Simulation-based matching inference with applications to DSGE models

Bertille Antoine [*] Lynda Khalaf [†] Maral Kichian [‡] Zhenjiang Lin [§]

July 22, 2019

Extended Abstract

PRELIMINARY

Please do not cite without authors' permission

[*]B. Antoine: Simon Fraser University, *Email address:* Bertille_Antoine@sfu.ca
[†]L. Khalaf: Carleton University, *Email address:* Lynda.Khalaf@carleton.ca
[‡]M. Kichian: University of Ottawa, *Email address:* Maral.Kichian@uOttawa.ca
[§]Z. Lin: University of Nottingham at Ningbo, *Email address:* Zhenjiang.Lin@nottingham.edu.cn

Abstract

Simulation-based matching methods are commonly used to estimate structural parameters through auxiliary statistics that summarize key features of the data and model implications. This paper develops a general procedure for inference with auxiliary statistics neither assuming identification of parameters of interest nor a one-to-one binding function. Specifically, the conditions underlying asymptotic validity of our simulators in conjunction with appropriate bootstraps are characterized beyond the strict and exact calibration of Dridi, Guay and Renault (2007).

Such settings include impulse-response (IR) matching for DSGE models, which we analyse in a laboratory environment and a variant of the DSGE model of Del Negro and Schorfheide (2008). In addition to usual Wald-type statistics that combine structural or reduced form IRs (Christiano, Eichenbaum and Evans (2005), Inoue and Killian (2013), Guerron-Quintana, Inoue and Killian (2017)), we analyze local projections (Jordà (2005), Jordà and Kozicki (2011)) through a factor-analytic measure of distance adapted from Bai (2013) which eschews the need to define a weighting matrix. We study exact and miscalibrated cases. Overall, our simulations illustrate the superiority of our proposed factor-analytic local projection approach to IR matching and associated asymptotic framework by documenting useful testable directions robust-to-calibration.

Keywords:

JEL Classification: .

1 Introduction

Simulation-based matching methods are commonly used to estimate structural parameters through auxiliary statistics that summarize key features of the data and/or model implications. This paper develops a general procedure for inference using auxiliary statistics without assuming identification of parameters of interest nor a one-to-one binding function. Specifically, our procedure relies on matching the chosen vector of auxiliary statistics to its counterpart obtained by simulation. Our simulator refers to a user-chosen (fully) parametric model which is used to simulate data compatible with the observables. It is generally misspecified as it depends on additional parameters that need to be calibrated. However, it is not assumed to coincide with the true model: we rather have in mind a parametric model that is *not too far* from our structural model of interest.

Our contributions are both methodological and practical. On the theoretical side, the conditions underlying the asymptotic validity of our inference procedure - and associated simulators in conjunction with double or triple bootstraps - extend the Partial Indirect Inference (PII hereafter) of Dridi, Guay and Renault (2007) in several important directions: (i) our auxiliary statistic (binding function) is not one-to-one; (ii) our inference procedure is identification-robust; (iii) our inference procedure remains valid beyond a strict (and exact) calibration of the additional parameters of the simulator. This allows us to consider environments where these additional (nuisance) parameters are incorrectly calibrated, so-called *miscalibration*. As such, our approach is also related to the issue of testing under misspecified alternatives as studied by Davidson and MacKinnon (1987), Saikkonen (1989), Bera and Yoon (1993), and more recently by Chernozhukov, Escanciano, Ichimura, Newey and Robins (2018).

On the practical side, we focus on impulse-response (IR) matching for DSGE models. We first consider Wald-type statistics that combine structural or reduced form IRs as done in Christiano, Eichenbaum and Evans (2005), Inoue and Killian (2013), Guerron-Quintana, Inoue and Killian (2017). In addition, we analyze impulse responses obtained by local projections (see Jordà (2005), Jordà and Kozicki (2011)) through a factor-analytic measure of distance adapted from Bai (2013) which has the advantage of circumventing the need for a weighting matrix.
In a laboratory environment, we consider a variant of the small-scale DSGE model of Del Negro and Schorfheide (2008) and study both exact and miscalibrated cases; such cases of miscalibration are particularly relevant for empirical macroeconomic policy analysis. Despite inevitable identification and sensitivity concerns, our results clearly document useful testable directions that remain robust to calibration. Overall, our simulations illustrate the superiority of our proposed factor-analytic local projection approach to IR matching and associated asymptotic framework.

The rest of the paper is organized as follows. Section 2 introduces our general framework and motivates our inference strategy through the structural (multivariate) linear regression model.

In sections 3 and 4, we present our new simulation-based inference procedure: respectively under a convenient simplified framework, and under a more realistic one. In section 5, we detail its practical implementation. In section 6, the relevance of our inference procedure is illustrated in a simulation study of a variant of the DSGE model of Del Negro and Schorheide (2008). Section 7 concludes. Graphs and tables of results as well as the proofs of our theoretical results are gathered in the Appendix.

2 Framework and Motivation

We start by introducing our general framework and motivating our inference strategy through the structural (multivariate) linear regression model.

Let θ_1 be a p_1-vector of structural parameters of interest. We propose a test of $\mathcal{H}_0 : \theta_1 = \theta_{1,0}$ (with $\theta_{1,0}$ known vector of size p_1) with the objective of inverting this test for inference on θ_1 or some known function of θ_1 to be defined below. θ_1 is our parameter of interest in the sense that it is freely varying, and we have in mind two cases of interest: (i) a parametric model where θ_1 is a subvector of the full vector of structural parameters; (ii) a partially parametric model where θ_1 may be defined through a set of identifying restrictions. In any case, our test relies on a (chosen) q-vector of auxiliary statistics denoted $\hat{g}(Y_T)$ where $Y_T = (y_1, \cdots, y_T)$ is the sample of observables of size T; and, more specifically, on matching the above vector of auxiliary statistics to its counterpart obtained by simulation. Our *simulator* refers to a user-chosen (fully) parametric model which is used to simulate data *compatible* with (i) or (ii). It is important to mention that our simulator is not assumed to coincide with the true model: we rather have in mind a parametric model that is not *too far* from our main model. For example, our simulator can be obtained as

- a parametric approximation - such as Quasi-maximum Likelihood - of the semi-parametric model of interest;
- an approximation - such as a linearization - of the main (nonlinear) model;
- as a reduced form model of our main structural model.

Definition 1 below formally introduces our simulator as a fully parametric model that can be simulated - but remains agnostic on the formal links between the chosen simulator and our underlying DGP. In the above-mentioned parametric case (i), our simulator requires p_2 additional parameters θ_2, whereas in the partially parametric case (ii), it may also require additional assumptions on the law of motion of the data. Hereafter, we denote the full vector of p structural parameters $\theta = (\theta_1', \theta_2')'$ (with $p = p_1 + p_2$).

Definition 1 *(Simulator)*
Our simulator is defined as the statistical model $(\tilde{\mathcal{Y}}, P_\theta)$. For a fixed vector $\overline{\theta}$ of size p, we can

simulate an analogue of the observables Y_T. The h-th simulated counterpart of a sample of size T is denoted $\tilde{Y}_{T,h}(\overline{\theta})$.

Given the simulated path $\tilde{Y}_{T,h}(\overline{\theta})$, the associated counterpart of the statistic $\hat{g}(Y_T)$ is defined as $\hat{g}(\tilde{Y}_{T,h}(\overline{\theta}))$ and denoted $\tilde{g}_{T,h}(\overline{\theta})$. Averaging over H simulations yields

$$\overline{\tilde{g}}_{T,H}(\overline{\theta}) \;=\; \frac{1}{H}\sum_{h=1}^{H}\tilde{g}_{T,h}(\overline{\theta})\,. \tag{1}$$

Our testing strategy relies on matching $\hat{g}(Y_T)$ with its simulated counterpart $\overline{\tilde{g}}_{T,H}(\theta)$ through a criterion function $Q(.)$. We consider inference procedures associated with different objective functions $Q(.)$ such as

- Wald or Score-type inference associated with some standardized squared norm of the difference between $\hat{g}(Y_T)$ and $\overline{\tilde{g}}_{T,H}(\theta)$.
- Likelihood Ratio-type inference associated with some ratio of $\hat{g}(Y_T)$ and $\overline{\tilde{g}}_{T,H}(\theta)$.

