, then all estimators are consistent and the infeasible and feasible versions have the same asymptotic properties so that valid asymptotic inference can be applied. This results basically hinges on the fact that a filtered version of the error process will fit the observationally equivalent invertible representation. The second element is that there is a direct relationship between the covariance matrix of the non-invertible and invertible representations, which implies that one is simply a scaling factor of the other. Hence, the infeasible estimator using one or the other version yields the same estimate with the same limit distribution, a consequence of the observational equivalence.

When the instruments are not exogenous, then all estimators are inconsistent. It is unknown to us what could work in that case since non-invertibility is not a testable hypothesis, at least working with the first two moments as done when dealing with linear estimators or with normally distributed errors. Maybe with non-normal errors, some non-linear methods could achieve consistency.

Consider now models of the "RE case". Suppose that some rational expectations model specifies that $y_t = \beta E(x_{t+2})$, with no other source of randomness. This is equivalent to $y_t = x_{t+2} + u_{t+2}$, where $u_{t+2} = \beta(\mathbb{E}(x_{t+2}) - x_{t+2})$. Then the errors are an $MA(1)$ process. Suppose that x_t is an $AR(2)$ process of the form $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + v_t$. Then standard arguments shows that $u_t = e_t + \varphi_1 e_{t-1}$, with e_t and $i.i.d.(0, \sigma_e^2)$ process. It certainly can be the case that $\varphi_1 > 1$. For instance, if x_t is an $AR(2)$ process with parameters $(1.34, -0.42)$.

As discussed in Appendix A.6, GMM and FF are consistent and provide valid asymptotic inference whether the instruments are exogenous or not. On the other hand, GLS-IV is inconsistent with non-exogenous instruments. It is consistent with exogenous instruments but offers no improvement over GMM and FF, i.e., all estimators have similar finite sample properties. Hence, this is a case where GLS-IV breaks down while GMM and FF provide reliable inference. The problem is worth consideration in the sense that GLS-IV should not be applied blindly. There are, however, several problems. First, one must be confident that the model is a well specified one that falls in the "RE case" and not affected by additional source of randomness that could be serially correlated as in the Phillips curve example discussed in Section 2. Second, as stated above, it is in general very difficult to ascertain whether the errors process is invertible. Here, with the overly simplified model, one could simply fit a process for x_t. But in more complex models, this simple approach may not be feasible. We have no reliable guidelines to offer other than simply state that in such RE models GLS-IV is not appropriate with non-exogenous instruments, while GMM and FF are.

8 Empirical Application

To show that the method of estimation can make substantial differences in important empirical applications, we consider estimating the forward-looking New Keynesian Phillips Curve outlined in Section 2. Theory suggests that the parameter of expected inflation, β, should be equal to the discount factor: any positive value between 0 and 1 is acceptable. In the presence of sticky prices, if firms expect higher inflation tomorrow they raise prices today. The slope of the Phillips curve, i.e., the parameter of the output gap, or the marginal cost of labor, λ, should be positive. The current empirical evidence is extensive and in disagreement. The initial works by Galí and Gertler (1999) and Sbordone (2002) were initially seen as successful in estimating the NKPC parameters but were later criticized. Mavroeidis et al. (2014) provides an extensive survey and report a myriad of estimates of the parameters obtained via GMM using a vast number of combinations of measures for inflation, the output gap, and the labor share series (equivalent to the marginal cost of labor under some conditions). They also considered other variations that we will not address here. The results showed parameter estimates scattered across the parameter space including many with positive values of β and/or negative ones for λ. They conclude saying "the literature has reached a limit on how much can be learned about the New-Keynesian Phillips curve from aggregate time series." We use a subset of the cases they considered by focusing on a single set of instruments (sensitivity analyses to using different instruments are present in the Appendix). For GMM, the results are similar, namely parameter estimates all over the map and often of the wrong sign. With GLS-IV, almost all estimates support the theory with the coefficient on expected inflation below one and those on the output gap (or marginal cost of labor) above zero, provided one uses level or detrended inflation and the same for the output gap or the marginal cost of labor, i.e., not level against detrended. Hence, the evidence is that the theory is just fine. The widely dispersed estimates obtained using GMM are likely simply caused by biases due to non-exogenous instruments and/or simply inefficiency.

8.1 Estimation of the New Keynesian Phillips curve

We now consider the NKPC in its general formulation, where we allow for a cost-push shock. Substituting the unobserved expectation with the actual realization into (1), we obtain

$$y_t = \beta y_{t+h} + \lambda x_t + u_t. \tag{24}$$

The error term u_t is a combination of the forecast error and the cost-push shock, given by $u_t = [\eta_t + \beta(\mathbb{E}_t[y_{t+h}] - y_{t+h})] = [\eta_t + \beta\nu_t]$. Rational expectations imply $\mathbb{E}_t[\nu_t] = 0$, but not $\mathbb{E}_t[u_t] = 0$. The majority of prior studies that addressed this problem used GMM type estimators selecting instruments based on the moment condition $\mathbb{E}[z_t \nu_t]$. For example, Galí and Gertler (1999) include as instruments four lags of inflation, labor share, output gap, 10-year minus 90-day yield spread, wage inflation, and commodity price inflation. These are effectively pre-determined with respect to the shocks ν_t. However, empirical data often reveal that inflation is persistent: cost-push shocks tend to have a lasting effect over multiple periods. Consequently, the residuals in the regression display autocorrelation, violating the assumption $\mathbb{E}_t[u_t] = 0$ and the moment condition on which GMM relies. The instruments are then plausibly non-exogenous with respect to past innovations in η_t, rendering GMM inconsistent. To address this issue, many researchers have introduced lags of inflation in the regression to mitigate residual autocorrelation. But this does not solve the problem.

Estimating the NKPC using GLS-IV is a more robust approach. It allows for the presence of serial correlation in the errors without the need to introduce lags. GLS-IV remains consistent whether the cost-push shocks are *i.i.d.* or serially correlated. In either case, it also remains efficient. This makes GLS-IV an obvious choice for estimating the NKPC.

To apply the feasible GLS-IV procedure, we need to modify the estimation algorithm slightly, to account for the fact that inflation is present both as a dependent variable and as a regressor. The reason is that when estimating equation (20), we do not identify the parameters ρ_i, but instead some non-linear transformation of various parameters. To see this, apply the filter $(1 - \sum_{i=1}^{k} \rho_i L^i)$ to the model in (1) with $h = 1$. Then,

$$\begin{aligned} y_t &= \textstyle\sum_{i=1}^{k} \rho_i L^i y_t + \beta \left(1 - \sum_{i=1}^{k} \rho_i L^i\right) y_{t+1} + \lambda \left(1 - \sum_{i=1}^{k} \rho_i L^i\right) x_t + e_{tk} \\ &= \frac{\beta}{1+\beta\rho_1} y_{t+1} + \textstyle\sum_{i=1}^{k-1} \dfrac{(\rho_i - \beta\rho_{1+i})}{1+\beta\rho_1} y_{t-i} + \dfrac{\rho_k}{1+\beta\rho_1} y_{t-k} \\ &+ \frac{\gamma}{1+\beta\rho_1} x_t - \textstyle\sum_{i=1}^{k} \dfrac{\gamma}{1+\beta\rho_1} \rho_i x_{t-i} + e_{tk}, \end{aligned} \tag{25}$$

where in the second equation we group all the same terms that come from the filtration of y_t and y_{t+1}. Equation (25) clearly shows that when estimating a regression of the form

$$y_t = \eta_1 y_{t+1} + \textstyle\sum_{i=1}^{k} \tau_i y_{t-i} + \delta_0 x_t - \sum_{i=1}^{k} \delta_i x_{t-i} + e_t,$$

there is no correspondence between the parameters α_i and ρ_i for $i = 1, ..., k$. For example α_1 would identify the ratio $(\rho_1 - \beta\rho_2)/(1 + \beta\rho_1)$. But one can use instead the ratio δ_i/δ_0 which corresponds to ρ_i for $i = 1, ..., k$. Such a method goes as far back as Wallis (1967). Following this intuition, we can modify Step 2 of the algorithm as follows: consider the regression

$$y_t = \eta_1 y_{t+1} + \textstyle\sum_{i=1}^{k_T} \tau_i y_{t-i} + \delta_0 x_t - \sum_{i=1}^{k_T} \delta_i x_{t-i} + e_{t,k}, \quad t = k_{\max} + 1, ..., T.$$

Assuming valid instruments with $E[w_t e_{t,k}] = 0$, we can use IV to estimate the parameters $\hat{\rho}_i^D = \hat{\delta}_i/\hat{\delta}_0$ $(i = 1, ..., k_T)$. Repeat for $k_T = 1, ..., k_{\max}$ and use BIC to select the lag length k_T^*. Then compute the IV estimates for β and λ from the OLS regression

$$(y_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t-j}) = \beta(y_{t+1} - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t+1-j}) - (x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})'\lambda + e_{tk}, \quad (t = k_T^* + 1, ..., T),$$

Note that in this modified Step 2, we cannot use lagged values of y_t (here inflation) as instruments as they are part of the regression. However, these are reintroduced when applying GLS-IV, once the values of ρ_i have been estimated. Similar arguments applies when multiple leads or lagged values of y_t are present as regressors, with minor modifications.

8.2 Data and Results

We use the data from Mavroeidis et al. (2014), which contains the exact specification of all variables and their original sources. These were obtain from the *Journal of Economic Literature* website at: https://www.aeaweb.org/articles?id=10.1257/jel.52.1.124. This dataset consists of quarterly aggregate time series for inflation, the output gap, and the labor share. Assuming a Cobb-Douglas production function, the latter is equivalent to the marginal cost of labor. Hence, we use also use this measure. There are nine series for the output gap, ten series for the labor share, and 19 series for inflation. The variables are summarized in Table 2 where the set of inflation variables is divided into measures in level and in gap. The data spans from q1-1950 to q4-2011. We keep the same sample period to allow for a fair comparisons with the results of Mavroeidis et al. (2014). These series are the results of various measurement techniques, filters, or transformations, resulting in a total of 1710 different possible combinations. We can categorize the various proxies for inflation into two primary groups: those measured in levels and those measured in “gaps” or some deviations from various fitted trend function.

We estimate a regression of the form (3) via GMM and GLS-IV, including a constant with $h = 4$, so that the expected inflation is used. For the specification we decided to follow as closed as possible Galí and Gertler (1999): we use the same set of instruments listed above, all the measures of inflation in levels, and the labor share series as a proxy for the marginal cost of labor. The results are displayed in the top two panels of Figure 9. In each plot we present the estimates of the parameter for the labor share, λ, on the vertical axis and the one associated with the future expected inflation β on the horizontal axis. The plots on the left show the results using GLS-IV, while those on the right the results using GMM. We can see that both the optimal GMM and the GLS-IV precisely identify λ, i.e.

the slope of the Phillips curve. According to theory, λ should be a positive value close to 0. For GLS-IV, the estimates vary between 0 and 0.12 with a median value of 0.03, a small but economically relevant value in line with the findings of Rotemberg and Woodford (1997), Hazell et al. (2022), Gagliardone et al. (2024). They obtain estimates of 0.024, 0.006, and 0.02, respectively, using alternative methods based no micro data. Our results show that a proper time series method can yield results in line with alternative micro-based techniques.

The estimates with GMM are, however, much more dispersed ranging from -0.07 to 0.20. However, only GLS-IV seems to correctly identify the parameter of the expected inflation, β, as it almost always yields values below 1, with a median estimate of 0.95, consistent with what the theory suggests. GMM tends to produce estimates well above this threshold, i.e., with pairs in the wrong quadrant.

The bottom panels of Figure 9 provide evidence contrary to some claims made by Galí and Gertler (1999). They motivate the use of the labor share as a forcing variable since using inflation in level and the output gap to estimate the NKPC leads to puzzling results. They used the labor share in level. Here, instead we have 8 out of 10 measures of the labor share in gap form. This is not common practice. The reason is that until the mid-90s, the labor share remained constant, hence the measure in level was stationary. Afterward, it declined so it is necessary to take the gap to detrend the series to make it stationary; see Gwin and VanHoose (2008) and King and Watson (2012). Rudd and Whelan (2007) previously raised similar arguments criticizing the widely accepted practice of "substituting labor's share of income in place of detrended output". For this reason, we have 8 out of 10 measures of the labor share in gap form. The bottom panels display the results using measures of the output gap with all the measures of inflation in gap form presented in Table 2. The results qualitatively agree with those using the labor share, with slight differences in the median estimates, which for GLS-IV are $\lambda = 0.1$ and $\beta = 0.92$, while the GMM estimates remain in a territory unsupported by theory, mostly the right quadrant again. The rationale of the use of the detrended inflation instead of the level when the output gap is the forcing variable is that it is necessary to apply a similar transformation to both variables to ensure consistency in with respect to the analytical approach. The appendix presents a sensitivity analysis using a different set of instruments used by Galí et al. (2001). The results are very similar to those in Figure 9. Our findings show, contrary to Galí and Gertler (1999), that by simply detrending inflation and using GLS-IV one can obtain reasonable estimates of the NKPC even when the output gap measures are used.

9 Conclusion

In the context of linear models with endogenous regressors and serially correlated errors, we propose an alternative estimator, the GLS-IV, and show that a) it is more efficient than the optimal GMM estimator and b) it preserves consistency with non-exogenous but predetermined instruments. When both are consistent such as with exogenous instruments or models of the "RE case", GLS-IV still remains more efficient. We propose a feasible implementation of the GLS-IV and show via simulations that it outperforms other methods. Our findings suggest that, both with exogenous and with predetermined instruments, the GLS-IV yields the lowest MSE for a wide range of specifications. The performance of the feasible GLS-IV is similar to the infeasible GLS-IV testifying for the robustness of our proposed implementation.

In our empirical application, we estimate the parameters of the New Keynesian Phillips Curve, comparing the results obtained using the optimal GMM and the GLS-IV. Contrary to what Mavroeidis et al. (2014) concludes, we find that GLS-IV provides reliable parameter estimates that align well with theoretical expectations, in contrast to results obtained with GMM. This suggests that by applying the proper estimator, there is still much to be learned from aggregate macroeconomic time series data. Our conclusion is that optimal GMM should not be applied indiscriminately, assuming its optimality in every contexts. Despite the general suboptimality of the GMM, a disproportionate amount of empirical work has been devoted to finding better instruments for GMM or slightly changing the model specifications. Our paper demonstrates that more emphasis should be put on adopting a long-neglected class of estimators based on the principles of efficiency dictated by the GLS method, rather than forcing optimality upon an estimator that is not. Our empirical application to the NKPC shows the benefits of using a more efficient estimator than GMM such as the GLS-IV.

References

Amemiya, T. (1966). "Specification analysis in the estimation of parameters of a simultaneous equation model with autoregressive residuals." *Econometrica*, 34, 283–306.

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59, 817–858.

Berk, K. N. (1974). "Consistent autoregressive spectral estimates." *The Annals of Statistics*, 2, 489–502.

Buse, A. and Moazzami, B. (1991). "Some evidence on the potential of Theil's generalized two stage least squares estimator." *Empirical Economics*, 16, 335–349.

Cumby, R. E., Huizinga, J., and Obstfeld, M. (1983). “Two-step two-stage least squares estimation in models with rational expectations.” *Journal of Econometrics*, 21, 333–355.

Durbin, J. (1970). “Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables.” *Econometrica*, 38, 410–421.

Fair, R. C. (1970). “The estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors.” *Econometrica*, 38, 507–516.

Fair, R. C. (1984). *Specification, Estimation, and Analysis of Macroeconometric Models.* Harvard University Press.

Flood, R. P. and Garber, P. M. (1980). “A pitfall in estimation of models with rational expectations.” *Journal of Monetary Economics*, 6, 433–435.

