A direct route to optimal parametric weighted least squares*

Saraswata Chaudhuri†

This version: August 5, 2020. Last version: August 28, 2019.

Abstract

A classical problem in econometrics is that if the weights are estimated based on a misspecified model for the conditional variance of the regression errors then the parametric weighted least square (WLS) estimator can be less precise than the ordinary LS (OLS) estimator. We argue that this problem arises because the weights are not estimated based on the "appropriate" objective function. Rather, it is based on an objective that, although appropriate when the parametric model is correct, is not necessarily appropriate under parametric misspecification. We solve this problem by proposing the appropriate objective that involves estimation of the unknown weights to directly minimize the asymptotic variance of the parameters of interest. We call this the Modified WLS (MWLS) principle of estimation, and the resulting weighted estimator the MWLS estimator. We show that the MWLS estimator is optimal in a class of estimators that includes OLS and WLS as special cases. Simulations under the design of a similar paper by Romano and Wolf (2017) demonstrate that the MWLS estimator performs much better than or as good as its competitors in terms of standard error, empirical mean squared error, empirical size and power even in small samples. We focus here on linear and nonlinear regression models to introduce the MWLS principle under the simplest setup. Extensions of this core principle to generic conditional moment restrictions models or minimization of the worst case approximate mean squared estimation/forecast error are being developed in followup papers.

JEL Classification: C12; C13; C21.

Keywords: asymptotic optimality; misspecification; nuisance parameters; weighted least squares

*Much older versions of the paper were circulated until May 29, 2019 under a longer title: "Efficient estimation when the misspecification of nuisance parameters does not affect the consistency of the estimator for the parameters of interest: An application to FGLS". I thank J. Galbraith, S. Goncalves, J-M. Dufour, J. MacKinnon, E. Renault, R. Startz, Y. Shin, K. Xu, V. Zinde-Walsh and various conference and seminar participants for their helpful comments.

†Department of Economics, McGill University & Cireq, Montreal. Email: saraswata.chaudhuri@mcgill.ca.

1 Introduction

It has been widely accepted since the seminal work of White (1980) that using the ordinary/nonlinear least squares (OLS/NLS) estimator with robust standard errors in linear/nonlinear regression models is preferable to using the corresponding classical parametric weighted LS (WLS) estimator.[1] This is because the user's parametric model for the unknown conditional variance of the regression error is very likely to be wrong, and then the asymptotic variance of the WLS estimator can actually be bigger than that of its OLS/NLS counterpart. Consequently, the use of WLS is rare nowadays.[2]

In a recent paper, "Resurrecting weighted least squares", Romano and Wolf (2017) challenge this practice (also see Leamer (2010)). Working with a linear regression model, Romano and Wolf (2017) propose an adaptive LS (ALS) estimator that is the OLS estimator if a pre-test cannot reject the conditional homoskedasticity of the regression error, but is the WLS estimator otherwise. They demonstrate that the ALS estimator can provide improvements over the OLS and WLS estimators. However, the asymptotic variance of the ALS estimator can still exceed that of the OLS estimator when the regression error is conditionally heteroskedastic and this conditional variance is parametrically misspecified. This is because the ALS estimator is not designed for any kind of asymptotic optimality in such cases. (There are followup proposals to address this issue for scalar parameters in DiCiccio et al. (2019) that we will also discuss and improve upon in our paper.)

In our paper, we wish to further advance this dissenting view of Romano and Wolf (2017) and others on the neglect of WLS by proposing a Modified parametric WLS (MWLS) estimator that takes a direct route to optimality and addresses, head-on, the well-known issues of non-optimality.

To make it clear at the outset what we mean by "optimal" and to describe the basic idea, it is useful to first focus on a linear regression model of an outcome variable y on p regressors X, i.e.,

$$y = X'\beta + u \quad \text{where} \quad E[u|X] = 0 \quad \text{almost surely in } X. \tag{1}$$

$\omega_0^2(X) := E[u^2|X]$ is an unknown positive function of X. Let $\omega^2(X;\gamma)$ be a positive function of X and γ where $\gamma \in \Gamma \subset \mathbb{R}^k$. $\omega^2(X;\gamma)$ is the user's attempt to parametrically model the unknown function $\omega_0^2(X)$. We maintain that the user's model $\omega^2(X;\gamma)$ may or may not be correct for $\omega_0^2(X)$.

[1] OLS and WLS may not identify the same parameters unless the regression model for the outcome variable y on the regressors X is correct for $E[y|X]$. Interest often lies in $E[y|X]$ and its partial derivatives rather than in the linear projection of y on X. Hence, being true to the classical parametric regression framework, in this paper we will maintain that the linear/nonlinear regression model for the outcome variable is correct for $E[y|X]$, and the competing estimators such as OLS/NLS, WLS, etc. for the parameters in that regression model have the same probability limit.

[2] To quote Angrist and Pischke (2010): "A legacy of White's (1980) paper on robust standard errors, one of the most highly cited from the period, is the near-death of generalized least squares in cross-sectional applied work."

Let $(y_i, X_i')_{i=1}^n$ be i.i.d. copies of (y, X'). We take the user's model $\omega^2(X;\gamma)$ as given and hence the dimension k of γ as fixed. We define the weighted-by-$\omega^{-1}(X;\gamma)$ LS estimator of β as a function of $\gamma \in \Gamma$ as:

$$\widehat{\beta}_n(\gamma) := \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega^2(X_i;\gamma)}X_iX_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega^2(X_i;\gamma)}X_iy_i\right).$$

To fix ideas, take any $j = 1, \ldots, p$ and let the j-th element β_j of β be the parameter of interest. The weighted-by-$\omega^{-1}(X;\gamma)$ LS estimator of β_j is the j-th element $\widehat{\beta}_{j,n}(\gamma)$ of $\widehat{\beta}_n(\gamma)$. Under standard conditions, its asymptotic variance for each $\gamma \in \Gamma$ is the (j,j)-th element $\Xi_{j,j}(\gamma)$ of the matrix:

$$\Xi(\gamma) := \left(E\left[\frac{1}{\omega^2(X;\gamma)}XX'\right]\right)^{-1} E\left[\frac{\omega_0^2(X)}{(\omega^2(X;\gamma))^2}XX'\right]\left(E\left[\frac{1}{\omega^2(X;\gamma)}XX'\right]\right)^{-1}.$$

To put $\widehat{\beta}_{j,n}(\gamma)$ and $\Xi_{j,j}(\gamma)$ in context, note that the WLS estimator of β_j plugs in for γ in $\widehat{\beta}_{j,n}(\gamma)$ some estimator $\widehat{\gamma}_n^{\text{wls}}$ that converges in probability to $\gamma^{\text{wls}} := \arg\min_{\gamma\in\Gamma} E\left[\left(u^2 - \omega^2(X;\gamma)\right)^2\right]$. It is well known that under standard assumptions both estimators $\widehat{\beta}_{j,n}(\widehat{\gamma}_n^{\text{wls}})$ and $\widehat{\beta}_{j,n}(\gamma^{\text{wls}})$ are asymptotically normal. Importantly, by virtue of the mean independence condition $E[u|X] = 0$ in (1), both estimators are asymptotically unbiased for β_j and have the same asymptotic variance $\Xi_{j,j}(\gamma^{\text{wls}})$.

To consolidate notation, we will henceforth denote the asymptotic variance of the weighted-by-$\omega^{-1}(X;\gamma)$ LS estimator of the parameter of interest by $\Sigma(\gamma)$. $\Sigma(\gamma) := \Xi_{j,j}(\gamma)$ in this example.

In our paper we consider "optimality" in terms of minimizing $\Sigma(\gamma)$ with respect to $\gamma \in \Gamma$. In this example, where β_j is the parameter of interest, this involves considering the class of estimators:[3]

$$\mathcal{C}(\beta_j, \omega^2(X;\gamma)) := \left\{ \widehat{\beta}_{j,n}(\widehat{\gamma}_n) \;\middle|\; \begin{array}{c} \text{where } \widehat{\gamma}_n \text{ is a possibly random sequence in } \Gamma \text{ such that} \\ \sqrt{n}(\widehat{\beta}_{j,n}(\widehat{\gamma}_n) - \beta_j) \xrightarrow{d} N\left(0, \Sigma(\gamma) := \Xi_{j,j}(\gamma)\right) \text{ for some } \gamma \in \Gamma \end{array} \right\}, \tag{2}$$

and then regarding any estimator $\widehat{\beta}_{j,n}(\widehat{\gamma}_n)$ in this class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ as "optimal" if its asymptotic variance equals:

$$\Sigma^* := \min_{\gamma\in\Gamma} \Sigma(\gamma). \tag{3}$$

(Σ^* exists by the extreme value theorem if $\Sigma(\gamma)$ is continuous in $\gamma \in \Gamma$ and Γ is compact in $\mathbb{R}^k$.)

The MWLS principle in our paper seeks the optimality characterized by (2) and (3). MWLS targets Σ^* and regards any value $\gamma^* \in \Gamma$ of the nuisance parameters γ for the user's model $\omega^2(X;\gamma)$

[3] $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ contains the OLS and WLS estimators; see Section 2.1. However, since we take the user's model $\omega^2(X;\gamma)$ as given and hence the dimension k of γ as fixed irrespective of n, the class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ rules out semiparametric WLS estimators as in Carroll (1982), Robinson (1987) and Newey (1994). The "implied model" $\omega^2(X;\gamma)$ in semiparametric WLS changes with sample size (e.g., the dimension k of γ increases with n as in Newey (1994)). This means that what follows in the sequel will be most useful when semiparametric WLS is not practical.

as optimal if $\Sigma(\gamma^*) = \Sigma^*$. The set of such optimal γ's may be non-singleton. However, since this set forms a level set for $\Sigma(\gamma)$ ($\Sigma(\gamma) = \Sigma^*$ as in (3)), a non-singleton set is not a problem if one obtains a point estimator $\widehat{\gamma}_n$ that converges in probability to this set sufficiently fast. We obtain such a $\widehat{\gamma}_n$ as the minimizer of some heteroskedasticity consistent (HC) sample analog of the unknown $\Sigma(\gamma)$. Plugging in this $\widehat{\gamma}_n$ in $\widehat{\beta}_n(\gamma)$ defined in (1) gives the MWLS estimator for the parameter of interest β_j as $\widehat{\beta}_{j,n}(\widehat{\gamma}_n)$. We present this same basic idea of MWLS under a general framework in the sequel.

MWLS does not search for parametric models $\omega^2(X;\gamma)$ to obtain a significant result. Rather, MWLS takes the user's/expert's parametric model $\omega^2(X;\gamma)$ as given. And, then, MWLS searches for an optimal γ-value for this given model. The MWLS estimate of the parameter of interest (e.g., $\widehat{\beta}_{j,n}(\widehat{\gamma}_n)$ above) thus obtained may or may not be significant, but is nevertheless the most precise one asymptotically that one could obtain given the user's choice of the parametric model $\omega^2(X;\gamma)$.

It should be noted that the general idea of targeting nuisance parameters to achieve some form of optimality of the estimator for the parameter of interest is very old. This idea has been practiced at least since the early days of the optimal design of experiments and sample surveys where the nuisance parameters are, respectively, the experiment-design and the sampling-design. Some of the directly relevant early works are cited in Section 3.4.[4] In fact, this idea should readily apply to any estimation framework where a possible parametric misspecification of the nuisance parameters does not affect the consistency of the estimator for the parameters of interest; see, e.g., Theorem 6.2 of Newey and McFadden (1994) for sufficient conditions.[5] Such conditions, known as orthogonality conditions, also ensure that the estimation of the unknown nuisance parameters γ does not affect the asymptotic variance of the estimator for the parameters of interest, as is reflected, e.g., in the definition of $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ in (2). These sufficient conditions hold in our paper under standard assumptions by virtue of the mean independence $E[u|X] = 0$ of the regression error u in (1).

We will now wrap up the Introduction with a road-map for the rest of our paper.

Section 2 discusses the intuition behind the optimality of MWLS by working under the same framework of the Introduction. We explain the precise mechanism by which this optimality works here and by virtue of which the MWLS estimator is always at least as precise asymptotically as the OLS, WLS, and other parametric estimators proposed recently in published and working papers.

[4] Numerous interesting papers in biostatistics, economics and statistics also studied similar optimality problems in contexts such as optimal design of experiments using large dimensional data, robust estimation using coarsened data, etc. The general idea is also related to that of finding the optimal estimating equations; see, e.g., Heyde (1997).

[5] More involved extension of this idea to cases where Theorem 6.2 of Newey and McFadden (1994) does not apply, as is, requires targeting certain local approximation of the worst-case mean squared error (e.g., Bonhomme and Weidner (2019)) of estimation or forecast. This is a topic of our multiple ongoing projects with various co-authors.

Then, Section 3 formally introduces the MWLS estimator under a general framework and establishes the first-order asymptotic properties of estimation and inference based on MWLS.

How much better is the precision of MWLS with respect to the other parametric estimators? It is not possible to give a general answer because this depends on the distribution of (u, X') and also on the user's choice of the parametric model $\omega^2(X;\gamma)$. Hence, to examine this we do a Monte Carlo experiment in Section 4 exactly under the design of Romano and Wolf (2017) and with their choices of $\omega^2(X;\gamma)$. We find that the gain in precision, decrease in mean squared error, and increase in power for inference due to MWLS can be very large compared to OLS, WLS, and all the estimators in the recently published work of Romano and Wolf (2017) and DiCiccio et al. (2019). Reassuringly, at the same time we do not find any major downside related to empirical bias and size for MWLS.

Section 5 concludes. Appendices A and B respectively contain proofs of the results and descriptive endnotes for Section 3. A supplemental appendix contains supplementary simulation results.

2 Optimality intuition: Comparison with other estimators

We continue the discussion of the MWLS estimators and others under the setup of the Introduction.

2.1 Optimality of MWLS: Comparison with OLS, WLS and the efficiency bound

(a) To compare with OLS, we first note that standard choices of $\omega^2(X;\gamma)$, see Wooldridge (2012), Romano and Wolf (2017), etc., always allow for the case of conditional homoskedasticity, i.e.,

$$\text{the existence of a } \gamma^{\text{hom}} \in \Gamma \text{ such that } \omega^2(X;\gamma^{\text{hom}}) \text{ does not depend on } X. \tag{4}$$

For example, if $\omega^2(X;\gamma) = \exp(X'\gamma)$, then $\gamma^{\text{hom}} = (\gamma_1, 0, \ldots, 0)'$ or $\gamma^{\text{hom}} = (0, \ldots, 0)'$ satisfies (4) depending on whether the first element of X is a constant (which is the convention that we follow) or if no element of X is a constant (signifying that the scale factor of the variance is isolated out).

Now, since $\widehat{\beta}_n(\gamma)$ in (1) is invariant to the scale of $\omega^2(X;\gamma)$, it follows that the OLS estimator of β_j is in fact $\widehat{\beta}_{j,n}(\gamma^{\text{hom}})$, i.e., a member of the class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ defined in (2) with $\widehat{\gamma}_n^{\text{ols}} = \gamma^{\text{hom}}$ for all n. Hence, Σ^* defined in (3) cannot exceed the asymptotic variance $\Sigma(\gamma^{\text{hom}})$ of OLS.

(b) Similarly, the WLS estimator $\widehat{\beta}_{j,n}(\widehat{\gamma}_n^{\text{wls}})$ for β_j is a member of the class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ defined in (2) with some $\widehat{\gamma}_n^{\text{wls}}$ that is characterized by its probability limit $\gamma^{\text{wls}} := \arg\min_{\gamma\in\Gamma} E\left[\left(u^2 - \omega^2(X;\gamma)\right)^2\right]$. Hence, Σ^* defined in (3) cannot exceed the asymptotic variance $\Sigma(\gamma^{\text{wls}})$ of WLS as well.

It is well known that under (1) and other standard conditions, the semiparametric efficiency bound for the estimation of β_j is Σ^{eff} which is the (j,j)-th element of $\left(E\left[\frac{1}{\omega_0^2(X)}XX'\right]\right)^{-1}$. It is well known that OLS cannot achieve this bound Σ^{eff} except by happenstance unless $\omega_0^2(X)$ in (1) is a constant, i.e., the regression error is conditionally homoskedastic. It is also well known that WLS cannot achieve this bound Σ^{eff} either except by happenstance unless the user's parametric model $\omega^2(X;\gamma)$ is correctly specified for $\omega_0^2(X)$, i.e.,[6,7]

$$\text{there exists a } \gamma_0 \in \Gamma \text{ such that } \omega^2(X;\gamma_0) = \omega_0^2(X). \tag{5}$$

The asymptotic variance Σ^* in (3) that MWLS targets will generally also be larger than the efficiency bound Σ^{eff} if (5) does not hold. Importantly, however, in these cases where (5) does not hold, which is the reality in practice, MWLS does the best that it could given the deficiency of the user's parametric model $\omega^2(X;\gamma)$. WLS cannot necessarily do that because it obtains $\widehat{\gamma}_n^{\text{wls}}$ by targeting γ^{wls} which, as we will show analytically in Section 2.3 below, is generally an inappropriate target unless (5) holds, i.e., unless the user's parametric model $\omega^2(X;\gamma)$ is correctly specified.

In summary, given the user's parametric model $\omega^2(X;\gamma)$ that, if misspecified, could rule out semiparametric efficiency Σ^{eff}, MWLS delivers the "second-best" solution while WLS cannot, and OLS, by construction, does not even try. This is the fundamental difference between the idea behind MWLS and that for the other weighted-by-$\omega^{-1}(X;\gamma)$ LS estimators. The early papers like Carroll and Ruppert (1982) or the recent papers like Romano and Wolf (2017) and Spady and Stouli (2019), always seek some "best fit" of the user's model $\omega^2(X;\gamma)$ to the true conditional variance $\omega_0^2(X)$. By contrast, MWLS does not actively seek such a "best fit" to the true conditional variance because if $\omega^2(X;\gamma)$ is misspecified then this does not necessarily lead to the smallest asymptotic variance for the estimator of the parameters of interest. This is why the asymptotic variance of estimators like WLS and ALS can, in theory, be larger than that of OLS when $\omega^2(X;\gamma)$ is misspecified. (There is no misspecification under conditional homoskedasticity; see footnote 7.) By contrast, MWLS enjoys an asymptotic variance that is, by definition, at least as good as that of the other estimators in $\mathcal{C}(\beta_j,\omega^2(X;\gamma))$. MWLS respects the user's model $\omega^2(X;\gamma)$ and directly leads to the optimum Σ^* given this model. Of course, $\Sigma^* = \Sigma^{\text{eff}}$ if, luckily, this model $\omega^2(X;\gamma)$ happens to be correct.

2.2 Application of the optimality discussion to compare with newer estimators

(c) First consider the ALS estimator proposed by Romano and Wolf (2017). ALS is the OLS estimator if a pre-test cannot reject the conditional homoskedasticity of the regression error, but

[6] Thanks to scale invariance, it is immaterial if $\omega^2(X;\gamma_0)$ and $\omega_0^2(X)$ are proportional instead of equal as in (5).

[7] Under conditional homoskedasticity, (5) always holds by virtue of (4), i.e., $\omega^2(X;\gamma)$ is always correct for $\omega_0^2(X)$ and, consequently, OLS and WLS have the same asymptotic variance Σ^*; see Remark 3.7 in Romano and Wolf (2017).

is the WLS estimator otherwise. Provided that the pre-test for conditional homoskedasticity is consistent against the user-specified (not necessarily correct) alternative, the asymptotic variance of ALS is the same as that of WLS, i.e., $\Sigma(\gamma^{\text{wls}})$, if the parametric model $\omega^2(X;\gamma)$ allows for (4). Therefore, the contrast between MWLS and WLS also applies to that between MWLS and ALS.

Next, consider the two estimators proposed in DiCiccio et al. (2019) for scalar parameters.

(d) Their first estimator, which we refer to as MINVAR, is the OLS estimator if the estimated asymptotic variance of OLS is smaller than that of WLS, and is the WLS estimator otherwise. The asymptotic variance of MINVAR is $\min(\Sigma(\gamma^{\text{hom}}), \Sigma(\gamma^{\text{wls}}))$, i.e., the minimum of the asymptotic variance of OLS and WLS. Hence, the asymptotic variance Σ^* that MWLS targets cannot exceed the asymptotic variance of MINVAR since, by (3), Σ^* cannot exceed $\Sigma(\gamma^{\text{hom}})$ or $\Sigma(\gamma^{\text{wls}})$.

(e) Their second estimator, which we refer to as LINCOM, interpolates between the OLS and WLS estimators. The interpolation-weights, say, $\widehat{\lambda}$ and $1-\widehat{\lambda}$ are estimated by minimizing the estimated asymptotic variance of the linear combination $\lambda \times OLS + (1-\lambda) \times WLS$. Under their assumptions, $\widehat{\lambda}$ converges in probability to $\lambda_0 := \arg\min_{\lambda\in[0,1]} \lambda^2\Sigma(\gamma^{\text{hom}}) + (1-\lambda)^2\Sigma(\gamma^{\text{wls}}) + 2\lambda(1-\lambda)\text{ACov}(OLS, WLS)$ where $\text{ACov}(OLS, WLS)$ denotes the asymptotic covariance of the OLS and WLS estimators. We cannot analytically compare the asymptotic variances of LINCOM and MWLS because the estimators belong in different classes. Therefore, we do a comparison using simulations under the design of Romano and Wolf (2017). We find that in all the 10 different DGPs in their paper, LINCOM performs very similar to WLS, ALS and MINVAR and hence performs either similar to or much worse than MWLS. Unfortunately, LINCOM is also computationally costly.[8]

What if one wants an estimator whose asymptotic variance can be analytically shown to never exceed that of LINCOM? This is easy. For example, we can build on DiCiccio et al. (2019) to define two such estimators as follows. Define the first estimator as MWLS if the estimated asymptotic variance of MWLS is smaller than that of LINCOM, and as LINCOM otherwise. Define the second

[8]The issue is the following. WLS, ALS, MINVAR, LINCOM, MWLS and similar estimators all involve estimation of the nuisance parameters γ. However, under the mean independence condition $E[u|X] = 0$ in (1) and other regularity conditions, the well-known orthogonality condition holds and the estimation of γ does not affect the asymptotic variance of the aforementioned estimators. LINCOM involves an additional nuisance parameter λ whose estimation should also not affect the asymptotic variance of LINCOM if $\widehat{\lambda} = \lambda_0 + o_p(1)$ since, under standard conditions, it easily follows that $\sqrt{n}(LINCOM(\widehat{\lambda}) - \beta_0) - \sqrt{n}(LINCOM(\lambda_0) - \beta_0)$ converges in probability to zero. DiCiccio et al. (2019) recognize that this is misleading in practice and can lead to badly downward-biased standard errors. They recommend bootstrap for inference to account for the estimation of λ and γ. They also recommend bootstrap to account for the estimation of γ when doing inference with WLS, ALS and MINVAR. However, under the simulation design of Romano and Wolf (2017) we find that using the estimated asymptotic variance only has mild effect on inference with the other estimators, while, confirming DiCiccio et al. (2019), using the estimated asymptotic variance is disastrous for LINCOM even in extremely large samples. For LINCOM, a non-bootstrap inference based on t-ratios using HC0-HC4 variance results in empirical size that is always about 100% when the nominal level is only 5%. Hence, following DiCiccio et al. (2019), LINCOM should always use bootstrap. This makes LINCOM computationally costly.

estimator as $\widehat{\tau} \times MWLS + (1-\widehat{\tau}) \times LINCOM$ where $\widehat{\tau}$ minimizes the estimated asymptotic variance of $\tau \times MWLS + (1-\tau) \times LINCOM$ with respect to $\tau \in [0,1]$. However, we will not take these routes. For two reasons. First, based on our experience with various simulations so far it seems that MWLS can stand on its own without needing such refinements. Second, such new estimators will inherit the misleading asymptotic properties of LINCOM unless one uses bootstrap like DiCiccio et al. (2019) that, on the other hand, will make them inherit the computational cost of LINCOM.

(f) In a recent working paper, Spady and Stouli (2019) consider the joint estimation of the conditional mean and variance models in a linear regression, and show that their estimator can perform at least as well as the OLS estimator under misspecification of the conditional variance. Their Corollary 2 imposes the mean independence of the regression error similar to (1) and arrives at an estimator that belongs in the class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ in (2). Hence, similar to the discussion in Section 2.1, it follows by the definition of Σ^* in (3) that the target asymptotic variance Σ^* of MWLS can never exceed that of the estimator in Spady and Stouli (2019) (also see Section 2.3).