Several comments are worth pointing out. First, it is important to mention that we could, alternatively, focus on testing directly restrictions on the population analogue of $\hat{g}(Y_T)$. However, this would require the definition of such population analogue, which is not always clear or well-defined in practice[1]. In addition, it would also yield an unrestricted and often non-structural interpretation of the parameter θ which may not be relevant in practice.
Second, the above simulator requires the knowledge of the entire vector θ, and not only the p_1 components that we are interested in testing. Correspondingly, we will distinguish between two null hypotheses: the full null $\mathcal{H}_0^F$ that tests the whole parameter θ,

$$\mathcal{H}_0^F : \theta = \theta_0\,. \tag{2}$$

Inference on the p_1-subvector of interest θ_1 will be obtained in a second stage by using projection techniques. Even though such setting is simple enough to provide intuition on our inference strategy, it is often undesirable and cumbersome in practice. This is why we will also consider a (partial) null hypothesis that solely focuses on θ_1,

$$\mathcal{H}_0 : \theta_1 = \theta_{1,0}\,. \tag{3}$$

Our framework under the partial null will be combined with a calibration of the remaining parameters θ_2. Calibration is used a lot in the macro models that we have in mind.
Third, the above simulator is in general misspecified with respect to (i) or (ii) (see e.g. Dridi, Guay and Renault (2007)). The validity of our inference strategy relies on the *adequacy* between the data-based statistic and its simulation-based counterpart as formalized in Assumption **??**

[1]See for example the DSGE model considered in our simulation study in section 6.

below.

We conclude this section by focusing our attention on the following structural linear model,

$$AY_t = BX_t + U_t \qquad \text{with } t = 1, \cdots, T\,, \tag{4}$$

where the observable dependent variable Y_t is $(n, 1)$, the observable explanatory variable X_t is $(k, 1)$ (and may contains lags of the dependent variables), A and B are matrices of unknown (structural) parameters of size (n, n) and (n, k), and the unknown error term U_t is $(n, 1)$. In such model, the matrix A cannot be estimated, and it is natural and standard to consider instead the following reduced form model,

$$Y_t = \Gamma' X_t + V_t \qquad \text{with } t = 1, \cdots, T\,, \tag{5}$$

where Y_t and X_t are the same observables as above, while Γ is the (new) matrix of unknown parameters of size (k, n) and V_t is the unknown error term of size $(n, 1)$. Our inference procedure requires three main ingredients:

(i) the auxiliary statistic using to summarize the information contained in the observables;

(ii) the simulator used to simulate data analogues;

(iii) the criterion function usd to compared the auxiliary statistic computed using the observables and the simulated data.

We now detail possible choices for each ingredient.

(i) The auxiliary statistic:

The auxiliary statistic is chosen as a convenient way to "summarize" some key features of the available information of the observables, $Z_T = \{(Y_t, X_t), t = 1, \cdots, T\}$: for example, one may focus on the OLS estimator of parameter Γ, that is

$$\hat{\Gamma}_T = (X'X)^{-1}X'Y \equiv \hat{g}_T(Z_T)$$

with Y the (T, n) matrix with t-th row Y_t' and X the (T, k) matrix with t-th row X_t'. Alternatively, other estimators may be considered, as well as other auxiliary statistics that may focus on different features of the observables such as the variance-covariance of the error term, or some impulse-response functions to only name a few.

(ii) The simulator:

The simulator is chosen as a parametric model that can be used to simulate an analogue of Z_T. For example, when X_t does not contain lags of Y_t, one may choose to condition on X, and then for given $\overline{\theta} = \mathtt{vec}(\overline{A}, \overline{B}, \overline{\Sigma}_u)$, we can generate $\tilde{Z}_{T,h}(\overline{\theta}) = \{(\tilde{Y}_{t,h}, X_t), t = 1, \cdots, T\}$ for $h = 1, \cdots, H$,

$$\overline{A}\tilde{Y}_{t,h} = \overline{B}X_t + U_t\,, \qquad U_t \sim \mathcal{N}(0, \overline{\Sigma}_u) \qquad \text{with } t = 1, \cdots, T\,. \tag{6}$$

The auxiliary statistic computed over each simulated path $\tilde{Z}_{T,h}$ is

$$\tilde{\Gamma}_{T,h} = (X'X)^{-1}X'\tilde{Y}_h = \hat{g}_T(\tilde{Z}_{T,h}(\overline{\theta})) \equiv \tilde{g}_{T,h}(\overline{\theta})\,.$$

(iii) The criterion function:
The criterion function $Q(.)$ compares the auxiliary statistic computed over the observed data and the simulated one, respectively $\hat{g}_T(Z_T)$ and $\overline{\tilde{g}}_{T,H}(\overline{\theta})$. For example, we may consider LR-type inference using the Wilks distance,

$$Q_{LR}(\hat{g}_T(Z_T), \overline{\tilde{g}}_{T,H}(\overline{\theta})) = Q_{LR}(\hat{\Gamma}_T, \overline{\tilde{\Gamma}}_{T,H}) = \frac{\det\left[(Y - X\overline{\tilde{\Gamma}}_{T,H})'(Y - X\overline{\tilde{\Gamma}}_{T,H})\right]}{\det\left[(Y - X\hat{\Gamma}_T)'(Y - X\hat{\Gamma}_T)\right]}$$

Alternatively, we may consider Wald-type inference using

$$Q_W(\hat{g}_T(Z_T), \overline{\tilde{g}}_{T,H}(\overline{\theta})) = T\left[\hat{g}_T(Z_T) - \overline{\tilde{g}}_{T,H}(\overline{\theta})\right]' \hat{\Sigma}_T^{-1}\left[\hat{g}_T(Z_T) - \overline{\tilde{g}}_{T,H}(\overline{\theta})\right]$$

where $\hat{\Sigma}_T$ is an estimator of the (asymptotic) variance covariance matrix of $[\hat{g}_T(Z_T) - \overline{\tilde{g}}_{T,H}(\overline{\theta})]$ which can be obtained by bootstrap as explained in the next section.
Many other criterion functions have been proposed in the literature, and we do not provide an exhaustive review of them in this paper.

3 Simulation-based matching inference under $\mathcal{H}_0^F$

`To be completed`

4 Simulation-based matching inference under $\mathcal{H}_0$

`To be completed`

5 Inference procedures

`To be completed`

6 Simulation study

We illustrate the reliability and the applicability of our proposed approach via various simulations performed in the context of the stylized DSGE model used in Fernandez-Villaverde, Rubio-Ramirez and Schorfheide (2016). The model consists of several sectors including households, intermediate and final goods producers, and a monetary authority. A Calvo assumption

is used to introduce nominal rigidity in prices, and firms that at a given time cannot re-optimize their prices adjust these by the steady-state inflation rate.

Denoting the log deviation of a variable w_t from its steady-state by $\hat{w}$, the log-linearized equilibrium conditions of the model for output, X_t, labor share, lsh_t, inflation, π_t and the interest rate, R_t, are given by:

$$\begin{aligned}
\hat{x}_t &= E_{t+1}[\hat{x}_{t+1}] - (\hat{R}_t - E[\hat{\pi}_{t+1}]) + E_t[z_{t+1}], && (7)\\
\hat{lsh}_t &= \hat{x}_t + \phi_t, && (8)\\
\hat{\pi}_t &= \beta E_t[\hat{\pi}_{t+1}] + \frac{(1-\zeta_p\beta)(1-\zeta_p)}{\zeta_p}(\hat{lsh}_t + \lambda_t), && (9)\\
\hat{R}_t &= \frac{1}{\beta}\hat{\pi}_t + \sigma_R\epsilon_{R,t}. && (10)
\end{aligned}$$

In the above equations, β is the stochastic discount rate and the probability with which a given firm is unable to re-optimize its price is given by ζ_p. In addition, four exogenous shocks influence the dynamics of the variables. These include a technology shock, z_t, a price markup shock, λ_t, a shock that affects the preference for leisure, ϕ_t, and a monetary policy shock, ϵ_t. Except for the monetary policy shock, which is assumed to be independently and identically normally distributed, with mean zero and variance 1, the remaining shocks are assumed to follow autoregressive processes. Thus, for each shock $i = z, \lambda, \phi$, the autoregression coefficient is ρ_i and the standard deviation is σ_i.