Gagliardone, L., Gertler, M., Lenzu, S., and Tielens, J. (2024). *Anatomy of the Phillips Curve: Micro Evidence and Macro Implications.* Unpublished Manuscript.

Galí, J. and Gertler, M. (1999). “Inflation dynamics: A structural econometric analysis.” *Journal of Monetary Economics*, 44, 195–222.

Galí, J., Gertler, M., and Lopez-Salido, J. D. (2001). “European inflation dynamics.” *European Economic Review*, 45, 1237–1270.

González-Coya, E. and Perron, P. (2024). “Estimation in the presence of heteroskedasticty of unknown form: A Lasso-based approach.” *Journal of Econometric Methods*, 13, 29–48.

Gwin, C. R. and VanHoose, D. D. (2008). “Alternative measures of marginal cost and inflation in estimations of new keynesian inflation dynamics.” *Journal of Macroeconomics*, 30, 928–940.

Hansen, L. P. (1982). “Large sample properties of generalized method of moments estimators.” *Econometrica*, 50, 1029–1054.

Hayashi, F. and Sims, C. (1983). “Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments.” *Econometrica*, 51, 783–798.

Hazell, J., Herreno, J., Nakamura, E., and Steinsson, J. (2022). “The slope of the Phillips curve: Evidence from U.S. states.” *The Quarterly Journal of Economics*, 137, 1299–1344.

King, R. G. and Watson, M. W. (2012). “Inflation and unit labor cost.” *Journal of Money, Credit and Banking*, 44, 111–149.

Mavroeidis, S., Plagborg-Møller, M., and Stock, J. H. (2014). “Empirical evidence on inflation expectations in the New Keynesian Phillips Curve.” *Journal of Economic Literature*, 52, 124–188.

McFadden, D. L. (1999). "Chapter 4: Instrumental variables." *Lecture Notes, Department of Economics, University of Berkeley.*

Moazzami, B. and Buse, A. (1990). "Limited information estimator of a dynamic structural equation with autocorrelated arrors: a correction and new evidence." *Journal of Statistical Computation and Simulation*, 34, 195–208.

Ng, S. and Perron, P. (2005). "A note on the selection of time series models." *Oxford Bulletin of Economics and Statistics*, 67, 115–134.

Perron, P. and González-Coya, E. (2023). "Feasible GLS for time series regression." *Manuscript, Department of Economics, Boston University.*

Prais, S. and Winsten, C. (1954). "Trend estimators and serial correlation." *Cowles Foundation, Discussion Paper 383.*

Romano, J. P. and Wolf, M. (2017). "Resurrecting weighted least squares." *Journal of Econometrics*, 197, 1–19.

Rotemberg, J. J. and Woodford, M. (1997). "An optimization-based econometric framework for the evaluation of monetary policy." *NBER Macroeconomics Annual*, 12, 297–346.

Rudd, J. B. and Whelan, K. (2007). "Modelling inflation dynamics: A critical review of recent research." *Journal of Money, Credit and Banking*, 39, 155–170.

Sargan, J. D. (1961). "The maximum likelihood estimation of economic relationships with autoregressive residuals." *Econometrica*, 29, 414–426.

Sbordone, A. M. (2002). "Prices and unit labor costs: a new test of price stickiness." *Journal of Monetary Economics*, 49, 265–292.

Schwarz, G. (1978). "Estimating the dimension of a model." *The Annals of Statistics*, 6, 461–464.

Theil, H. (1961). *Economic Forecasts and Policy.* North-Holland.

Wallis, K. F. (1967). "Lagged dependent variables and serially correlated errors: A reappraisal of three-pass least squares." *The Review of Economics and Statistics*, 49, 555–567.

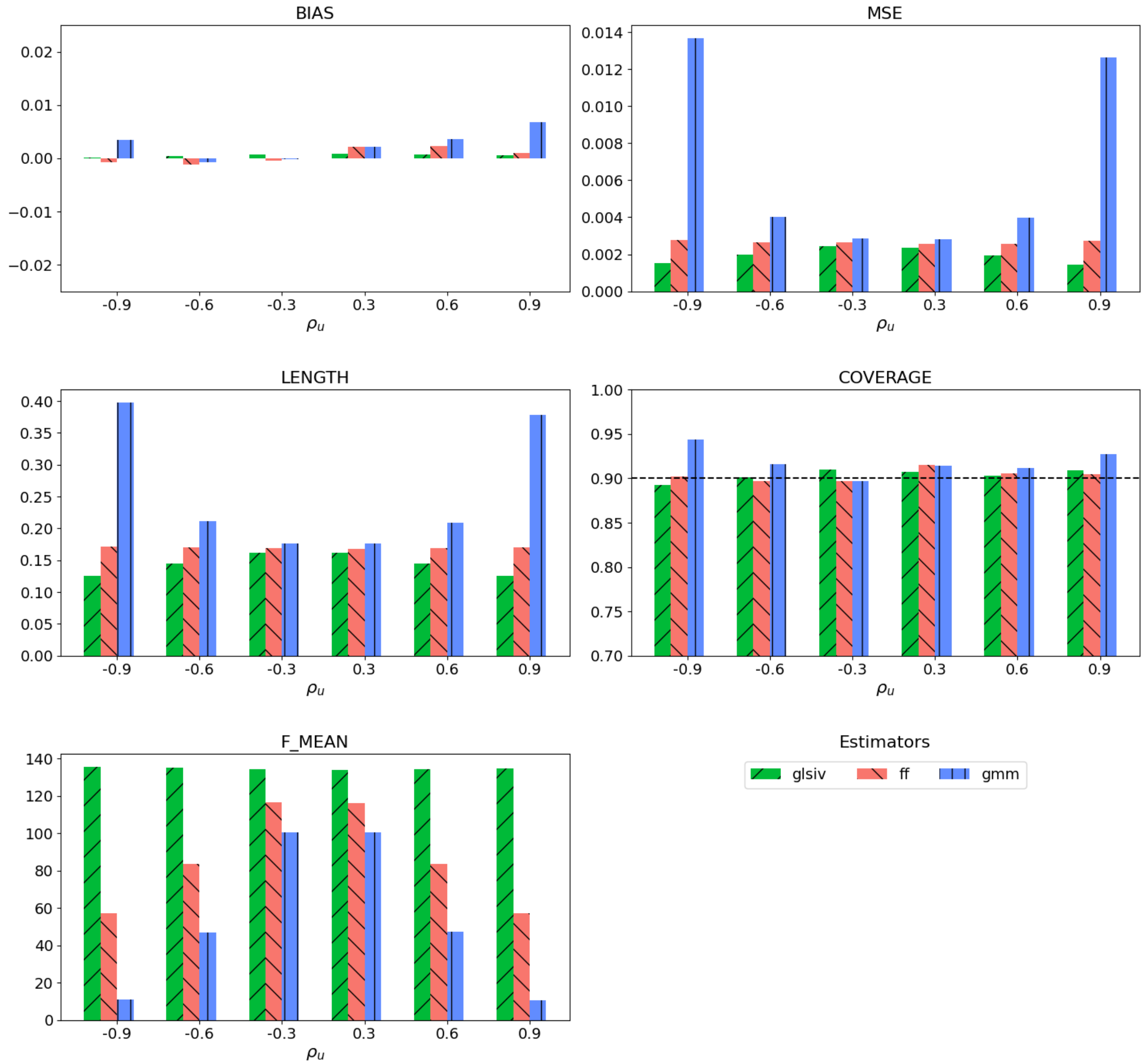
White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, 48, 817–838.

White, H. (1982). "Instrumental variables regression with independent observations." *Econometrica*, 50, 483–499.

Table 1: Parameter Descriptions

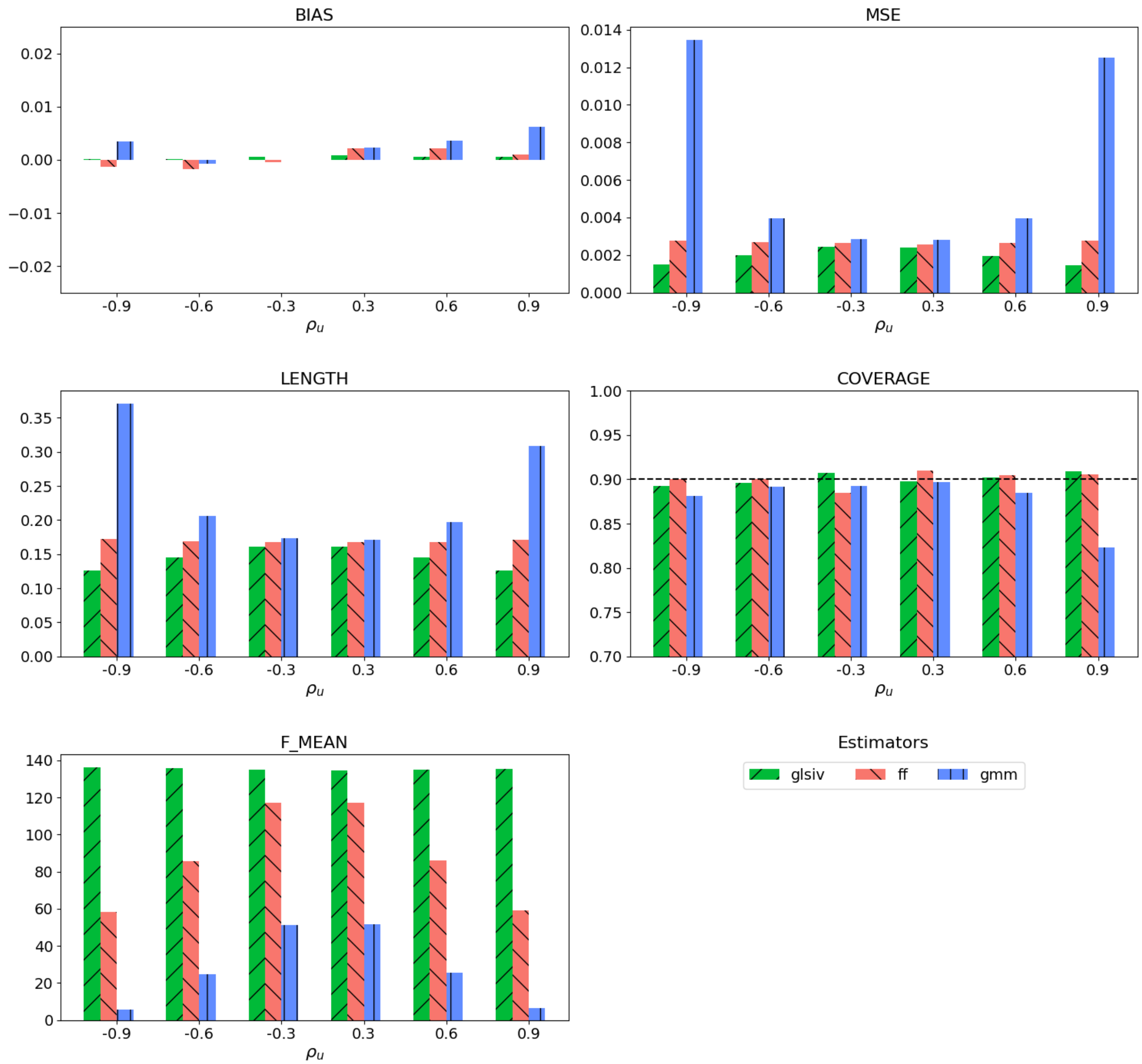
Parameter	**Value**	**Description**
N	1000	Number of replications
T	200	Sample size
$k_{\max}$	12	Maximum number lags considered in the estimation of the $AR(k)$ process
β	(1, 1)	Parameters to estimate
γ	1	Parameter governing the strength of the instrument
μ	1	Parameter in the data generating process of x_t
σ_ϵ^2	1	Variance of ϵ_t
σ_v^2	1	Variance of v_t
$\sigma_{z_1}^2$	5	Variance of ω_1
$\sigma_{z_2}^2$	5	Variance of ω_2
ϕ	0.5	Covariance between ϵ and v

Figure 1: Simulations with Ω known and $u \sim AR(1)$; exogenous instruments



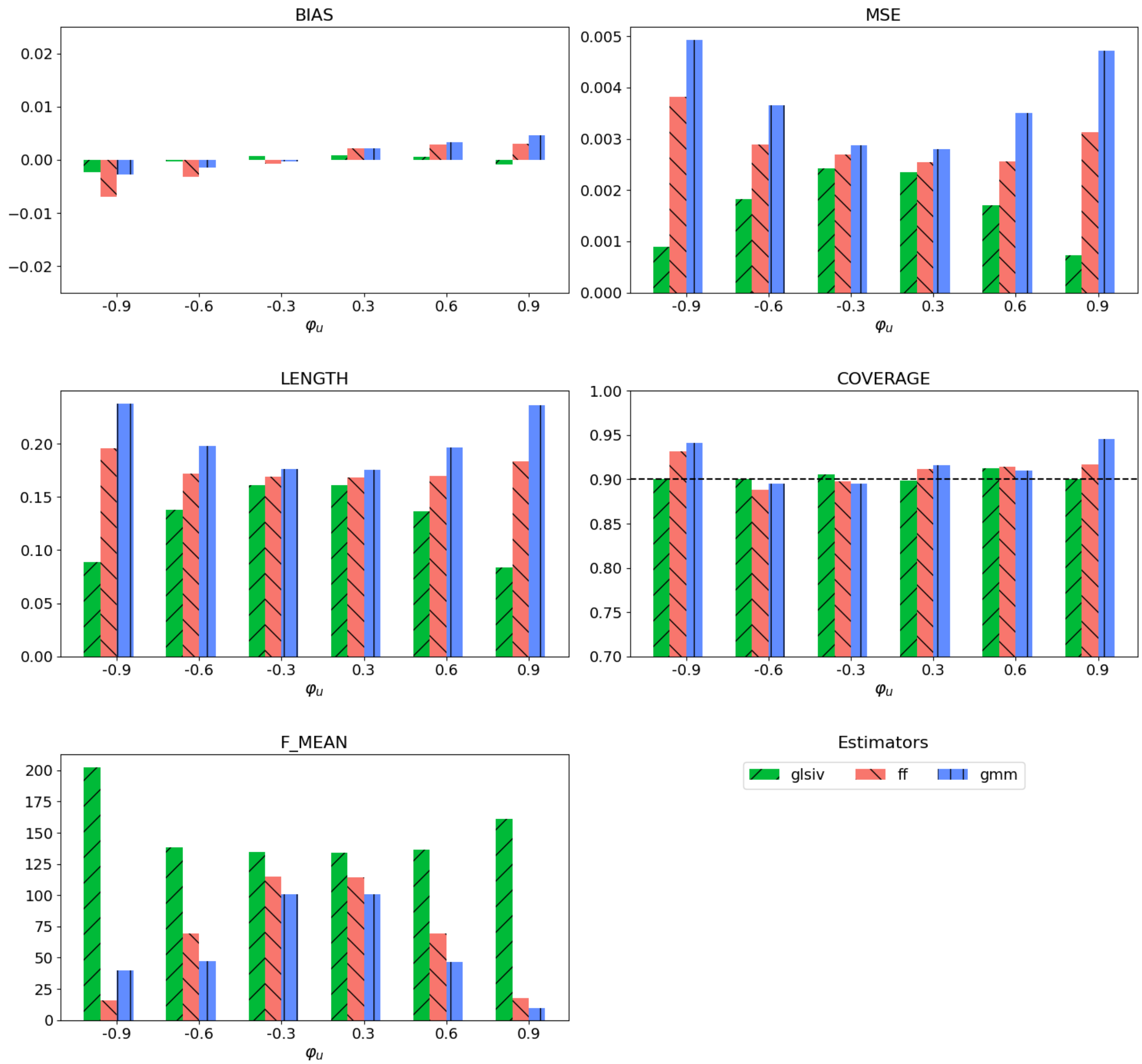
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 2: Simulations with Ω unknown and $u \sim AR(1)$; exogenous instruments