2.3 Digging deeper into the characterization of the minimizer of $\Sigma(\gamma)$

It is worth characterizing explicitly the channel through which MWLS attains the optimal asymptotic variance in (3) while the other estimators may not/cannot when the parametric model $\omega^2(X;\gamma)$ is misspecified for the true conditional variance $\omega_0^2(X)$. To emphasize the key point transparently, consider the simplest case of a single regressor and no intercept. Then, for notational simplicity, write $\Sigma(\gamma) = C(\gamma)/B^2(\gamma)$ where $B(\gamma) = E\left[\frac{1}{\omega^2(X;\gamma)}X^2\right]$ and $C(\gamma) = E\left[\frac{\omega_0^2(X)}{(\omega^2(X;\gamma))^2}X^2\right]$. Taking logarithm and assuming differentiability, the first-order condition for minimizing $\Sigma(\gamma)$ implies that:

$$\frac{\partial}{\partial\gamma}\log\left(\Sigma(\gamma^*)\right) = 0 \quad \Rightarrow \quad \frac{\partial}{\partial\gamma}\log\left(\frac{C(\gamma^*)}{B(\gamma^*)^2}\right) = 0 \quad \Rightarrow \quad \frac{\partial}{\partial\gamma}\log(B(\gamma^*)) = \frac{1}{2}\frac{\partial}{\partial\gamma}\log(C(\gamma^*))$$

where γ^* is a minimizer of $\Sigma(\gamma)$, i.e., $\Sigma(\gamma^*) = \Sigma^*$ in (3). Then, it follows after some algebra that:

$$E\left[\frac{1}{\omega^2(X;\gamma^*)}\frac{X^2}{\omega^2(X;\gamma^*)}\frac{\partial}{\partial\gamma}\log\left(\omega^2(X;\gamma^*)\right)\left\{\omega_0^2(X) - \frac{C(\gamma^*)}{B(\gamma^*)}\omega^2(X;\gamma^*)\right\}\right] = 0. \tag{6}$$

As it will turn out below, the unique feature of (6) is the "tilting" of $\omega^2(X;\gamma^*)$ by $C(\gamma^*)/B(\gamma^*)$.

First consider the case of a correctly specified $\omega^2(X;\gamma)$ as in (5), i.e., $\omega^2(X;\gamma_0) = \omega_0^2(X)$ for some $\gamma_0 \in \Gamma$. Hence $B(\gamma_0) = C(\gamma_0)$. Therefore, $B(\gamma^*) = C(\gamma^*)$ at $\gamma^* = \gamma_0$, and hence the tilting term $C(\gamma^*)/B(\gamma^*) = 1$, which means that, under correct specification (5), the condition in (6) holds if:

$$E\left[\frac{1}{\omega^2(X;\gamma^*)}\frac{X^2}{\omega^2(X;\gamma^*)}\frac{\partial}{\partial\gamma}\log\left(\omega^2(X;\gamma^*)\right)\left\{\omega_0^2(X) - \omega^2(X;\gamma^*)\right\}\right] = 0.$$

This latter condition holds tautologically irrespective of the terms outside the braces in the expectation because at $\gamma^* = \gamma_0$ we have $\omega^2(X;\gamma^*) = \omega^2(X;\gamma_0) = \omega_0^2(X)$ where the last equality follows by (5). Hence, under correct specification, the optimal γ, i.e., the minimizer of the asymptotic variance $\Sigma(\gamma) = C(\gamma)/B(\gamma)^2$, leads to this same well-known result of targeting $\omega_0^2(X)$ by $\omega^2(X;\gamma)$.

Now, consider the case of a misspecified $\omega^2(X;\gamma)$. Here the novelty of (6) will come into play. The presence of the tilting term $C(\gamma^*)/B(\gamma^*)$ in (6) means that, under misspecification, MWLS is not necessarily looking for a "best fit" (e.g., least squares) of the user's model $\omega^2(X;\gamma)$ to the true conditional variance $\omega_0^2(X)$. That is not necessarily the case for the other estimators in the class $\mathcal{C}(\beta_j, \omega^2(X;\gamma))$ in (2). Consider two examples covering all the relevant estimators in the class (2):

(Ex1) The WLS estimator (and hence the estimators based on it that we noted in Section 2.2) characterizes its target γ^{wls} by the "best fit" (un-weighted least squares) of $\omega^2(X;\gamma)$ to $\omega_0^2(X)$ as:

$$\gamma^{\text{wls}} = \arg\min_{\gamma\in\Gamma} E\left[(u^2 - \omega^2(X;\gamma))^2\right] \quad \Rightarrow \quad E\left[\frac{\partial}{\partial\gamma}\omega^2(X;\gamma^{\text{wls}})\left\{\omega_0^2(X) - \omega^2(X;\gamma^{\text{wls}})\right\}\right] = 0. \tag{7}$$

(Ex2) Imposing $E[u|X] = 0$ in the first-order condition for minimizing the criterion in Spady and Stouli (2019)'s equation (2.2) with respect to γ implies that their "best fit" of $\omega^2(X;\gamma)$ to $\omega_0^2(X)$ is characterized by a $\gamma^{SS} \in \Gamma$ that satisfies (SS stands for Spady and Stouli):

$$E\left[\frac{\partial}{\partial\gamma}\omega(X;\gamma^{SS})\frac{\{\omega_0^2(X) - \omega^2(X;\gamma^{SS})\}}{\omega^2(X;\gamma^{SS})}\right] = 0. \tag{8}$$

((8) uses our notation and ignores the fact that Spady and Stouli (2019) take $\omega^2(X;\gamma) = \omega^2(X'\gamma)$, i.e., as a linear index. (8) is otherwise the same as equation (3.9) of their Corollary 2.)

A key observation is that there is no tilting term like $C(\gamma)/B(\gamma)$ in (7) and (8). Consequently, irrespective of whether the user's parametric model $\omega^2(X;\gamma)$ is correct or misspecified for $\omega_0^2(X)$, the γ^{wls} in (7) and the γ^{SS} in (8) equate to zero the expectation of the respective weighted difference $\{\omega_0^2(X) - \omega^2(X;\gamma)\}$. By contrast, for MWLS, the γ^* in (6) equates to zero the expectation of the weighted difference $\{\omega_0^2(X) - C(\gamma)/B(\gamma)\omega^2(X;\gamma)\}$ where $\omega^2(X;\gamma)$ is tilted by $C(\gamma)/B(\gamma)$.

As noted above, this distinction with (6) would be immaterial under correct specification (5), i.e., if $\omega_0^2(X) = \omega^2(X;\gamma_0)$ for some $\gamma = \gamma_0$. In this case, all the methods are equivalent in the sense that $\gamma^* = \gamma_0$ (with $C(\gamma^*)/B(\gamma^*) = 1$), $\gamma^{\text{wls}} = \gamma_0$ and $\gamma^{SS} = \gamma_0$ respectively solve (6), (7) and (8).

However, this distinction does matter when the model $\omega^2(X;\gamma)$ is misspecified for $\omega_0^2(X)$. The tilting term in (6) arises directly from the fact that our target γ^* minimizes the asymptotic variance

$\Sigma(\gamma) = C(\gamma)/B(\gamma)^2$. Its absence from (7) and (8) means that γ^{wls} and γ^{SS} that are used by these other methods cannot possibly minimize this asymptotic variance $\Sigma(\gamma)$ except by happenstance.

Thus, the unique feature in the characterization (6) of γ^* is the tilting term $C(\gamma^*)/B(\gamma^*)$. It plays a fundamental role for the optimality of γ^* and, hence, that of MWLS by switching between one and not-one automatically depending on the data, the true conditional variance $\omega_0^2(X)$, and the user's model $\omega^2(X;\gamma)$, and with the sole purpose of minimizing the asymptotic variance $\Sigma(\gamma)$.

3 The modified weighted least squares (MWLS) estimator

3.1 Framework and preliminaries

Now, generalizing the setup of Sections 1 and 2, consider the possibly nonlinear regression model:

$$\begin{aligned} y &= g(X;\beta_0) + u \qquad (9) \\ \text{where } \; E[u|X] &= 0 \;\text{ almost surely in } X. \qquad (10) \end{aligned}$$

$g(X;\beta)$ is a known function of X and $\beta \in \mathcal{B} \subset \mathbb{R}^p$; e.g., $g(X;\beta) = X'\beta$. β_0 is the true value of β.

Our parameter of interest $h(\beta)$ is a scalar. For example, it can be the individual elements of β, their sums, differences or, more generally, any non-random smooth function of β. $h_0 = h(\beta_0)$ denotes the true value of $h(\beta)$. We will now consider MWLS-based estimation and inference on h_0.

Suppose that we have i.i.d. observations $(y_i, X_i')_{i=1}^n$ on (y, X'). Then, under standard conditions (see Chamberlain (1987)), the semiparametric efficiency bound for the estimation of h_0 is:

$$\Sigma^{\text{eff}} := H\left(E\left[\frac{1}{\omega_0^2(X)}G'(X)G(X)\right]\right)^{-1}H' \qquad (11)$$

where we use the notation:

$$\omega_0^2(X) := E[u^2|X], \;\; G(X;\beta) := \frac{\partial}{\partial\beta'}g(X;\beta), \;\; H(\beta) := \frac{\partial}{\partial\beta'}h(\beta), \;\; G(X) := G(X;\beta_0), \;\; H := H(\beta_0).$$

If $\omega_0^2(X)$ were known then under standard regularity conditions (see Theorem 5.23 and Example 5.27 in van der Vaart (1998)) one could obtain an asymptotically efficient estimator of h_0 as:

$$\widehat{h}_n^{\text{inf}} := h(\widehat{\beta}_n^{\text{inf}}) \;\text{ where }\; \widehat{\beta}_n^{\text{inf}} := \arg\min_\beta \frac{1}{n}\sum_{i=1}^n \frac{1}{\omega_0^2(X_i)}(y_i - g(X_i;\beta))^2.$$

This estimator is infeasible if $\omega_0^2(X)$ is unknown. Adaptation of semiparametric WLS as in Carroll

(1982), Robinson (1987) and Newey (1994) by plugging in nonparametric estimators of $\omega_0^2(X_i)$ in the above objective function gives analogous estimators that are asymptotically equivalent to $\widehat{h}_n^{\inf}$.

In our paper, however, we take the user's parametric model $\omega^2(X;\gamma)$ for $\omega_0^2(X)$ seriously. Generalizing (1), we now define the weighted-by-$\omega^{-1}(X;\gamma)$ estimator of h_0 as a function of $\gamma \in \Gamma$ as:

$$\widehat{h}_n(\gamma) := h(\widehat{\beta}_n(\gamma)) \quad \text{where} \quad \widehat{\beta}_n(\gamma) := \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\omega^2(X_i;\gamma)}(y_i - g(X_i;\beta))^2. \tag{12}$$

Generalizing the class of linear estimators in (2) to nonlinear estimators $\widehat{h}_n(\widehat{\gamma}_n)$ where $\widehat{\gamma}_n \in \Gamma$ is possibly a random sequence, we will be concerned with optimality as in (3) in terms of asymptotic variance. The conventional weighted and un-weighted NLS estimators of h_0 belong in this class.

It is well known that a misspecified parametric model $\omega^2(X;\gamma)$ does not in general affect the consistency of such parametric estimators for h_0 under standard conditions and (10). This is what allows us to take the user-specified parametric model $\omega^2(X;\gamma)$ seriously for the MWLS estimation.

Before proceeding, it is useful to list here the standard notation that we use. $|.|$, $\|.\|$, $\xrightarrow{d}$, $O_p(1)$, and $o_p(1)$ respectively denote the absolute value, the Euclidean norm, convergence in distribution, bounded in probability, and convergent in probability to 0. We will also use the following notation:

$$\psi(y,X;\beta,\gamma) := \frac{(y-g(X;\beta))^2}{\omega^2(X;\gamma)},\ \psi_\beta(y,X;\beta,\gamma) := \frac{\partial}{\partial\beta}\psi(y,X;\beta,\gamma),\ \psi_{\beta\beta}(y,X;\beta,\gamma) := \frac{\partial}{\partial\beta'}\psi_\beta(y,X;\beta,\gamma),$$
$$\bar{\psi}_n(\beta,\gamma) := \frac{1}{n}\sum_{i=1}^{n}\psi(y_i,X_i;\beta,\gamma),\ \bar{\psi}_{\beta,n}(\beta,\gamma) := \frac{1}{n}\sum_{i=1}^{n}\psi_\beta(y_i,X_i;\beta,\gamma),\ \bar{\psi}_{\beta\beta,n}(\beta,\gamma) := \frac{1}{n}\sum_{i=1}^{n}\psi_{\beta\beta}(y_i,X_i;\beta,\gamma),$$
$$B_n(\gamma) := E[\bar{\psi}_{\beta\beta,n}(\beta_0,\gamma)],\ B(\gamma) := \lim_{n\to\infty} B_n(\gamma),\ C_n(\gamma) := \mathrm{Var}\left(\sqrt{n}\bar{\psi}_{\beta,n}(\beta_0,\gamma)\right),\ \text{and } C(\gamma) := \lim_{n\to\infty} C_n(\gamma).$$

Given the i.i.d. data, it follows under (10) that $B_n(\gamma) = B(\gamma) = E[\psi_{\beta\beta}(y,X;\beta_0,\gamma)] = E\left[G'(X)G(X)/\omega^2(X;\gamma)\right]$ and $C_n(\gamma) = C(\gamma) = \mathrm{Var}\left(\psi_\beta(y,X;\beta_0,\gamma)\right) = E\left[\omega_0^2(X)G'(X)G(X)/(\omega^2(X;\gamma))^2\right]$, when they exist, for all $\gamma \in \Gamma$ and all $n \geq 1$. Finally, to accommodate for the possible nonlinearity of the regression function $g(X;\beta)$ in β, we generalize the definitions of $\Xi(\gamma)$ and $\Sigma(\gamma)$ from Sections 1 and 2 as:

$$\Xi(\gamma) := B^{-1}(\gamma)C(\gamma)B^{-1}(\gamma) \quad \text{and} \quad \Sigma(\gamma) := H\Xi(\gamma)H'. \tag{13}$$

To avoid pathological cases we will henceforth maintain that $\Xi(\gamma)$ is finite and positive definite for $\gamma \in \Gamma$. More precisely, we will maintain throughout that the parameter space Γ for γ satisfies:

$$\inf_{\gamma\in\Gamma}(\text{smallest eigen value of } \Xi(\gamma)) > 0 \text{ and } \sup_{\gamma\in\Gamma}(\text{largest eigen value of } \Xi(\gamma)) < \infty. \tag{14}$$

Lemma 1 *Let (9) and (10) hold. Consider the estimators $\widehat{\beta}_n(\gamma)$ and $\widehat{h}_n(\gamma)$ in (12) for any $\gamma \in \Gamma$.*

(i) Let: (a) $\sup_{\beta\in\mathcal{B}} |\bar{\psi}_n(\beta,\gamma) - E[\psi(y,X;\beta,\gamma)]| = o_p(1)$; and (b) $\inf_{\beta\in\mathcal{B}:|\beta-\beta_0|\geq\epsilon} E[\psi(y,X;\beta,\gamma)] > E[\psi(y,X;\beta_0,\gamma)]$ for each $\epsilon > 0$. Then $\widehat{\beta}_n(\gamma) = \beta_0 + o_p(1)$. Additionally, let (c) $h(\beta)$ be continuous at $\beta = \beta_0$. Then $\widehat{h}_n(\gamma) := h(\widehat{\beta}_n(\gamma)) = h_0 + o_p(1)$.

(ii) Let: (d) $B(\gamma)$, $B^{-1}(\gamma)$ and $C(\gamma)$ be finite; (e) standard continuity and convergence conditions, including $\widehat{\beta}_n(\gamma) = \beta_0 + o_p(1)$, be satisfied such that the following expansion holds around β_0:

$$\sqrt{n}\bar{\psi}_{\beta,n}(\widehat{\beta}_n(\gamma),\gamma) = \sqrt{n}\bar{\psi}_{\beta,n}(\beta_0,\gamma) + B_n(\gamma)\sqrt{n}(\widehat{\beta}_n(\gamma) - \beta_0) + o_p(1)$$

Then $\sqrt{n}(\widehat{\beta}_n(\gamma) - \beta_0) = -B^{-1}(\gamma)\sqrt{n}\bar{\psi}_{\beta,n}(\beta_0,\gamma) + o_p(1) \xrightarrow{d} N(0,\Xi(\gamma))$ where $\Xi(\gamma)$ is defined in (13). Additionally, let: (f) the expansion $\sqrt{n}(\widehat{h}_n(\gamma) - h_0) = \sqrt{n}(h(\widehat{\beta}_n(\gamma)) - h_0) = H\sqrt{n}(\widehat{\beta}_n(\gamma) - \beta_0) + o_p(1)$ hold. Then $\sqrt{n}(\widehat{h}_n(\gamma) - h_0) \xrightarrow{d} N(0,\Sigma(\gamma))$ where $\Sigma(\gamma)$ is defined in (13).

Remark: The well-known result in Lemma 1 is presented as a reference to facilitate the discussion below. It is difficult to provide useful primitive conditions for the well-separability assumption (b) with a generic $g(X;\beta)$. Primitive conditions for the other high level assumptions are well known.

The key quantity that Lemma 1 provides for our paper is the asymptotic variance function $\Sigma(\gamma)$ as defined in (13). Note that $\Sigma(\gamma) - \Sigma^{\text{eff}} \geq 0$ where Σ^{eff} is the efficiency bound as defined in (11). This is because $E\left[G'(X)G(X)/\omega_0^2(X)\right] - \Xi^{-1}(\gamma) = E[e(X;\gamma)e'(X;\gamma)]$ is positive semi-definite where:

$$e(X;\gamma) \;:=\; \frac{1}{\omega_0(X)}G'(X) \;-\; B(\gamma)C^{-1}(\gamma)\frac{\omega_0^2(X)}{\omega^2(X;\gamma)}\frac{1}{\omega_0(X)}G'(X)$$

is the residual from the population regression of $\frac{1}{\omega_0(X)}G'(X)$ on $\frac{\omega_0^2(X)}{\omega^2(X;\gamma)}\frac{1}{\omega_0(X)}G'(X)$. If the user's parametric model $\omega^2(X;\gamma)$ is correct for $\omega_0^2(X)$, i.e., if $\omega^2(X;\gamma_0) = \omega_0^2(X)$ for some γ_0 as in (5), then $B(\gamma_0) = C(\gamma_0)$, and hence $e(X;\gamma_0) = 0$. Naturally, in this case, $\Xi(\gamma_0) = B^{-1}(\gamma_0) = C^{-1}(\gamma_0) = \left(E\left[G'(X)G(X)/\omega_0^2(X)\right]\right)^{-1}$ and $\Sigma(\gamma_0) = \Sigma^{\text{eff}}$ defined in (11). However, the point of our paper and the related papers in the recent literature is that the user's model $\omega^2(X;\gamma)$ may not be correct.

Therefore, appealing to our extensive discussion in Sections 1 and 2 on the "second-best" optimality sought by MWLS, the MWLS principle in this general context will seek the optimum:

$$\Sigma^* := \min_{\gamma\in\Gamma} \Sigma(\gamma). \tag{15}$$

This general definition of Σ^* for a general parameter of interest $h(\beta)$ in a nonlinear regression model supersedes the definition of Σ^* in (3) that focused on individual coefficients in a linear regression.

3.2 Estimation of $\Sigma(\gamma)$ and its minimizer with respect to γ

The key step of the MWLS principle of estimation is to find an "optimal" value of γ that minimizes the function $\Sigma(\gamma)$. In practice, this has to be done based on the sample analog of $\Sigma(\gamma)$, and one can use any of the HC versions for this purpose; see, e.g., MacKinnon (2012). Appealing to the structure of $\Xi(\gamma)$ and $\Sigma(\gamma)$ in (13) and the result of Lemma 1, we will now define these sample analogs as the HC versions HC0-HC3. First, for each $\beta \in \mathcal{B}$ and $\gamma \in \Gamma$ and each $i = 1, \ldots, n$ define:

$$\widetilde{u}_i(\beta,\gamma) := \frac{y_i - g(X_i;\beta)}{\sqrt{\omega^2(X_i;\gamma)}}, \quad \widetilde{G}_i(\beta,\gamma) := \frac{G(X_i;\beta)}{\sqrt{\omega^2(X_i;\gamma)}} \quad \text{and} \quad \widehat{B}_n(\gamma;\beta) := \frac{1}{n}\sum_{i=1}^{n} \widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)$$

for simpler notation. Then, for each HC version that is generically denoted as $\widehat{C}_n(\gamma;\beta)$, define:

$$\widehat{\Xi}_n(\gamma;\beta) := \widehat{B}_n^{-1}(\gamma;\beta)\widehat{C}_n(\gamma;\beta)\widehat{B}_n^{-1}(\gamma;\beta) \quad \text{and} \quad \widehat{\Sigma}_n(\gamma;\beta) := H(\beta)\widehat{\Xi}_n(\gamma;\beta)H'(\beta) \tag{16}$$

where, letting $q_{i,n}(\gamma;\beta) := \widetilde{G}_i(\beta,\gamma)\left(\sum_{i=1}^{n} \widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)\right)^{-1}\widetilde{G}_i'(\beta,\gamma)$ denote the leverage for the i-th observation, we allow the generic $\widehat{C}_n(\gamma;\beta)$ to take any of the following four HC forms:

$$HC0: \ \widehat{C}_n^{(0)}(\gamma;\beta) := \frac{1}{n}\sum_{i=1}^{n} \widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma), \quad HC1: \ \widehat{C}_n^{(1)}(\gamma;\beta) := \frac{n}{n-p}\widehat{C}_n^{(0)}(\gamma;\beta),$$

$$HC2: \ \widehat{C}_n^{(2)}(\gamma;\beta) := \frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)}{1-q_{i,n}(\gamma;\beta)}, \quad HC3: \ \widehat{C}_n^{(3)}(\gamma;\beta) := \frac{1}{n}\sum_{i=1}^{n} \frac{\widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)}{(1-q_{i,n}(\gamma;\beta))^2}.$$

Lemma 2 *Consider $\Xi(\gamma)$ and $\Sigma(\gamma)$ defined in (13), and $\widehat{\Xi}_n(\gamma;\beta)$ and $\widehat{\Sigma}_n(\gamma;\beta)$ defined in (16). Let: (a) $E[\sup_{\gamma\in\Gamma}|\widetilde{u}_i(\beta_0,\gamma)|^4] < \infty$; (b) $E[\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\beta_0,\gamma)\|^4] < \infty$; (c) $E[\sup_{\beta\in\mathcal{B}:|\beta-\beta_0|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4] < \infty$ for some $\epsilon > 0$; (d) $E[\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)]$ be finite, continuous and nonsingular at $\beta = \beta_0$ uniformly in $\gamma \in \Gamma$; and (e) $H(\beta)$ be finite and continuous at $\beta = \beta_0$. Then, for any estimator $\widehat{\beta}$ such that $\widehat{\beta} - \beta_0 = o_p(1)$, we have: $\sup_{\gamma\in\Gamma}\|\widehat{\Xi}_n(\gamma;\widehat{\beta}) - \Xi(\gamma)\| = o_p(1)$ and $\sup_{\gamma\in\Gamma}|\widehat{\Sigma}_n(\gamma;\widehat{\beta}) - \Sigma(\gamma)| = o_p(1)$ for all the four HC versions HC0-HC3 of $\widehat{\Xi}_n(\gamma;\widehat{\beta})$ and $\widehat{\Sigma}_n(\gamma;\widehat{\beta})$ defined in (16).*

Remarks: 1. We maintain the assumptions uniformly in γ in this lemma to ensure that the sample objective function does not deviate far from $\Sigma(\gamma)$ anywhere in Γ. 2. As in Lemma 1(ii), this lemma also implicitly maintains differentiability at β_0. 3. Existence of the fourth moment assumptions in (a) and (b) are standard, while that in (c) is maintained in the given form for expositional ease in dealing with the nonlinearity of $g(X;\beta)$. 4. Assumptions (d) and (e) enable the continuous mapping theorem. The finiteness and nonsingularity in assumption (d) almost (but not quite) follow from assumption (b) and (14). So we impose them explicitly. 5. $\widehat{\beta}$ can be the un-weighted NLS estimator of β since Lemma 1(i) holds for this estimator at $\gamma = \gamma^{\text{hom}}$ by virtue of (4).