The unknown structural parameters of the model are collected in the vector

$$\theta = [\beta, \gamma, \lambda, \pi^*, \zeta_p, \rho_\phi, \rho_\lambda, \rho_z, \sigma_\phi, \sigma_\lambda, \sigma_z, \sigma_R]',$$

where γ is the growth rate of technology, λ is the steady-state markup charged by the intermediate goods producers, and π^* is the steady-state inflation rate. The steady-states for the interest rate and for the labor share can be obtained from the expressions $\bar{R} = \pi^*\gamma/\beta$, and, $\bar{lsh} = 1/(1+\lambda)$, respectively.

To simplify the exposition, and without loss of generality, in our simulations we will focus on estimating only one parameter of the model, namely ζ_p. We calibrate the remaining structural parameters to values taken from Fernadez-Villaverde *et al.* (2016) and Del Negro and Schorfheide (2008); see Table 1.

As explained above, our approach allows us to make use of different functions of the model parameters. In this section, we focus on three such functions:

(i) VAR-based reduced-form impulse response functions (RIRF),

(ii) VAR-based Cholesky-orthogonalized structural impulse response functions (SIRF),

(iii) impulse responses based on Jorda's local projection method.

Table 1: Calibrated Parameter Values

Parameter	Value
stochastic discount rate	$\beta_0 = 0.98$
growth rate of technology	$\gamma_0 = 1.005$
steady-state intermediate goods markup	$\lambda_0 = 0.15$
steady-state inflation rate	${\pi_0}^* = 1.005$
ρ_z	$\rho_{z,0} = 0.13$
ρ_λ	$\rho_{\lambda,0} = 0.88$
ρ_ϕ	$\rho_{\phi,0} = 0.30$
σ_z	$\sigma_{z,0} = 1.50$
σ_λ	$\sigma_{\lambda,0} = 0.50$
σ_ϕ	$\sigma_{\phi,0} = 3.00$
σ_R	$\sigma_{R,0} = 1.00$

In each case, we consider the dynamic impacts up to four quarters of a 1-unit increase in the monetary policy shock jointly on the four endogenous variables of our model. To examine the distances between data and population, several methods are available. In the case of the impulse responses, these hinge on the choice of the weighting matrix; we consider population weights (diagonal and non-diagonal), as well as non-parametric non-diagonal bootstrap weights. For the local projection approach of Jorda (2005), we consider statistics proposed for the SURE setting, including the likelihood ratio (LR), a statistic adapted from Bai (2013), as well as the maximum root statistic. In addition, we consider the simple average of LR and Bai's statistic and a scenario where the moving average errors are corrected by recursively including the residuals.

6.1 Simulation results on Changing Nulls

We first present simulation results where we test $\theta = \theta_0$ when, for one parameter of the model, the null is different than its "true" value.

6.1.1 Monte Carlo Test for ρ_ϕ

We focus first on the autoregression parameter of the process that describes the dynamics of the shock to leisure. While the true value is 0.30, we compute the size of the test when testing a null hypothesis for ρ_ϕ when it takes a different value in a grid (0.1 to 0.99). Figure 1 shows

the results for different DGPs for ρ_ϕ equals 0, 0.1, 0.2, ..., up to 0.9 when Bai's statistics is used along with Jorda's local projection in SURE form, whereas Figures 2 and 3 show the corresponding results with SIRF and RIRF matching respectively, using the WALD statistic with the non-diagonal population weight applying the Monte Carlo (MC) test method.

The results indicate that the test has the most power with Bai's statistic when Jorda's method is used, and that this power is present primarily when the null posits a parameter value that is larger than the true one.

6.1.2 The Impact of Miscalibrations

Typically, several DSGE parameters are calibrated, drawing on values available in other studies in the literature. Since several alternatives may be available to calibrate the value of a particular parameter, it is important to examine the impact of using the wrong calibrated value on the properties of the test. We consider, in turn, miscalibration in several parameters of our model.

First, suppose that ρ_λ, the autoregression parameter of the markup-cost shock process, is miscalibrated. We consider two types of miscalibrations:

- Case 1 (high persistence): the true value of ρ_λ is 0.88 and we mistakenly calibrate it to another value;
- Case 2 (low persistence): the true value of ρ_λ is 0.1 and we mistakenly calibrate it to another value.

Figure 4 displays the results for Case 1 using MC method. We can see that the MC method is very sensitive to miscalibrations, but such sensitivity is very asymmetric. If the truth is $\rho_\lambda = 0.88$, the "left-hand-side" mistakes are very costly. For example, the black and red curves show that when ρ_λ is mistakenly calibrated to 0, 0.2, 0.4 or 0.6, then the associated rejection frequencies are equal to 1. As a result, the true parameter value $\rho_\phi = 0.30$ is always rejected! However, the "right-hand-side" mistakes are less costly: e.g. if ρ_λ is mistakenly set to 0.90, then the size of the test is maintained. Finally, the green curve shows that mistake of the unit root type is costly as well.

Next, we consider miscalibration in ρ_ϕ. Figure 5 considers Case 1 with miscalibrated ρ_ϕ with SIRF matching using the WALD statistic (with the non-diagonal population weights). Figure 6 considers Case 1 with miscalibrated ρ_ϕ with RIRF matching using the WALD statistic (with the non-diagonal population weight). We find that both SIRF and RIRF matching with WALD are not sensitive to miscalibration.

Given that both ρ_ϕ and ρ_λ are none-structural, it would be interesting to also study the effect of miscalibration on a structural parameter. We thus focus next on ζ_p, the Calvo parameter. Figure 9 considers Case 1 for ζ_p using the MC method, while Figure 10 considers Case 2 for ζ_p using MC method. We find that the outcomes are quite comparable to those pertaining to the previous two non-structural parameters.

Finally, we consider the impact of miscalibration on the variance of one of our model shocks. Figure 11 shows the rejection frequencies for ζ_p with miscalibration on σ_λ at the 5% significance level for small samples using the MC method. Once again, we observe similar outcomes.

6.1.3 Non-parametric Bootstrap P value for ρ_ϕ

In a second stage of the Monte Carlo Test, we construct simulated statistics based on the non-parametric bootstrap (the number of replications is 999). We resample the data with replacement from the true DGP. The black dashed/solid lines in Figure 8 show severe over-rejections with and without misspecification when the sample size $T = 100$. As the sample size increases to 500, the size stabilizes to 0.05. Again, we find that the non-parametric bootstrap method is not sensitive to miscalibrations.

6.2 Simulation results with Fixed Null and Changing Alternative

We next consider simulation results where the null fixes all the calibrated parameters to their (correct) calibrated values as displayed in Table 1, and where the "true" ζ_p is assumed to be 0.65. We then vary the value of one of the parameters under the alternative hypothesis and study rejection frequencies at the 5% significance level with the above-described methods. The sample size is set to $T = 100$. Figure 12 plots these rejection frequencies for different options for the parameter that is allowed to vary under the alternative. Thus, for instance, the top left-hand-side graph in the plot depicts the rejection frequencies when ζ_p is varied under the alternative while its null value is 0.65 (and where all the other parameters are fixed at their calibrated null values).

All methods are size-corrected (by simulation), implying that the power results are formally comparable. We find that, among the considered auxiliary statistics, the following by far outperforms all others: SURE local projection regression estimates, where the distance between data and population is assessed using the measure adapted from Bai (2013; yellow line in Figure 12). One reason for this success may be that while its basis is multivariate, the method is not a Wald-type measure of distance, and it thus eschews the need to define a weighting matrix.

We find, in particular, that the method has the power to discriminate against Calvo parameter values that are lower than the assumed true value. In other words, the method has the power to shed light on the extent of nominal rigidity that is present in the data which is still the subject of some debate in the literature. Similarly, the method is informative on the parameters driving the technology shock (that is, both on the autoregressive parameter and the variance of the shock).