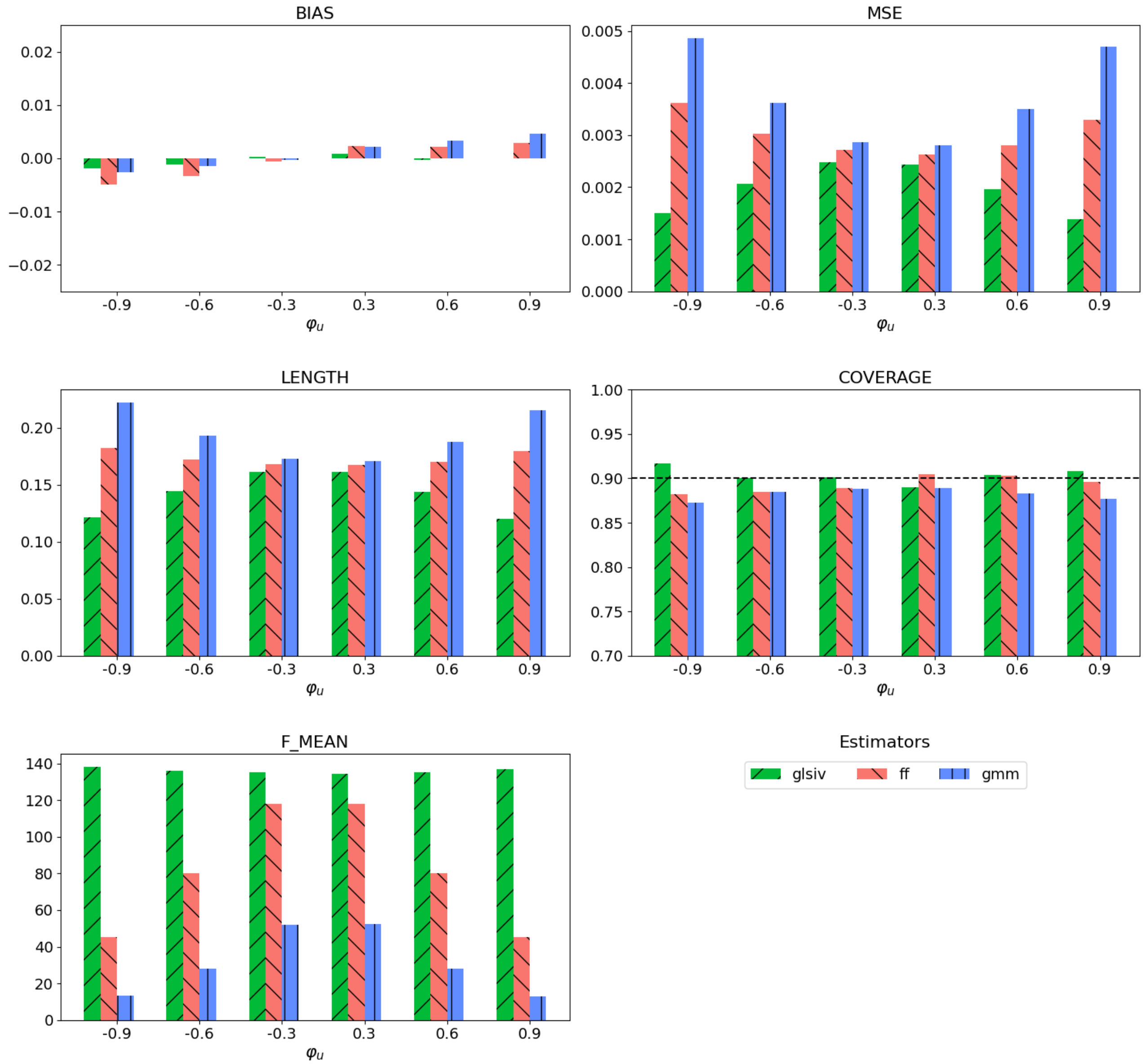
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 3: Simulations with Ω known, $u \sim MA(1)$; exogenous instruments



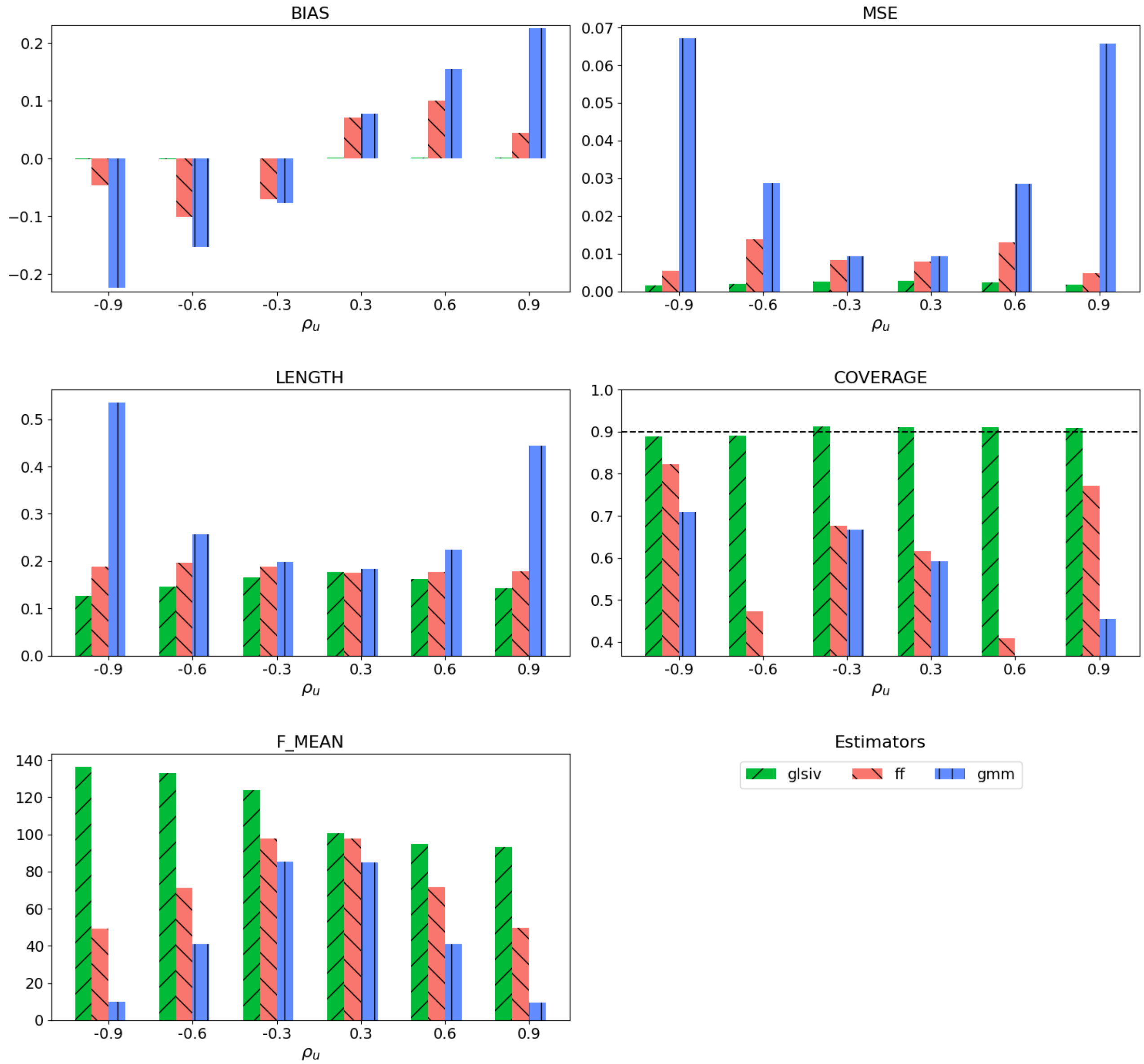
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 4: Simulations with Ω unknown, $u \sim MA(1)$; exogenous instruments



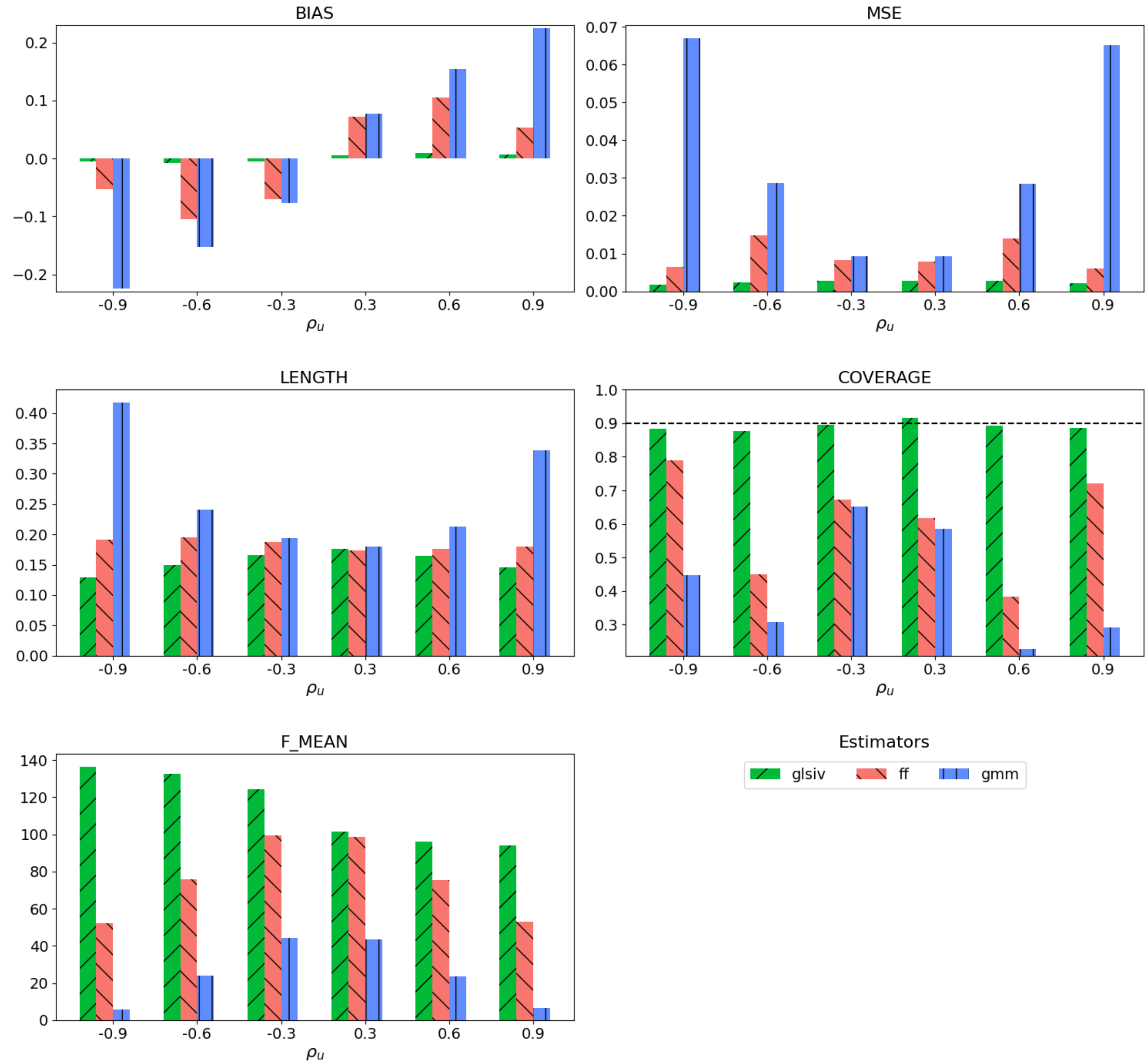
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 5: Simulations with Ω known and $u \sim AR(1)$; predetermined instruments



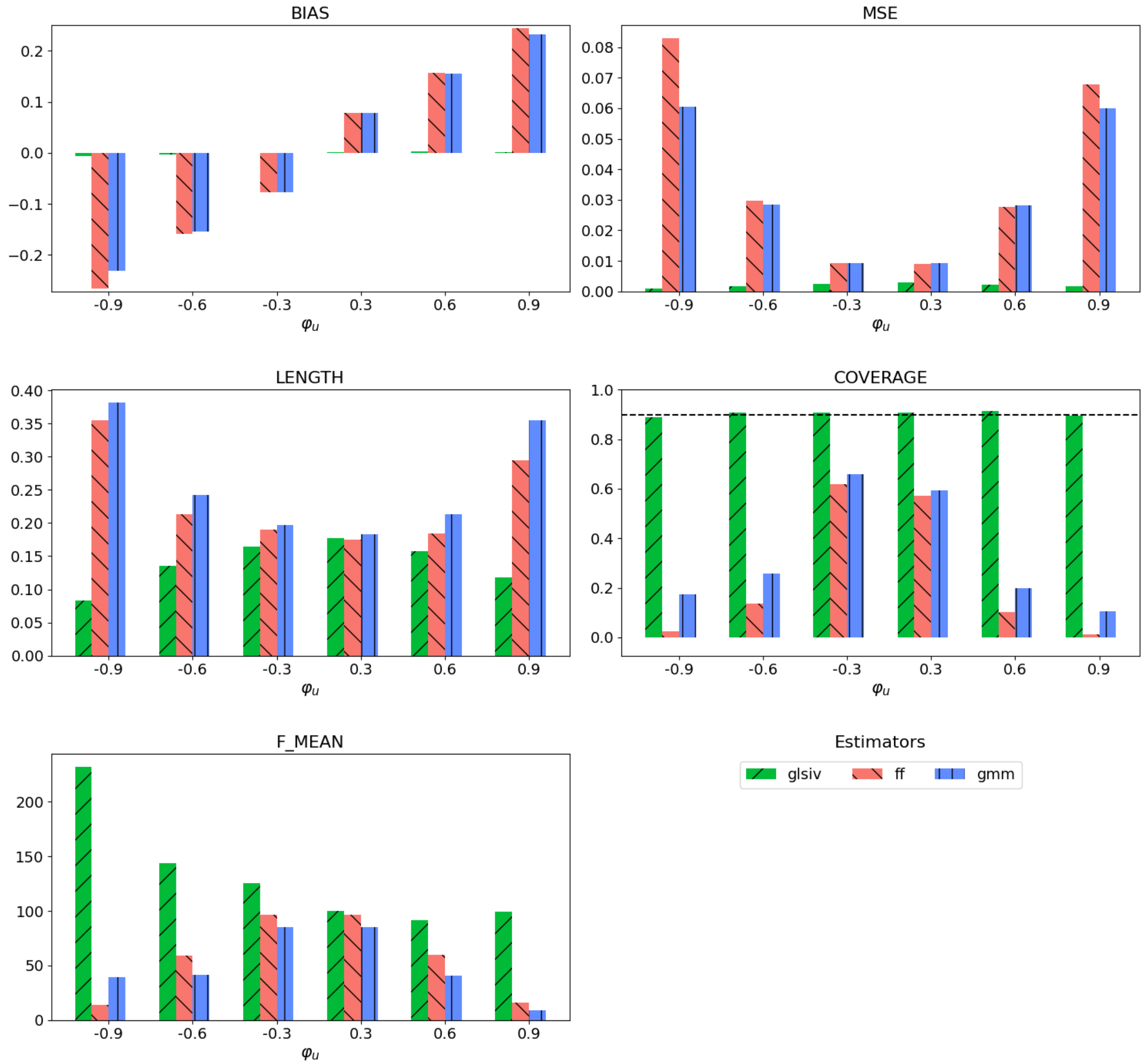
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 6: Simulations with Ω unknown and $u \sim AR(1)$; predetermined instruments