Now, based on this initial estimator $\widehat{\beta}$ for β that subsequently leads to the estimator $\widehat{\Sigma}_n(\gamma; \widehat{\beta})$ for $\Sigma(\gamma)$ as in Lemma 2, we define an estimator for the minimizer(s) of $\Sigma(\gamma)$ as $\widehat{\gamma}_n(\widehat{\beta})$ such that:

$$\widehat{\gamma}_n(\widehat{\beta}) := \arg\min_{\gamma \in \Gamma} \widehat{\Sigma}_n(\gamma; \widehat{\beta}). \tag{17}$$

A minimizer of $\Sigma(\gamma)$ exists by the extreme value theorem if $\Sigma(\gamma)$ is continuous in γ and if the parameter space Γ is compact in $\mathbb{R}^k$. However, while it can be easily avoided if it relates to the scale factor of $\omega^2(X; \gamma)$, the minimizer of $\Sigma(\gamma)$ may not be unique in general, i.e., the set of minimizers

$$\Gamma^* := \{\gamma \in \Gamma \mid \Sigma(\gamma) = \Sigma^* \text{ where } \Sigma^* = \min_{g \in \Gamma} \Sigma(g) \text{ as defined in (15)}\} \tag{18}$$

may not be singleton. Hence, $\widehat{\gamma}_n(\widehat{\beta})$ may not converge in probability to any point. This will not hinder the practical usefulness of our results since $\Sigma(\gamma) = \Sigma^*$ for all $\gamma \in \Gamma^*$. More precisely, as we will see, a non-singleton Γ^*: (i) precludes any asymptotic equivalence result for the MWLS estimator of h_0 because it obscures the reference point (probability limit point for $\widehat{\gamma}_n(\widehat{\beta})$) for such asymptotic equivalence, but (ii) does not preclude the asymptotic normal distribution with variance Σ^* for the MWLS estimator of h_0. Therefore, by virtue of (ii), we will be able to conduct inference, and for that matter optimal inference (in the sense of utilizing the optimal asymptotic variance Σ^*), on h_0 using MWLS in the usual manner irrespective of a singleton or non-singleton Γ^*.

What does matter for our results is that $\widehat{\gamma}_n(\widehat{\beta})$ gets close in probability to the set Γ^* in terms of a suitable distance at a suitably fast rate. While a rate of $n^{-1/4}$ would be good enough, under standard assumptions of parametric estimation and abstracting from nonstandard tail behavior, Lemma 3 below establishes that one could actually obtain the parametric rate.

The distance measure that we consider in our paper for any $\gamma \in \Gamma$ from the set Γ^* in (18) is:

$$d(\gamma, \Gamma^*) = \inf_{\bar{\gamma} \in \Gamma^*} \|\gamma - \bar{\gamma}\|. \tag{19}$$

Lemma 3 *Consider the set Γ^* defined in (18), and the estimator $\widehat{\gamma}_n(\widehat{\beta})$ defined in (17) that is based on some estimator $\widehat{\beta}$, e.g., the un-weighted NLS estimator, satisfying $\widehat{\beta} = \beta_0 + o_p(1)$. If $\Gamma = \Gamma^*$ then $d(\widehat{\gamma}_n(\widehat{\beta}), \Gamma^*) = 0$ trivially. Now consider the nontrivial case that $\Gamma \neq \Gamma^*$, i.e., let $\Gamma \setminus \Gamma^*$ be nonempty. Let $\Gamma^*_\delta := \{\gamma \in \Gamma \,|\, d(\gamma, \Gamma^*) \leq \delta\} \subseteq \Gamma$ denote the δ-expansion of Γ^* in Γ for any $\delta \geq 0$.*

(i) Let: (a) $\sup_{\gamma\in\Gamma}|\widehat{\Sigma}_n(\gamma;\widehat{\beta})-\Sigma(\gamma)|=o_p(1)$*; and (b) for all* $\delta>0$*, there exist* $\epsilon(\delta)>0$ *such that* $\inf_{\gamma\in\Gamma^*_\delta\setminus\Gamma^*}\Sigma(\gamma)-\Sigma^*\geq\epsilon(\delta)$*. Then* $d(\widehat{\gamma}_n(\widehat{\beta}),\Gamma^*)=o_p(1)$*.*

(ii) Let: (c) $d(\widehat{\gamma}_n(\widehat{\beta}),\Gamma^*)=o_p(1)$*; (d)* $\sup_{\gamma\in\Gamma^*_\delta}\sqrt{n}|\widehat{\Sigma}_n(\gamma;\widehat{\beta})-\Sigma(\gamma)|=O_p(1)$ *for some* $\delta>0$*; and (e)* $\inf_{\gamma\in\Gamma^*_\delta\setminus\Gamma^*}\ [(\Sigma(\gamma)-\Sigma^*)-\kappa\times d(\gamma,\Gamma^*)]\geq 0$ *for some* $\delta>0$ *and* $\kappa>0$*. Then* $\sqrt{n}d(\widehat{\gamma}_n(\widehat{\beta}),\Gamma^*)=O_p(1)$*.*

Remarks: 1. Assumptions (a) and (b) in Lemma 3(i) are standard conditions similar to those in Lemma 1(i). Assumption (a) follows from Lemma 2. Assumption (b) ensures that Γ^* is well-separated from the other γ's in Γ. It is difficult to provide useful sufficient conditions for assumption (b). 2. Assumption (c) restates the result of Lemma 3(i). Assumption (d) strengthens assumption (a) locally by specifying the $\sqrt{n}$ rate at which $d(\widehat{\gamma}_n(\widehat{\beta}),\Gamma^*)$ may converge.[9] Finally, assumption (e) delivers this convergence by strengthening assumption (b) locally so that Γ^* is locally identified.

3.3 MWLS estimator of $h(\beta)$: Algorithm, asymptotic properties, and inference

3.3.1 Algorithm to obtain the MWLS estimator of $h(\beta)$

The algorithm for MWLS is exactly the same as that for WLS but with one crucial difference that concerns how the parameter γ in the user's parametric model $\omega^2(X;\gamma)$ for the conditional variance $\omega_0^2(X)$ is estimated. In the conventional WLS, γ is estimated by a least squares fit of $\omega^2(X_i;\gamma)$ to the squared residual $(y_i-g(X_i;\widehat{\beta}))^2$ based on the un-weighted LS/NLS estimator $\widehat{\beta}$ of β. In MWLS, γ is estimated by minimizing with respect to γ the $\widehat{\Sigma}_n(\gamma;\widehat{\beta})$ that is obtained based on the un-weighted LS/NLS estimator $\widehat{\beta}$ of β. Formally, the MWLS algorithm is as follows.

- **Step 1:** Obtain any preliminary estimator $\widehat{\beta}$; e.g., $\widehat{\beta}$ is the un-weighted LS/NLS estimator depending on whether $g(X;\beta)$ is linear or nonlinear in β.[10] Then, based on the user's preference for the HC0-HC3 version, obtain $\widehat{\Sigma}_n(\gamma;\widehat{\beta})$ exactly as described in (16) as a function of γ. Then, obtain $\widehat{\gamma}_n(\widehat{\beta})$ by minimizing $\widehat{\Sigma}_n(\gamma;\widehat{\beta})$ with respect to γ as described in (17), i.e.,

$$\widehat{\gamma}_n(\widehat{\beta}):=\arg\min_{\gamma\in\Gamma}\widehat{\Sigma}_n(\gamma;\widehat{\beta}).$$

- **Step 2:** Using the $\widehat{\gamma}_n(\widehat{\beta})$ obtained in Step 1, now follow (12) exactly to obtain the MWLS estimator for the parameter of interest $h(\beta)$ as $\widehat{h}_n(\widehat{\gamma}_n(\widehat{\beta})):=h(\widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta})))$ where $\widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta}))$ is

[9] Lemma 2 established assumption (a). We presented its proof in full details to ensure that if $\sqrt{n}(\widehat{\beta}-\beta_0)=O_p(1)$, which it will be for standard choices such as un-weighted NLS, then exactly the same proof will work to establish assumption (d) under the additional condition that $\sup_{\in\Gamma^*_\delta}\sqrt{n}\|\frac{1}{n}\sum_{i=1}^n\widetilde{u}_i^2(\beta_0,\gamma)\widetilde{G}_i'(\beta_0,\gamma)\widetilde{G}_i(\beta_0,\gamma)-\Sigma(\gamma)\|=O_p(1)$.

[10] It is worth remembering from (4) and (12) that the un-weighted LS/NLS estimator $\widehat{\beta}$ of β satisfies $\widehat{\beta}=\widehat{\beta}_n(\gamma^{\text{hom}})$.

the weighted-by-$\omega^{-1}(X;\widehat{\gamma}_n(\widehat{\beta}))$ LS/NLS estimator of β, i.e.,

$$\widehat{h}_n(\widehat{\gamma}_n(\widehat{\beta})) := h(\widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta}))) \quad \text{where} \quad \widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta})) := \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\omega^2(X_i;\widehat{\gamma}_n(\widehat{\beta}))}(y_i - g(X_i;\beta))^2.$$

One might choose to stop the algorithm here, and our experience so far is that this works well. Alternatively, one might add a Step 3 that repeats Steps 1 and 2 by using $\widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta}))$ as the (updated) preliminary estimator $\widehat{\beta}$. One might even iterate on Step 3. Our presentation of Lemma 2 and Lemma 3, as it is, can handle the simple algorithm based on Steps 1 and 2 only, or its iterative variants noted above. One might also set up the algorithm to simultaneously obtain estimates of β and γ by solving their respective profiled estimating equations. However, as is well known more generally: Step 3, or iterations of Step 3, or the simultaneous estimation of β and γ do not provide improvement in the first-order asymptotics over the simple implementation involving Steps 1 and 2 only. Higher order improvements leading to better finite-sample properties might be possible, but analytical exploration of this topic and re-sampling methods in this context is left for future work.

3.3.2 Asymptotic properties of the MWLS estimator of $h(\beta)$

We will now present the asymptotic properties of the MWLS estimator of $h(\beta)$. The MWLS estimators obtained in Step 2 or, if the user prefers, in Step 3 or iterations of it, differ in only one aspect — that is the estimator of γ that they use. For example, the MWLS estimator in Step 2 uses the estimator $\widehat{\gamma}_n(\widehat{\beta})$; the MWLS estimator in Step 3 uses the estimator $\widehat{\gamma}_n(\widehat{\beta}_n(\widehat{\gamma}_n(\widehat{\beta})))$; and so on and so forth for the MWLS estimators obtained in the further iterations. However, none of these matter for the consistency, asymptotic normality and asymptotic optimality properties of the MWLS estimator as long as the estimator of γ, denote it generically by $\widehat{\gamma}_n$, satisfies $\sqrt{n}d(\widehat{\gamma}_n,\Gamma^*) = O_p(1)$, i.e., the result of Lemma 3(ii), or even $n^{1/4}d(\widehat{\gamma}_n,\Gamma^*) = o_p(1)$. Therefore, to consider the different estimators of γ under a unified framework we present below the asymptotic properties of an MWLS estimator that is generically denoted as $\widehat{h}_n(\widehat{\gamma}_n)$ to signify the generic nature of the $\widehat{\gamma}_n$ used by it.

The conditions maintained to obtain the asymptotic properties of $\widehat{h}_n(\widehat{\gamma}_n)$ are similar to those in Lemma 1 except that they are now suitably strengthened because Lemma 1 only considered a given and known γ whereas we use estimated γ here. Accordingly, define the following quantities:

$$\begin{aligned}
\psi_{\beta\gamma}(y,X;\beta,\gamma) &:= \frac{\partial}{\partial\gamma'}\psi_\beta(y,X;\beta,\gamma) = -\left\{\frac{\partial}{\partial\gamma'}\frac{1}{\omega^2(X_i;\gamma)}\right\}G'(X_i;\beta)(y_i - g(X_i;\beta)),\\
\bar{\psi}_{\beta\gamma,n}(\beta,\gamma) &:= \frac{1}{n}\sum_{i=1}^{n}\psi_{\beta\gamma}(y_i,X_i;\beta,\gamma), \quad D_n(\gamma) := E[\bar{\psi}_{\beta\gamma,n}(\beta_0,\gamma)], \text{ and } D(\gamma) := \lim_{n\to\infty} D_n(\gamma).
\end{aligned}$$

By virtue of the i.i.d. data and the mean independence $E[u|X] = 0$ of the regression error in (10), it follows that: $D_n(\gamma) = D(\gamma) = -E\left[\left\{\frac{\partial}{\partial \gamma'}\frac{1}{\omega^2(X;\gamma)}\right\} G'(X;\beta)E\left[(y - g(X;\beta_0))|X\right]\right] = 0$ for all $\gamma \in \Gamma$ and all $n \geq 1$. We also recall from our discussion earlier that the i.i.d. data give: $B_n(\gamma) = B(\gamma)$, $C_n(\gamma) = C(\gamma)$ and, thereby, $\Sigma_n(\gamma) := HB_n^{-1}(\gamma)C_n(\gamma)B_n^{-1}(\gamma)H' = \Sigma(\gamma)$ for all $\gamma \in \Gamma$ and all $n \geq 1$.

Theorem 4 below presents the asymptotic properties of MWLS. Any nonstandard aspect in the presentation is due to the non-singleton nature of Γ^* that we have allowed for to give the user more freedom in choosing their model $\omega^2(X;\gamma)$. If Γ^* is singleton, i.e., $\Gamma^* = \{\gamma^*\}$ for a fixed $\gamma^* \in \Gamma$ with $\Sigma(\gamma^*) = \Sigma^*$, then we are back to the conventional case and our presentation reduces to the standard presentation of two-step estimation as in, e.g., Section 6 of Newey and McFadden (1994).

Theorem 4 *Consider the estimators* $\widehat{\beta}_n(\widehat{\gamma}_n)$ *and* $\widehat{h}_n(\widehat{\gamma}_n) := h(\widehat{\beta}_n(\widehat{\gamma}_n))$ *where* $\widehat{\gamma}_n$ *is a sequence of generic estimators obtained in Step 1 of the MWLS algorithm or iterations of the algorithm. Let* $\gamma_n \in$ *closure*(Γ^*) *in* Γ *be such that* $\|\widehat{\gamma}_n - \gamma_n\| = d(\widehat{\gamma}_n, \Gamma^*)$ *for all* $n \geq 1$. *Let* $\Gamma^*_\delta := \{\gamma \in \Gamma \,|d(\gamma, \Gamma^*) \leq \delta\} \subseteq \Gamma$ *denote the* δ*-expansion of* Γ^* *in* Γ *for any* $\delta \geq 0$. *Let (9) and (10) hold.*

(i) Let: (a) $d(\widehat{\gamma}_n, \Gamma^*) = o_p(1)$; *(b)* $\sup_{\beta \in \mathcal{B}, \gamma \in \Gamma^*_\delta} |\bar{\psi}_n(\beta, \gamma) - E[\psi(y, X; \beta, \gamma)]| = o_p(1)$ *for some* $\delta > 0$; *and (c)* $\inf_{\gamma \in \Gamma^*_\delta}\left(\inf_{\beta \in \mathcal{B}:|\beta - \beta_0| \geq \epsilon} E[\psi(y, X; \beta, \gamma)] - E[\psi(y, X; \beta_0, \gamma)]\right) > 0$ *for each* $\epsilon > 0$ *and some* $\delta > 0$. *Then* $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$. *Additionally, let (d)* $h(\beta)$ *be continuous at* $\beta = \beta_0$. *Then* $\widehat{h}_n(\widehat{\gamma}_n) := h(\widehat{\beta}_n(\widehat{\gamma}_n)) = h_0 + o_p(1)$.

(ii) Let: (e) suitable continuity and convergence conditions, including $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$ *and* $\sqrt{n}d(\widehat{\gamma}_n, \Gamma^*) = O_p(1)$, *be satisfied such that the following expansion holds around* (β_0, γ_n):

$$\begin{aligned}\sqrt{n}\bar{\psi}_{\beta,n}(\widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n) &= \sqrt{n}\bar{\psi}_{\beta,n}(\beta_0, \gamma_n) + B_n(\gamma_n)\sqrt{n}(\widehat{\beta}_n(\widehat{\gamma}_n) - \beta_0) + D_n(\gamma_n)\sqrt{n}(\widehat{\gamma}_n - \gamma_n) + o_p(1) \\ &= \sqrt{n}\bar{\psi}_{\beta,n}(\beta_0, \gamma_n) + B_n(\gamma_n)\sqrt{n}(\widehat{\beta}_n(\widehat{\gamma}_n) - \beta_0) + o_p(1);\end{aligned}$$

(f) the expansion $\sqrt{n}(\widehat{h}_n(\widehat{\gamma}_n) - h_0) = H\sqrt{n}(\widehat{\beta}_n(\widehat{\gamma}_n) - \beta_0) + o_p(1)$ *hold; (g) a suitable central limit theorem give* $\Sigma_n^{-1/2}(\gamma_n)HB_n^{-1}(\gamma_n)\sqrt{n}\bar{\psi}_{\beta,n}(\beta_0, \gamma_n) \xrightarrow{d} N(0, 1)$ *where* $\Sigma_n^{1/2}(\gamma)\Sigma_n^{1/2'}(\gamma) = \Sigma_n(\gamma)$; *and (h)* $\lim_{n\to\infty} \Sigma_n(\gamma_n) = \Sigma^*$ *defined in (15). Then* $\sqrt{n}(\widehat{h}_n(\widehat{\gamma}_n) - h_0) \xrightarrow{d} N(0, \Sigma^*)$.

(iii) Let assumptions (a)-(e) of Lemma 2 hold but with Γ *replaced by* $\Gamma^*_\delta \subseteq \Gamma$ *for some* $\delta \geq 0$. *Let assumptions (a) and (h) of Theorem 4 (i) and (ii) respectively hold. Let* $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$. *Then* $\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n)) = \Sigma^* + o_p(1)$ *for the HC0-HC3 versions of* $\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n))$ *in (16).*

Remarks: The assumptions in Theorem 4, like those in Lemma 1, are stated at a very high level.[11] Theorem 4 (i) and (ii) give consistency, asymptotic normality and optimality of the MWLS estimator. Theorem 4(iii) gives consistency of the estimator for the asymptotic variance of MWLS. The result of Theorem 4(iii) continues to hold if $\widehat{\beta}_n(\widehat{\gamma}_n)$ is replaced by any estimator $\widetilde{\beta}_n = \beta_0 + o_p(1)$.

It is however important to note that two very familiar results are missing from Theorem 4. First, there is no result on the asymptotic distribution of $\widehat{\beta}_n(\widehat{\gamma}_n)$. Second, there is no result on the asymptotic equivalence for $\widehat{h}_n(\widehat{\gamma}_n)$, i.e., no result like $\sqrt{n}(\widehat{h}_n(\widehat{\gamma}_n) - \widehat{h}_n(\gamma)) = o_p(1)$ for some fixed γ.

Let us quickly recall why both these familiar results are missing. A possibly non-singleton Γ^* means that $\widehat{\gamma}_n$ may not converge in probability to any given point in Γ^* but can only come close in probability to the set Γ^*. Now, while $\Sigma(\gamma) = \Sigma^*$ for $\gamma \in \Gamma^*$ by the definition in (18) and $D(\gamma) = 0$ for $\gamma \in \Gamma$ by the mean independence of the regression error u in (10), others like $B(\gamma)$, $C(\gamma)$ and hence $\Xi(\gamma)$ may not be constant for $\gamma \in \Gamma^*$. Hence, although $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$ as in Theorem 4(i), in general, $\widehat{\beta}_n(\widehat{\gamma}_n)$ does not possess the conventional distributional properties. Again, because of the general non-convergence of $\widehat{\gamma}_n$ to any given point in Γ^*, it is not possible to find a reference point γ such that an asymptotic equivalence for $\widehat{h}_n(\widehat{\gamma}_n)$ like $\sqrt{n}(\widehat{h}_n(\widehat{\gamma}_n) - \widehat{h}_n(\gamma)) = o_p(1)$ holds.

Nevertheless, the good news is that since $\widehat{\gamma}_n$ comes close in probability at a desired rate to the set Γ^* and since $\Sigma(\gamma) = \Sigma^*$ for $\gamma \in \Gamma^*$, Theorem 4(ii) can still show that $\widehat{h}_n(\widehat{\gamma}_n)$ is asymptotically unbiased and normal with optimal variance Σ^*. Therefore, appealing to Theorem 4(iii), one can test hypotheses on $h(\beta)$ and construct confidence intervals for $h(\beta)$ in the usual way as follows.

3.3.3 Wald-inference based on the MWLS estimator of $h(\beta)$

One can reject the null hypothesis $K_{\text{Null}} : h(\beta) = h_{\text{Null}}$ against the alternative $K_{\text{Alt}} : h(\beta) \neq h_{\text{Null}}$ at the level $\alpha \in (0, 1)$ by a Wald test if:

$$|T_n(h_{\text{Null}})| > z_{1-\alpha/2} \quad \text{where} \quad T_n(h_{\text{Null}}) := \frac{\widehat{h}_n(\widehat{\gamma}_n) - h_{\text{Null}}}{\sqrt{\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n))/n}} \tag{20}$$

is the usual t-ratio, z_c is c-th quantile ($c \in (0, 1)$) of the $N(0, 1)$ distribution ($\Phi(z_c) = c$ where $\Phi(.)$ is the distribution function of $N(0, 1)$), and $\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n))$ is any of the four HC estimators for Σ^*.

Similarly, one can construct a Wald confidence interval with nominal asymptotic level $100(1 - \alpha)\%$ for $h(\beta)$ as:

$$CI_n = \left[\widehat{h}_n(\widehat{\gamma}_n) - z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n))/n}, \quad \widehat{h}_n(\widehat{\gamma}_n) + z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_n(\widehat{\gamma}_n; \widehat{\beta}_n(\widehat{\gamma}_n))/n} \right]. \tag{21}$$

[11] Appendix B contains descriptive endnotes related to Theorem 4. Appendix B.1 sketches the standard steps giving the expansion in assumption (e). Appendix B.2 discusses the cost of the compactness of Γ to us; also see footnote 3 of Romano and Wolf (2017). Appendix B.3 discusses the possibly non-singleton Γ^* and its relation to the literature.

One could similarly test against one-sided alternatives and construct one-sided confidence intervals.

Corollary 5 *The following results hold under the conditions of Theorem 4 (ii)-(iii).*

(i) If $h_0 = h_{Null} + \mu/\sqrt{n}$ for some finite scalar μ, then the probability with which the Wald test defined in (20) rejects the null hypothesis $K_{Null} : h(\beta) = h_{Null}$ against the alternative hypothesis $K_{Alt} : h(\beta) \neq h_{Null}$ converges to $\Phi(z_{\alpha/2} - \mu/\sqrt{\Sigma^*}) + \Phi(z_{\alpha/2} + \mu/\sqrt{\Sigma^*})$ as $n \to \infty$.

(ii) The true asymptotic level of the Wald confidence interval CI_n defined in (21) that is obtained by inverting the Wald test in (20) is $100(1-\alpha)\%$.

In the Monte Carlo experiment in Section 4, we conduct estimation and inference based on the MWLS principle by exactly following the description in Section 3.3.1 and Section 3.3.3 respectively.

3.4 What if the parameter of interest $h(\beta)$ is a vector?

$h(\beta)$ is a scalar in most empirical studies since interest generally lies in the regression coefficients, their sums, differences, etc. individually. Empirical research generally reports estimators, standard errors, t-ratios, etc., i.e., quantities that correspond to distinct but inherently scalar $h(\beta)$'s (e.g., each coefficient is a distinct scalar $h(\beta)$ in a regression output table). In light of this observation, one can always report distinct scalar $h(\beta)$-specific MWLS estimators, standard errors, t-ratios, etc.

However, while these scalar $h(\beta)$-specific estimators are individually optimal in the sense of (15), they are generally not useful for joint-inference if one wishes, e.g., to test that these distinct $h(\beta)$'s are jointly zero (say, $\beta_1 = \beta_2 = 0$), or to obtain joint confidence sets for these distinct $h(\beta)$'s. Therefore, let us now consider a vector $h(\beta)$ so that MWLS can be used for optimal joint-inference.