The above simulation studies demonstrate that Bai's statistic with Jorda's projection method generally outperforms the other considered options. In addition, the method is sensitive to several miscalibrated parameters of the model, although this sensitivity is asymmetric and con-

siderably more present when calibrating at higher values than the truth. Since a considerable amount of dynamics is injected into DSGE models via its 'external' sources, namely the model's shock processes, our simulation studies indicate the importance of properly pinning down the parameter values of these processes for testing θ (which, in our simplified case, consists here on the the Calvo parameter ζ_p). Indeed, if the latter are to be calibrated, they should be fixed at the highest values economically-acceptable so that the test has the most power.

7 Conclusion

To be completed

References

[1] J. Bai, *Fixed-effects dynamic panel models, a factor analytical method*, Econometrica **81** (2013), 285314.

[2] A. Bera and M. Yoon, *Specification testing with locally misspecified alternatives*, Econometric Theory **9** (1993), 649–658.

[3] M. Bergamelli, A. Bianchi, L. Khalaf, and G. Urga, *Combining p-values to test for multiple structural breaks in cointegrated regressions*, Journal of Econometrics **forthcoming** (2019).

[4] V. Chernozhukov, J.C. Escanciano, H. Ichimura, W.K. Newey, and J.M. Robins, *Locally robust semiparametric estimation*, Cemmap working paper, Centre for Microdata Methods and Practice (2018).

[5] L.J. Christiano, M. Eichenbaum, and C.L. Evans, *Nominal rigidities and the dynamic effects of a shock to monetary policy*, Journal of Political Economy **113** (2005), 1–45.

[6] R. Davidson and J. MacKinnon, *Implicit alternatives and the local power of test statistics*, Econometrica **55** (1987), 1305–1329.

[7] R. Dridi, A. Guay, and E. Renault, *Indirect inference and calibration of dynamic stochastic general equilibrium models*, Journal of Econometrics **136** (2007), 397430.

[8] J. Fernandez-Villaverde, J.F. Rubio-Ramirez, and F. Schorfheide, *Solution and estimation methods for DSGE models*, Handbook of Macroeconomics **2** (2016).

[9] P. Guerron-Quintana, A. Inoue, and L. Kilian, *Impulse response matching estimators for DSGE models*, Journal of Econometrics **196** (2017), 144–155.

[10] A. Inoue and L. Kilian, *Inference on impulse response functions in structural VAR models*, Journal of Econometrics **177** (2013), 1–13.

[11] O. Jordà, *Estimation and inference of impulse responses by local projections*, American Economic Review **95** (2005), 161–182.

[12] O. Jordà and S. Kozicki, *Estimation and inference by the method of projection minimum distance: An application to the new keynesian hybrid phillips curve*, International Economic Review **52** (2011), no. 2, 461–487.

[13] M. Del Negro and F. Schorfheide, *Forming priors for dsge models (and how it affects the assessment of nominal rigidities)*, Journal of Monetary Economics **55** (2008), no. 7, 1191–1208.

[14] P. Saikonnen, *Asymptotic relative efficiency of the classical tests under misspecification*, Journal of Econometrics **42** (1989), 351–369.

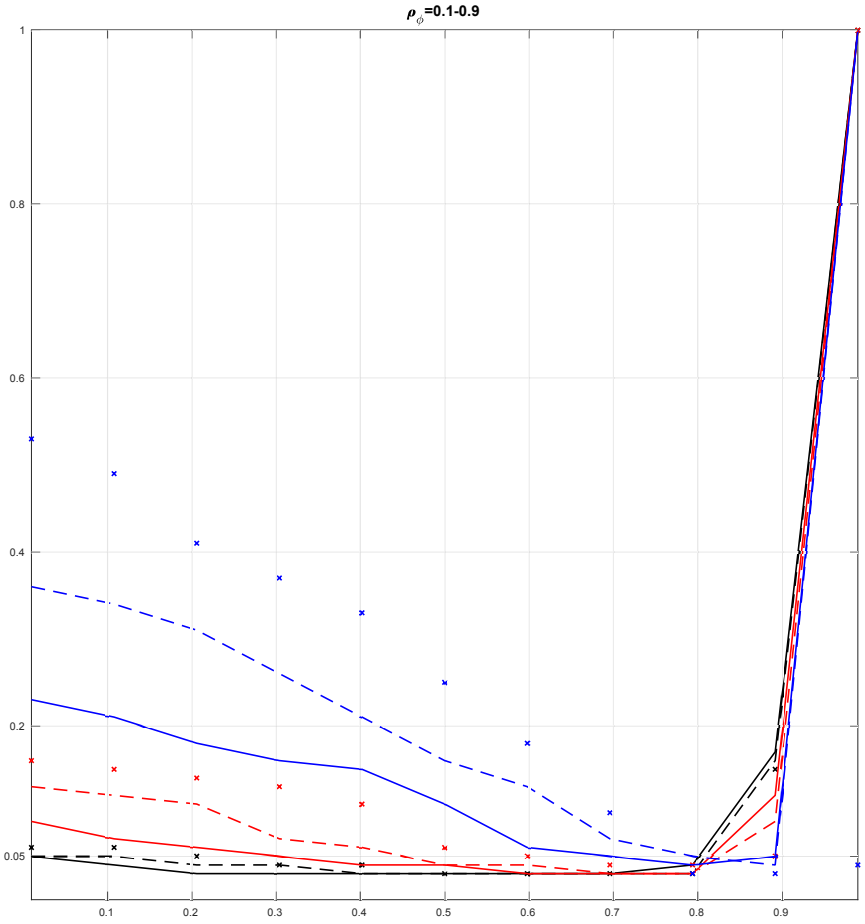

Figure 1: Changing Nulls rejection frequencies for ρ_ϕ at the 5% significance level for small samples with MC method.

Note: Different DGPs for ρ_ϕ.Solid Black: ρ_ϕ=0.1; Dashed Black: ρ_ϕ=0.2; Black Crosses: ρ_ϕ=0.3; Solid Red: ρ_ϕ=0.4; Dashed Red: ρ_ϕ=0.5; Red Crosses: ρ_ϕ=0.6 Solid Blue: ρ_ϕ=0.7; Dashed Blue: ρ_ϕ=0.8; Blue Crosses: ρ_ϕ=0.9. Sample Size: $T = 100$.

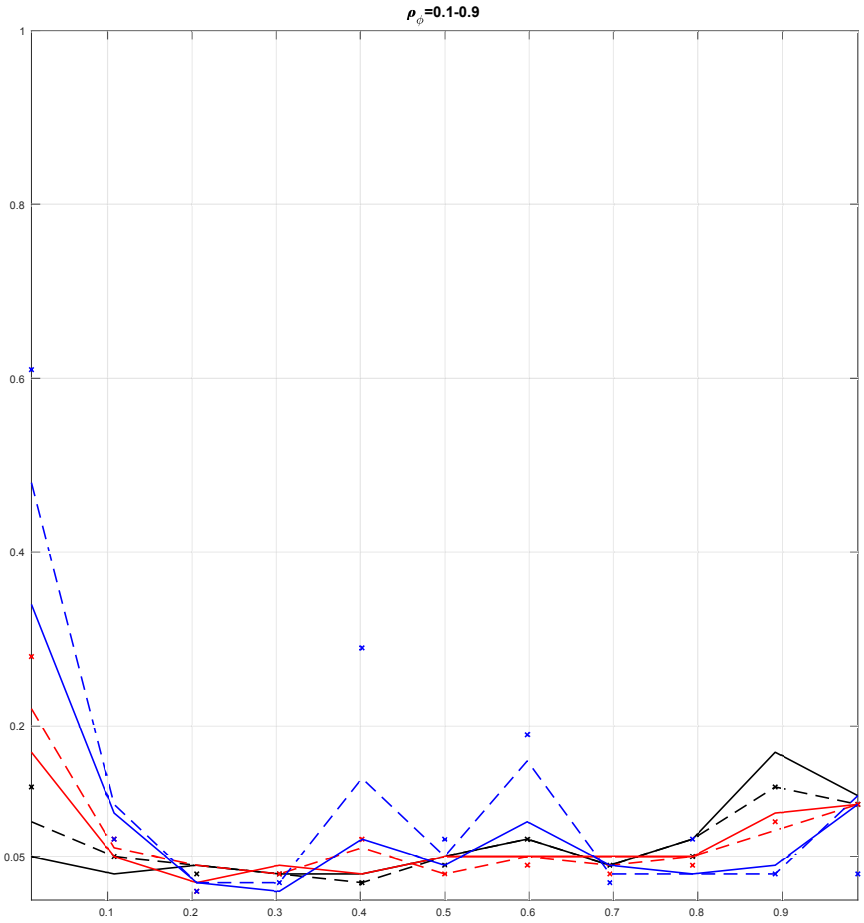

Figure 2: Changing Nulls rejection frequencies for ρ_ϕ at the 5% significance level for small samples with Structural IRF matching using the WALD statistic with the non-diagonal population weight.