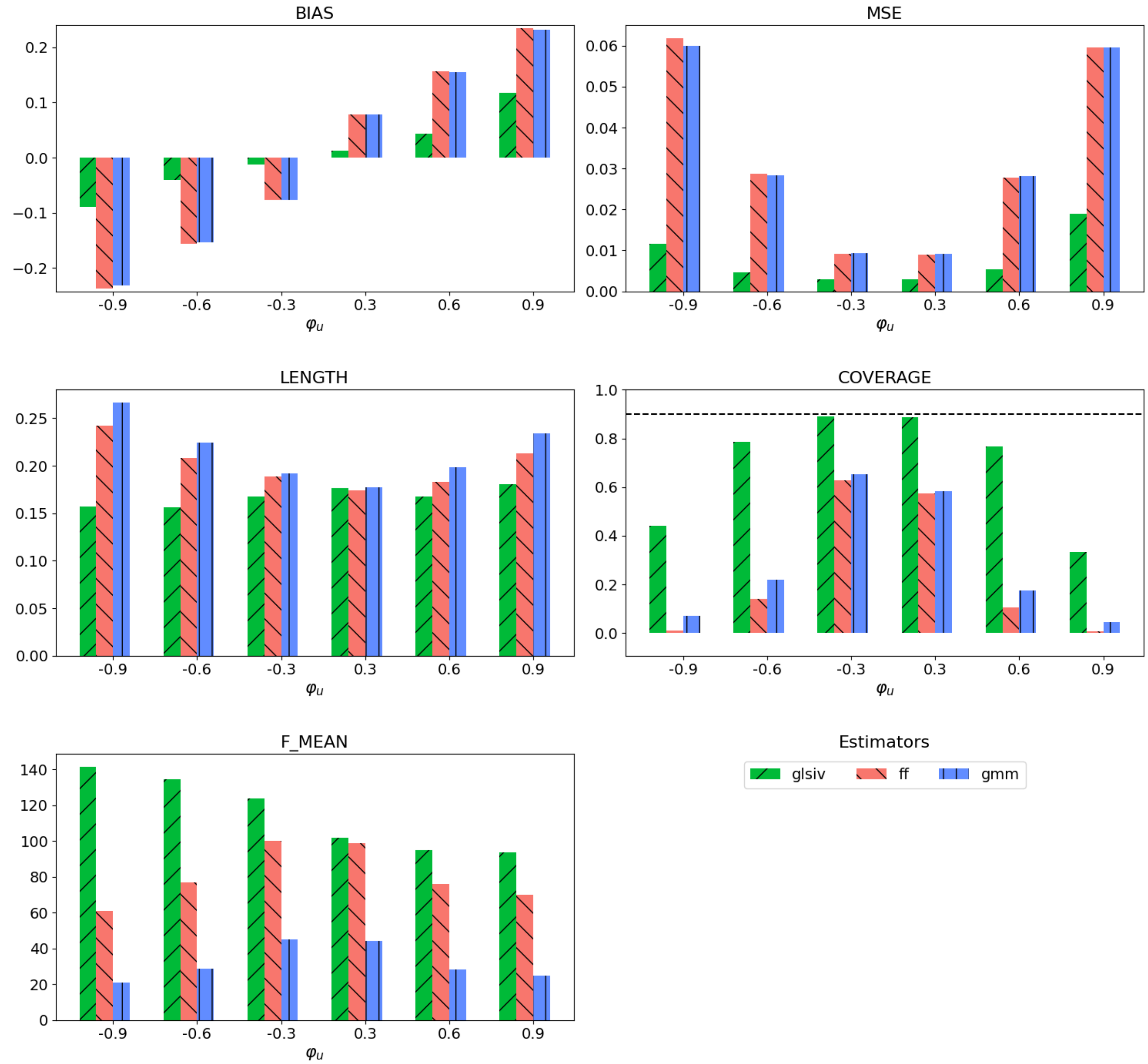
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$

Figure 7: Simulations with Ω known, $u \sim MA(1)$; predetermined instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure 8: Simulations with Ω unknown, $u \sim MA(1)$; predetermined instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Table 2: Summary of the variables used

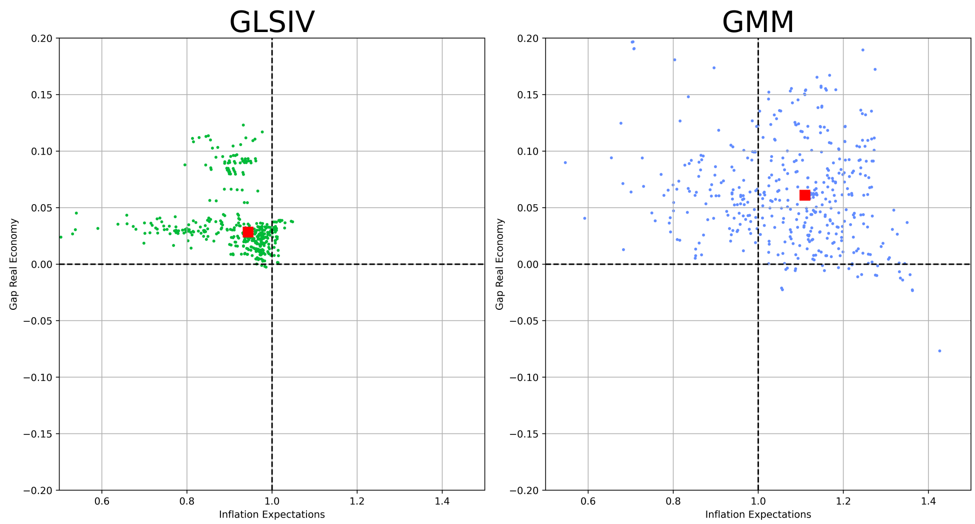
Inflation level	Inflation Gap
GDP Deflator	Filtered GDP Deflator Gap
CPI	Smoothed GDP Deflator Gap
Chained GDP Deflator	Filtered CPI Gap
GNP Deflator	Smoothed CPI Gap
Chained GNP Deflator	SPF-based CPI Gap
NFB GDP Deflator	Filtered Core CPI Gap
PCE	Smoothed Core CPI Gap
Core PCE	Filtered PCE Gap
Core CPI	Smoothed PCE Gap
	Filtered Core PCE Gap
	Smoothed Core PCE Gap

Labor Share	Output Gap
NFB	CBO difference
NFB Cointegration Relation	HP Filter GDP Gap ($\lambda = 10000$)
HP filtered NFB gap	BK filter GDP Gap
BK filtered NFB gap	Linear Detrended GDP Gap
Linearly Detrended NFB Gap	Quadratic Detrended GDP Gap
Quadratically Detrended NFB Gap	Real-time HP filtered GDP Gap ($\lambda = 10000$)
Real-time NFB HP Gap ($\lambda = 1600$)	Real-time BK Filter GDP Gap
Real-time NFB BK Gap	Real-time Linear Detrended GDP Gap
Real-time NFB Linear Detrended Gap	Real-time Quadratic Detrended GDP Gap
Real-time NFB Quadratic Detrended Gap	

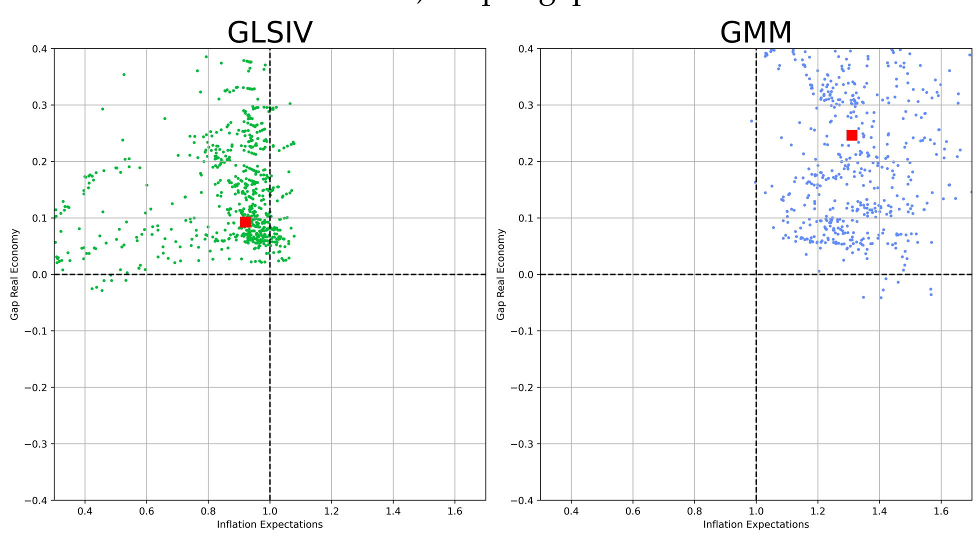
Abbreviations: CPI (Consumer Price Index), SPF (Survey of Professional Forecasters), NFB (Non-Farm Business), PCE (Personal Consumption Expenditures), CBO (Congressional Budget Office), HP (Hodrick-Prescott), BK (Baxter-King). Fore more details, see Mavroeidis et al. (2014).

Figure 9: Empirical Results for the NKPC

a) labor share



b) output gap



The four plots present the estimates of the coefficient of the forcing variable λ on the vertical axis and the estimates of the expected inflation coefficient β on the horizontal one. The plots on the left show the results for GLS–IV, while the two on the right display the estimates using GMM. The upper plots use labor share as a forcing variable and inflation measured in level. The bottom plots use output gap as a forcing variable and inflation measured in gap. Every point in a plot corresponds to a different combination of regressors and instruments using the available measures of inflation, labor share and output gap.

GLS–IV for Time Series Regressions with Application to the "New Keynesian Phillips Curve"
by Marco Olivari and Pierre Perron

Appendix for Online Publication

In this supplement, we present a variety of additional results to illustrate the robustness of the results and discuss in more details some issues discussed in the text.

A.1 Details about Argument 2 of Section 5

We show that GLS-IV is the only estimator that reduces to GLS which we know is optimal with non-spherical errors. Start by considering the GMM and replacing $Z = X$ Because we have as many instruments as regressors ($\ell = k$), GMM reduce to standard IV, so that

$$\begin{aligned}\beta_{GMM} &= \left[X'Z(Z'\Omega Z)^{-1}Z'X\right]^{-1} X'Z(Z'\Omega Z)^{-1}Z'y \\ &= \left[X'X(X'\Omega X)^{-1}X'X\right]^{-1} X'X(X'\Omega X)^{-1}X'y \\ &\neq \beta_{GLS},\end{aligned}$$

so that GMM is consistent but not efficient. The FF estimator also does not reduce to GLS, namely

$$\begin{aligned}\beta_{FF} &= \left[X'F'Z(Z'Z)^{-1}Z'FX\right]^{-1} X'F'Z(Z'Z)^{-1}Z'Fy \\ &= \left[X'F'X(X'X)^{-1}X'FX\right]^{-1} X'F'X(X'X)^{-1}X'Fy \\ &= \left[X'F'X\right]^{-1} X'F'y \neq \beta_{GLS}.\end{aligned}$$

Moving now to GLS-IV, if we use X as instruments we revert to GLS, since

$$\begin{aligned}\beta_{GLS-IV} &= \left[X'^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'^{-1}X\right]^{-1} X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'^{-1}y \\ &= \left[X'^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\right]^{-1} X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'^{-1}y \\ &= \left[X'\Omega^{-1}X\right]^{-1} X'\Omega^{-1}y \equiv \beta_{GLS}.\end{aligned}$$

A.2 MA(1) case with Different Strength of Exogeneity

In Section 7.3, we reported simulation for the case with $MA(1)$ errors and non-exogenous instruments with parameter $\alpha = 1$. This implies a very strong correlation, which led to the GLS-IV estimator having liberal coverage rate. Here, we show that with a smaller vale of $\alpha = 0.25$, the coverage rates for GLS-IV is fine. Also it remains by far superior to GMM and FF when considering bias, MSE and length of the confidence intervals. The simulation design is exactly the same, only the value of α changes. The results are presented in Figure

A.1 for the case of Ω unknown (the case with Ω known is similar). The results show no bias for GLS-IV and biased estimates for FF and GMM. The MSE and length of the confidence intervals are smallest with GLS-IV. But now the coverage rate of GLS-IV is very near the nominal 90% level, while that of FF and GMM are still below. This shows some sensitivity of the coverage rate of GLS-IV to the extent of non-exogeneity between the instruments and past innovations. When the strength of the correlation is moderate ($\alpha = 0.25$) everything is fine. However, there is some deterioration when the strength of the correlation is large ($\alpha = 1$) as reported in the main text. It remains an open question as what to be done to achieve a precise coverage rate for any permissible value of α.

A.3 Sensitivity Analysis to Other Error Processes

In the main text, we used simple $AR(1)$ and $MA(1)$ processes. We here present results for a broader class of processes. the design is exactly the same as in Section 7.3, except that we vary the specification for the structure of the error process. We only report the case for Ω unknown as the results assuming Ω omega known are similar. We consider the following:

- $MA(3)$: $u_t = e_t + \varphi_{u1}e_{t-1} + \varphi_{u2}e_{t-2} + \varphi_{u3}e_{t-3}$. We consider the following sets of values for $(\varphi_{u1}, \varphi_{u2}, \varphi_{u3})$: $(0.9, 0.3, 0.2)$, $(-0.5, 0.9, 0.2)$, $(0, 0.3, 0.2)$, $(0.4, -0.3, 0.2)$. The results are presented in Figure A.2 for exogenous instruments and Figure A.3 for non-exogenous but pre-determined instruments.

- $AR(2)$: $u_t = \rho_{u1}u_{t-1} + \rho_{u2}u_{t-2} + e_t$. We consider the following sets of values for (ρ_{u1}, ρ_{u2}): $(1.34, -0.42)$, $(0.5, -0.3)$, $(-0.5, 0.3)$, $(0.0, 0.3)$, $(0.5, 0.3)$. The results are presented in Figure A.4 for exogenous instruments and Figure A.5 for non-exogenous but pre-determined instruments.

- $ARMA(1,1)$: $u_t = \rho_u u_{t-1} + e_t + \varphi_u e_{t-1}$. We consider the following sets of values for (ρ_u, φ_u): $(0.5, -0.4)$, $(0.2, -0.4)$, $(0.5, -0.4)$, $(0.8, -0.4)$, $(0.8, 0.5)$. The results are presented in Figure A.6 for exogenous instruments and Figure A.7 for non-exogenous but pre-determined instruments.

The results lead to qualitatively similar conclusions: GLS-IV has lower bias, MSE and length of the confidence intervals in essentially all cases. The differences are larger when considering non-exogenous instruments compared to when the instruments are exogenous; e.g., most of the cases with $ARMA(1,1)$ errors. The superiority of GLS-IV can be especially large in some cases; e.g., an $AR(2)$ with parameters $(1.34, -0.42)$. GLS-IV also has better coverage rates, though some cases show that the 90% nominal is not achieved. In some cases, all methods fails with respect to the coverage rate. This is especially the case for MA processes and non-exogenous regressors. As before reducing the extent of the correlation between the instrument and the past innovation to $\alpha = 0.25$ yields substantially better results for GLS-IV.

A.4 Sensitivity Analysis to Different Instruments

In this section, we report the empirical results with an alternative set of instruments, namely those used by Galí et al. (2001). These are four lags of inflation and two lags of the labor share, the output gap, and the wage inflation. The results are presented in the bottom panels of Figure A.8 when using the labor share and in Figure A.9 when using the output gap. Again, note that in Steps 2-3, we do not use the lagged inflation instruments as they are part of the regression. However, these are reintroduced when applying GLS-IV, once the values of ρ_i have been estimated.