If the parameter of interest $h(\beta)$ is a vector, e.g., $h(\beta) = \beta$, then an "optimal" γ that minimizes the asymptotic variance of $h(\widehat{\beta}_n(\gamma))$ for $\gamma \in \Gamma$ (c.f. (15)) would be defined as $\gamma^* \in \Gamma$ such that:

$$\Sigma(\gamma) - \Sigma(\gamma^*) \text{ is positive semi-definite for all } \gamma \in \Gamma. \tag{22}$$

However, unlike in the scalar case, such a γ^* may not exist unless $\omega^2(X;\gamma)$ is correctly specified (even if Γ is compact in $\mathbb{R}^k$ and $\Sigma(\gamma)$ is continuous in $\gamma \in \Gamma$). Therefore, to talk about the MWLS principle here, one would first need to make the assumption that there exists a γ^* satisfying (22).

As before, γ^* may not be unique, i.e., there may exist $\gamma \in \Gamma$ such that $\gamma \neq \gamma^*$ but $\Sigma(\gamma) = \Sigma(\gamma^*) =: \Sigma^*$. Moreover, a direct optimization in the sense of (22) is generally difficult to operationalize. One could, however, take into account of the non-uniqueness of γ^* in (22) and at the same time try to circumvent the difficulty of the direct optimization by defining Γ^* (c.f. (18)) as:

$$\Gamma^* = \left\{ \gamma \in \Gamma \,\middle|\, \mathrm{Trace}(\Sigma(\gamma)) - \min_{\bar{\gamma} \in \Gamma} \mathrm{Trace}(\Sigma(\bar{\gamma})) = 0 \right\}. \tag{23}$$

Γ^* is nonempty if $\mathrm{Trace}(\Sigma(\gamma))$ is continuous in $\gamma \in \Gamma$ and Γ is compact in $\mathbb{R}^k$. Γ^* is the set of "optimal" γ's in the sense of (22) provided that a γ^* satisfying (22) exists. Hence, the difficult direct optimization problem in (22) reduces to a simple problem of minimizing the trace of $\Sigma(\gamma)$. (See Lemma 6 in Appendix B.4 for the details concerning the equivalence of these two problems.)

Now, the MWLS estimator of $h(\beta)$ can be obtained by following the algorithm in Section 3.3.1 except that the estimation of γ in Step 1 needs to be done by minimizing the trace of a sample analog of $\Sigma(\gamma)$ with respect to $\gamma \in \Gamma$. The asymptotic properties of the MWLS estimator of $h(\beta)$ do not depend on whether $h(\beta)$ is a scalar or a vector and, with obvious changes to accommodate that $H(\beta) := \partial h(\beta)/\partial \beta'$ is now a matrix and not a vector, our previous analysis also applies here.

There is one caveat. Unlike in the case of a scalar $h(\beta)$, it is now difficult to provide useful sufficient conditions for the existence of a γ^* defined in (22). ($\gamma^* = \gamma_0$ exists if $\omega^2(X;\gamma)$ is correct as in (5). Existence is an issue only if $\omega^2(X;\gamma)$ is wrong.) If a γ^* does not exist, then the set Γ^* defined in (23) is simply the set of the so-called A-optimal γ's in the terminology of the design of experiments literature; see Elfving (1952) and Chernoff (1953). The notion of A-optimality is not going to be attractive in empirical work unless backed by the existence of γ^* that elevates it to the more robust notion of optimality in (22). Moreover, without the existence of a γ^*, the generally non-singleton set Γ^* would form a level set only for $\mathrm{Trace}(\Sigma(\gamma))$ but not necessarily a level set for $\Sigma(\gamma)$. Hence, the asymptotic properties of the MWLS estimator of $h(\beta)$, which requires estimating the unknown optimal γ by minimizing the trace of a sample analog of $\Sigma(\gamma)$, could be invalid.[12]

4 Monte Carlo experiment

We will now conduct a Monte Carlo experiment to investigate how well the asymptotic results discussed so far are reflected in finite samples, and how well or poorly estimation and inference based on MWLS perform relative to those based on OLS, WLS, and the other related methods proposed in the recently published papers by Romano and Wolf (2017) and DiCiccio et al. (2019).

Since our paper is inspired by Romano and Wolf (2017), we follow their simulation design here.

[12]The above discussion is also applicable to the other notions of optimality such as the D-optimality of Wald (1943), the E-optimality of Ehrenfeld (1956), the L-optimality (generalization of A-optimality) due to Karlin and Studden (1966) and Federov (1971), Kiefer (1974)'s general optimality of which the A/L, D and E optimalities are limiting cases, etc. The empirical attractiveness of the interpretability of these notions of optimality are also dependent on the existence of an optimal γ^*, and they may not agree unless $h(\beta)$ is a scalar or, more generally, unless γ^* exists.

4.1 Simulation design and implementation

Romano and Wolf (2017)'s design takes $p = 2$ in (1): more precisely, $y = \beta_1 X_{(1)} + \beta_2 X_{(2)} + u$, with $X_{(1)} = 1$, $X = (X_{(1)}, X_{(2)})'$, $\beta = (\beta_1, \beta_2)'$, $\beta_0 = (0, 0)'$, and the error $u = \omega_0(X)Z$ where $Z \sim N(0, 1)$ is independent of $X_{(2)} \sim U(1, 4)$, and considers the following $4 + 2 + 2 + 2 = 10$ cases:

DGP 1: (a) $\omega_0^2(X) = 1$; (b) $\omega_0^2(X) = X_{(2)}$; (c) $\omega_0^2(X) = X_{(2)}^2$; (d) $\omega_0^2(X) = X_{(2)}^4$.

DGP 2: (a) $\omega_0^2(X) = \big(\log(X_{(2)})\big)^2$; (b) $\omega_0^2(X) = \big(\log(X_{(2)})\big)^4$.

DGP 3: (a) $\omega_0^2(X) = \exp\Big(.1(X_{(2)} + X_{(2)}^2)\Big)$; (b) $\omega_0^2(X) = \exp\Big(.15(X_{(2)} + X_{(2)}^2)\Big)$.

DGP 4: (a) $\omega_0^2(X) = \begin{cases} 1 & \text{if } X_{(2)} < 2 \\ 2 & \text{if } 2 \le X_{(2)} < 3 \\ 3 & \text{if } X_{(2)} \ge 3 \end{cases}$; (b) $\omega_0^2(X) = \begin{cases} 1 & \text{if } X_{(2)} < 2 \\ 2^2 & \text{if } 2 \le X_{(2)} < 3 \\ 3^2 & \text{if } X_{(2)} \ge 3 \end{cases}$.

We consider two choices for the parameter of interest $h(\beta)$ — β_1 and β_2 — separately. We consider six types of estimation methods discussed in Section 2.1 and 2.2: (i) OLS, (ii) WLS, (iii) ALS, (iv) MINVAR, (v) LINCOM and (vi) MWLS. Following Romano and Wolf (2017), we consider two parametric models $\omega^2(X; \gamma)$ for which we report all the simulation results separately:

$$\text{Model 1: } \omega^2(X; \gamma) := \exp(\gamma_1 + \gamma_2 \log(X_{(2)})) \quad \text{and} \quad \text{Model 2: } \omega^2(X; \gamma) := \exp(\gamma_1 + \gamma_2 X_{(2)}).$$

Model 1 is correct for $\omega_0^2(X)$ with $\gamma_2 = 0, 1, 2, 4$ under DGPs 1(a)-1(d) respectively. Model 2 is correct with $\gamma_2 = 0$ for $\omega_0^2(X)$ under DGP 1(a) (i.e., conditional homoskedasticity). Hence, we recall from Section 2 that: (1) all methods (i)-(vi) reach the efficiency bound under DGP 1(a) with both Models 1 and 2; and (2) methods (ii)-(vi) with Model 1 also reach the efficiency bound under DGP 1 (b)-(d). Both Models 1 and 2 are, however, misspecified for $\omega_0^2(X)$ under the other DGPs.

The ALS, MINVAR and LINCOM estimators depend on the WLS estimator. We compute the WLS estimator as in Romano and Wolf (2017) using the same truncation parameter as they did. We also follow Romano and Wolf (2017) to compute their ALS estimator. We follow DiCiccio et al. (2019) and use the HC3 version of the variance estimator to compute their MINVAR and LINCOM estimators. (DiCiccio et al. (2019) call MINVAR and LINCOM as Min and Optimal respectively.)

We compute the MWLS estimator by exactly following Steps 1 and 2 of the algorithm in Section 3.3.1. We tried all four HC versions of the variance estimators that were minimized to estimate the γ for the MWLS estimator. For brevity, we report here the results for the MWLS estimator computed based only on the HC1 and HC3 versions that are commonly used in practice. These two versions of MWLS are referred to as MWLS-HC1 and MWLS-HC3 respectively in the sequel.

Asymptotic variance of all estimators is estimated using all the HC versions. However, since the HC3 version is generally preferred (see MacKinnon (2012)) and was also used by Romano and Wolf (2017) and DiCiccio et al. (2019), we report these results and those on inference only for HC3.

4.2 Simulation results

For all the 10 DGPs of Romano and Wolf (2017), for each parameter of interest β_1 and β_2, for each estimator, for each of Model 1 and Model 2, and for samples sizes $n = 25, 50, 100, 200$ and 400, we report based on 25,000 Monte Carlo trials: (a) the empirical bias in Tables 7-8; (b) the ratio of empirical mean squared error (MSE) with respect to that of MWLS-HC1 in Tables 9-10; (c) the Monte Carlo variance (MCVar) in Tables 11-12; (d) the average of the variance based on the asymptotic variance formula (ASVar), i.e., the estimated HC3 variance divided by n and averaged over 25,000 Monte Carlo trials, in Tables 13-14; (e) the empirical size (null-rejection probability) of two-sided 5% Wald tests using t_{n-2} critical value as in Romano and Wolf (2017) in Tables 15-16; (f) the empirical size of two-sided 5% Wald tests using the asymptotic $N(0, 1)$ critical value in Tables 17-18; and (g) the empirical size-corrected power of the two-sided 5% Wald tests in Figures 10-44.

We do not report ASVar and the empirical size for LINCOM in these tables because its ASVar is misleadingly small and, as a result, the empirical size for the LINCOM-based test is always close to 100%. This is not a new observation and, as we noted earlier in footnote 8, DiCiccio et al. (2019) recommend to always use inference based on bootstrap to avoid this problem for LINCOM.

Our complete set of results involves too many tables and figures. Hence, we take the following route to give the reader a complete picture of the results while keeping the presentation brief.

- Since the qualitative results in most of these tables and figures are similar, we discuss the main takeaway message in Section 4.2.1 and in a supplemental appendix we collect all these tables and figures, i.e., Tables 7–18 and Figures 10-44, containing the complete set of results for all the DGPs. We hope that the descriptive table of contents in the supplemental appendix will help the reader to navigate through these numerous tables and figures without much difficulty.
- In Section 4.2.2 we report in Tables 1-5 and Figures 1-9 the results for DGP 2 where the benefit of using MWLS is the most prominent. We focus only on Model 2 for brevity since the results are similar for Models 1 and 2 and since it appears to us that, compared to Model 1, the use of Model 2 is more prevalent in economics. To corroborate DiCiccio et al. (2019) and follow up on footnote 8, we include a column in Tables 3 and 6 for LINCOM's ASVar, but do not delve on it as DiCiccio et al. (2019)'s proposal of using bootstrap already solves this issue.

4.2.1 The main takeaway message from the complete set of simulation results

First, as seen from Tables 7 and 8, the bias of all estimators is small and there is no clear winner.[13]

Second, consider the ratio of the MSE of OLS, WLS, ALS, MINVAR, LINCOM and MWLS-HC3 with respect to that of MWLS-HC1 for estimating β_1 and β_2 as reported in Tables 9 and 10. There is ample evidence to prefer the other estimators over OLS in terms of MSE. Since both Models 1 and 2 allow for conditional homoskedasticity (with $\gamma_2 = 0$), the MSEs of all these estimators under DGP 1(a) are essentially the same as that of OLS when the sample size is 100 or more. WLS, ALS, MINVAR and LINCOM perform very similarly, which we also observe for aspects other than MSE. The MSEs of MWLS-HC1 and MWLS-HC3 are always competitive with that of the others. Under DGP 2, MWLS-HC1 and MWLS-HC3 are preferable by a big margin to all the other estimators.

Third, Tables 11-12 and Tables 13-14 respectively report the MCVar and ASVar of all the estimators. Our observations related to MSE also apply to the case of MCVar and ASVar. MCVar is generally a more reliable (but infeasible) measure of the true variability of estimators. It turns out here that ASVar, which is based on a feasible formula, is also similar to the corresponding MCVar. In fact ASVar and MCVar are essentially equal here when the sample size is 100 or more.

Fourth, consider the empirical size of the two-sided 5% Wald test based on all the estimators. Tables 15-16 and Tables 17-18 report in percentage the empirical size when the critical value used is, respectively, from the t_{n-2} distribution as in Romano and Wolf (2017) and from $N(0,1)$ as in the standard first-order asymptotics. There is negligible over-rejection of the truth when $h(\beta) = \beta_2$ is the parameter of interest. This is true even when the sample size is very small, e.g., $n = 25, 50$. In fact, when $h(\beta) = \beta_2$, the maximum empirical size for MWLS-HC1 or MWLS-HC3 based on Models 1 and 2 respectively is 6.4% and 6.1% with t_{n-2} critical value, and 7.5% and 7.2% with $N(0,1)$ critical value. On the other hand, when $h(\beta) = \beta_1$ is the parameter of interest, MWLS based on Model 1 (and to a lesser extent, Model 2) can lead to empirical size as large as 10.3% when the sample size is 25. The problem of over-rejection vanishes quickly as sample size increases.

Fifth, consider the empirical size-corrected power of these two-sided Wald tests. These power curves for testing β_1 and β_2 respectively are plotted in Figures 10-14 and Figures 15-19 under DGP 1, in Figures 20-24 and Figures 25-29 under DGP 2, in Figures 30-34 and Figures 35-39 under DGP 3, and in Figures 40-44 and Figures 45-49 under DGP 4. In the case of conditional homoskedasticity,

[13] The results reported here are very similar to those in the last version Chaudhuri (2019). The difference is in the presentation. Here, we: (i) used the HC3 instead of the HC0 version for inference, (ii) reported as precision gain the reduction in MSE instead of MCVar or ASVar, (iii) compared OLS, WLS, ALS, MINVAR, LINCOM, MWLS-HC1 and MWLS-HC3 instead of OLS, WLS, ALS and MWLS-HC0, and (iv) reported empirical bias and empirical power.

i.e., DGP 1(a), the power of the test based on OLS is almost indistinguishable from that of the tests based on the other estimators. On the other hand, as the degree of heteroskedasticity increases under each DGP ((a) is less than (b) and so on), the powers of the tests based on the other estimators start to dominate that of the OLS-based test, often by a big margin. With the increase in the sample size, according to which the figures are arranged, we observe two interesting phenomena: (1) the powers of the two MWLS-based tests become indistinguishable, and (2) the power of MWLS is either very similar to or dominates, sometimes by a big margin, the power of all its competitors.

4.2.2 The benefit of MWLS is the most prominent under DGP 2

Finally, we wish to highlight here the magnitude of the gain in precision and its quantitative implications on inference resulting from the use of MWLS. DGP 2 is where this gain is the most prominent. While this gain is already prominent under DGP 2(b), to illustrate it even more prominently we also consider an extension of DGP 2, namely, DGP 2(c) where $\omega_0^2(X) = \left(\log(X_{(2)})\right)^6$.

The results are very similar for Models 1 and 2, and, for brevity, we only report the results for Model 2 here since the use of Model 2 seems more prevalent in economics.[14]

We observe the following for the empirical MSE. First, Figure 1 indicates that the MSE of OLS can be enormous compared to that of MWLS-HC1. Second, Figure 2 indicates that the MSE can be reduced a lot by accounting for heteroskedasticity as do WLS, ALS, MINVAR and LINCOM. Nevertheless, the MSE of these estimators also can still be very much (even twice) bigger than that of MWLS-HC1. Third, supporting their asymptotic equivalence, Figure 2 also indicates that the difference between MWLS-HC3 and MWLS-HC1 vanishes, albeit slowly, as sample size increases.

We observe the following for the empirical size-corrected power in Figures 5-9. First, the equivalence of MWLS-HC1 and MWLS-HC3 continues to hold if the sample size is not too small. Second, and importantly, there is huge gains in power over OLS, WLS, ALS, MINVAR and LINCOM when we use MWLS. Given that MWLS does not much over-reject the truth here (see Figures 3-4), the evidence from Figures 5-9 makes a compelling case for the use of MWLS since the goal of inference is to reject the false hypotheses as frequently as possible controlling for the rejection of the truth.

Overall, we find that: (1) MWLS competes well in all aspects under all these DGPs, and (2) if a misspecified model $\omega^2(X;\gamma)$ leaves room for precision gains then MWLS can be very appealing because, by virtue of its construction, MWLS is best equipped to deliver that gain in precision.

[14] Let us quickly note the similarities in the simulation results under DGP 2 and the other DGPs. First, there is no noticeable difference between WLS, ALS, MINVAR and LINCOM. Second, under DGP 2 also, the empirical bias is small (see Tables 1 and 4), MCVar and ASVar are quite similar especially when the sample size is not too small (see Tables 2-3 and 5-6), and the empirical size is not too upward biased (see Figures 3-4), etc. for all the estimators.

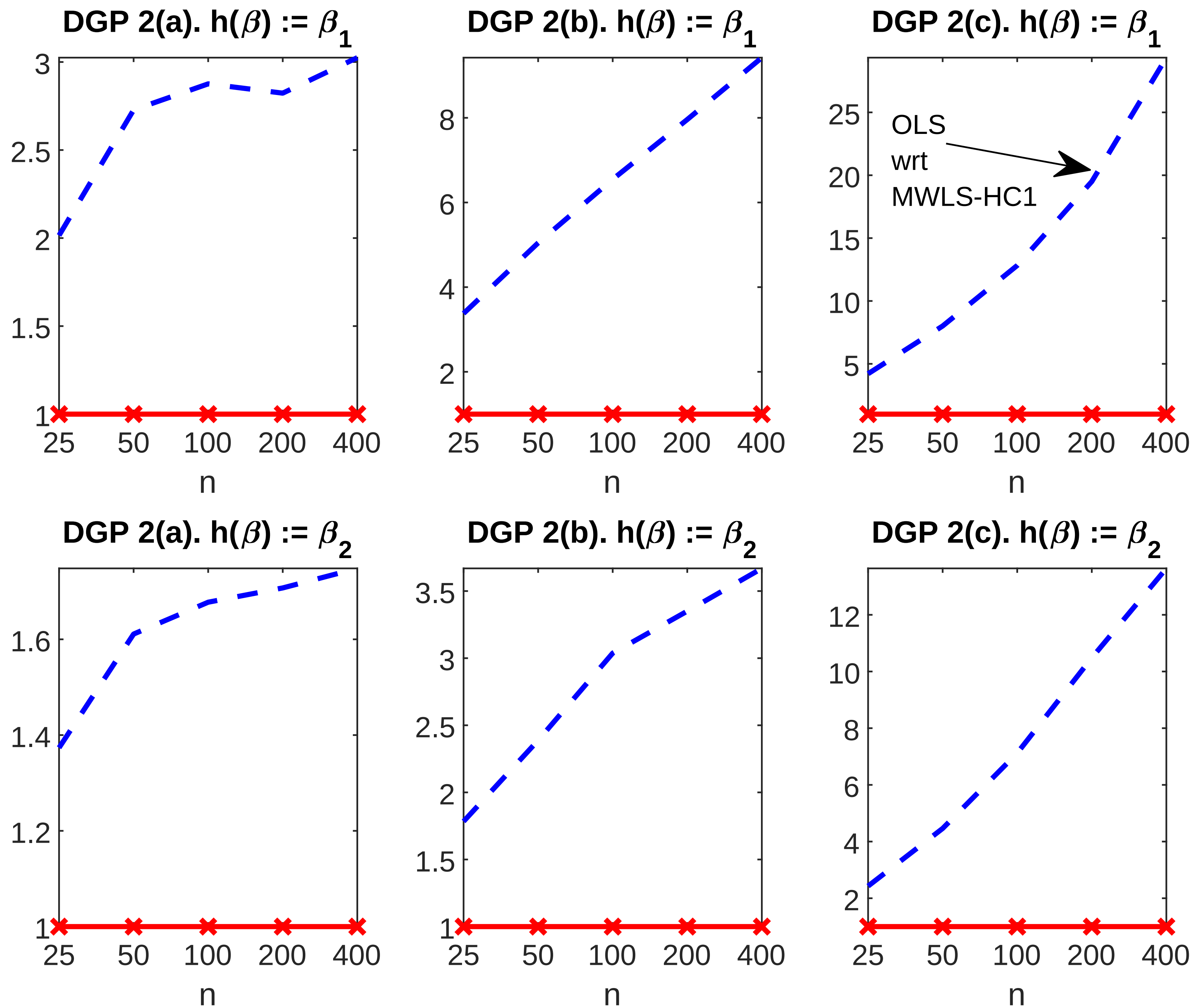

Figure 1: DGP 2 and Model 2: The blue (- -) line is the ratio of MSE of OLS with respect to (wrt) MWLS-HC1 plotted against sample size n. The red horizontal (x-) line is drawn at 1 for convenience of readers and to emphasize that the MSE of OLS is indeed large relative to that of MWLS-HC1.