Note: Different DGPs for ρ_ϕ.Solid Black: ρ_ϕ=0.1; Dashed Black: ρ_ϕ=0.2; Black Crosses: ρ_ϕ=0.3; Solid Red: ρ_ϕ=0.4; Dashed Red: ρ_ϕ=0.5; Red Crosses: ρ_ϕ=0.6 Solid Blue: ρ_ϕ=0.7; Dashed Blue: ρ_ϕ=0.8; Blue Crosses: ρ_ϕ=0.9. Sample Size: $T = 100$.

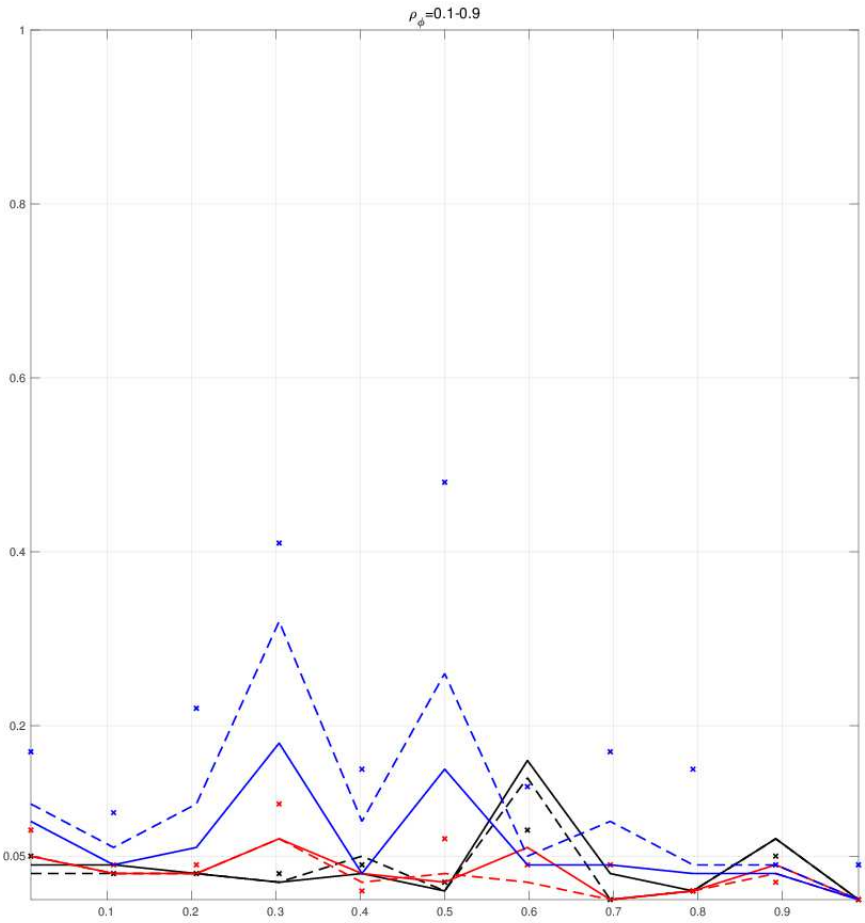

Figure 3: Changing Nulls rejection frequencies for ρ_ϕ at the 5% significance level for small samples with Reduced IRF matching using the WALD statistic with the non-diagonal population weight.

Note: Different DGPs for ρ_ϕ.Solid Black: ρ_ϕ=0.1; Dashed Black: ρ_ϕ=0.2; Black Crosses: ρ_ϕ=0.3; Solid Red: ρ_ϕ=0.4; Dashed Red: ρ_ϕ=0.5; Red Crosses: ρ_ϕ=0.6 Solid Blue: ρ_ϕ=0.7; Dashed Blue: ρ_ϕ=0.8; Blue Crosses: ρ_ϕ=0.9. Sample Size: $T = 100$.

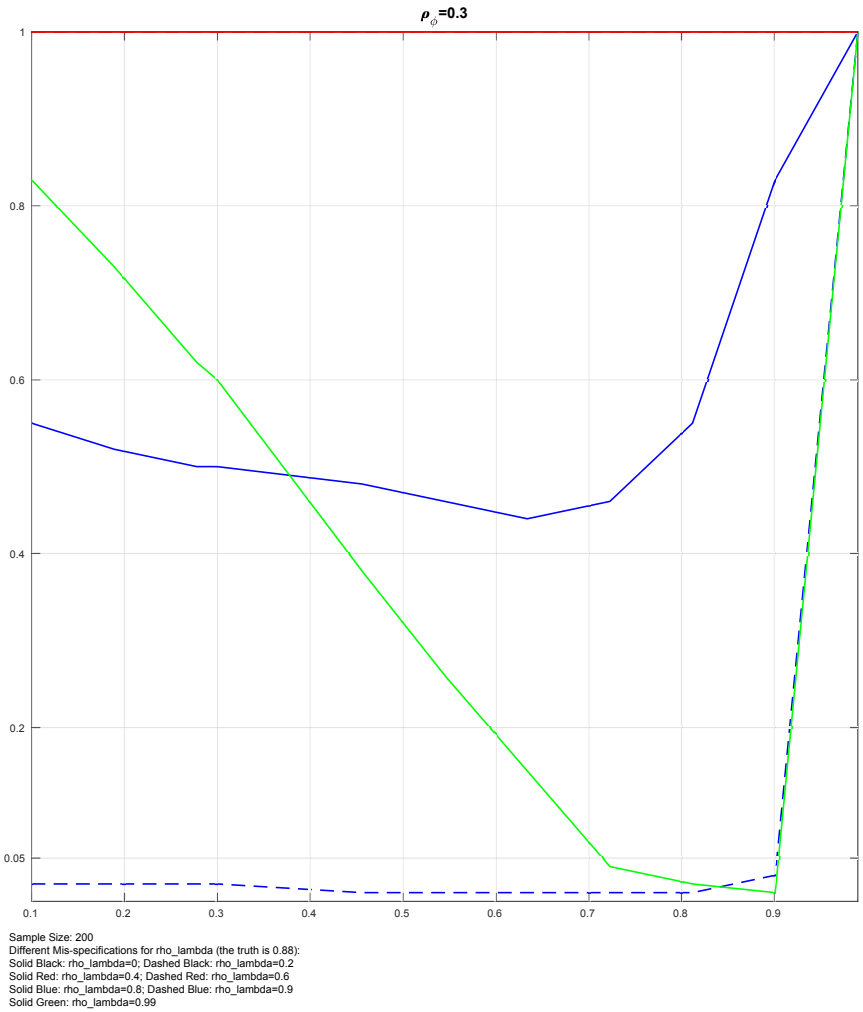

Figure 4: Misspecification Case 1 for ρ_ϕ at the 5% significance level for larger samples with the MC method.

Note: Different Misspecifications for ρ_λ (the truth is 0.88): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.2; Solid Red: ρ_λ=0.4; Dashed Red: ρ_λ=0.6; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.99. Sample Size: $T = 200$.