The results are in the same direction as those in the main text but not as definitive. When using the labor share, the GLS-IV estimates are again less dispersed than the GMM estimates. But we have estimates of β, the coefficient on expected inflation above one in both cases. The median of the GLS-IV estimates is slightly above one, while the median of the GMM estimates is slightly below one. In most cases, the estimates of γ, the coefficient associated with the labor shares are above zero. When using the output gap, the results are broadly in line with those reported in the main text. Almost all estimates of γ are above zero whether using GLS-IV or GMM. However the estimate of β are spread out above one using GMM, while they are more often below one when using GLS-IV. The medians of the estimates of β and γ are similar to those reported in the main text: slightly below one for β and about 0.1 for γ, consistent with theoretical predictions.

A.5 Non-invertible MA processes in the General Case

In this section, we present simulation results for the general case. The DGP is exactly as specified in Section 7.3, except that we concentrate on the $MA(1)$ case with parameters taking values

$$\varphi_u = \{-1.9, -1.6, -1.3, 1.3, 1.6, 1.9\}.$$

Hence, all values are greater than one in absolute value so that the process is non-invertible. We consider two cases with exogenous ($\alpha = 0$) and non-exogenous instruments ($\alpha = 1$). Recall that here the instruments are correlated with the lagged value of the innovation, namely e_{t-1}. Since the results assuming Ω known or not are the same, we present only the results for the latter case. These are in Figures A.10 for the case with exogenous instruments and in Figure A.11 for the case with non-exogenous but pre-determined instruments.

When the instruments are exogenous, the results are similar to those reported in Figure 4 for the invertible $MA(1)$ case. The main differences are: a) slightly larger bias, including for GLS-IV; b) much improved coverage rates for GMM but now with a larger length of the confidence intervals. The relative ranking in terms of MSE remains the same with GLS-IV having the smallest value across the parameter space. The intuition for this result is as follows. Any non-invertible MA process has an observationally equivalent representation in terms of an invertible process. GLS will fit this process adequately. Then inference can proceed as usual. The same applies to FF.

Things are very different when the instrument are not exogenous. The bias and MSE are much larger for all estimates, though those for GLS-IV are still the smallest. What is striking is the complete breakdown in inference. The coverage rates of the confidence intervals are near zero for all estimators, making a comparisons of the length of the confidence intervals meaningless.

Some explanations for this issue can be understood using the following arguments. To keep things simple and still get an idea of the main issues, we consider $MA(1)$ errors given by

$$u_t = \varphi_u e_{t-1} + e_t = (1 + \varphi_u L)e_t,$$

with $|\varphi_u| > 1$. This error process u_t is observationally equivalent to

$$v_t = \varphi_u^{-1}\epsilon_{t-1} + \epsilon_t = (1 + \varphi_u^{-1}L)\epsilon_t,$$

where $\epsilon_t = \varphi_u e_t$. Let

$$\Omega_{MA}(\varphi_u) = \sigma_e^2\Gamma = \sigma_e^2 \begin{pmatrix} 1+\varphi_u^2 & \varphi_u & 0 & \cdots & 0 \\ \varphi_u & 1+\varphi_u^2 & \varphi_u & \cdots & 0 \\ 0 & \varphi_u & 1+\varphi_u^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\varphi_u^2 \end{pmatrix}$$

be the covariance matrix of u_t. Given that $1 + \varphi_u^2 = \varphi_u^2(1 + 1/\varphi_u^2)$ and that $\varphi_u^2 = \varphi_u^2(1/\varphi_u^2)$, we have

$$\Omega_{MA}(\varphi_u) = \varphi_u^2\Omega_{MA}(1/\varphi_u) \tag{A.1}$$

Equation (A.1) is key for understanding the breakdown in inference for GLS-IV, FF, and GMM. Consider first GLS-IV. We have

$$\begin{aligned} & \hat{\beta}_{GLS-IV} \\ = \; & \left[X'\Omega_{MA}^{-1}(\varphi_u)Z\left(Z'\Omega_{MA}^{-1}(\varphi_u)Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u)X\right]^{-1} \\ & \times X'\Omega_{MA}^{-1}(\varphi_u)Z\left(Z'\Omega_{MA}^{-1}(\varphi_u)Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u)y \\ = \; & \left[X'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\left(Z'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u^{-1})X\right]^{-1} \\ & \times X'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\left(Z'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u^{-1})y \\ = \; & \beta + \left[X'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\left(Z'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u^{-1})X\right]^{-1} \\ & \times X'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\left(Z'\Omega_{MA}^{-1}(\varphi_u^{-1})Z\right)^{-1}Z'\Omega_{MA}^{-1}(\varphi_u^{-1})u \end{aligned}$$

since the proportionality factor φ_u^2 cancels out. Hence, consistency holds provided

$$\mathbb{E}[Z'\Omega_{MA}^{-1}(\varphi_u^{-1})u] = 0, \tag{A.2}$$

where $u = (u_0, ..., u_T)'$. This is equivalent to

$$\mathbb{E}[Z_t^* e_t^*] = \mathbb{E}\left[\left((1+\varphi_u^{-1}L)^{-1}Z_t\right)\left((1+\varphi_u^{-1}L)^{-1}u_t\right)\right] = 0,$$

where

$$e_t^* = (1+\varphi_u^{-1}L)^{-1}(1+\varphi_u L)e_t = e_t + \varphi_u(\varphi_u^{-1} - \varphi_u)\textstyle\sum_{j=1}^{\infty}(-\varphi_u)^{-j}e_{t-j}$$

and $Z_t^* = \sum_{j=0}^{\infty}(\varphi_u)^{-j}Z_{t-j}$ It is clear then that condition A.2 is not satisfied with non exogenous instruments, i.e., $\mathbb{E}(Z_t^* e_t^*) \neq 0$.

Consider now FF. It uses an upper triangular matrix $F(\varphi_u)$ such that $F'(\varphi_u)F(\varphi_u) = \Omega_{MA}(\varphi_u)$. From the arguments above, it is clear that $F(\varphi_u) = F(\varphi_u^{-1})$, since

$$F'(\varphi_u)F(\varphi_u) = \Omega_{MA}(\varphi_u) = \varphi_u^2\Omega_{MA}(1/\varphi_u) = F'(\varphi_u^{-1})F(\varphi_u^{-1}). \tag{A.3}$$

Then, with $P_Z = Z\left(Z'Z\right)^{-1}Z'$, the usual orthogonal projection matrix,

$$\begin{aligned}
\hat{\beta}_{FF} &= [X'F'(\varphi_u)P_Z F(\varphi_u)X]^{-1}X'F'(\varphi_u)P_Z F(\varphi_u)y \\
&= \left[X'F'(\varphi_u^{-1})P_Z F(\varphi_u^{-1})X\right]^{-1}X'F'(\varphi_u^{-1})P_Z F(\varphi_u^{-1})y \\
&= \beta + \left[X'F'(\varphi_u^{-1})P_Z F(\varphi_u^{-1})X\right]^{-1}X'F'(\varphi_u^{-1})P_Z F(\varphi_u^{-1})u
\end{aligned} \tag{A.4}$$

since again the proportionality factor φ_u cancels out. Hence, the condition for consistency is

$$\mathbb{E}[Z'F'(\varphi_u^{-1})u] = 0 \tag{A.5}$$

This is equivalent to

$$\mathbb{E}[Z_t\dot{e}_t] = \mathbb{E}\left[Z_t\left((1+\varphi_u^{-1}L^{-1})^{-1}u_t\right)\right] = 0,$$

where

$$\mathbb{E}[Z_t\dot{e}_t] = \mathbb{E}\left[Z_t\left((1+\varphi_u^{-1}L^{-1})^{-1}u_t\right)\right] = 0,$$

with

$$\begin{aligned}
\dot{e}_t &= (1+\varphi_u^{-1}L^{-1})^{-1}(1+\varphi_u L)e_t \\
&= \textstyle\sum_{j=0}^{\infty}\left(-\varphi_u^{-1}\right)^j u_{t+j} = u_t - \varphi_u^{-1}u_{t+1} + \varphi_u^{-2}u_{t+2} + ... \\
&= \varphi_u e_{t-1} + e_t - \varphi_u^{-1}(\varphi_u e_t + e_{t+1}) + \varphi_u^{-2}(\varphi_u e_{t+1} + e_{t+2}) + ... \\
&= \varphi_u e_{t-1},
\end{aligned} \tag{A.6}$$

so that $\mathbb{V}(\dot{e}_t) = \varphi_u^2\sigma_e^2$. This means that condition A.5 is not satisfied with non exogenous instruments since one can have $\mathbb{E}(Z_t e_{t-1}) \neq 0$. When the instruments are exogenous, the limit distribution of the infeasible FF estimator is

$$\sqrt{T}(\hat{\beta}_{FF} - \beta) \xrightarrow{d} N(0, [X'F(\varphi_u)P_Z F(\varphi_u)X]^{-1}).$$

Since the feasible version will fit the invertible representation given the observational equivalence, the limit distribution of the feasible version will be, in view of (A.6),

$$\begin{aligned} & \sqrt{T}(\hat{\beta}_{FF} - \beta) \xrightarrow{d} N(0, \varphi_u^2 plim_{T\to\infty}\sigma^2 \left[X'F'(\varphi_u^{-1})P_Z F(\varphi_u^{-1})X\right]^{-1}) \\ = \quad & N(0, plim_{T\to\infty}\sigma^2 \left[X'F'(\varphi_u)P_Z F(\varphi_u)X\right]^{-1}) \end{aligned}$$

Hence, the two are equivalent and the feasible version have provides the correct asymptotic confidence intervals.

Moving now to GMM, we have

$$\begin{aligned} \hat{\beta}_{GMM} &= \left[X'Z(Z'\Omega Z)^{-1}Z'X\right]^{-1} X'Z(Z'\Omega Z)^{-1}y \\ &= \beta + \left[X'Z(Z'\Omega Z)^{-1}Z'X\right]^{-1} X'Z(Z'\Omega Z)^{-1}u. \end{aligned}$$

The condition for consistency is then simply

$$\mathbb{E}[Z'u] = 0.$$

This is equivalent to

$$\mathbb{E}[Z_t u_t] = \mathbb{E}\left[Z_t(1 - \varphi_u L)e_t\right] = \mathbb{E}\left[Z_t(e_t - \varphi_u e_{t-1})\right] = 0,$$

which, in general, requires exogenous instruments. When this is the case, the limit distribution of the feasible and infeasible version of GMM coincide and is given by

$$\sqrt{T}(\hat{\beta}_{FF} - \beta) \xrightarrow{d} N(0, plim_{T\to\infty}\sigma^2[X'Z(Z'\Omega Z)^{-1}Z'X^{-1}]).$$

A.6 Non-invertible MA Processes in the RE Case

Again, to keep things simple, we discuss the case of rational expectations model by looking at an application related to the estimation of the Taylor Rule. This is a case where the rational expectation hypothesis can generate errors that follows a moving average process that are uncorrelated with contemporaneous instruments. We will see that the Forward Filter and GMM remain consistent, while GLS-IV is not.

The Taylor Rule is widely used as an approximation of the monetary policy rule and states that the interest rate in each period r_t is a function of expected inflation $\mathbb{E}_t(\pi_{t+2})$ ad output gap g_t as follows

$$r_t = \beta_0 + \beta_1 \mathbb{E}_t(\pi_{t+2}) + \beta_2 g_t$$

Here, we specify a two-periods ahead expected inflation so that the expectation errors follow a $MA(1)$ process. We assume that inflation follows an $AR(2)$ process given by:

$$\pi_{t+2} = \phi_1 \pi_{t+1} + \phi_2 \pi_t + e_{t+2} + \gamma \eta_t, \tag{A.7}$$

where e_{t+2} and η_t are two error terms. The need of the latter will become clear later. Using the rational expectation hypothesis, we can rewrite A.7 as follows

$$r_t = \beta_0 + \beta_1 \pi_{t+2} + \beta_2 g_t + u_t \tag{A.8}$$

where the error follows an $MA(1)$ process given by

$$u_t = -\beta_1 (e_{t+2} + \phi_1 e_{t+1}).$$

We assume for simplicity that $e_t,\ g_t, \eta_t$ are jointly drawn from the following multivariate normal distribution

$$\begin{pmatrix} e_t \\ g_t \\ \eta_t \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_\eta^2 \end{bmatrix} \right).$$

We note that π_{t+2} is correlated with u_t in (A.8). We specify the instrument z_t as follows

$$z_t = \eta_t + \alpha e_{t-1},$$

where η_t guarantees that the relevance condition is satisfied and α controls the exogeneity of the instrument. Note that whenever $\alpha = 0$, then u_t and z_t are uncorrelated. This implies that GLS-IV, FF, and GMM are all consistent. However, when $\alpha \neq 0$ multiple cases need to be considered. Within the set of parameters ϕ_1, ϕ_2 such that inflation is stationary (i.e., $-(1-\phi_2) < \phi_1 < 1 - \phi_2$ and $\phi_2 > -1$) if $|\phi_1| < 1$ then u_t is invertible and all estimators are all consistent. If (ii) $|\phi_1| > 1$ then u_t is not invertible and GLS-IV is not consistent while FF an GMM are. To understand why in case (ii) FF is consistent we rewrite the condition A.5 for the rational expectation case as

$$\mathbb{E}[Z' F(\phi_1^{-1}) u] = 0, \tag{A.9}$$

where $u = (u_2, ..., u_{T+2})'$. This is equivalent to

$$\mathbb{E}[Z_t \dot{e}_{t+2}] = \mathbb{E}\left[Z_t \left((1 + \phi_1^{-1} L^{-1})^{-1} u_{t+2} \right) \right] = 0,$$

and

$$\begin{aligned} \dot{e}_{t+2} &= \beta_1 (1 + \phi_1^{-1} L^{-1})^{-1} (1 + \phi_1 L) e_{t+2} \\ &= \beta_1 (\phi_1 e_{t+1} + e_{t+2} - \phi_1^{-1} (\phi_1 e_{t+2} + e_{t+3}) + \phi_1^{-2} (\phi_1 e_{t+3} + e_{t+4}) + ...) \\ &= \beta_1 \phi_1 e_{t+1}. \end{aligned}$$

This means that condition A.9 is satisfied with non exogenous instruments ($\alpha \neq 0$). A similar argument for GMM shows that

$$\mathbb{E}[Z_t u_{t+2}] = \mathbb{E}\left[Z_t \beta_1 (1 - \phi_1 L) e_{t+2} \right] = \mathbb{E}\left[Z_t \beta_1 (e_{t+2} - \phi_1 e_{t+1}) \right] = 0,$$

requiring only pre-determined instruments. Following the arguments in the previous subsection, the limit distribution of GMM and FF are the same and as stated previously. For GLS-IV, the transformation will involve the observationally equivalent representation with $\alpha(L) = (1 + \phi^{-1}L)$. A researcher using the invertible model would not recover the true structural shocks, but rather

$$\begin{aligned}(1+\phi_1^{-1}L)^{-1}u_{t+k} &= (1+\phi_1^{-1}L)^{-1}(1+\phi_1 L)e_{t+k} \\ &= e_{t+k} + (\phi_1^{-1} - \phi_1)(1+\phi_1^{-1}L)^{-1}e_{t+k-1} \\ &= e_{t+k} + (\phi_1^{-1} - \phi_1)\textstyle\sum_{i=0}^{\infty}(-\phi_1^{-1}L)^i e_{t+k-1} \\ &= e_{t+k} + (\phi_1^{-1} - \phi_1)\textstyle\sum_{i=0}^{\infty}(-\phi_1)^{-i} e_{t+k-1-i}.\end{aligned}$$

The problem is with the second term, which involves all past values of the innovations, thereby requiring exogenous instruments for GLS-IV to remain consistent.