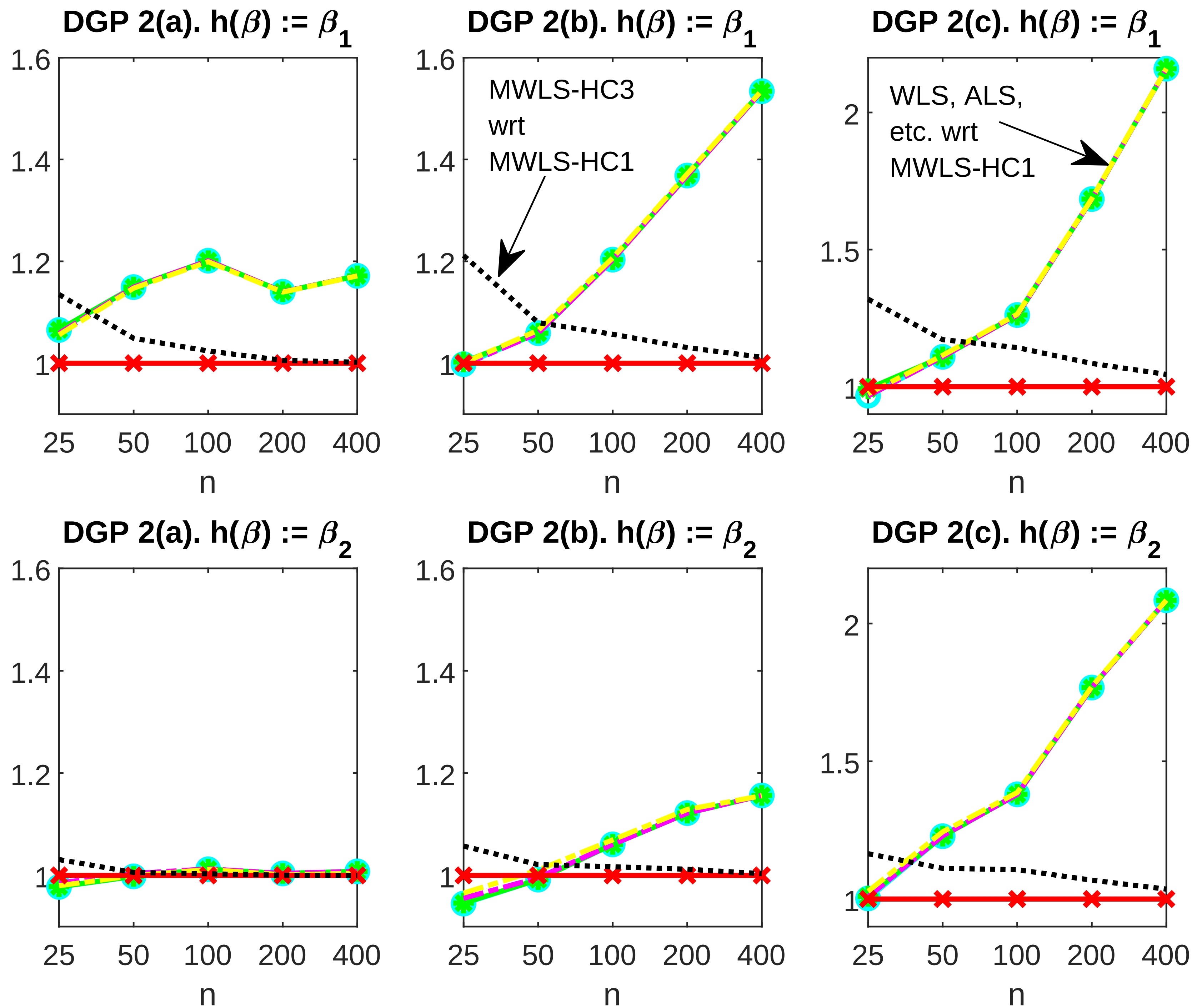

Figure 2: DGP 2 and Model 2: Ratio of MSE of estimators with respect to (wrt) MWLS-HC1 against sample size n. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC3: black (..) line. The red (x-) line is drawn at 1 for convenience of readers and to emphasize that the MSEs of these other estimators are indeed large relative to that of MWLS-HC1.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
	25	-.0008	.0009	.0009	.0010	.0011	.0015	.0013
	50	.0007	-.0006	-.0006	-.0006	-.0005	-.0011	-.0010
2(a)	100	.0001	.0007	.0007	.0007	.0007	.0006	.0006
	200	-.0004	-.0002	-.0002	-.0002	-.0002	.0001	.0000
	400	-.0001	-.0002	-.0002	-.0002	-.0002	-.0001	-.0001
	25	-.0054	-.0035	-.0034	-.0035	-.0036	-.0037	-.0041
	50	.0001	-.0001	-.0001	-.0001	-.0001	.0001	.0001
2(b)	100	-.0013	-.0007	-.0007	-.0007	-.0007	-.0004	-.0005
	200	-.0008	.0001	.0001	.0001	.0001	.0004	.0003
	400	.0008	.0005	.0005	.0005	.0005	.0001	.0001
	25	-.0025	-.0011	-.0012	-.0011	-.0013	-.0013	-.0014
	50	-.0007	-.0002	-.0002	-.0002	-.0002	-.0002	-.0003
2(c)	100	.0011	.0008	.0008	.0008	.0008	.0008	.0008
	200	-.0003	.0001	.0001	.0001	.0001	-.0002	-.0001
	400	.0002	-.0004	-.0004	-.0004	-.0004	-.0002	-.0002

Table 1: Estimated bias of estimators for $h(\beta) := \beta_1$ based on Model 2 for $\omega^2(X;\gamma)$.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
	25	.1591	.0841	.0841	.0835	.0833	.0789	.0896
	50	.0808	.0341	.0341	.0341	.0340	.0296	.0311
2(a)	100	.0478	.0200	.0200	.0200	.0200	.0166	.0170
	200	.0234	.0094	.0094	.0094	.0094	.0083	.0083
	400	.0128	.0050	.0050	.0050	.0050	.0042	.0042
	25	.2067	.0611	.0613	.0610	.0612	.0612	.0741
	50	.1067	.0224	.0224	.0224	.0225	.0212	.0228
2(b)	100	.0643	.0118	.0118	.0118	.0118	.0098	.0104
	200	.0320	.0055	.0055	.0055	.0055	.0040	.0041
	400	.0172	.0028	.0028	.0028	.0028	.0018	.0018
	25	.3389	.0778	.0798	.0777	.0784	.0804	.1061
	50	.1810	.0251	.0251	.0251	.0252	.0226	.0265
2(c)	100	.1049	.0103	.0103	.0103	.0103	.0082	.0093
	200	.0526	.0045	.0045	.0045	.0045	.0027	.0029
	400	.0293	.0022	.0022	.0022	.0022	.0010	.0010

Table 2: Estimated variance (MCVar) of estimators for $h(\beta) := \beta_1$ based on Model 2 for $\omega^2(X;\gamma)$.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
	25	.2192	.1060	.1060	.1058	.0002	.1034	.1022
	50	.0931	.0349	.0349	.0349	1.4×10^{-5}	.0298	.0303
2(a)	100	.0512	.0201	.0201	.0201	2.0×10^{-6}	.0167	.0168
	200	.0241	.0095	.0095	.0095	2.4×10^{-7}	.0084	.0084
	400	.0128	.0049	.0049	.0049	3.1×10^{-8}	.0043	.0043
	25	.2754	.0667	.0667	.0667	.0001	.0649	.0702
	50	.1244	.0227	.0227	.0227	9.0×10^{-6}	.0198	.0208
2(b)	100	.0690	.0117	.0117	.0117	1.7×10^{-6}	.0093	.0096
	200	.0330	.0056	.0056	.0056	1.4×10^{-7}	.0040	.0042
	400	.0177	.0028	.0028	.0028	1.7×10^{-8}	.0019	.0019
	25	.4479	.0787	.0787	.0787	.0001	.0751	.0900
	50	.2056	.0236	.0236	.0236	9.4×10^{-6}	.0178	.0204
2(c)	100	.1135	.0102	.0102	.0102	1.0×10^{-6}	.0070	.0078
	200	.0544	.0045	.0045	.0045	1.2×10^{-7}	.0025	.0026
	400	.0295	.0022	.0022	.0022	1.4×10^{-8}	.0010	.0010

Table 3: Estimated variance (ASVar) of estimators for $h(\beta) := \beta_1$ based on Model 2 for $\omega^2(X;\gamma)$.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
2(a)	25	.0019	.0022	.0022	.0022	.0021	.0022	.0022
	50	.0002	.0000	.0000	.0001	.0001	.0003	.0003
	100	-.0005	-.0007	-.0007	-.0006	-.0006	-.0006	-.0006
	200	-.0003	-.0001	-.0001	-.0001	-.0001	.0000	.0000
	400	.0001	.0001	.0001	.0001	.0002	.0001	.0001
2(b)	25	-.0018	-.0008	-.0008	-.0009	-.0009	-.0010	-.0011
	50	.0010	.0009	.0009	.0009	.0009	.0011	.0010
	100	-.0004	.0001	.0001	.0001	.0001	.0000	-.0001
	200	.0009	.0008	.0008	.0008	.0008	.0007	.0006
	400	.0009	.0000	.0000	.0000	.0000	-.0002	-.0002
2(c)	25	.0026	.0010	.0010	.0010	.0010	.0006	.0008
	50	.0010	.0007	.0007	.0007	.0007	.0006	.0006
	100	.0001	.0004	.0004	.0004	.0004	.0004	.0005
	200	-.0008	-.0004	-.0004	-.0004	-.0004	-.0004	-.0004
	400	-.0001	-.0002	-.0002	-.0002	-.0002	-.0003	-.0003

Table 4: Estimated bias of estimators for $h(\beta) := \beta_2$ based on Model 2 for $\omega^2(X;\gamma)$.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
2(a)	25	.0323	.0230	.0230	.0231	.0230	.0235	.0242
	50	.0185	.0115	.0115	.0115	.0115	.0115	.0116
	100	.0110	.0066	.0066	.0066	.0066	.0065	.0066
	200	.0058	.0034	.0034	.0034	.0034	.0034	.0034
	400	.0030	.0017	.0017	.0017	.0017	.0017	.0017
2(b)	25	.0523	.0277	.0277	.0280	.0283	.0293	.0310
	50	.0293	.0122	.0122	.0122	.0124	.0123	.0125
	100	.0174	.0061	.0061	.0061	.0061	.0057	.0058
	200	.0093	.0031	.0031	.0031	.0031	.0028	.0028
	400	.0049	.0016	.0016	.0016	.0016	.0013	.0013
2(c)	25	.0897	.0370	.0373	.0372	.0380	.0369	.0430
	50	.0503	.0138	.0138	.0139	.0140	.0113	.0125
	100	.0298	.0058	.0058	.0058	.0058	.0042	.0046
	200	.0154	.0026	.0026	.0026	.0026	.0015	.0016
	400	.0082	.0013	.0013	.0013	.0013	.0006	.0006

Table 5: Estimated variance (MCVar) of estimators for $h(\beta) := \beta_2$ based on Model 2 for $\omega^2(X;\gamma)$.

DGP	n	OLS	WLS	ALS	MINVAR	LINCOM	MWLS-HC1	MWLS-HC3
2(a)	25	.0422	.0276	.0276	.0275	4.3×10^{-5}	.0265	.0266
	50	.0210	.0123	.0123	.0122	4.8×10^{-6}	.0118	.0118
	100	.0119	.0068	.0068	.0068	6.8×10^{-7}	.0065	.0065
	200	.0060	.0035	.0035	.0035	8.6×10^{-8}	.0034	.0034
	400	.0031	.0018	.0018	.0018	1.1×10^{-8}	.0017	.0017
2(b)	25	.0672	.0319	.0319	.0317	4.9×10^{-5}	.0303	.0311
	50	.0333	.0129	.0129	.0129	5.0×10^{-6}	.0119	.0122
	100	.0186	.0063	.0063	.0063	6.2×10^{-7}	.0057	.0058
	200	.0095	.0031	.0031	.0031	7.7×10^{-8}	.0027	.0028
	400	.0050	.0015	.0015	.0015	9.5×10^{-9}	.0013	.0013
2(c)	25	.1168	.0402	.0402	.0402	6.3×10^{-5}	.0364	.0405
	50	.0571	.0139	.0139	.0139	5.5×10^{-6}	.0102	.0112
	100	.0316	.0058	.0058	.0058	5.8×10^{-7}	.0039	.0043
	200	.0159	.0026	.0026	.0026	6.5×10^{-8}	.0015	.0015
	400	.0084	.0013	.0013	.0013	7.8×10^{-9}	.0006	.0006

Table 6: Estimated variance (ASVar) of estimators for $h(\beta) := \beta_2$ based on Model 2 for $\omega^2(X;\gamma)$.

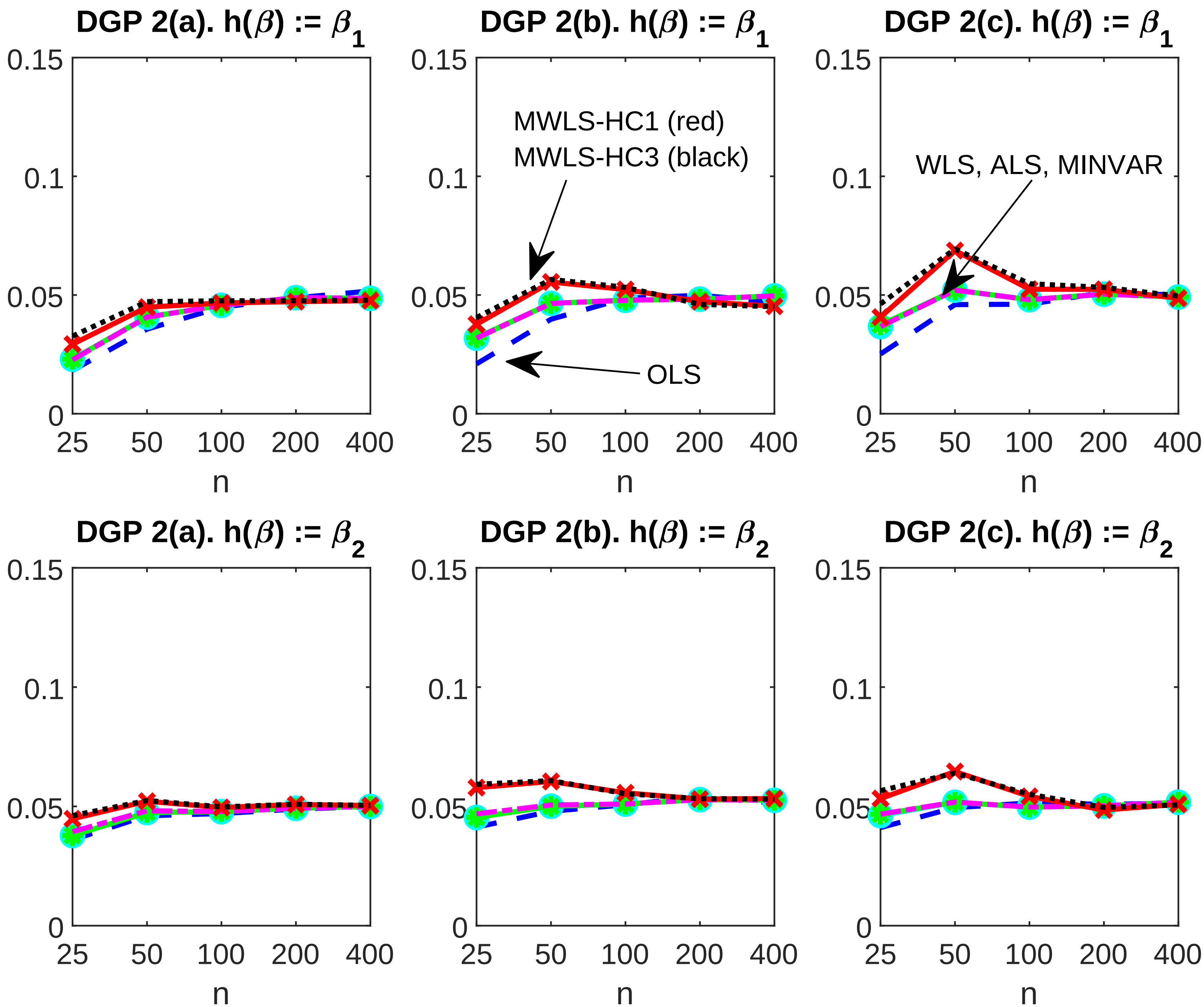

Figure 3: DGP 2 and Model 2: Empirical size of two-sided 5% Wald test plotted against sample size n. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line. These Wald tests here use critical value from the t_{n-2} distribution following Romano and Wolf (2017).

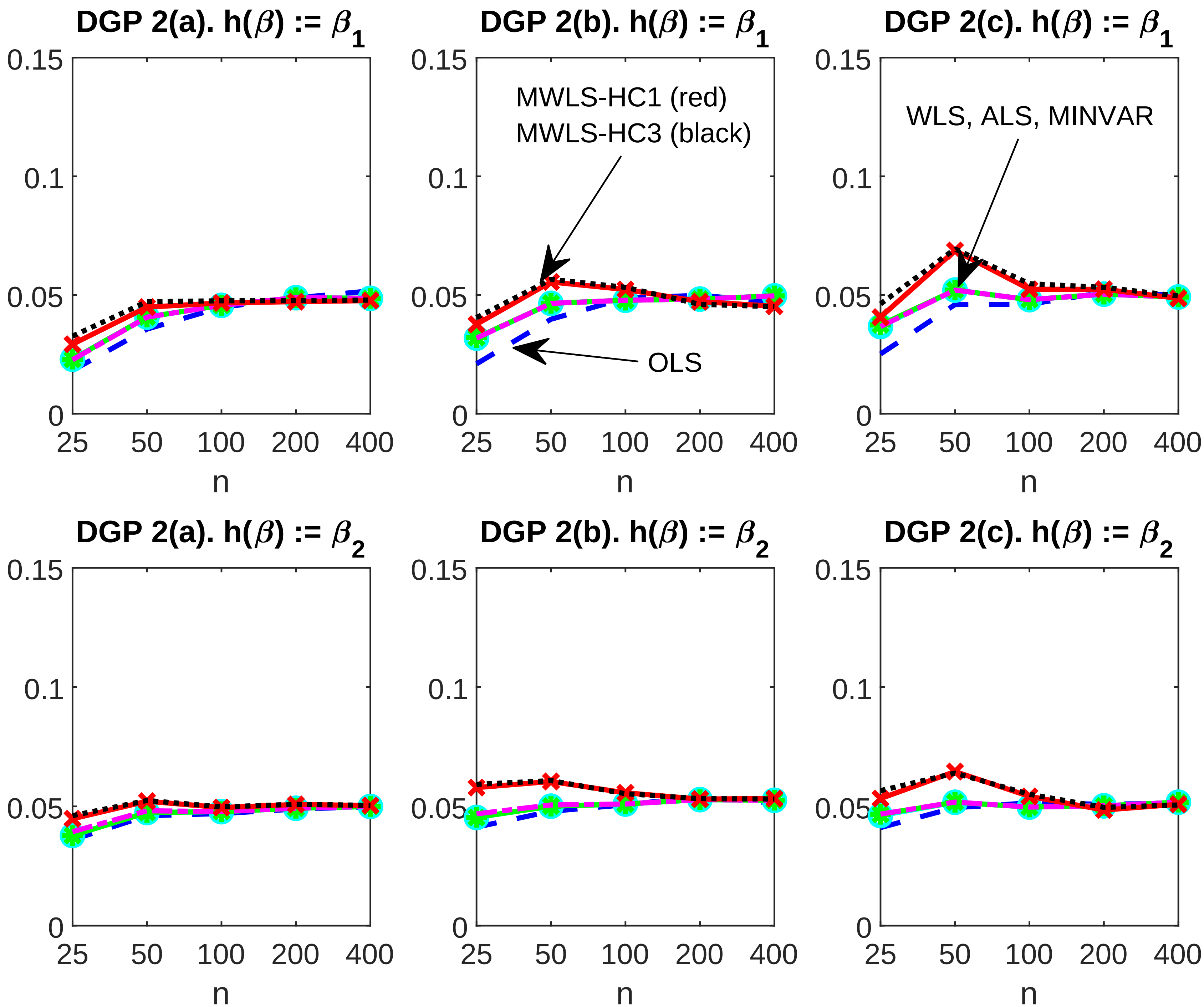

Figure 4: DGP 2 and Model 2: Empirical size of two-sided 5% Wald test plotted against sample size n. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line. These Wald tests here use critical value from the $N(0, 1)$ distribution as in (20) of our paper.

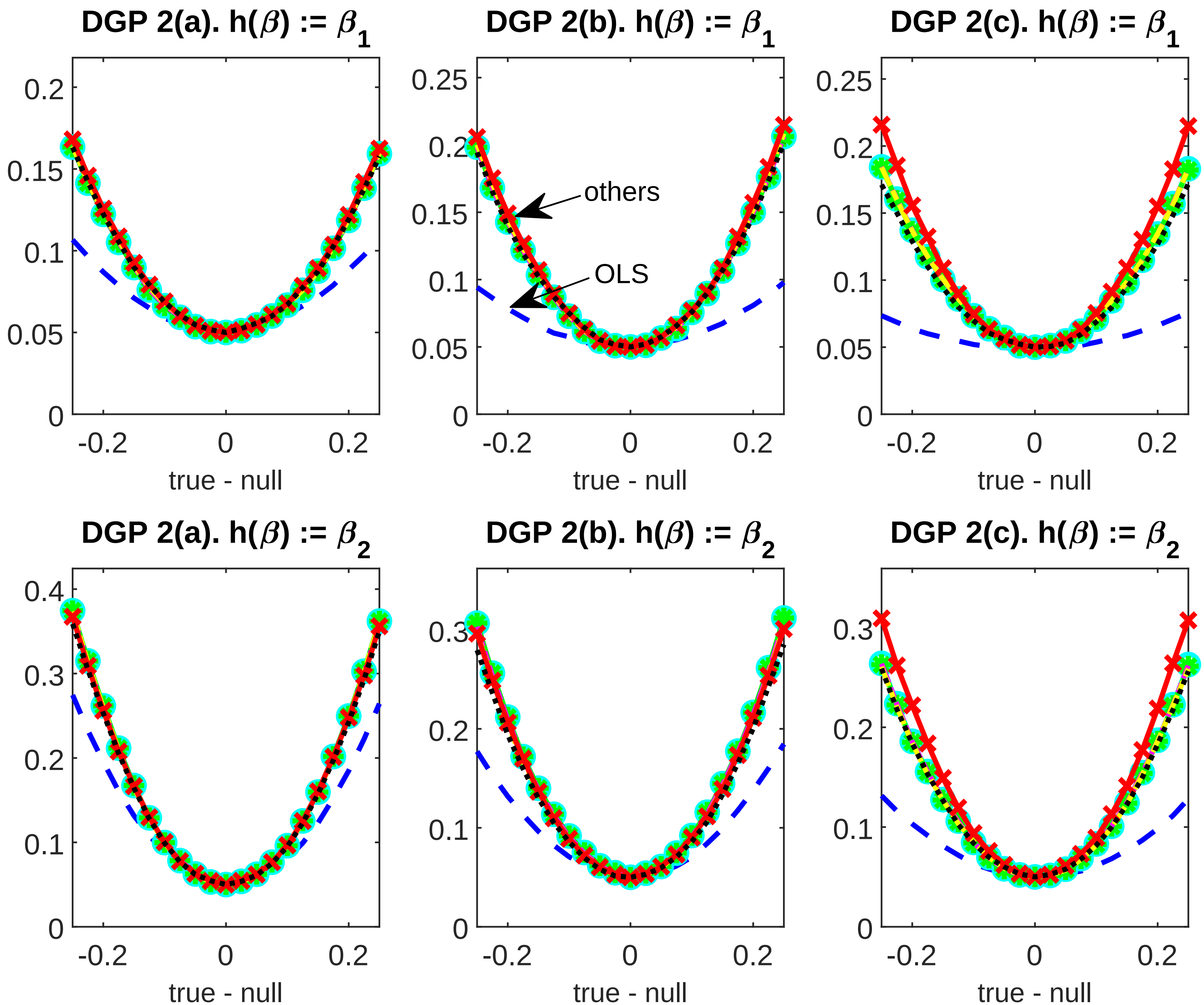

Figure 5: DGP 2 and Model 2: Empirical size-corrected power of two-sided 5% Wald test plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

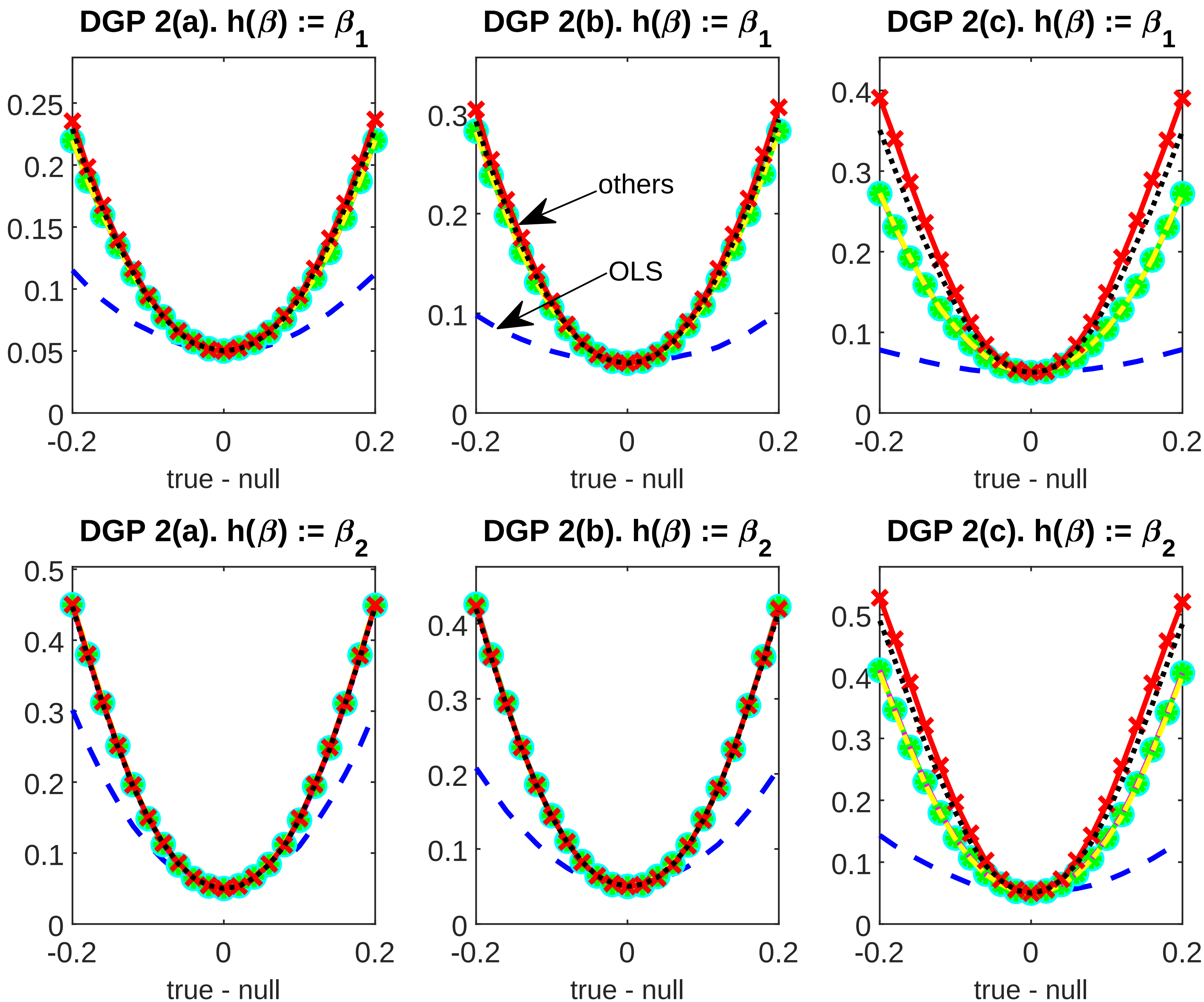

Figure 6: DGP 2 and Model 2: Empirical size-corrected power of two-sided 5% Wald test plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