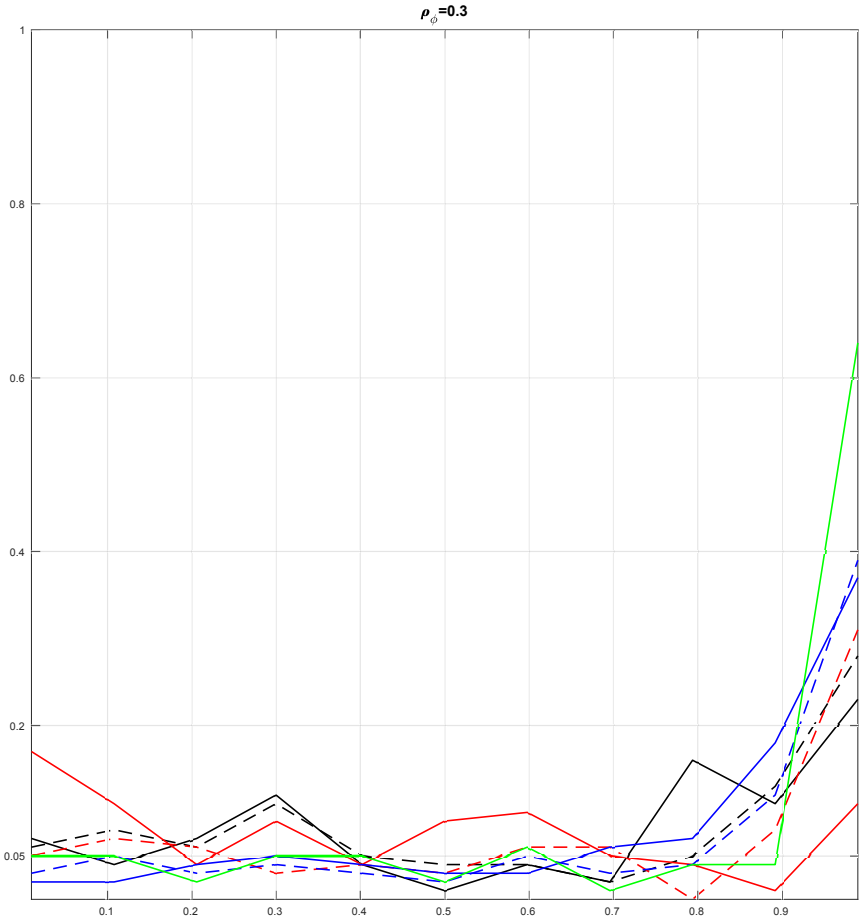

Figure 5: Misspecification Case 1 for ρ_ϕ at the 5% significance level for small samples with the Structural IRF matching using the WALD statistic with the non-diagonal population weight. Note: Different Misspecifications for ρ_λ (the truth is 0.88): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.2; Solid Red: ρ_λ=0.4; Dashed Red: ρ_λ=0.6; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.99. Sample Size: $T = 100$.

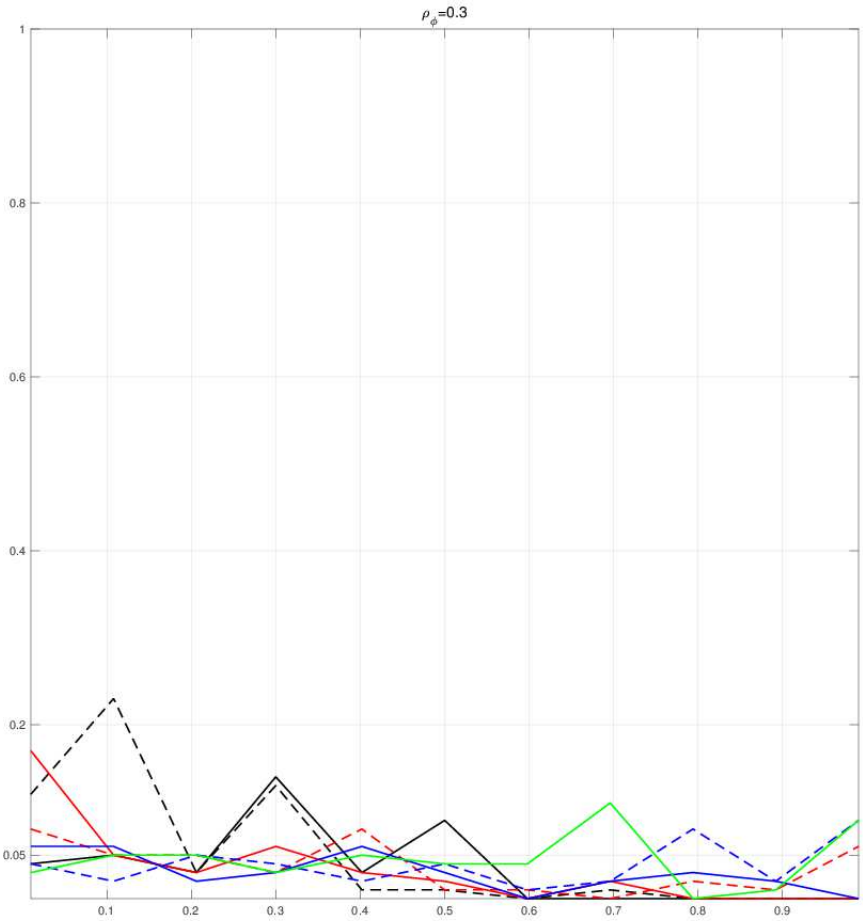

Figure 6: Misspecification Case 1 for ρ_ϕ at the 5% significance level for small samples with the Reduced IRF matching using the WALD statistic with the non-diagonal population weight. Note: Different Misspecifications for ρ_λ (the truth is 0.88): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.2; Solid Red: ρ_λ=0.4; Dashed Red: ρ_λ=0.6; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.99. Sample Size: $T = 100$.

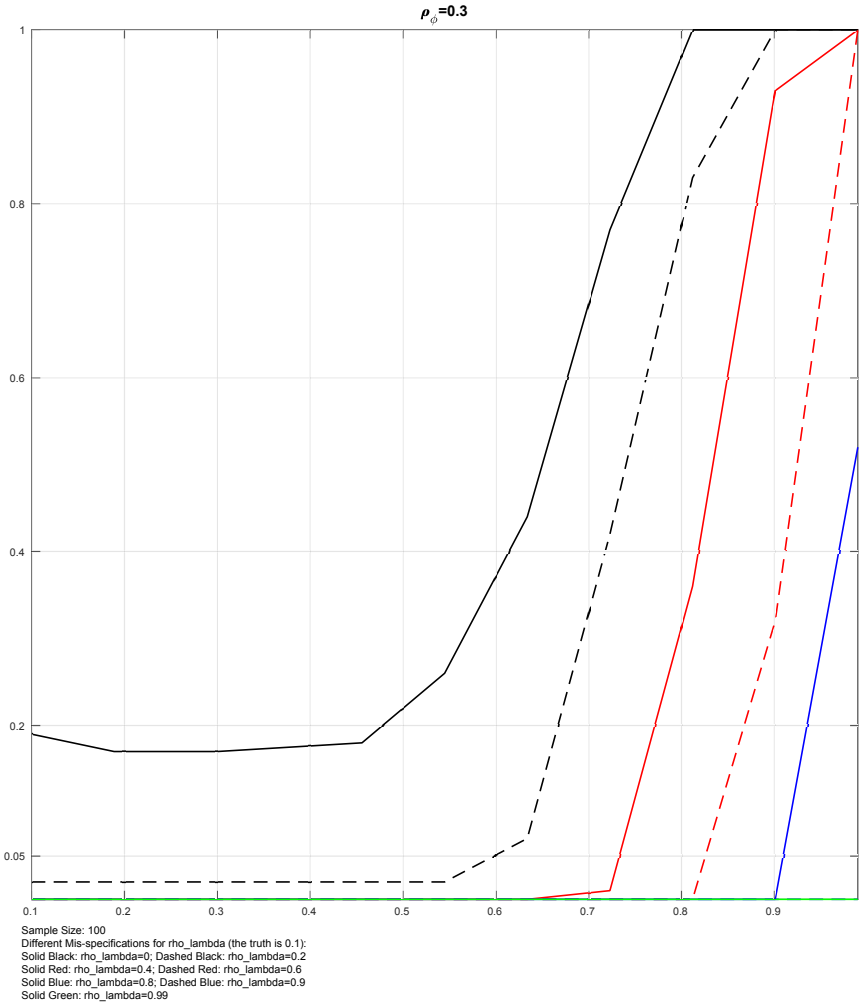

Figure 7: Misspecification Case 2 for ρ_ϕ at the 5% significance level for small samples with the MC method.