To describe the results above, we simulate the Taylor Rule and run the following regression:

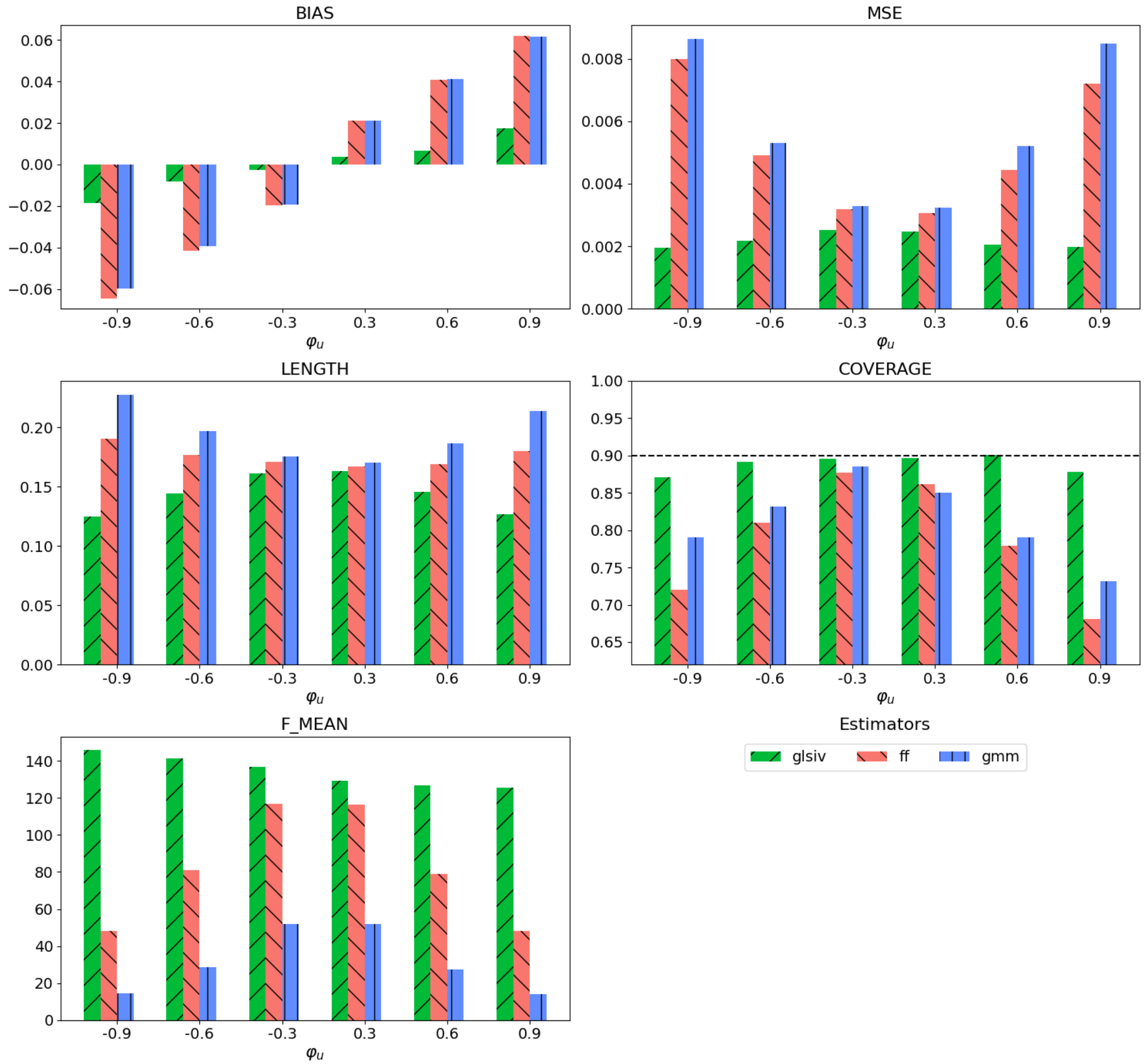
$$\begin{aligned}r_t &= \beta_0 + \beta_1 \pi_{t+2} + \beta_2 g_t + u_t, \\ \pi_{t+2} &= \Pi Z_t + \nu_t,\end{aligned}$$

where $\Pi = (\Pi_1, \Pi_2)$ and $Z_t = (z_t, z_{t-1})$. We set the parameters to the following values: $T = 1000$, $N = 500$, $j = 1$, $\beta = (1, 1, 1)$, $\sigma_e = 1$, $\sigma_v = 1$, $\sigma_\eta = 1$, and $\gamma = 1$.

Tables A-1 and A-2 report results where $\phi = [1.2, -0.6]$ with $\alpha = 1$ (non-exogenous instruments). The former assumes Ω is known and we can see that the GMM and FF estimators are unbiased GLS-IV is not. Qualitatively similar results are found when Ω is unknown and the feasible versions are used.

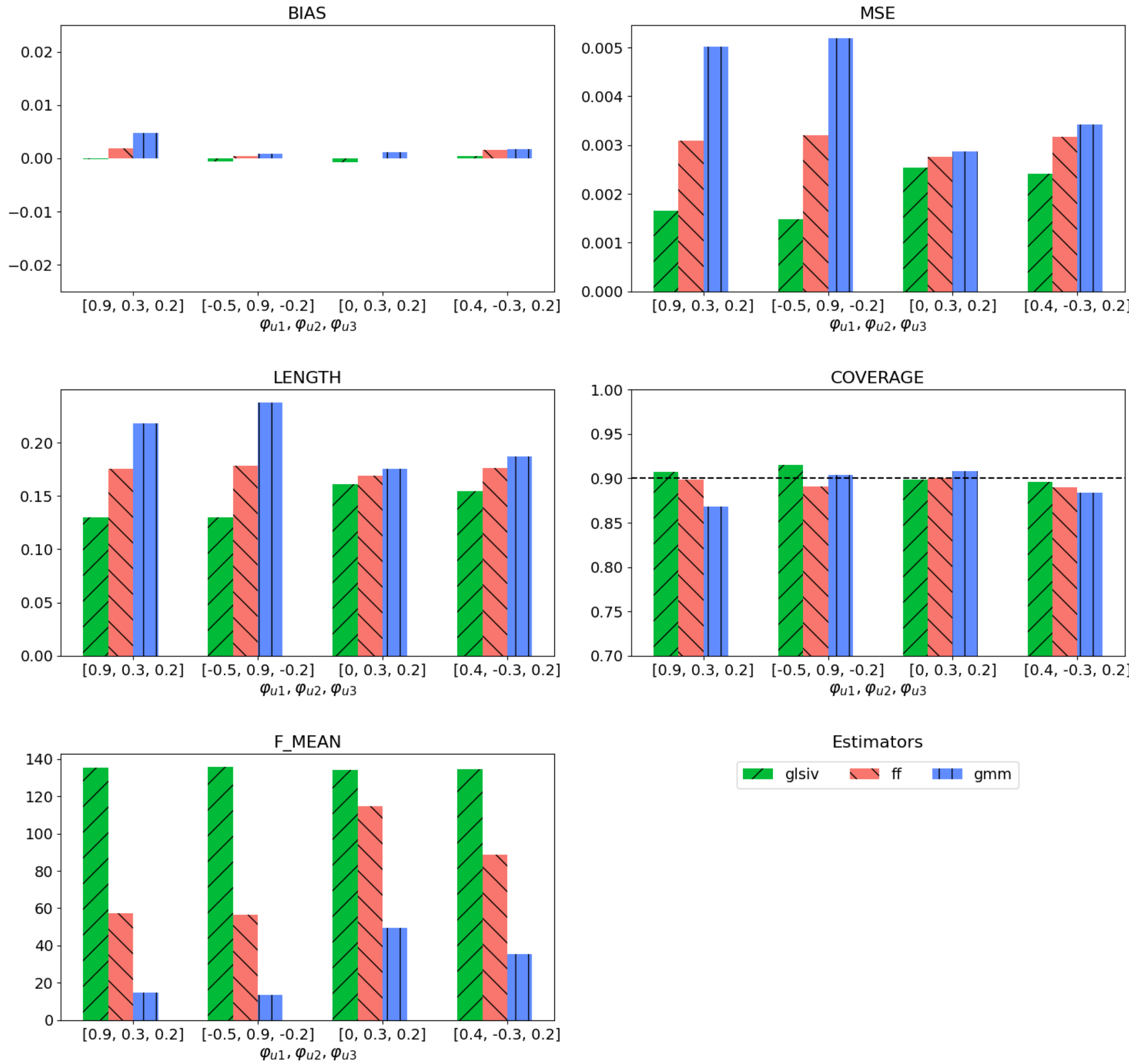
Tables A-3 and A-4 report similar results with $\alpha = 0$ (exogenous instruments). The former assumes Ω is known. Then, all estimators unbiased and perform similarly. The same conclusions apply to the case with the feasible versions of the estimators that assume Ω unknown. Again, all estimators are unbiased, however GLS-IV performs poorly in terms of MSE. Overall, the coverage rates are near the nominal 90% significance level, though the length of the confidence intervals are smaller with GMM or FF compared to GLS-IV.

Figure A–1: Simulations with Ω unknown, $u \sim MA(1)$; predetermined instruments ($\alpha = 0.25$)



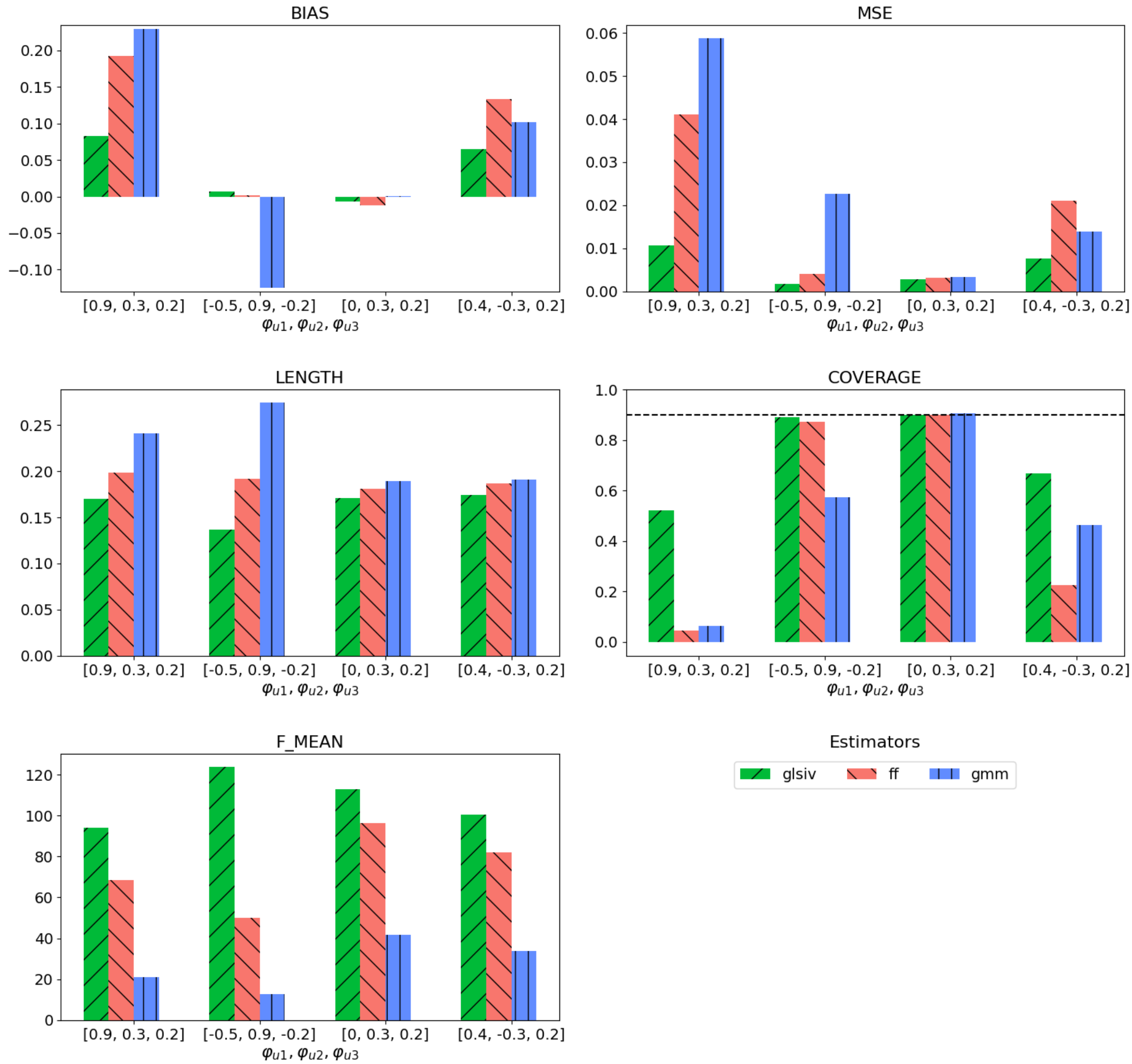
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the autoregressive parameter are considered: $\{-0.9, -0.6, -0.3, 0.3, 0.6, 0.9\}$.

Figure A–2: Simulations with Ω unknown, $u \sim MA(3)$; exogenous instruments



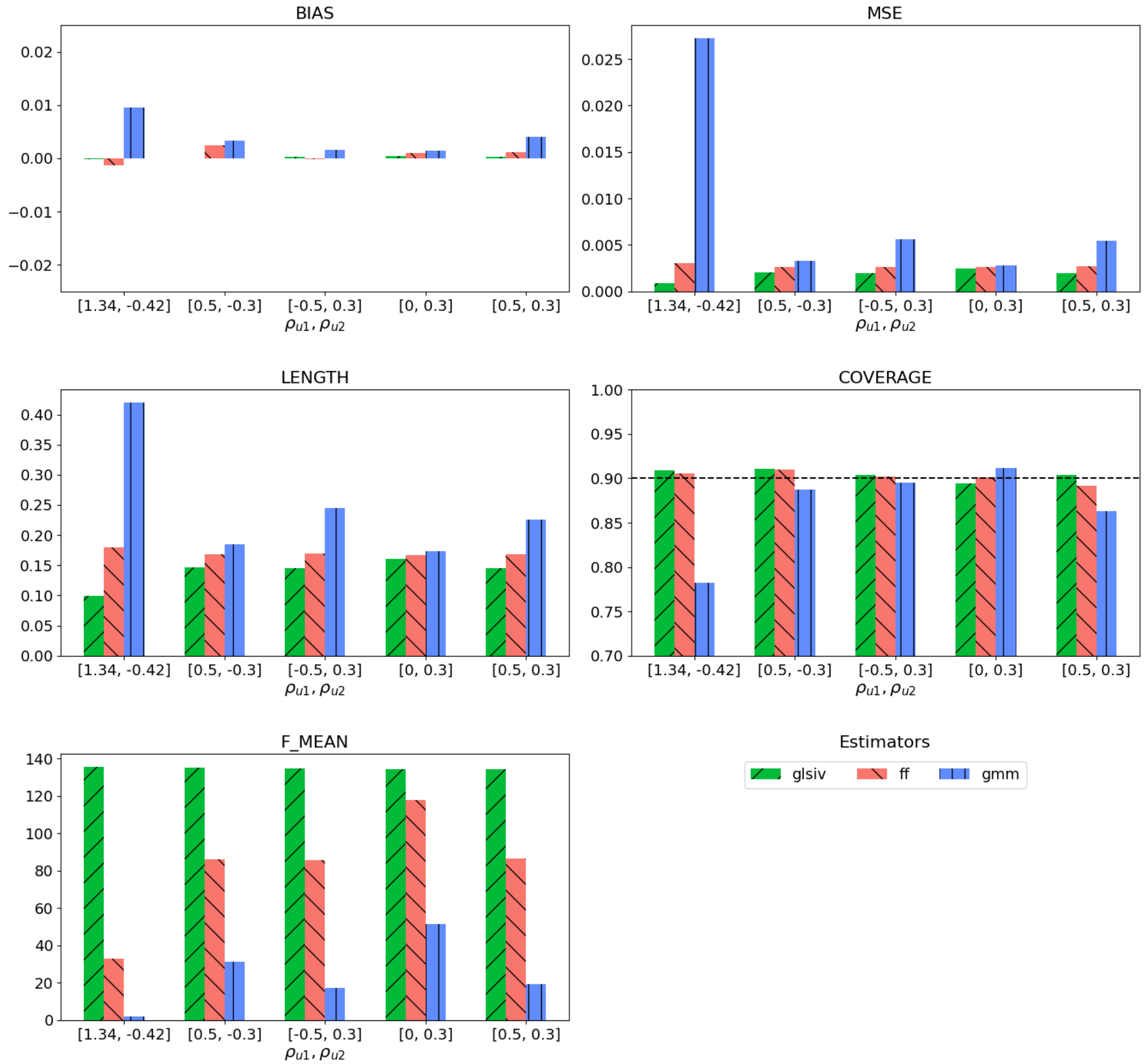
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the moving average parameter are considered: $\{(0.9, 0.3, 0.2), (-0.5, 0.9, -0.2), (0, 0.3, 0.2), (0.4, -0.3, 0.2)\}$.