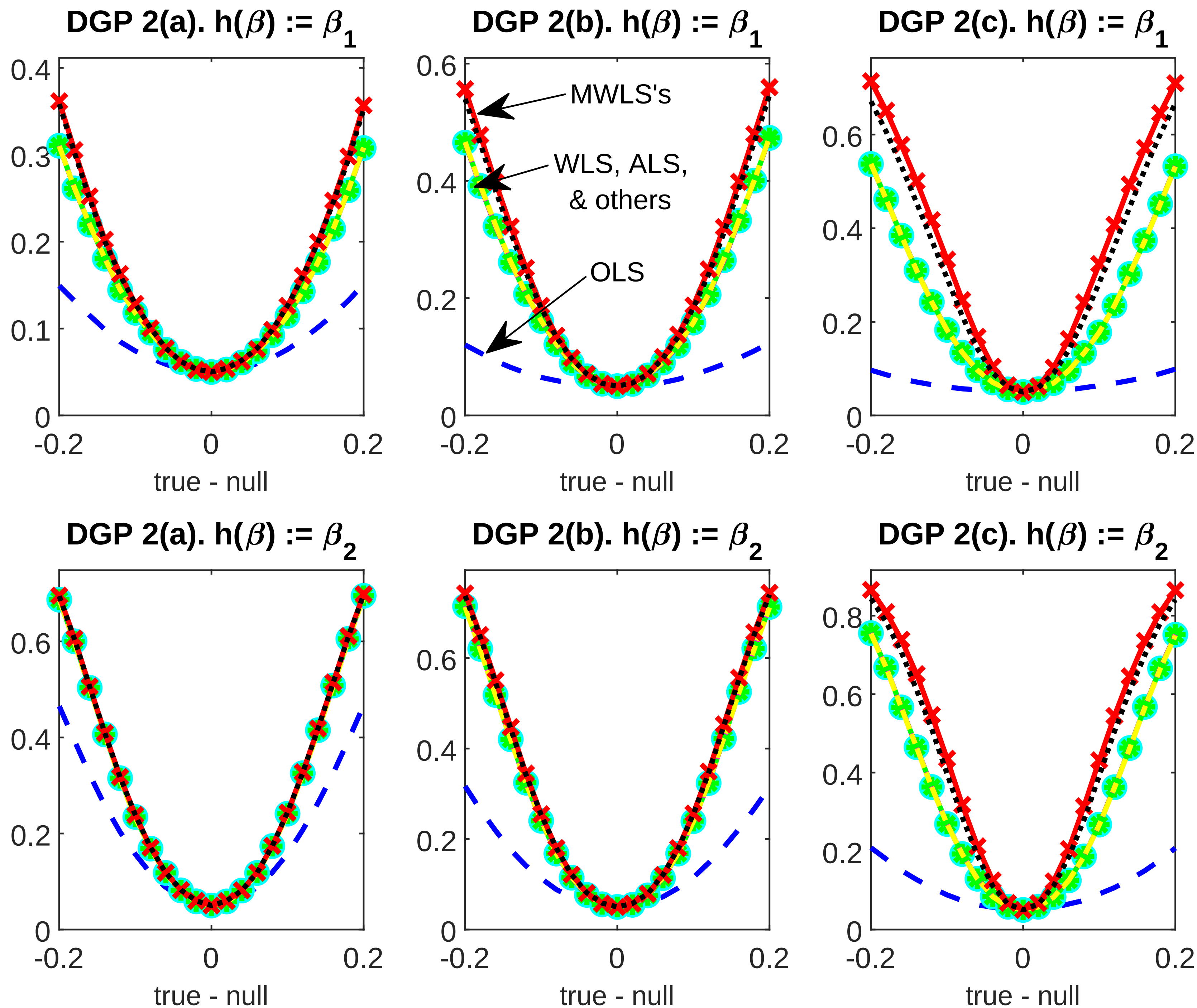

Figure 7: DGP 2 and Model 2: Empirical size-corrected power of two-sided 5% Wald test plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

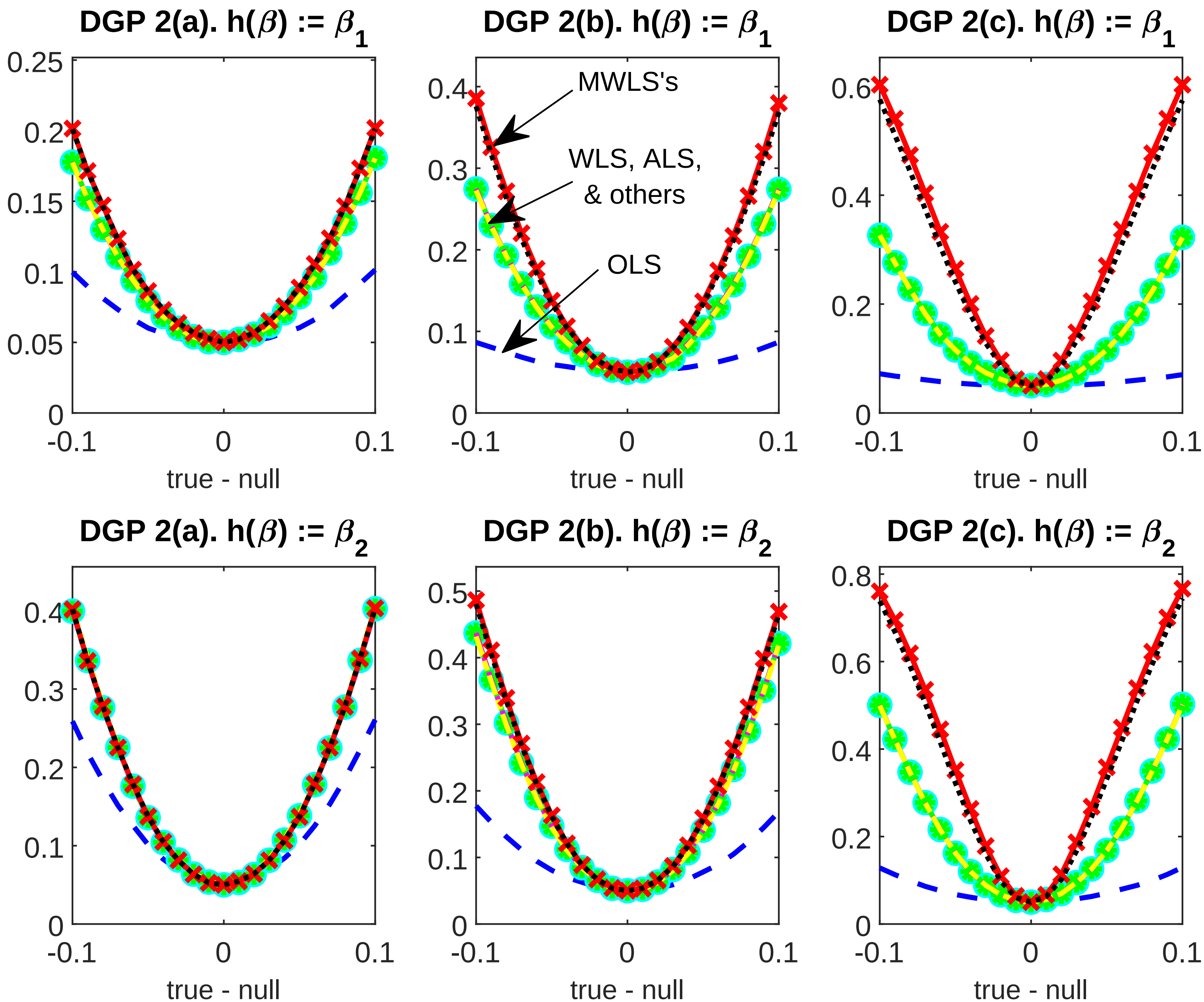

Figure 8: DGP 2 and Model 2: Empirical size-corrected power of two-sided 5% Wald test plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

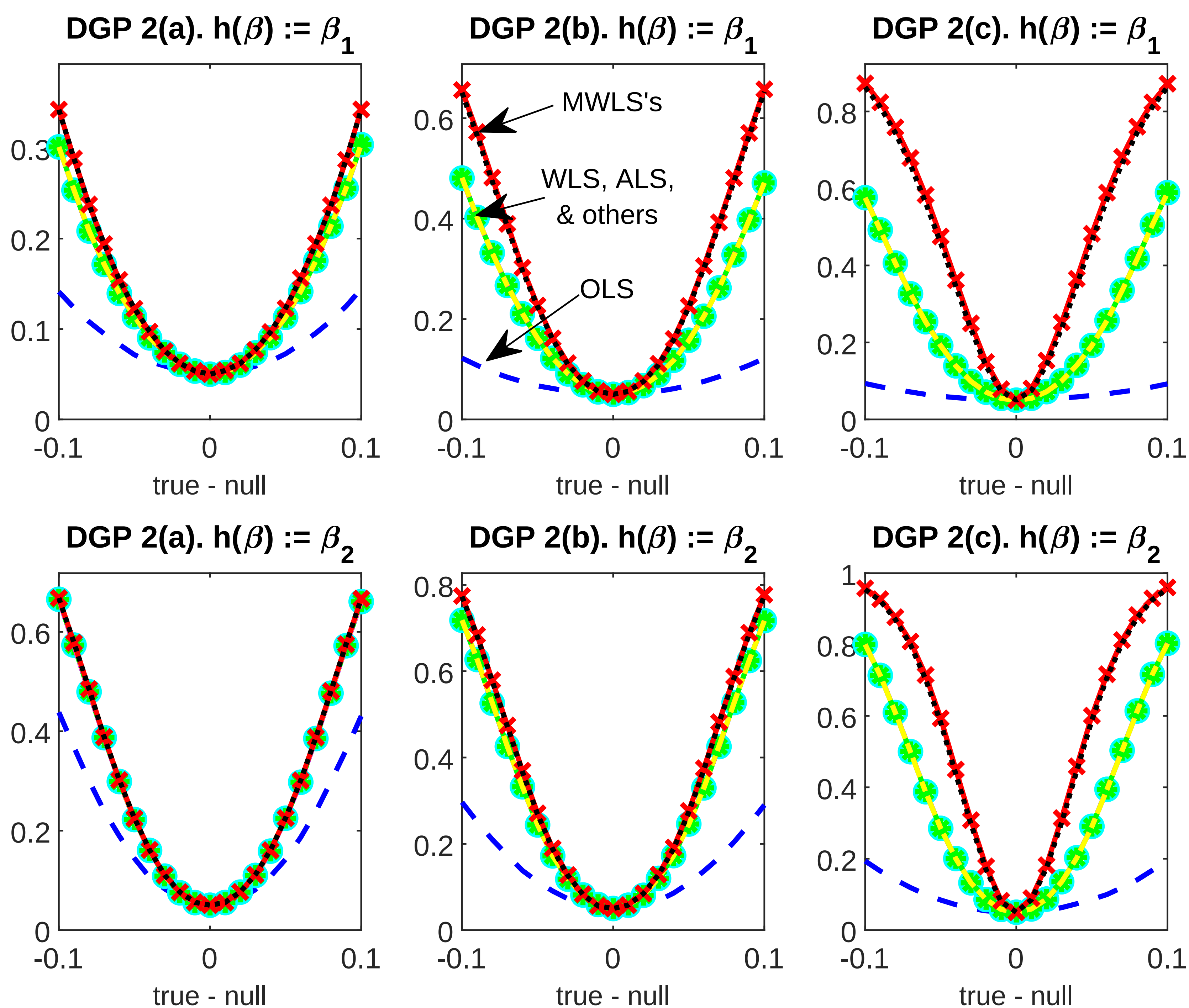

Figure 9: DGP 2 and Model 2: Empirical size-corrected power of two-sided 5% Wald test plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

5 Conclusion

Our paper was inspired by Romano and Wolf (2017) who challenged the conventional wisdom of preferring OLS estimation with robust standard errors over (parametric) WLS estimation. OLS estimation has always been attractive because, unlike WLS, it does not require the user to postulate parametric models for the conditional variance of the regression error. Hence, WLS estimation naturally fell out of favor among practitioners following the seminal work of White (1980).

Practitioners could only be encouraged to use WLS if, along with the well-known efficiency gains that result from WLS in some cases, it can also be guaranteed that WLS will never suffer any efficiency loss relative to OLS in other cases. However, this is not possible. It is well known that the asymptotic variance of WLS can be more than that of OLS if the postulated parametric model for the conditional variance is misspecified. Unfortunately, in spite of its desirable performance otherwise, the ALS estimator proposed by Romano and Wolf (2017) did not solve this problem. The MINVAR and LINCOM estimators in DiCiccio et al. (2019) solved this problem but perhaps at the expense of intensive computation. Therefore, the question remained: Can we do better?

Our paper showed that we can do better by taking a direct route to optimality. We called this the MWLS principle. Given the user's model $\omega^2(X;\gamma)$ — correct or wrong, sophisticated or naive — the asymptotic variance of MWLS cannot exceed that of OLS, WLS, all the related newly proposed estimators (that we know of), and any weighted-by-$\omega^{-1}(X;\gamma)$ LS estimator. MWLS achieves the efficiency bound if, luckily, the user's model $\omega^2(X;\gamma)$ is correct. We provided a thorough discussion of MWLS and its superior asymptotic properties. A Monte Carlo experiment demonstrated the excellent performance of MWLS even in reasonably small samples under the simulation design of Romano and Wolf (2017) where MWLS delivered truly noteworthy gains in precision and power.

We conclude by emphasizing again that the key feature that we exploited in proposing the MWLS estimator is that: A parametric misspecification of the nuisance parameters in this estimation framework does not affect the consistency of the estimator for the parameters of interest. This feature gave us the freedom to tailor the definition of the "optimal" nuisance parameters according to our objective for the estimation of the parameter of interest. While of course this key feature was always well known, it seems that the related literature never exploited this feature in our way. There are various other frameworks that also enjoy or "nearly" enjoy this feature. The same idea should, in principle, be applicable there. An exploration of this idea under these other frameworks and the search for optimality even in the higher-order properties are left for our ongoing research.

A Appendix A: Proofs for the results in Section 3

Proof of Lemma 1:

(i) The proof of Theorem 5.7 in van der Vaart (1998) directly applies to give $\widehat{\beta}_n(\gamma) = \beta_0 + o_p(1)$. Then $\widehat{h}_n(\gamma) = h_0 + o_p(1)$ follows by assumption (c) using the continuous mapping theorem.

(ii) The well-understood high level assumptions in the statement of the lemma directly lead to the desired results from the expansion of the first order condition of the sample minimization problem once we note that assumption (d) and the fact that $(y, X_i')_{i=1}^n$ are i.i.d. copies of (y, X') imply that: $\sqrt{n}\bar{\psi}_{\beta,n}(\beta_0,\gamma) \xrightarrow{d} N(0, \Xi(\gamma))$ where $\Xi(\gamma) := B^{-1}(\gamma)C(\gamma)B^{-1}(\gamma)$. ∎

Proof of Lemma 2: Although well known (see, e.g., Hansen (2020)), we write a detailed proof here to emphasize the asymptotic equivalence of MWLS based on the different HC versions in spite of their differences in very small samples. We begin by noting that $\widehat{\beta} \in \{\beta : |\beta - \beta_0| < \epsilon\}$ with probability approaching 1, and write the rest of the proof conditional on this event. Assumptions (b), (d) and consistency of $\widehat{\beta}$ imply that $\sup_{\gamma\in\Gamma} \|\widehat{B}_n(\gamma;\widehat{\beta}) - B(\gamma)\| = o_p(1)$. Similarly, assumptions (e) and consistency of $\widehat{\beta}$ imply that $H(\widehat{\beta}) - H = o_p(1)$. So, the key for the proof is to show that for the HC0-HC3 versions of $\widehat{C}_n(\gamma;\beta)$ we have $\sup_{\gamma\in\Gamma} \|\widehat{C}_n^{(j)}(\gamma;\widehat{\beta}) - C(\gamma)\| = o_p(1)$ for $j = 0, 1, 2, 3$.

First, consider the HC0 version $\widehat{C}_n^{(0)}(\gamma;\widehat{\beta})$ of $\widehat{C}_n(\gamma;\beta)$ and note that by the triangle inequality:

$$\sup_{\gamma\in\Gamma}\left\|\widehat{C}_n^{(0)}(\gamma;\widehat{\beta}) - C(\gamma)\right\| \le \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^n \widetilde{u}_i^2(\gamma)\widetilde{G}_i'(\gamma)\widetilde{G}_i(\gamma) - C(\gamma)\right\| + T_{1,n} + T_{2,n} \tag{24}$$

where we write $\widetilde{u}_i(\beta_0,\gamma)$ and $\widetilde{G}_i(\beta_0,\gamma)$ as $\widetilde{u}_i(\gamma)$ and $\widetilde{G}_i(\gamma)$ respectively for brevity, and where:

$$T_{1,n} = \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^n(\widetilde{u}_i^2(\widehat{\beta},\gamma) - \widetilde{u}_i^2(\gamma))\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)\right\|,\ T_{2,n} = \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^n \widetilde{u}_i^2(\gamma)(\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma) - \widetilde{G}_i'(\gamma)\widetilde{G}_i(\gamma))\right\|.$$

Appealing to condition that $\widehat{\beta} = \beta_0 + o_p(1)$, we will use the following expressions obtained by mean-value expansion (row-by-row) repeatedly, with mean-values generically denoted by $\bar{\beta}$, $\bar{\bar{\beta}}$, etc.

$$\begin{aligned}
\widetilde{G}_i(\widehat{\beta},\gamma) &= \widetilde{G}_i(\beta_0,\gamma) + (\widehat{\beta} - \beta_0)'\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right) \text{ and} && (25)\\
\widetilde{u}_i^2(\widehat{\beta},\gamma) - \widetilde{u}_i^2(\gamma) &= \frac{1}{\omega^2(X_i;\gamma)}\left[\left(g(X_i;\widehat{\beta}) - g(X_i;\beta_0)\right)^2 - 2u_i\left(g(X_i;\widehat{\beta}) - g(X_i;\beta_0)\right)\right]\\
&= \frac{1}{\omega^2(X_i;\gamma)}\left[(\widehat{\beta} - \beta_0)'G'(X_i;\bar{\bar{\beta}})G(X_i;\bar{\bar{\beta}})(\widehat{\beta} - \beta_0) - 2u_iG(X_i;\bar{\beta})(\widehat{\beta} - \beta_0)\right]\\
&= (\widehat{\beta} - \beta_0)'\widetilde{G}_i'(\bar{\bar{\beta}},\gamma)\widetilde{G}_i(\bar{\bar{\beta}},\gamma)(\widehat{\beta} - \beta_0) - 2\widetilde{u}_i(\gamma)\widetilde{G}_i(\bar{\beta},\gamma)(\widehat{\beta} - \beta_0). && (26)
\end{aligned}$$

Since expansion of the terms in $T_{2,n}$ is less messy let us consider it first. Using (25), we can write:

$$
\begin{aligned}
T_{2,n} &= \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_i^2(\gamma)(\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)-\widetilde{G}_i'(\gamma)\widetilde{G}_i(\gamma))\right\| \\
&= \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_i^2(\gamma)\left[\left\{\widetilde{G}_i(\beta_0,\gamma)+(\widehat{\beta}-\beta_0)'\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)\right\}'\right.\right. \\
&\qquad\qquad \left.\left.\times\left\{\widetilde{G}_i(\beta_0,\gamma)+(\widehat{\beta}-\beta_0)'\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)\right\}-\widetilde{G}_i'(\gamma)\widetilde{G}_i(\gamma)\right]\right\| \\
&= \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_i^2(\gamma)\left[\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)'(\widehat{\beta}-\beta_0)\widetilde{G}_i(\beta_0,\gamma)+\widetilde{G}_i'(\beta_0,\gamma)(\widehat{\beta}-\beta_0)'\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)\right.\right. \\
&\qquad\qquad \left.\left.+\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)'(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)'\left(\partial\widetilde{G}_i(\bar{\beta},\gamma)/\partial\beta\right)\right]\right\| \\
&\le \frac{2}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\Gamma}\widetilde{u}_i^2(\gamma)\right)\left(\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|\right)\left(\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|\right)\|\widehat{\beta}-\beta_0\| \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\Gamma}\widetilde{u}_i^2(\gamma)\right)\left(\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^2\right)\|\widehat{\beta}-\beta_0\|^2 \\
&\le 2\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|^4\right)^{1/4}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4\right)^{1/4} \\
&\quad \times\|\widehat{\beta}-\beta_0\| \\
&\quad +\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4\right)^{1/2}\|\widehat{\beta}-\beta_0\|^2 \\
&= 2\left(E^{1/2}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{1/4}[\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|^4]+o_p(1)\right) \\
&\qquad\qquad \times\left(E^{1/4}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\| \\
&\quad +\left(E^{1/2}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{1/2}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\|^2 \\
&= O_p(\|\widehat{\beta}-\beta_0\|)
\end{aligned}
$$

by assumptions (a), (b) and (c), the triangle and Cauchy-Schwartz/Holder inequalities, and the law of large numbers. We wrote $E^{1/l}[.]$ to denote $(E[.])^{1/l}$ for $l=2,4$ in the second last equality.