Note: Different Misspecifications for ρ_λ (the truth is 0.1): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.2; Solid Red: ρ_λ=0.4; Dashed Red: ρ_λ=0.6; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.99. Sample Size: $T = 100$.

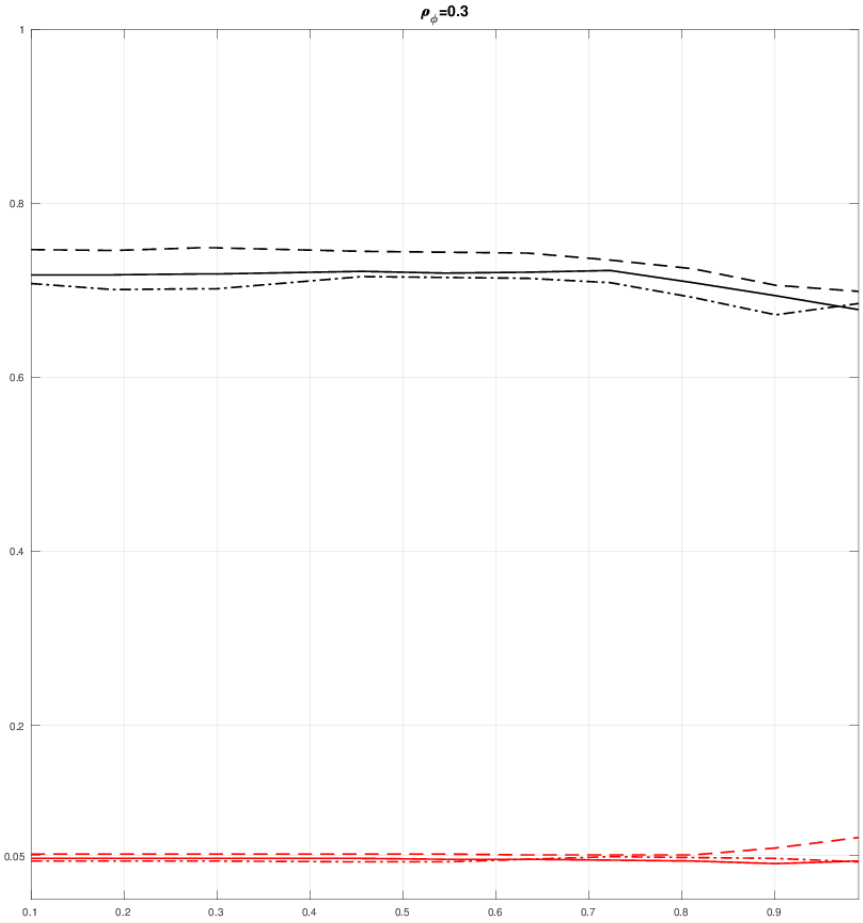

Figure 8: Misspecification Case 1 and non Misspecification combined figures for ρ_ϕ at the 5% significance level with the NP Bootstrap method.
Note: The true value for ρ_λ is 0.88. Solid Black: no misspecification, $T = 100$; Dashed Black: misspecification, ρ_λ is calibrated to 0.75, $T = 100$; Dash-dot Black: misspecification, ρ_λ is calibrated to 0.95, $T = 100$; Solid Red: no misspecification, $T = 500$; Dashed Red: misspecification, ρ_λ is calibrated to 0.75, $T = 500$; Dash-dot Red: misspecification, ρ_λ is calibrated to 0.95, $T = 500$.

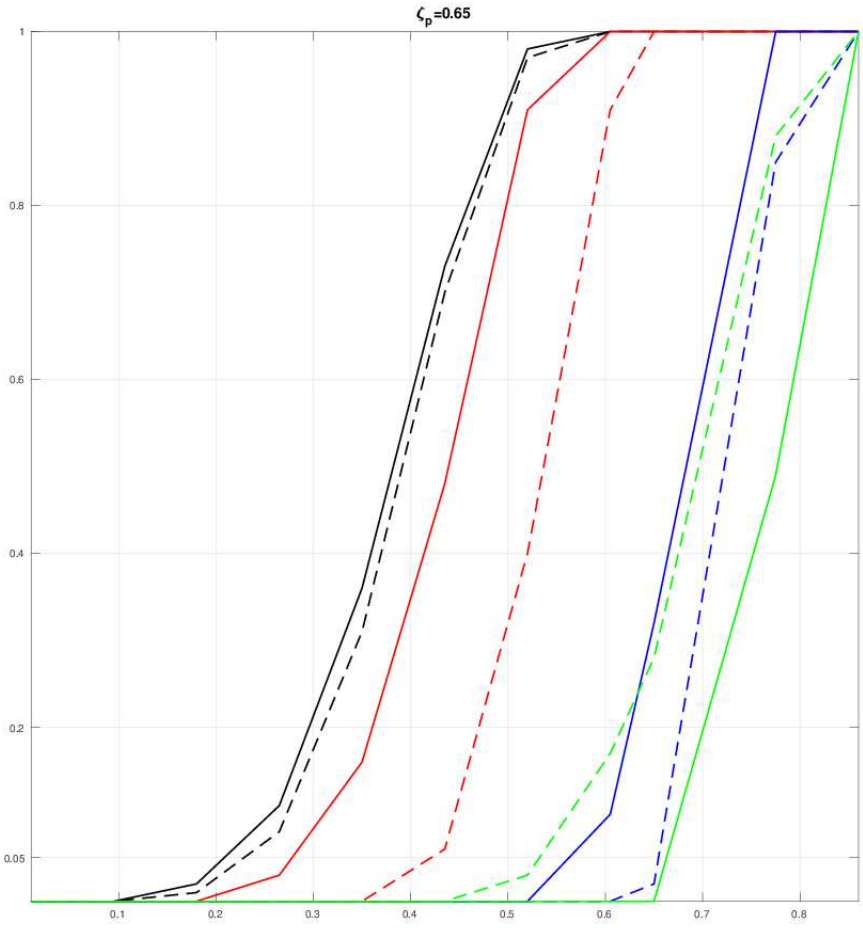

Figure 9: Misspecification Case 1 for ζ_p at the 5% significance level for small samples with the MC method.

Note: Different Misspecifications for ρ_λ (the truth is 0.88): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.05; Solid Red: ρ_λ=0.2; Dashed Red: ρ_λ=0.5; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.95; Dashed Green: ρ_λ=0.99. Sample Size: $T = 100$.

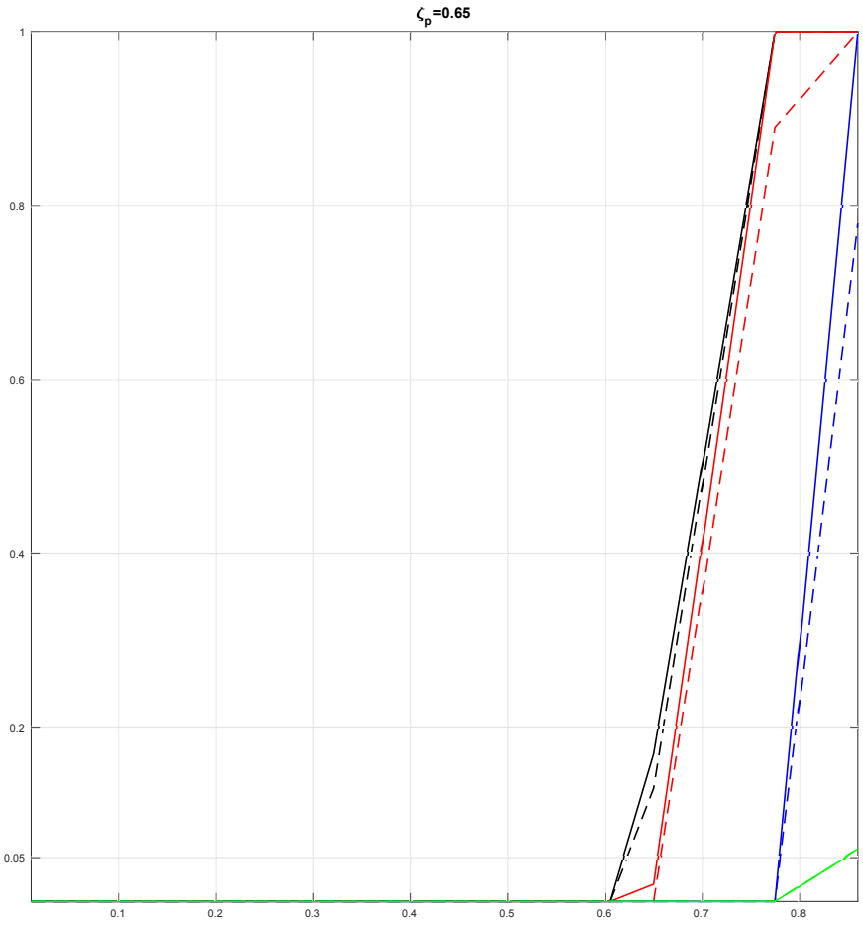

Figure 10: Misspecification Case 2 for ζ_p at the 5% significance level for small samples with the MC method.