Figure A–3: Simulations with Ω unknown, $u \sim MA(3)$; predetermined instruments



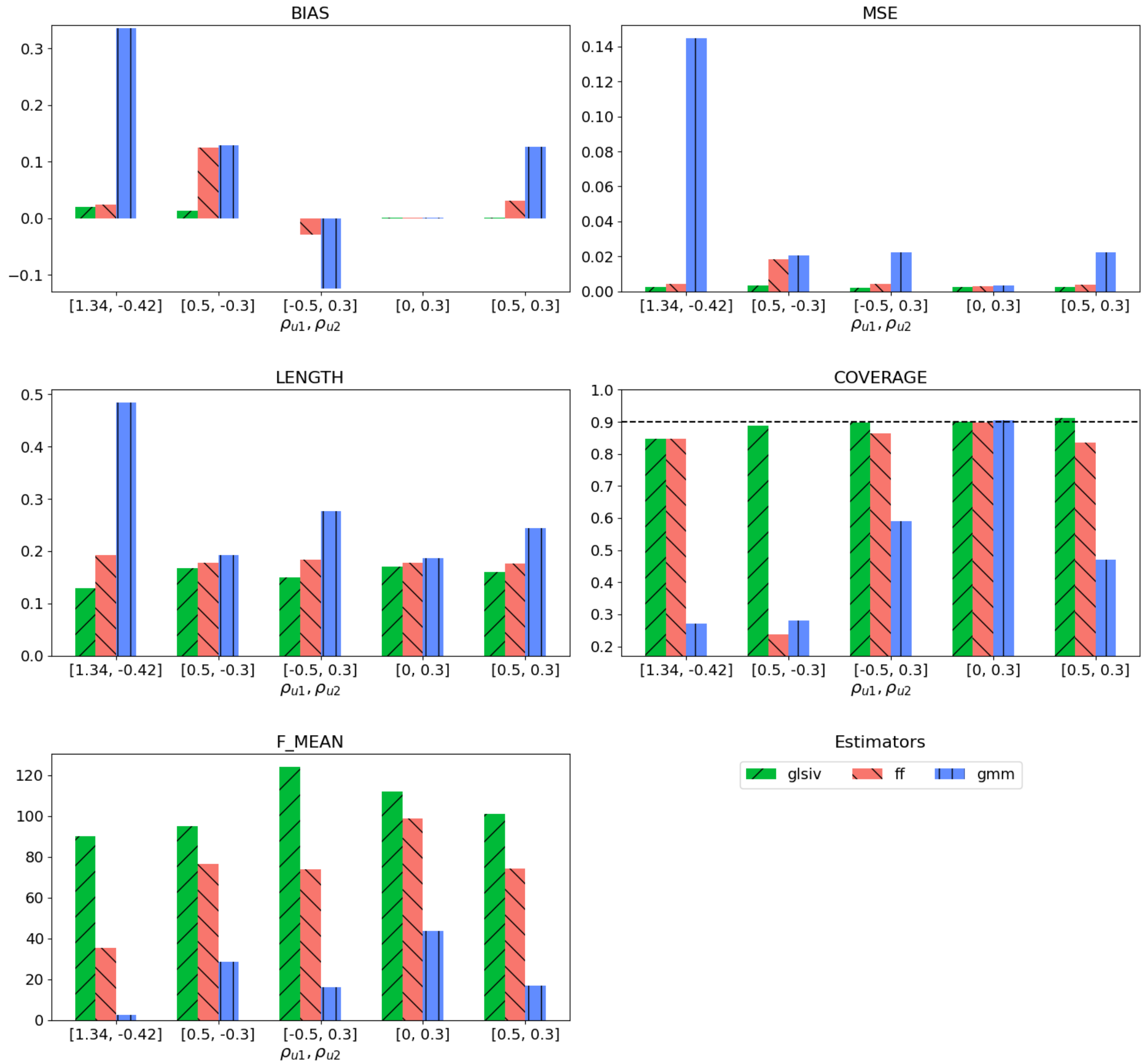
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the moving average parameter are considered: $\{(0.9, 0.3, 0.2), (-0.5, 0.9, -0.2), (0, 0.3, 0.2), (0.4, -0.3, 0.2)\}$.

Figure A–4: Simulations with Ω unknown, $u \sim AR(2)$; exogenous instruments



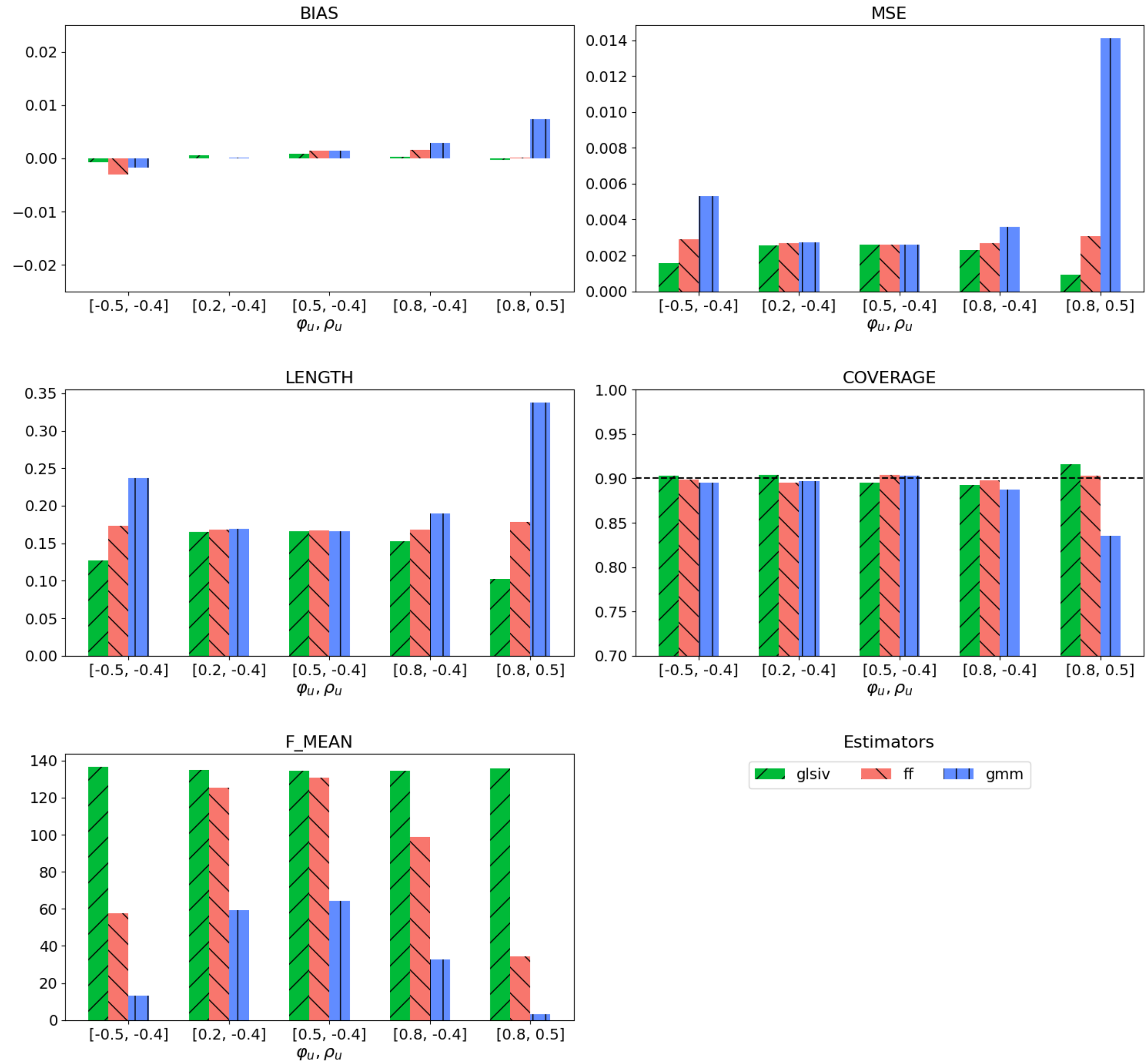
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the autoregressive parameters are considered: $\{(1.34, -0.42), (0.5, -0.3), (-0.5, 0.3), (0, 0.3), (0.5, 0.3)\}$.

Figure A–5: Simulations with Ω unknown, $u \sim AR(2)$; predetermined instruments



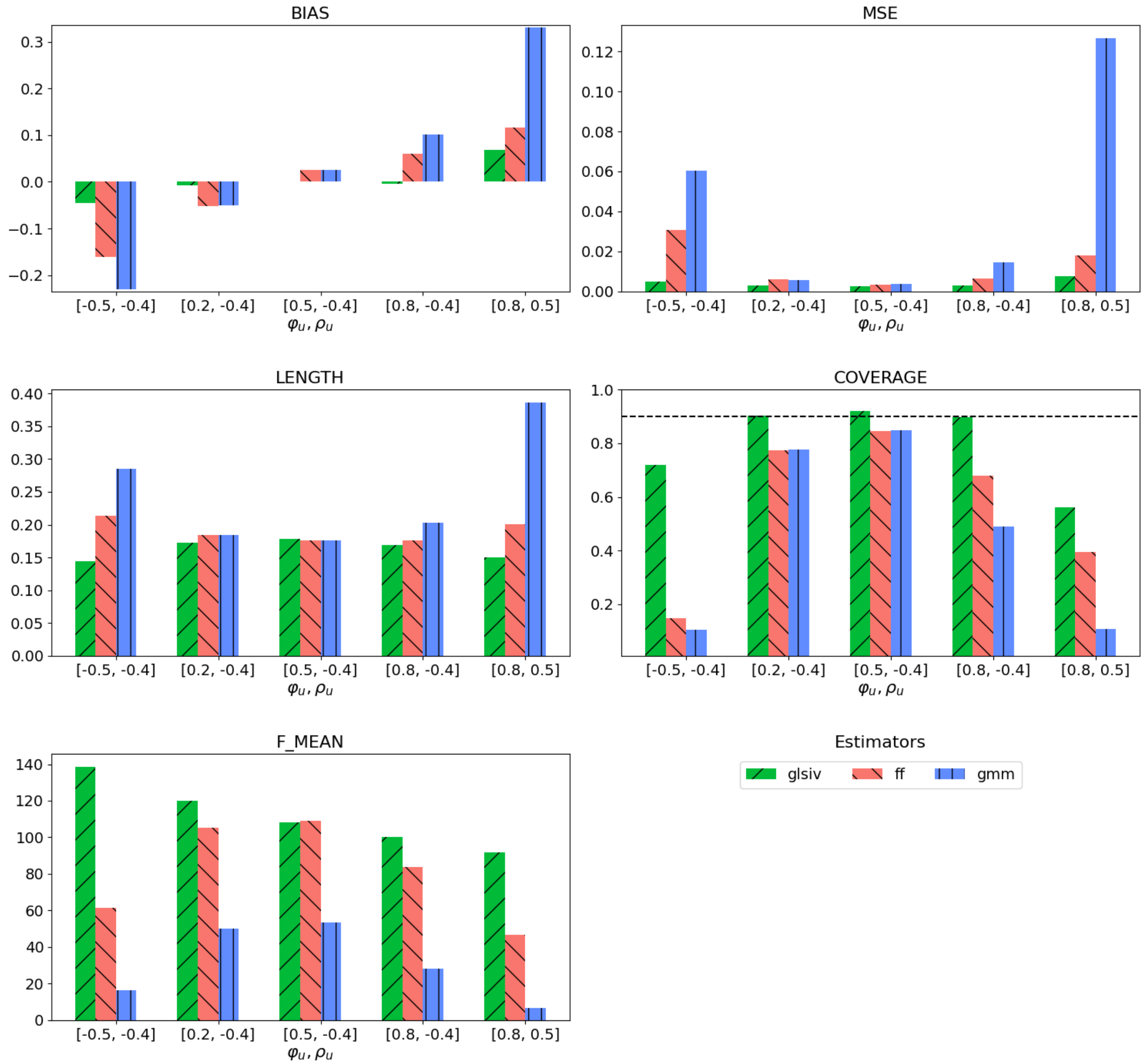
Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the autoregressive parameters are considered: $\{(1.34, -0.42), (0.5, -0.3), (-0.5, 0.3), (0, 0.3), (0.5, 0.3)\}$.

Figure A–6: Simulations with Ω unknown, $u \sim ARMA(1,1)$; exogenous instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the autoregressive and moving average parameters (ρ_u, ϕ_u) are considered: $\{(-0.5, -0.4), (0.2, -0.5), (0.5, -0.4), (0.5, 0.5), (0.8, -0.4), (0.8, 0.5)\}$.

Figure A–7: Simulations with Ω unknown, $u \sim ARMA(1,1)$; predetermined instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 4 different values of the autoregressive and moving average parameters (ρ_u, ϕ_u) are considered: $\{(-0.5,-0.4),(0.2,-0.5),(0.5,-0.4),(0.5,0.5),(0.8,-0.4),(0.8,0.5)\}$.