Now, we consider $T_{1,n}$ and by using (26) and the triangle inequality we obtain that:

$$
\begin{aligned}
T_{1,n} &= \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}(\widetilde{u}_i^2(\widehat{\beta},\gamma)-\widetilde{u}_i^2(\gamma))\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)\right\| \\
&\le \sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}(\widehat{\beta}-\beta_0)'\widetilde{G}_i'(\bar{\bar{\beta}},\gamma)\widetilde{G}_i(\bar{\bar{\beta}},\gamma)(\widehat{\beta}-\beta_0)\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)\right\| \\
&\qquad +2\sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_i(\gamma)\widetilde{G}_i(\bar{\bar{\beta}},\gamma)(\widehat{\beta}-\beta_0)\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)\right\| \\
&= T_{11,n}+T_{12,n}
\end{aligned}
$$

where $T_{11,n}$ and $T_{12,n}$ respectively denote the first term and the second term after the inequality above. We will need to further use the mean-value expansion in (25) separately on $\widetilde{G}_i(\widehat{\beta},\gamma)$ and $\widetilde{G}_i(\bar{\bar{\beta}},\gamma)$ with mean-value denoted by $\bar{\beta}$ and $\tilde{\beta}$ respectively. Since the expansions contain too many terms we will temporarily ignore γ in the notation and use the following notation for brevity:

$$a_i := \widetilde{u}_i(\gamma), \quad b_i := \widetilde{G}_i(\beta_0;\gamma), \quad \text{and} \quad d_i(\beta) := \frac{\partial}{\partial\beta}\widetilde{G}_i(\beta;\gamma).$$

Consider $T_{11,n}$. By the appropriate mean-value expansions as noted above, we obtain that:

$$\begin{aligned}
& (\widehat{\beta}-\beta_0)'\widetilde{G}_i'(\bar{\bar{\beta}},\gamma)\widetilde{G}_i(\bar{\bar{\beta}},\gamma)(\widehat{\beta}-\beta_0)\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma) \\
= \; & (\widehat{\beta}-\beta_0)'\left(b_i(\gamma)+(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})\right)'\left(b_i+(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})\right)(\widehat{\beta}-\beta_0) \\
& \qquad \times\left(b_i+(\widehat{\beta}-\beta_0)'d_i(\bar{\beta})\right)'\left(b_i+(\widehat{\beta}-\beta_0)'d_i(\bar{\beta})\right) \\
= \; & (\widehat{\beta}-\beta_0)'\Big[b_i'b_i(\widehat{\beta}-\beta_0)b_i'b_i+b_i'b_i(\widehat{\beta}-\beta_0)b_i'(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +b_i'b_i(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)b_i \\
& \qquad +b_i'b_i(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +b_i'(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)b_i'b_i \\
& \qquad +b_i'(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)b_i'(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +b_i'(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)b_i \\
& \qquad +b_i'(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)b_i(\widehat{\beta}-\beta_0)b_i'b_i \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)b_i(\widehat{\beta}-\beta_0)b_i'(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)b_i(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)b_i \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)b_i(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)b_i'b_i \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)b_i'(\widehat{\beta}-\beta_0)'d_i(\bar{\beta}) \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)b_i \\
& \qquad +d_i'(\tilde{\beta})(\bar{\bar{\beta}}-\beta_0)(\bar{\bar{\beta}}-\beta_0)'d_i(\tilde{\beta})(\widehat{\beta}-\beta_0)d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)'d_i(\bar{\beta})\Big].
\end{aligned}$$

Therefore, reverting to the original notation, recalling that the definition of the mean values implies that $\|\bar{\beta}-\beta_0\|\leq\|\widehat{\beta}-\beta_0\|$ and $\|\tilde{\beta}-\beta_0\|\leq\|\bar{\bar{\beta}}-\beta_0\|\leq\|\widehat{\beta}-\beta_0\|=o_p(1)$, grouping the similar terms in the last equation together[15], and following similar steps as were used to deal with $T_{2,n}$, we obtain that:

$$
\begin{aligned}
& T_{11,n} \\
\leq\; & \left(E[\sup_{\gamma\in\Gamma} \|\widetilde{G}_i(\gamma)\|^4] + o_p(1) \right) \|\widehat{\beta}-\beta_0\|^2 \\
& +4 \left(E^{3/4}[\sup_{\gamma\in\Gamma} \|\widetilde{G}_i(\gamma)\|^4] + o_p(1) \right) \left(E^{1/4}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma} \|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4] + o_p(1) \right) \|\widehat{\beta}-\beta_0\|^3 \\
& +6 \left(E^{1/2}[\sup_{\gamma\in\Gamma} \|\widetilde{G}_i(\gamma)\|^4] + o_p(1) \right) \left(E^{1/2}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma} \|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4] + o_p(1) \right) \|\widehat{\beta}-\beta_0\|^4 \\
& +4 \left(E^{1/4}[\sup_{\gamma\in\Gamma} \|\widetilde{G}_i(\gamma)\|^4] + o_p(1) \right) \left(E^{3/4}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma} \|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4] + o_p(1) \right) \|\widehat{\beta}-\beta_0\|^5 \\
& + \left(E[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma} \|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4] + o_p(1) \right) \|\widehat{\beta}-\beta_0\|^6 \\
=\; & O_p(\|\widehat{\beta}-\beta_0\|^2).
\end{aligned}
$$

Now for $T_{12,n}$ we will follow the same approach as for $T_{11,n}$. To do this, we note that:

$$
\begin{aligned}
& \widetilde{u}_i(\gamma)\widetilde{G}_i(\bar{\bar{\beta}},\gamma)(\widehat{\beta}-\beta_0)\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma) \\
=\; & a_i \left(b_i + (\bar{\bar{\beta}}-\beta_0)' d_i(\tilde{\beta}) \right) (\widehat{\beta}-\beta_0) \left(b_i + (\widehat{\beta}-\beta_0)' d_i(\bar{\beta}) \right)' \left(b_i + (\widehat{\beta}-\beta_0)' d_i(\bar{\beta}) \right) \\
=\; & a_i b_i (\widehat{\beta}-\beta_0) b_i' b_i + a_i b_i (\widehat{\beta}-\beta_0) b_i' (\widehat{\beta}-\beta_0)' d_i(\bar{\beta}) \\
& + a_i b_i (\widehat{\beta}-\beta_0) d_i'(\bar{\beta})(\widehat{\beta}-\beta_0) b_i \\
& + a_i b_i (\widehat{\beta}-\beta_0) d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)' d_i(\bar{\beta}) \\
& + a_i (\bar{\bar{\beta}}-\beta_0)' d_i(\tilde{\beta})(\widehat{\beta}-\beta_0) b_i' b_i \\
& + a_i (\bar{\bar{\beta}}-\beta_0)' d_i(\tilde{\beta})(\widehat{\beta}-\beta_0) b_i' (\widehat{\beta}-\beta_0)' d_i(\bar{\beta}) \\
& + a_i (\bar{\bar{\beta}}-\beta_0)' d_i(\tilde{\beta})(\widehat{\beta}-\beta_0) d_i'(\bar{\beta})(\widehat{\beta}-\beta_0) b_i \\
& + a_i (\bar{\bar{\beta}}-\beta_0)' d_i(\tilde{\beta})(\widehat{\beta}-\beta_0) d_i'(\bar{\beta})(\widehat{\beta}-\beta_0)(\widehat{\beta}-\beta_0)' d_i(\bar{\beta}).
\end{aligned}
$$

Therefore, reverting to the original notation, recalling that the definition of the mean values implies that $\|\bar{\beta}-\beta_0\| \leq \|\widehat{\beta}-\beta_0\|$ and $\|\tilde{\beta}-\beta_0\| \leq \|\bar{\bar{\beta}}-\beta_0\| \leq \|\widehat{\beta}-\beta_0\| = o_p(1)$, grouping the similar terms in the last equation together[16], and following similar steps as were used to deal with $T_{11,n}$, we obtain

[15] Terms (1), (2, 3, 5 and 9), (4, 6, 7, 10, 11 and 13), (8, 12, 14 and 15) and (16) from the right hand side of the last equation are represented respectively in terms 1, 2, 3, 4 and 5 on the right hand side of the current equation.

[16] Terms (1), (2, 3 and 5), (4, 6 and 7) and (8) from the right hand side of the last equation are represented respectively in terms 1, 2, 3 and 4 on the right hand side of the current equation.

that:

$$\begin{aligned}
T_{12,n} &\leq 2\left(E^{1/4}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{3/4}[\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\| \\
&\quad +6\left(E^{1/4}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{1/2}[\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|^4]+o_p(1)\right) \\
&\qquad \times\left(E^{1/4}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\|^2 \\
&\quad +6\left(E^{1/4}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{1/4}[\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\gamma)\|^4]+o_p(1)\right) \\
&\qquad \times\left(E^{1/2}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\|^3 \\
&\quad +\left(E^{1/4}[\sup_{\gamma\in\Gamma}\widetilde{u}_i^4(\gamma)]+o_p(1)\right)\left(E^{3/4}[\sup_{\beta\in\mathcal{B}:\|\beta-\beta_0\|<\epsilon,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^4]+o_p(1)\right)\|\widehat{\beta}-\beta_0\|^4 \\
&= O_p(\|\widehat{\beta}-\beta_0\|).
\end{aligned}$$

Therefore, $T_{1,n}\leq T_{11,n}+T_{12,n}=O_p(\|\widehat{\beta}-\beta_0\|)$ and hence $T_{1,n}+T_{2,n}=O_p(\|\widehat{\beta}-\beta_0\|)$.

Finally, noting that: $\sup_{\gamma\in\Gamma}\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{u}_i^2(\gamma)\widetilde{G}_i'(\gamma)\widetilde{G}_i(\gamma)-C(\gamma)\right\|=o_p(1)$ by assumptions (a) and (b), it follows from (24) that for the HC0 version: $\sup_{\gamma\in\Gamma}\left\|\widehat{C}_n^{(0)}(\gamma;\widehat{\beta})-C(\gamma)\right\|=o_p(1)$.

Now, since the HC1 version $\widehat{C}_n^{(1)}(\gamma;\widehat{\beta})$ is simply $n/(n-p)$ times the HC0 version of $\widehat{C}_n^{(0)}(\gamma;\widehat{\beta})$, and since p is finite, it follows that for the HC1 version as well: $\sup_{\gamma\in\Gamma}\left\|\widehat{C}_n^{(1)}(\gamma;\widehat{\beta})-C(\gamma)\right\|=o_p(1)$.

Finally, to show the same result for the HC2 and HC3 versions, note that the differences between the HC2 and HC0 versions and HC3 and HC0 versions respectively for the key term $\widehat{C}_n(\gamma;\widehat{\beta})$ are:

$$\begin{aligned}
\sup_{\gamma\in\Gamma}\|\widehat{C}_n^{(2)}(\gamma;\widehat{\beta})-\widehat{C}_n^{(0)}(\gamma;\widehat{\beta})\| &\leq \frac{1}{n}\sum_{i=1}^n\sup_{\gamma\in\Gamma}\left\|\widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)\frac{q_{i,n}(\gamma;\beta)}{1-q_{i,n}(\gamma;\beta)}\right\| \\
&\leq \left(\frac{1}{n}\sum_{i=1}^n\sup_{\gamma\in\Gamma}\left\|\widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)\right\|\right)\max_{1\leq i\leq n}\sup_{\gamma\in\Gamma}\frac{q_{i,n}(\gamma;\beta)}{1-q_{i,n}(\gamma;\beta)} \\
&= O_p(1)\times O_p\left(\max_{1\leq i\leq n}\sup_{\gamma\in\Gamma}\frac{q_{i,n}(\gamma;\beta)}{1-q_{i,n}(\gamma;\beta)}\right) \qquad (27)
\end{aligned}$$

by the results just obtained above. Similarly, noting that the leverage term $q_{i,n}(\gamma;\widehat{\beta})$ is $\in[0,1]$:

$$\begin{aligned}
\sup_{\gamma\in\Gamma}\|\widehat{C}_n^{(3)}(\gamma;\widehat{\beta})-\widehat{C}_n^{(0)}(\gamma;\widehat{\beta})\| &\leq \frac{1}{n}\sum_{i=1}^n\sup_{\gamma\in\Gamma}\left\|\widetilde{u}_i^2(\beta,\gamma)\widetilde{G}_i'(\beta,\gamma)\widetilde{G}_i(\beta,\gamma)\frac{2q_{i,n}(\gamma;\beta)-q_{i,n}^2(\gamma;\beta)}{(1-q_{i,n}(\gamma;\beta))^2}\right\| \\
&= O_p(1)\times O_p\left(\max_{1\leq i\leq n}\sup_{\gamma\in\Gamma}\frac{2q_{i,n}(\gamma;\beta)}{(1-q_{i,n}(\gamma;\beta))^2}\right). \qquad (28)
\end{aligned}$$

Therefore, the key to the required results is to show that $\max_{1\leq i\leq n}\sup_{\gamma\in\Gamma}q_{i,n}(\gamma;\widehat{\beta})=o_p(1)$. For this, exactly following the same steps as before by using (25) and assumptions (b) and (c) gives:

$$\sup_{\gamma\in\Gamma} \|\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma) - E[\widetilde{G}_i'(\beta_0,\gamma)\widetilde{G}_i(\beta_0,\gamma)]\| = o_p(1).$$

Similarly, since $\max_{1\le i\le n}\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\beta,\gamma)\| = o_p(n^{1/4})$ and $\max_{1\le i\le n}\sup_{\gamma\in\Gamma}\|\widetilde{G}_i(\beta,\gamma)\|^2 = o_p(n^{1/2})$ by assumption (b) and i.i.d. data, and $\max_{1\le i\le n}\sup_{\beta:|\beta-\beta_0\|,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\| = o_p(n^{1/4})$ and $\max_{1\le i\le n}\sup_{\beta:|\beta-\beta_0\|,\gamma\in\Gamma}\|\partial\widetilde{G}_i(\beta,\gamma)/\partial\beta\|^2 = o_p(n^{1/2})$ by assumption (c) and i.i.d. data, it follows by using (25) and exactly the same steps as above that:

$$\max_{1\le i\le n}\sup_{\gamma\in\Gamma}\frac{1}{n}|\widetilde{G}_i(\widehat{\beta},\gamma)\widetilde{G}_i'(\widehat{\beta},\gamma)| = n^{-1}o_p(n^{1/2}) = o_p(n^{-1/2}).$$

Now, letting $\lambda_{\max}(A)$ denote the maximum eigen value of any square matrix A, and by using assumption (d) and the two conditions just obtained in the above two displays, we get that:

$$\begin{aligned}
\sup_{\gamma\in\Gamma} q_{i,n}(\gamma;\widehat{\beta}) &\le \sup_{\gamma\in\Gamma}\lambda_{\max}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{G}_i'(\widehat{\beta},\gamma)\widetilde{G}_i(\widehat{\beta},\gamma)\right)^{-1}\right)\left(\frac{1}{n}\widetilde{G}_i(\widehat{\beta},\gamma)\widetilde{G}_i'(\widehat{\beta},\gamma)\right)\\
\Rightarrow \max_{1\le i\le n}\sup_{\gamma\in\Gamma} q_{i,n}(\gamma;\widehat{\beta}) &\le \sup_{\gamma\in\Gamma}\lambda_{\max}\left(\left(E[\widetilde{G}_i'(\beta_0,\gamma)\widetilde{G}_i(\beta_0,\gamma)]\right)^{-1}\right)\max_{1\le i\le n}\sup_{\gamma\in\Gamma}\frac{1}{n}|\widetilde{G}_i(\widehat{\beta},\gamma)\widetilde{G}_i'(\widehat{\beta},\gamma)|\\
&\quad + o_p(1)o_p(n^{-1/2})\\
&= (\text{finite number})\times o_p(n^{-1/2}) + o_p(1)o_p(n^{-1/2}) = o_p(n^{-1/2}).
\end{aligned}$$

Therefore, by virtue of (27) and (28), this also proves the result for the HC2 and HC3 versions. ■

Proof of Lemma 3: To avoid clutter we write $\widehat{\gamma}_n(\widehat{\beta})$ as $\widehat{\gamma}_n$ and $\widehat{\Sigma}_n(\gamma;\widehat{\beta})$ and $\widehat{\Sigma}_n(\gamma)$ in this proof.

(i) Take any $\delta > 0$. Assumption (b) implies that $P(d(\widehat{\gamma}_n,\Gamma^*) > \delta) \le P(\Sigma(\widehat{\gamma}_n) - \Sigma^* \ge \epsilon(\delta))$. Hence, the result will follow if $P(\Sigma(\widehat{\gamma}_n) - \Sigma^* \ge \epsilon(\delta)) \to 0$ for any $\epsilon(\delta) > 0$. We show it now.

First, note that $\widehat{\Sigma}_n(\widehat{\gamma}_n) \le \Sigma^* + o_p(1)$. To see this, take any element $\gamma^* \in \Gamma^*$ and note that $\widehat{\Sigma}_n(\widehat{\gamma}_n) \le \widehat{\Sigma}_n(\gamma^*) \le |\widehat{\Sigma}_n(\gamma^*) - \Sigma(\gamma^*)| + |\Sigma(\gamma^*)| \le \sup_{\gamma\in\Gamma}|\widehat{\Sigma}_n(\gamma) - \Sigma(\gamma)| + \Sigma^* = o_p(1) + \Sigma^*$ where the first inequality follows by the definition of $\widehat{\gamma}_n$; the second inequality follows by the triangle inequality; the third inequality follows by the definition of supremum and since $|\Sigma(\gamma^*)| = \Sigma(\gamma^*) = \Sigma^*$; and the equality at the end follows by assumption (a).

Hence, $\Sigma(\widehat{\gamma}_n) - \Sigma^* \le \Sigma(\widehat{\gamma}_n) - \widehat{\Sigma}_n(\widehat{\gamma}_n) + o_p(1) \le \sup_{\gamma\in\Gamma}|\Sigma(\gamma) - \widehat{\Sigma}_n(\gamma)| + o_p(1) = o_p(1)$ where the first inequality follows by the result $\widehat{\Sigma}_n(\widehat{\gamma}_n) \le \Sigma^* + o_p(1)$ shown above; the second inequality by the definition of supremum; and the equality at the end by assumption (a). Note that $\Sigma(\widehat{\gamma}_n) - \Sigma^* \le o_p(1)$ implies that $\Sigma(\widehat{\gamma}_n) - \Sigma^* = o_p(1)$ since, by definition, $\Sigma(\widehat{\gamma}_n) \ge \Sigma^*$. Hence, $P(\Sigma(\widehat{\gamma}_n) - \Sigma^* \ge \epsilon(\delta)) \to 0$ for any $\epsilon(\delta)$. Therefore, $d(\widehat{\gamma}_n,\Gamma^*) = o_p(1)$.

(ii) Assumption (c), i.e., the result of part (i), implies that $\widehat{\gamma}_n \in \Gamma_\delta^*$ with probability approaching one. Hence we write the rest of the proof conditional on this event. Note that assumption (e) implies that: $d(\widehat{\gamma}_n, \Gamma^*) \leq \frac{1}{\kappa}(\Sigma(\widehat{\gamma}_n) - \Sigma^*)$. Therefore, the result follows if $\sqrt{n}(\Sigma(\widehat{\gamma}_n) - \Sigma^*) = O_p(1)$. We show it now by exactly following the steps of part (i) by additionally using the result of part (i) that, as noted above, allows us now to appeal to the rate-condition in assumption (d).

First, note that $\widehat{\Sigma}_n(\widehat{\gamma}_n) \leq \Sigma^* + O_p(1/\sqrt{n})$. To see this, take any element $\gamma^* \in \Gamma^*$ and note that $\widehat{\Sigma}_n(\widehat{\gamma}_n) \leq \widehat{\Sigma}_n(\gamma^*) \leq |\widehat{\Sigma}_n(\gamma^*) - \Sigma(\gamma^*)| + |\Sigma(\gamma^*)| \leq \sup_{\gamma \in \Gamma^*_\delta} |\widehat{\Sigma}_n(\gamma) - \Sigma(\gamma)| + \Sigma^* = O_p(1/\sqrt{n}) + \Sigma^*$ where the first inequality follows by the definition of $\widehat{\gamma}_n$; the second inequality follows by the triangle inequality; the third inequality follows by the definition of supremum and since $|\Sigma(\gamma^*)| = \Sigma(\gamma^*) = \Sigma^*$; and the equality at the end follows by assumption (d).

Hence, $\Sigma(\widehat{\gamma}_n) - \Sigma^* \leq \Sigma(\widehat{\gamma}_n) - \widehat{\Sigma}_n(\widehat{\gamma}_n) + O_p(1/\sqrt{n}) \leq \sup_{\gamma \in \Gamma} |\Sigma(\gamma) - \widehat{\Sigma}_n(\gamma)| + O_p(1/\sqrt{n}) = O_p(1/\sqrt{n})$ where the first inequality follows by the result $\widehat{\Sigma}_n(\widehat{\gamma}_n) \leq \Sigma^* + O_p(1/\sqrt{n})$ shown above; the second inequality by the definition of supremum; and the equality at the end by assumption (d). Note that $\Sigma(\widehat{\gamma}_n) - \Sigma^* \leq O_p(1/\sqrt{n})$ implies that $\Sigma(\widehat{\gamma}_n) - \Sigma^* = O_p(1/\sqrt{n})$ since, by definition, $\Sigma(\widehat{\gamma}_n) \geq \Sigma^*$. Since $\kappa > 0$, this result implies that $\sqrt{n} d(\widehat{\gamma}_n, \Gamma^*) \leq \frac{1}{\kappa}\sqrt{n}(\Sigma(\widehat{\gamma}_n) - \Sigma^*) = O_p(1)$. ■

Proof of Theorem 4: Assumption (a) is maintained for both parts (i) and (ii) of the theorem. This means that for any $\delta > 0$, we have $\widehat{\gamma}_n \in \Gamma^*_\delta$ with probability approaching 1. Therefore, the rest of the proof proceeds by conditioning on this event that $\widehat{\gamma}_n \in \Gamma^*_\delta$ for any $\delta > 0$.

(i) Take any $\epsilon > 0$. By assumption (c), $|\widehat{\beta}_n(\widehat{\gamma}_n) - \beta_0| > \epsilon$ implies that $E[\psi(y, X; \widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n)] - E[\psi(y, X; \beta_0, \widehat{\gamma}_n)] > \eta$ for some $\eta > 0$. Hence, the result that $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$ will follow if $P\left(E[\psi(y, X; \widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n)] - E[\psi(y, X; \beta_0, \widehat{\gamma}_n)] < \eta\right) \to 1$ for all $\eta > 0$. We will show this now.

Take any $\eta > 0$, define $A_n(\epsilon, \delta) := \sup_{\beta: |\beta - \beta_0| \geq \epsilon, \gamma \in \Gamma^*_\delta} |\bar{\psi}_n(\beta, \gamma) - E[\psi(y, X; \beta, \gamma)]|$, and note that:

$$\begin{aligned}
E[\psi(y, X; \widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n)] &\leq \bar{\psi}_n(\widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n) + |\bar{\psi}_n(\widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n) - E[\psi(y, X; \widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n)]| \\
&\leq \bar{\psi}_n(\widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n) + A_n(\epsilon, \delta) \\
&\leq \bar{\psi}_n(\beta_0, \widehat{\gamma}_n) + A_n(\epsilon, \delta) \quad \text{[by the definition of } \widehat{\beta}_n(\widehat{\gamma}_n)] \\
&\leq E[\psi(y, X; \beta_0, \widehat{\gamma}_n)] + |\bar{\psi}_n(\beta_0, \widehat{\gamma}_n) - E[\psi(y, X; \beta_0, \widehat{\gamma}_n)]| + A_n(\epsilon, \delta) \\
&\leq E[\psi(y, X; \beta_0, \widehat{\gamma}_n)] + 2A_n(\epsilon, \delta).
\end{aligned}$$

On the other hand, $P(A_n(\epsilon, \delta) < \eta/2) \to 1$ by assumption (b). This gives, by using the inequality in the above display, that $P\left(E[\psi(y, X; \widehat{\beta}_n(\widehat{\gamma}_n), \widehat{\gamma}_n)] - E[\psi(y, X; \beta_0, \widehat{\gamma}_n)] < \eta\right) \to 1$. Hence, $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$. Therefore, now $\widehat{h}_n(\widehat{\gamma}_n) := h(\widehat{\beta}_n(\widehat{\gamma}_n)) = h_0 + o_p(1)$ follows directly by assumption (d).

(ii) The high level assumptions in the theorem directly lead to the final result from the expansion of the first order condition of the sample minimization problem; see, e.g., Appendix B.1 below.