Note: Different Misspecifications for ρ_λ (the truth is 0.1): Solid Black: ρ_λ=0; Dashed Black: ρ_λ=0.05; Solid Red: ρ_λ=0.2; Dashed Red: ρ_λ=0.5; Solid Blue: ρ_λ=0.8; Dashed Blue: ρ_λ=0.9; Solid Green: ρ_λ=0.95; Dashed Green: ρ_λ=0.99. Sample Size: $T = 100$.

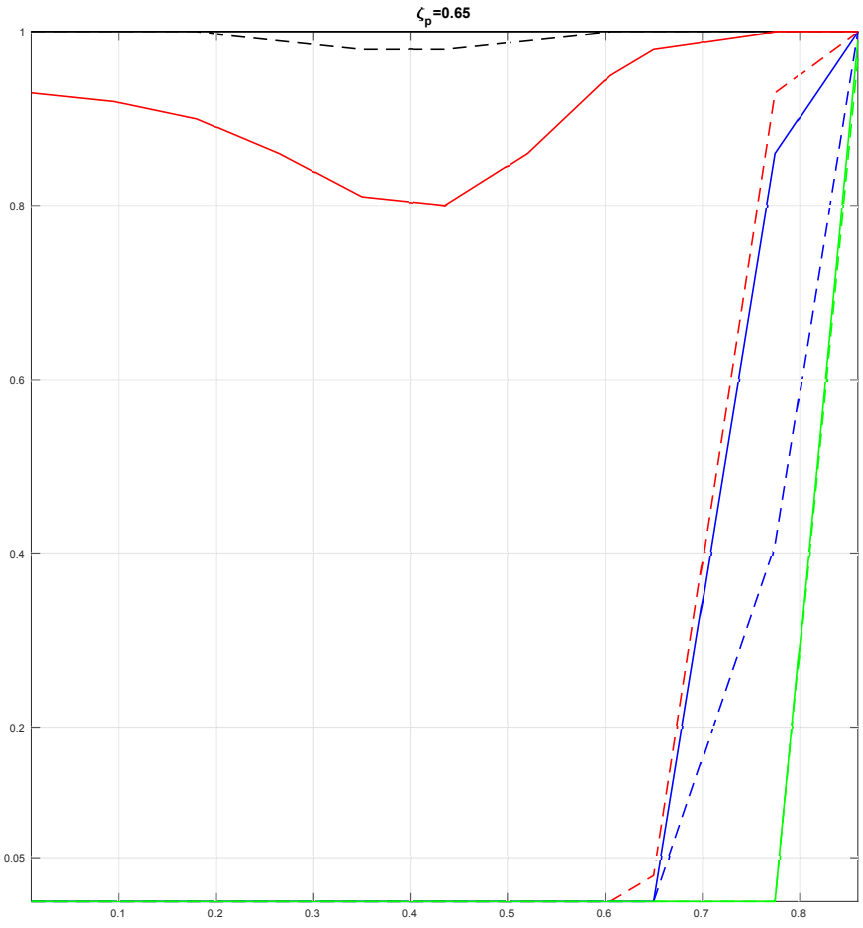

Figure 11: Rejection frequencies for ζ_p with misspecification on σ_λ at the 5% significance level for small samples with the MC method.

Note: Different Misspecifications for σ_λ (the truth is 0.50): Solid Black: σ_λ=0.05; Dashed Black: σ_λ=0.2; Solid Red: σ_λ=0.25; Dashed Red: σ_λ=0.5; Solid Blue: σ_λ=0.55; Dashed Blue: σ_λ=0.7; Solid Green: σ_λ=0.95; Dashed Green: σ_λ=0.99. Sample Size: $T = 100$.

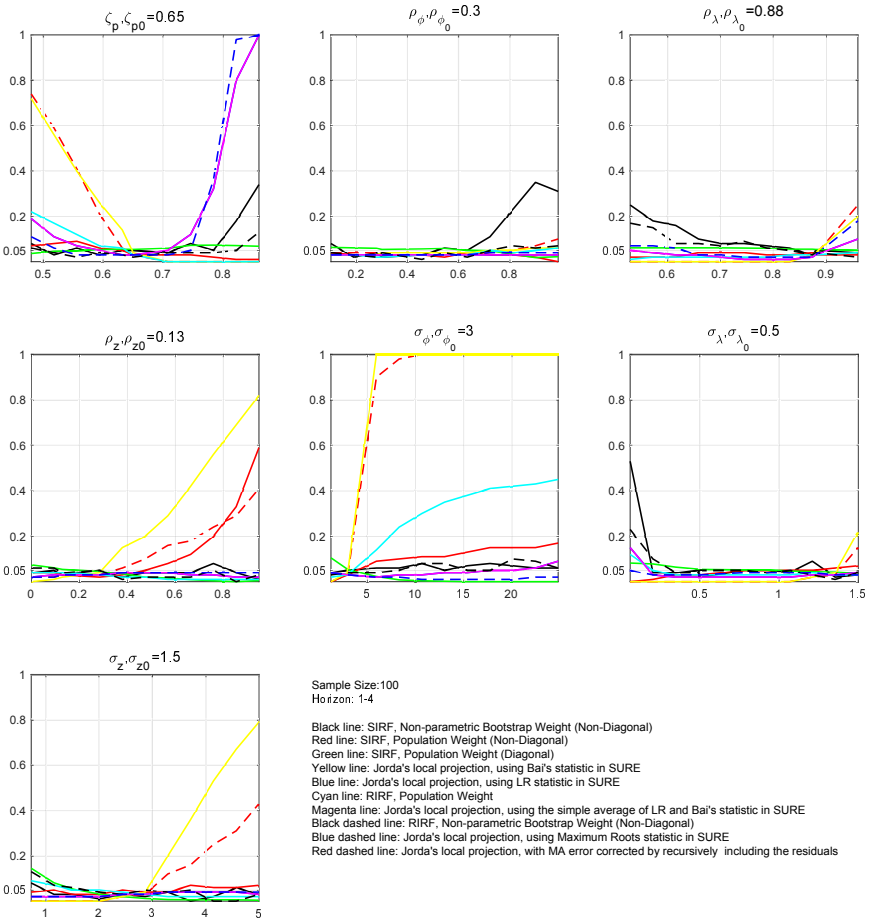

Figure 12: Comparison of rejection frequencies for different methods at the 5% significance level for $T = 100$.

Note: Horizon: 1-4; Black line: SIRF, Non-parametric Bootstrap Weight (Non-Diagonal); Red line: SIRF, Population Weight (Non-Diagonal); Green line: SIRF, Population Weight (Diagonal); Yellow line: Jorda's local projection, using Bai's statistic in SURE; Blue line: Jorda's local projection, using LR statistic in SURE; Cyan line: RIRF, Population Weight; Magenta line: Jorda's local projection, using the simple average of LR and Bai's statistic in SURE; Black dashed line: RIRF, Non-parametric Bootstrap Weight (Non-Diagonal); Blue dashed line: Jorda's local projection, using Maximum Roots statistic in SURE; Red dashed line: Jorda's local projection, with MA error corrected by recursively including the residuals.