Figure A–8: Galí, Gertler and López-Salido instrument set; labor share

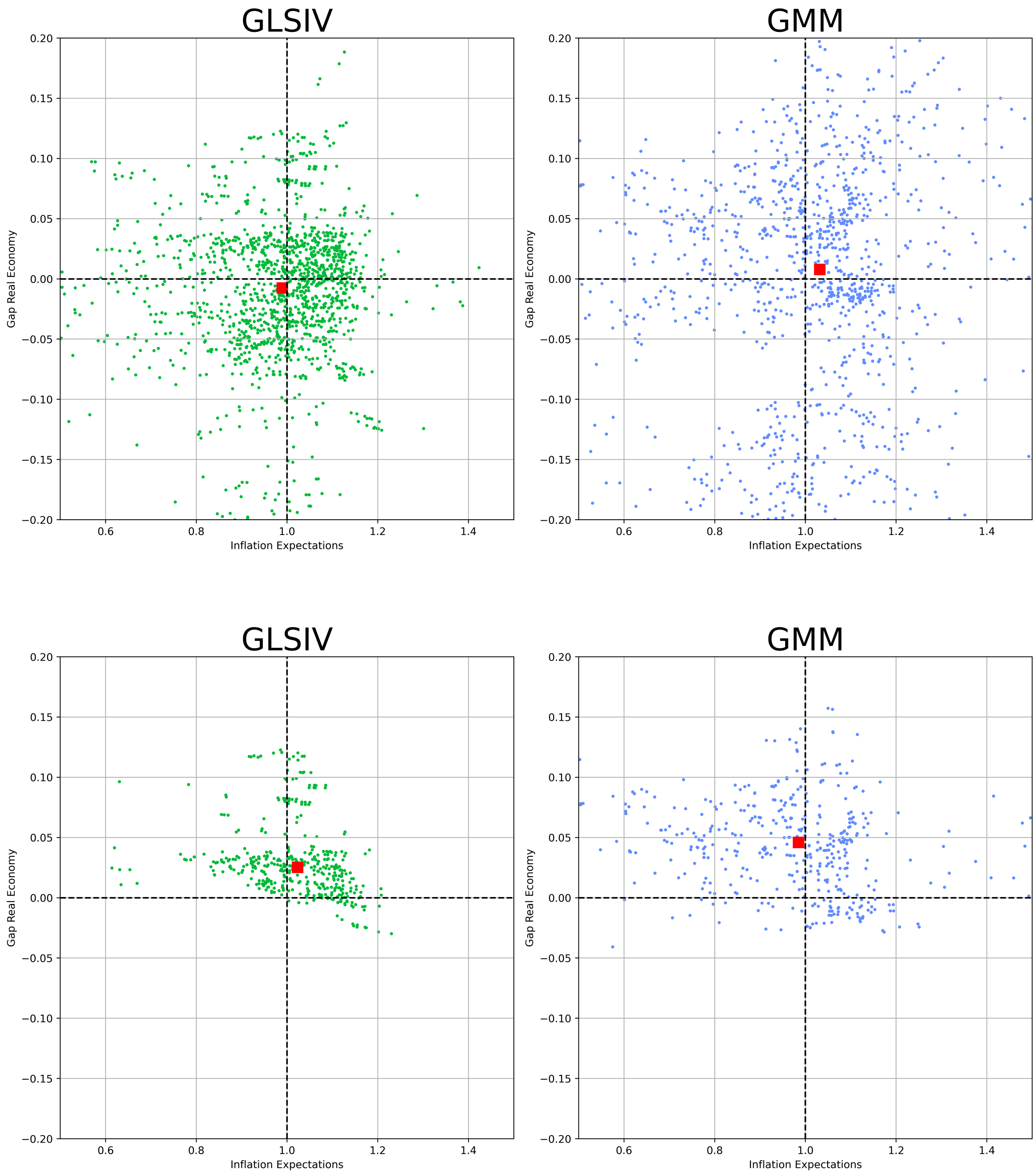


Figure A–9: Galí, Gertler and López-Salido instrument set; output gap

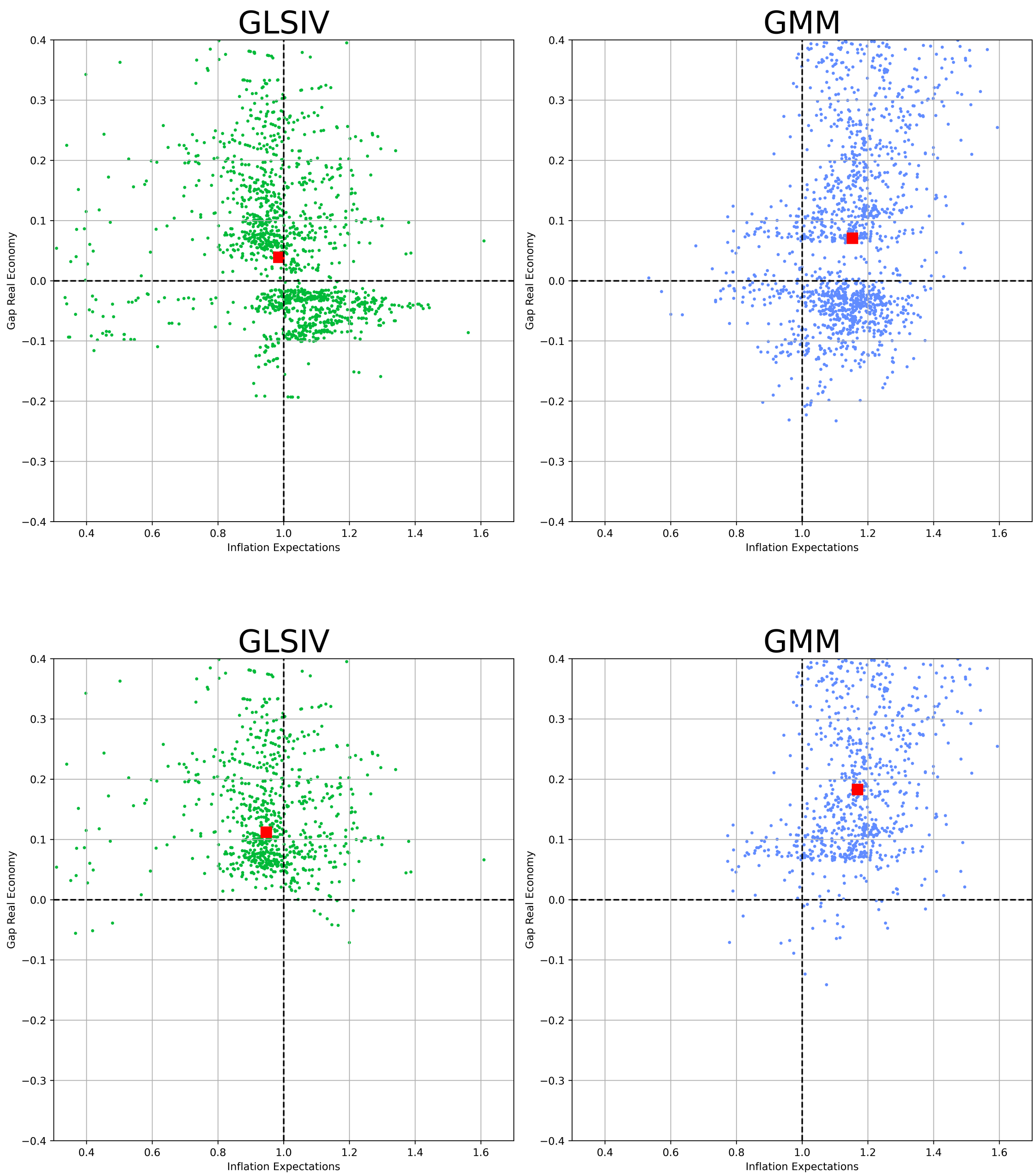
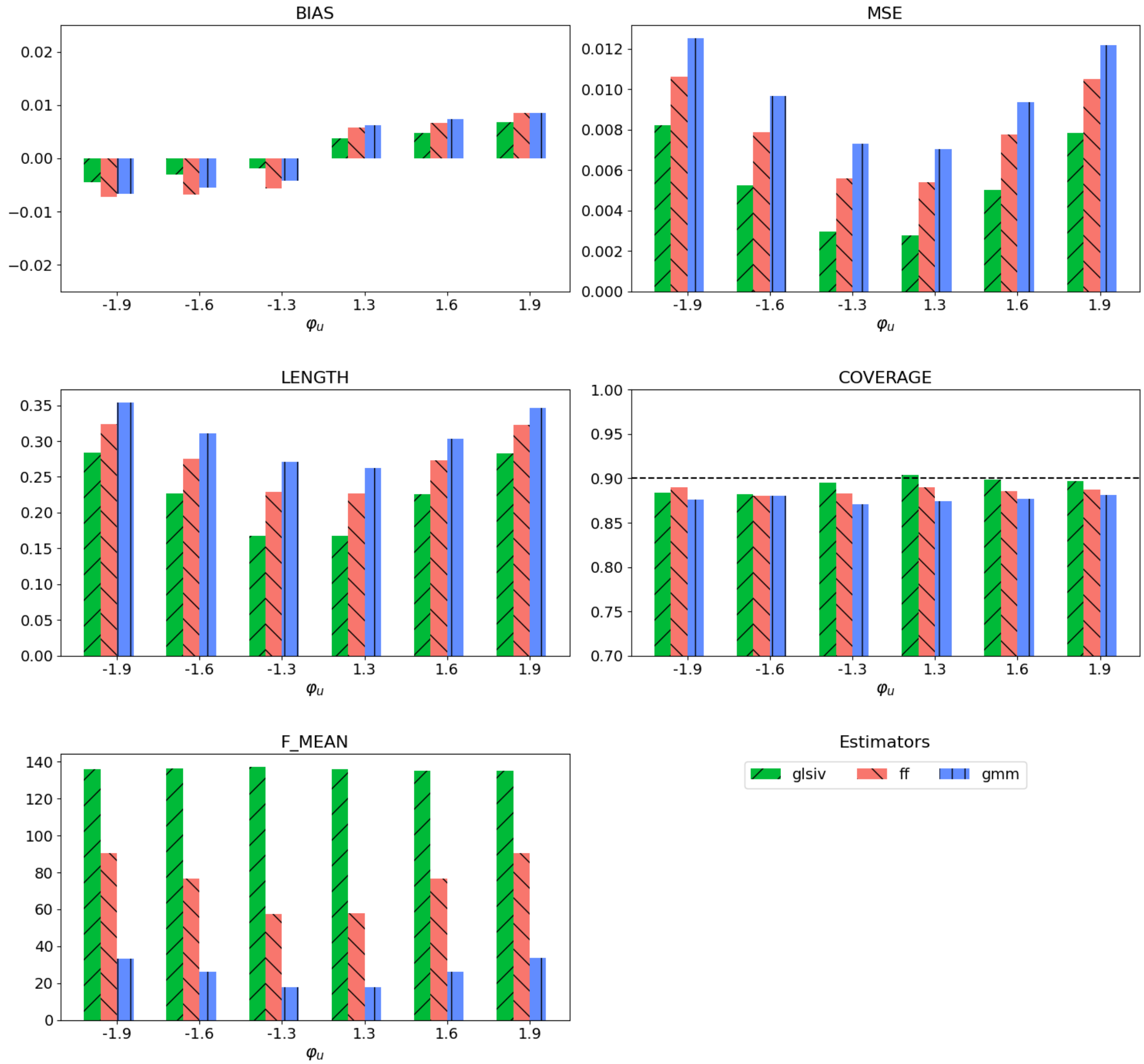
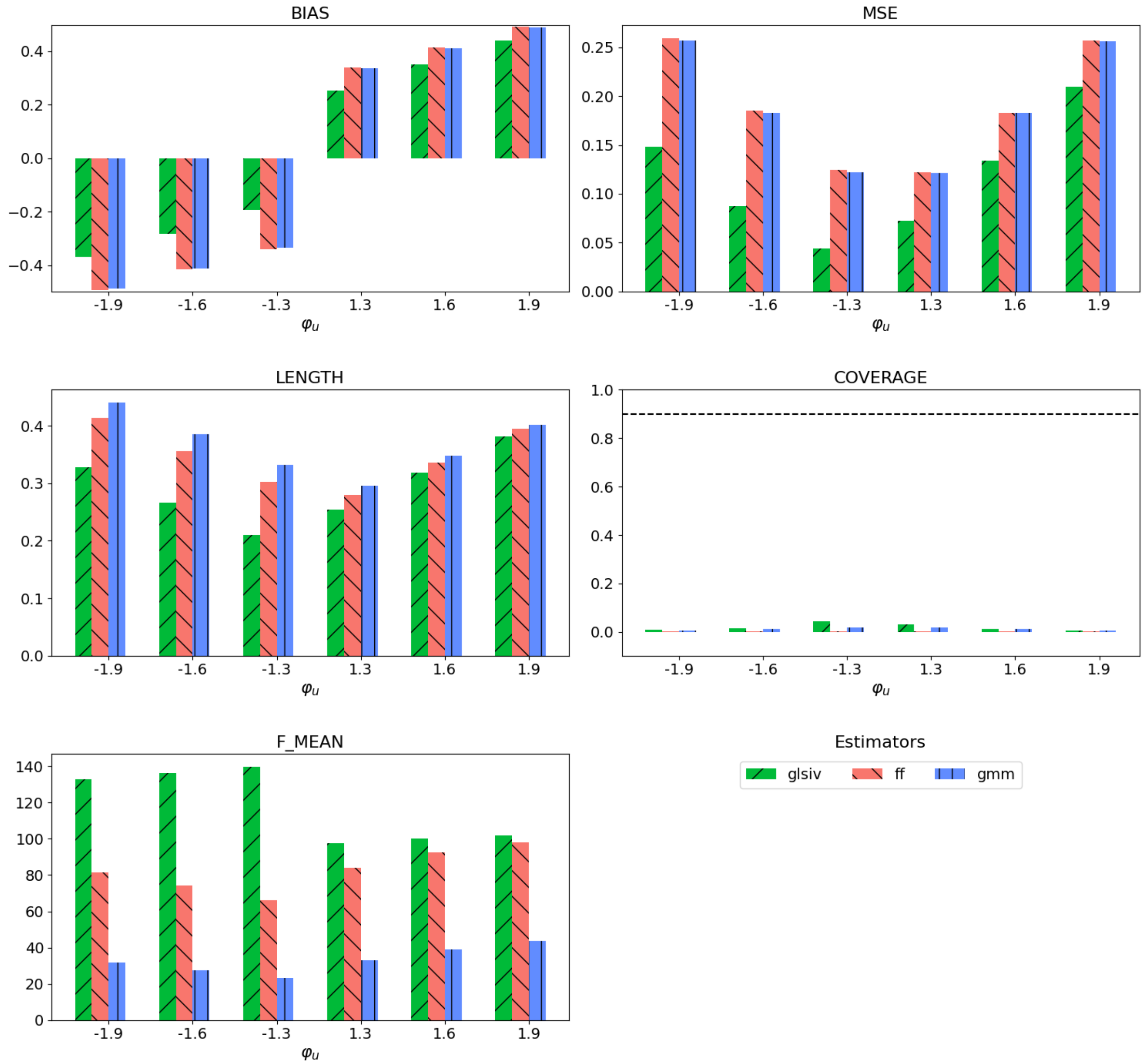


Figure A–10: Simulations with Ω unknown, $u \sim MA(1)$ (non-invertible); exogenous instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the moving average parameter are considered: $\{-1.9, -1.6, -1.3, 10.3, 1.6, 1.9\}$.

Figure A–11: Simulations with Ω unknown, $u \sim MA(1)$ (non-invertible); predetermined instruments



Note: The plots show the value of the (BIAS), mean square error (MSE), the average length of the confidence interval (LENGTH), the exact coverage rate of the confidence interval for a nominal 90% level (COVERAGE), and the mean value of the F statistic of the first stage (F_MEAN). The different colors represent different estimators, namely GLS–IV, FF, and GMM. On the horizontal axis, 6 different values of the moving average parameter are considered: $\{-1.9, -1.6, -1.3, 10.3, 1.6, 1.9\}$.

Table A–1: Summary Statistics for $\hat{\beta}_1$ with Ω unknown, $\alpha = 1$

	Coverage	Bias	MSE	MAE	Std. Error	Length
glsiv	0.01	-0.697	0.633	0.697	0.15	0.49
ff	0.90	0.006	0.003	0.046	0.06	0.19
gmm	0.91	0.002	0.003	0.045	0.06	0.19

Table A–2: Summary Statistics for $\hat{\beta}_1$ with Ω known, $\alpha = 1$

	Coverage	Bias	MSE	MAE	Std. Error	Length
glsiv	0.05	-0.726	0.601	0.735	0.14	0.47
ff	0.89	0.005	0.003	0.045	0.06	0.19
gmm	0.91	0.002	0.003	0.045	0.06	0.19

Table A–3: Summary Statistics for $\hat{\beta}_1$ with Ω known, $\alpha = 0$

	Coverage	Bias	MSE	MAE	Std. Error	Length
glsiv	0.90	0.002	0.004	0.052	0.06	0.21
ff	0.91	0.004	0.004	0.048	0.06	0.19
gmm	0.90	0.003	0.004	0.047	0.06	0.19

Table A–4: Summary Statistics for $\hat{\beta}_1$ with Ω unknown, $\alpha = 0$

	Coverage	Bias	MSE	MAE	Std. Error	Length
glsiv	0.89	-0.042	0.138	0.107	0.13	0.44
ff	0.89	0.005	0.004	0.048	0.06	0.19
gmm	0.90	0.004	0.004	0.048	0.06	0.19