(iii) The result follows because $|\widehat{\Sigma}_n(\widehat{\gamma}_n;\widehat{\beta}_n(\widehat{\gamma}_n)) - \Sigma^*| \leq |\widehat{\Sigma}_n(\widehat{\gamma}_n;\widehat{\beta}_n(\widehat{\gamma}_n)) - \Sigma_n(\widehat{\gamma}_n)| + |\Sigma_n(\widehat{\gamma}_n) - \Sigma_n(\gamma_n)| + |\Sigma_n(\gamma_n) - \Sigma^*| \leq \sup_{\gamma\in\Gamma_\delta^*} |\widehat{\Sigma}_n(\gamma;\widehat{\beta}_n(\widehat{\gamma}_n)) - \Sigma_n(\gamma)| + o_p(1) = o_p(1)$ for the following reasons. The first inequality follows by the triangle inequality. The second inequality follows because: (1) $P(\widehat{\gamma}_n \in \Gamma_\delta^*) \to 1$ for any $\delta > 0$ by assumption (a) of Theorem 4(i); (2) $\Sigma(\gamma)$ is continuous in $\gamma \in \Gamma$ and, by virtue of the i.i.d data, $\Sigma_n(\gamma) = \Sigma(\gamma)$ for $\gamma \in \Gamma$ and all n; and (3) assumption (h) of Theorem 4(ii) holds. Finally, the last equality follows because $\widehat{\beta}_n(\widehat{\gamma}_n) = \beta_0 + o_p(1)$ and assumptions (a)-(e) of Lemma 2 (with Γ_δ^* in place of Γ) give that $\sup_{\gamma\in\Gamma_\delta^*} |\widehat{\Sigma}_n(\gamma;\widehat{\beta}_n(\widehat{\gamma}_n)) - \Sigma_n(\gamma)| = o_p(1)$. ■

Proof of Corollary 5: Under the conditions of Theorem 4 (ii)-(iii), it follows that $T_n(h_{\text{Null}}) \xrightarrow{d} N(\mu/\sqrt{\Sigma^*}, 1)$ by Slutsky's theorem. Then, the final results in both (i) and (ii) follow by the definition of the convergence in distribution. ■

B Appendix B: Descriptive endnotes for Section 3

B.1 Expansion with respect to γ in assumption (e) for Theorem 4

The high level conditions related to the asymptotic expansion with respect to γ in Theorem 4 would follow under standard regularity conditions if a differentiability condition such as the following is assumed: There exist a sequence (triangular array) of $k \times 1$ vectors $\nabla_\gamma(X_i;\gamma_n)$ and a sequence (triangular array) of positive scalars $\Delta(X_i;\gamma_n)$ such that the following inequality holds with probability one for all $i = 1,\ldots,n$ and for all n large enough and some $\delta > 0$:

$$\sup_{\gamma\in\Gamma:\|\gamma-\gamma_n\|<\delta} \left\{ \left| \frac{1}{\omega^2(X_i;\gamma)} - \frac{1}{\omega^2(X_i;\gamma_n)} - \nabla_\gamma(X_i;\gamma_n)'(\gamma-\gamma_n) \right| - \frac{1}{2}\|\gamma-\gamma_n\|^2 \Delta(X_i;\gamma_n) \right\} \leq 0. \quad (29)$$

To see how (29) leads to the terms involving the expansion with respect to γ be $o_p(1)$ in assumption (e) for Theorem 4, consider a linear model to abstract from the expansion related to β and thus keep focus on the key terms. Then, $\psi_\beta(y_i, X_i;\beta_0,\gamma) = X_i u_i/\omega^2(X_i;\gamma)$ (where $u_i = y_i - X_i'\beta_0$), and:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i u_i}{\omega^2(X_i;\widehat{\gamma}_n)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i u_i}{\omega^2(X_i;\gamma_n)} + E\left[\frac{1}{n}\sum_{i=1}^n X_i u_i \nabla_\gamma(X_i;\gamma_n)'\right]\sqrt{n}(\widehat{\gamma}_n-\gamma_n) + R_{1,n} + R_{2,n}, \quad (30)$$

$$\text{where } R_{1,n} := \left[\frac{1}{n}\sum_{i=1}^n \left(X_i u_i \nabla_\gamma(X_i;\gamma_n)' - E\left[X_i u_i \nabla_\gamma(X_i;\gamma_n)'\right]\right)\right]\sqrt{n}(\widehat{\gamma}_n-\gamma_n),$$

$$\text{and } R_{2,n} := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i \left[\frac{1}{\omega^2(X_i;\widehat{\gamma}_n)} - \frac{1}{\omega^2(X_i;\gamma_n)} - \nabla_\gamma(X_i;\gamma_n)'(\widehat{\gamma}_n-\gamma_n)\right].$$

Now, note that $R_{1,n} = o_p(1)$ by a suitable uniform law of large numbers because $\sqrt{n}(\widehat{\gamma}_n - \gamma_n) = O_p(1)$. On the other hand, $R_{2,n} = o_p(1)$ because:

$$
\begin{aligned}
|R_{2,n}| &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \|X_i u_i\| \times \left| \frac{1}{\omega^2(X_i;\widehat{\gamma}_n)} - \frac{1}{\omega^2(X_i;\gamma_n)} - \nabla_\gamma(X_i;\gamma_n)'(\widehat{\gamma}_n - \gamma_n) \right| \\
&\leq \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \|X_i u_i\| \times |\Delta(X_i;\gamma_n)| \times \|\widehat{\gamma}_n - \gamma_n\|^2 \\
&\leq \left(\frac{1}{2n} \sum_{i=1}^{n} \|X_i u_i \Delta(X_i;\gamma_n)\| \right) \left(n^{1/4} \|\widehat{\gamma}_n - \gamma_n\| \right)^2 = o_p(1),
\end{aligned}
$$

where the first inequality follows by the Cauchy-Schwartz inequality; the second inequality by (29); the third inequality by noting that $\Delta(X_i;\gamma_n)$ is a triangular array of positive scalar random variables; whereas the last equality would follow if the term inside the first parentheses on the left hand side of the last equality is $O_p(1)$, i.e., uniformly tight. Therefore, (30) now matches with the right hand side of the first equality in assumption (e) for Theorem 4, with $D_n(\gamma_n) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i u_i \nabla_\gamma(X_i;\gamma_n)'\right]$. Recalling that $\lim_{n\to\infty} D_n(\gamma) =: D(\gamma) = 0$ for $\gamma \in \Gamma$, the right hand side of the second equality in assumption (e) for Theorem 4 would now follow if $\lim_{n\to\infty} D_n(\gamma_n) = 0$.

B.2 Compactness of Γ in $\mathbb{R}^k$ to ensure the existence of optimum

Our primary purpose for the compactness assumption of Γ was to provide a familiar sufficient condition (along with the continuity of $\Sigma(\gamma)$) for the existence of an "optimal" γ that minimizes $\Sigma(\gamma)$ with respect to $\gamma \in \Gamma$ irrespective of the correctness of the user's model. In a general context, compactness is not needed when the optimum can be directly shown to exist, e.g., if the objective function is convex or if the "true" parameter value is naturally defined as the parameter value that is associated with the probability distribution from which a given sample is drawn. However, compactness is generally a standard assumption in nonlinear estimations (at least since Jenrich (1969)), although the stated purpose in general is to ensure that the sample objective function is uniformly consistent for the population objective function. In the context of WLS, chapter 12 of Wooldridge (2010) maintains compactness while Romano and Wolf (2017) do not need to explicitly maintain compactness since they use a log-linear model $\omega^2(X;\gamma)$ and hence automatically obtain a closed-form expression for the solution γ of the first step minimization problem of WLS as in (7).

The question is: What is the cost of artificially imposing compactness? To answer this question, note that in the Euclidean space, i.e., in our context, this imposes that Γ is closed and bounded.

Relaxing closedness would actually be useful for us because we could then directly avoid the problem of differentiability at the boundary; see above Theorem 4. Therefore, the cost of the compactness assumption manifests through the imposition of boundedness. (Statistical softwares always impose boundedness; e.g., Matlab can only handle numbers smaller than "realmax" in absolute value.) Hence, the cost of assuming a compact Γ in our paper is that it potentially undersells the MWLS estimator since we are perhaps settling for a sub-optimal minimizer of $\Sigma(\gamma)$, and thereby a sub-optimal level set Γ^*, and hence a conservative variance bound Σ^* that the MWLS estimator achieves.

B.3 Multiple optima, i.e., non-singleton Γ^*, and relation to the literature

It is worth noting that the issue of a possibly set-identified nuisance parameter (non-singleton Γ^*) can also arise in conventional WLS or its variants. The literature typically assumes away this issue; e.g., assumption (3.11) in Romano and Wolf (2017) essentially maintain the existence of a unique pseudo-true value for the nuisance parameters.[17] This is perfectly fine in Romano and Wolf (2017) since they ultimately use a log-linear $\omega^2(X;\gamma)$. Log-linearity will lead to a globally convex objective function to define the pseudo-true value as the minimizer and hence will guarantee its existence and, under a standard full-rank condition, its uniqueness. However, without more assumptions, a general choice of $\omega^2(X;\gamma)$ by the user may not guarantee the uniqueness of the pseudo-true value that is defined as the solution to the relevant optimization problem of WLS:

$$\min_{\gamma\in\Gamma} E\left[\left(u^2 - \omega^2(X;\gamma)\right)^2\right].$$

(Existence will follow, as in our paper, if $\omega^2(X;\gamma)$ is continuous in $\gamma \in \Gamma$ and Γ is compact in $\mathbb{R}^k$.)

Without uniqueness, the solutions to the above optimization problem do not necessarily form a level set for $\Sigma(\gamma)$. Therefore, conventional WLS or its variants cannot deal with the case of non-unique solutions. (Also see the last paragraph of our Section 3.4.) By contrast, the relevant optimization problem for MWLS is designed to construct a level set Γ^* for $\Sigma(\gamma)$ (i.e., $\Sigma(\gamma) = \Sigma^*$ for $\gamma \in \Gamma^*$). Hence, our results on the asymptotic distribution of MWLS and inference based on it in Theorem 4 and Corollary 5 respectively, resembling their conventional form, follow naturally.

Lastly, we note that non-singleton Γ^*'s, i.e., not-point-identified nuisance parameters, abound

[17] Similar assumptions of uniqueness are typically also made for the parameters of interest; see, e.g., Theorem 2.1, Corollary 2.2 and the discussion following it in White (1981) in the context of NLS. Of course, we have also assumed uniqueness of β_0. Recently Chen et al. (2011) relax the uniqueness assumptions for the parameters of interest and nuisance parameters for the purpose of sensitivity analysis in likelihood-based inference; also see the references therein to the vast literature in statistics on finite mixture models where the non-uniqueness problem arises naturally.

in econometrics; see, e.g., Hansen (1991) (in particular, his Assumption 1(v) to see the connection) and the references therein to the literature on testing of hypotheses when the nuisance parameter is unidentified under the null hypothesis. More recently, the weak identification literature studies inference that is robust to weakly identified nuisance parameters; see, e.g., Andrews (2017), Andrews (2018) and the references therein. Our treatment of nuisance parameters, however, differs from both these literature. Unlike the former, that concerns with the so-called "sup" statistics, we actually estimate the nuisance parameters γ in a standard manner by searching for the optimal γ. Unlike the latter, we cannot fix as reference a unique "true value" γ^* of the nuisance parameters unless Γ^* is actually singleton. This inability to fix a unique true value makes our treatment of γ resemble one strand of the partial identification literature; e.g., Chernozhukov et al. (2007). There is, however, one difference: We are not interested in the entire identified set Γ^*. We only wish that the estimator $\widehat{\gamma}_n$ comes close to the set Γ^* in probability in terms of the one-sided Hausdorff distance as in (19).

B.4 The equivalence of the two optimization problems in Section 3.4

Define the set of minimizers of $\Sigma(\gamma)$ in the general sense of optimality as in (22) as:

$$\bar{\Gamma}^* := \left\{ \gamma \in \Gamma \,\middle|\, \Sigma(\gamma^\dagger) - \Sigma(\gamma) \text{ is positive semi-definite for all } \gamma^\dagger \in \Gamma \right\}.$$

$\bar{\Gamma}^*$ is nonempty if and only if $\gamma^* \in \Gamma$ satisfying (22) exists. On the other hand, the trace of $\Sigma(\gamma)$ minimizing set, i.e., the A-optimal set, Γ^* was defined in (23) as:

$$\Gamma^* = \left\{ \gamma \in \Gamma \,\middle|\, \text{Trace}(\Sigma(\gamma)) - \min_{\gamma^\dagger \in \Gamma} \text{Trace}(\Sigma(\gamma^\dagger)) = 0 \right\}.$$

Γ^* is nonempty if, e.g., $\text{Trace}(\Sigma(\gamma))$ is continuous in $\gamma \in \Gamma$ and Γ is compact in $\mathbb{R}^k$.

Lemma 6 *Let $\Sigma(\gamma)$ be positive definite for $\gamma \in \Gamma$, Trace$(\Sigma(\gamma))$ be continuous in $\gamma \in \Gamma$ and Γ be compact in $\mathbb{R}^k$. Then $\bar{\Gamma}^* = \Gamma^*$ if and only if γ^* defined in (22) exists.*

Proof of Lemma 6: The "only if" part is trivial because $\bar{\Gamma}^*$ is empty unless γ^* satisfying (22) exists, whereas Γ^* is nonempty under the condition of the lemma. Therefore, now consider the "if" part, i.e., $\bar{\Gamma}^* = \Gamma^*$ if γ^* defined in (22) exists. We prove this by showing that $\bar{\Gamma}^* \subseteq \Gamma^*$ and $\Gamma^* \subseteq \bar{\Gamma}^*$.

Step 1 [$\gamma \in \bar{\Gamma}^* \Rightarrow \gamma \in \Gamma^*$]: Take any $\gamma \in \bar{\Gamma}^*$, which is possible since the existence of γ^* gives that $\bar{\Gamma}^*$ is nonempty. We will now show that $\gamma \in \Gamma^*$. Suppose that this is not true. Then there exists a $\gamma^\dagger \in \Gamma^*$ such that $\text{Trace}(\Sigma(\gamma)) - \text{Trace}(\Sigma(\gamma^\dagger)) > 0$. This implies that at least one diagonal element of $\Sigma(\gamma)$ is strictly greater than the corresponding diagonal element of $\Sigma(\gamma^\dagger)$. This implies that $\Sigma(\gamma^\dagger) - \Sigma(\gamma)$ cannot be positive semi-definite. Hence $\gamma \notin \bar{\Gamma}^*$, which is a contradiction. Hence, $\gamma \in \Gamma^*$.

Step 2 [$\gamma \in \Gamma^* \Rightarrow \gamma \in \bar{\Gamma}^*$]: Take any $\gamma \in \Gamma^*$. We will now show that $\gamma \in \bar{\Gamma}^*$. Suppose that this is not true. Note that, $\gamma \in \Gamma^*$ implies that $\text{Trace}(\Sigma(\gamma^\dagger) - \Sigma(\gamma)) \geq 0$ for all $\gamma^\dagger \in \Gamma$ and hence for all $\gamma^\dagger \in \bar{\Gamma}^* \subseteq \Gamma$. However, since both $\Sigma(\gamma)$ and $\Sigma(\gamma^\dagger)$ are positive definite, this will imply that for $\gamma^\dagger \in \bar{\Gamma}^*$: (i) either at least one diagonal element of $\Sigma(\gamma) - \Sigma(\gamma^\dagger)$ is negative, or (ii) $\Sigma(\gamma) = \Sigma(\gamma^\dagger)$. Note that, (i) cannot happen by the definition of $\bar{\Sigma}^*$ in which $\gamma^\dagger$ belongs. Therefore, only (ii) can happen. However, (ii) happens means that $\gamma \in \bar{\Gamma}^*$, which contradicts our supposition. Hence, $\gamma \in \bar{\Gamma}^*$. ■

C Bibliography

Andrews, D. W. K. (2017). Identification-Robust Subvector Inference. Yale University Working Paper.

Andrews, I. (2018). Robust Two-Step Confidence Sets, and the Trouble with the First Stage F-Statistic. *Review of Economics and Statistics*, 100: 337–348.

Angrist, J. D. and Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspective*, 24: 3–30.

Bonhomme, S. and Weidner, M. (2019). Minimizing sensitivity to model misspecification. Working paper.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *The Annals of Statistics*, 10: 1224–1233.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *The Annals of Statistics*, 10: 429–441.

Chamberlain, G. (1987). Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics*, 34:305–334.

Chaudhuri, S. (2019). A direct route to optimal parametric weighted least squares. Technical report, McGill University.

Chen, X., Tamer, E., and Torgivitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. Cowles Foundation Discussion Paper No. 1836.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24: 586–602.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75: 1243–1284.

DiCiccio, C. J., Romano, J. P., and Wolf, M. (2019). Improving weighted least squares inference. *Econometrics and Statistics*, 10: 96–119.

Ehrenfeld, S. (1956). Complete class theorems in experimental design. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 57–67. University of California Press.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Annals of Mathematical Statistics*, 23: 255–262.

Federov, V. V. (1971). Design of experiments for linear optimality criteria. *Theory of Probability and its Applications*, 16: 189–195.

Hansen, B. E. (1991). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. Technical report, WP No: 296, University of Rochester.

Hansen, B. E. (2020). Econometrics. Online Textbook.

Heyde, C. C. (1997). *Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation*. Springer.

Jenrich, R. L. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics*, 40:633–643.

Karlin, S. and Studden, W. J. (1966). Optimal experimental designs. *Annals of Mathematical Statistics*, 37: 783–815.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2: 849–879.

Leamer, E. E. (2010). Tantalus on the Road to Asymptotia. *Journal of Economic Perspective*, 24: 31–46.

MacKinnon, J. G. (2012). Thirty Years of Heteroskedasticity-Robust Inference. In Chen, X. and Swanson, N. R., editors, *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pages 437–461. Springer.

Newey, W. K. (1994). Series Estimation of Regression Functionals. *Econometric Theory*, 10: 1–28.

Newey, W. K. and McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 36, pages 2212–2245. Elsevier Science Publisher.

Robinson, P. M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica*, 55: 875–891.

Romano, J. P. and Wolf, M. . (2017). Resurrecting Weighted Least Squares. *Journal of Econometrics*, 197: 1–19.

Spady, R. and Stouli, S. (2019). Simultaneous Mean-Variance Regression. Working paper.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Wald, A. (1943). On the efficient design of statistical investigations. *Annals of Mathematical Statistics*, 14: 134–140.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heterogeneity. *Econometrica*, 48:817–838.

White, H. (1981). Consequence and Detection of Misspecified Nonlinear Regression Models. *Journal of the American Statistical Association*, 76:419–433.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press.

Wooldridge, J. M. (2012). *Introductory Econometrics.* South-Western, Mason, Ohio.

Complete set of simulation results for: A direct route to optimal parametric weighted least squares

Saraswata Chaudhuri[18]

List of Tables

List of Figures

[18]Department of Economics, McGill University, Montreal, Canada. Email: saraswata.chaudhuri@mcgill.ca.

			Model 1 for $\omega^2(X;\gamma)$						Model 2 for $\omega^2(X;\gamma)$					
DGP	n	OLS	WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3	WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3
1(a)	25	-.0004	-.0016	-.0018	-.0016	-.0014	-.0030	-.0022	-.0011	-.0012	-.0012	-.0011	-.0026	-.0012
	50	.0025	.0018	.0018	.0020	.0019	.0017	.0016	.0018	.0018	.0021	.0020	.0018	.0018
	100	-.0009	-.0010	-.0010	-.0008	-.0009	-.0008	-.0008	-.0012	-.0012	-.0010	-.0010	-.0009	-.0009
	200	.0005	.0003	.0003	.0003	.0003	.0002	.0002	.0004	.0004	.0003	.0003	.0002	.0002
	400	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006
1(b)	25	.0028	.0037	.0041	.0024	.0030	-.0002	.0011	.0035	.0041	.0024	.0024	-.0001	.0006
	50	.0026	.0027	.0025	.0026	.0027	.0019	.0019	.0024	.0023	.0025	.0026	.0019	.0021
	100	-.0009	.0003	.0006	.0005	.0003	.0002	.0001	.0004	.0005	.0005	.0003	.0003	.0002
	200	-.0002	-.0004	-.0004	-.0004	-.0004	-.0003	-.0003	-.0005	-.0005	-.0005	-.0004	-.0004	-.0004
	400	-.0030	-.0020	-.0020	-.0020	-.0020	-.0017	-.0017	-.0022	-.0022	-.0022	-.0022	-.0019	-.0020
1(c)	25	-.0080	-.0009	-.0015	-.0013	-.0012	.0013	-.0006	-.0006	-.0007	-.0008	-.0011	.0034	.0000
	50	.0094	.0087	.0086	.0087	.0088	.0085	.0088	.0097	.0098	.0100	.0097	.0095	.0095
	100	.0018	.0006	.0006	.0005	.0009	.0008	.0009	.0006	.0006	.0006	.0008	.0007	.0008
	200	.0004	.0008	.0008	.0008	.0010	.0014	.0013	.0005	.0005	.0005	.0007	.0010	.0010
	400	-.0008	-.0018	-.0018	-.0018	-.0017	-.0017	-.0017	-.0016	-.0016	-.0016	-.0015	-.0016	-.0016
1(d)	25	-.0133	.0025	.0025	.0058	.0015	-.0048	-.0088	.0018	.0018	.0011	.0014	-.0030	-.0059
	50	.0017	.0033	.0033	.0033	.0029	.0040	.0039	.0029	.0029	.0029	.0027	.0035	.0033
	100	-.0005	.0038	.0038	.0038	.0039	.0037	.0035	.0041	.0041	.0041	.0042	.0045	.0042
	200	-.0008	.0025	.0025	.0025	.0022	.0020	.0020	.0039	.0039	.0039	.0036	.0035	.0035
	400	-.0031	-.0033	-.0033	-.0033	-.0034	-.0037	-.0036	-.0025	-.0025	-.0025	-.0028	-.0026	-.0026
2(a)	25	-.0008	.0008	.0008	.0010	.0010	.0013	.0009	.0009	.0009	.0010	.0011	.0015	.0013
	50	.0007	-.0005	-.0005	-.0005	-.0004	-.0007	-.0006	-.0006	-.0006	-.0006	-.0005	-.0011	-.0010
	100	.0001	.0005	.0005	.0005	.0005	.0003	.0003	.0007	.0007	.0007	.0007	.0006	.0006
	200	-.0004	-.0001	-.0001	-.0001	-.0001	.0001	.0001	-.0002	-.0002	-.0002	-.0002	.0001	.0000
	400	-.0001	-.0002	-.0002	-.0002	-.0002	-.0001	-.0001	-.0002	-.0002	-.0002	-.0002	-.0001	-.0001
2(b)	25	-.0054	-.0034	-.0032	-.0034	-.0034	-.0038	-.0042	-.0035	-.0034	-.0035	-.0036	-.0037	-.0041
	50	.0001	-.0002	-.0002	-.0002	-.0002	-.0003	-.0003	-.0001	-.0001	-.0001	-.0001	.0001	.0001
	100	-.0013	-.0007	-.0007	-.0007	-.0007	-.0006	-.0006	-.0007	-.0007	-.0007	-.0007	-.0004	-.0005
	200	-.0008	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0004	.0003
	400	.0008	.0004	.0004	.0004	.0004	.0002	.0002	.0005	.0005	.0005	.0005	.0001	.0001
3(a)	25	.0038	.0086	.0068	.0085	.0082	.0068	.0073	.0089	.0078	.0079	.0083	.0086	.0075
	50	-.0034	-.0015	-.0017	-.0020	-.0021	-.0011	-.0014	-.0013	-.0014	-.0019	-.0017	-.0011	-.0014
	100	.0008	.0005	.0005	.0006	.0005	.0005	.0004	.0005	.0008	.0007	.0006	.0002	.0002
	200	-.0042	-.0029	-.0029	-.0028	-.0029	-.0027	-.0027	-.0027	-.0027	-.0025	-.0027	-.0024	-.0024
	400	-.0039	-.0027	-.0027	-.0027	-.0027	-.0027	-.0027	-.0027	-.0027	-.0027	-.0026	-.0026	-.0026
3(b)	25	-.0146	-.0133	-.0142	-.0144	-.0148	-.0161	-.0152	-.0139	-.0143	-.0146	-.0147	-.0155	-.0143
	50	-.0055	-.0042	-.0036	-.0036	-.0048	-.0054	-.0053	-.0041	-.0038	-.0042	-.0044	-.0050	-.0048
	100	.0043	.0029	.0029	.0023	.0026	.0018	.0020	.0028	.0028	.0026	.0025	.0017	.0018
	200	-.0027	-.0007	-.0007	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0003	-.0004
	400	-.0008	.0001	.0001	.0001	.0001	.0000	-.0001	-.0002	-.0002	-.0002	-.0001	-.0002	-.0003
4(a)	25	.0039	.0072	.0066	.0069	.0071	.0064	.0062	.0073	.0069	.0070	.0068	.0067	.0064
	50	-.0015	-.0013	-.0015	-.0016	-.0012	-.0010	-.0009	-.0010	-.0010	-.0014	-.0010	-.0009	-.0009
	100	.0025	.0018	.0018	.0019	.0021	.0020	.0021	.0018	.0018	.0019	.0021	.0021	.0022
	200	.0005	.0008	.0008	.0005	.0006	.0006	.0006	.0009	.0009	.0006	.0006	.0007	.0007
	400	.0000	.0001	.0001	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0001	.0001
4(b)	25	-.0104	-.0088	-.0104	-.0093	-.0102	-.0121	-.0104	-.0094	-.0108	-.0099	-.0102	-.0105	-.0100
	50	-.0012	-.0030	-.0032	-.0031	-.0033	-.0035	-.0035	-.0034	-.0035	-.0034	-.0035	-.0040	-.0039
	100	.0004	-.0017	-.0018	-.0018	-.0017	-.0015	-.0014	-.0017	-.0017	-.0014	-.0017	-.0018	-.0016
	200	.0044	.0025	.0025	.0024	.0028	.0028	.0028	.0026	.0026	.0025	.0028	.0027	.0028
	400	.0014	.0019	.0019	.0019	.0017	.0016	.0016	.0016	.0016	.0016	.0016	.0014	.0014

Table 7: Estimated bias of estimators for $h(\beta) := \beta_1$.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$						Model 2 for $\omega^2(X;\gamma)$					
			WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3	WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3
1(a)	25	-.0002	.0006	.0005	.0003	.0004	.0004	.0003	.0005	.0004	.0002	.0003	.0004	.0005
	50	.0006	.0007	.0007	.0008	.0007	.0008	.0008	.0007	.0007	.0008	.0007	.0008	.0008
	100	-.0007	-.0009	-.0009	-.0008	-.0008	-.0007	-.0007	-.0009	-.0009	-.0008	-.0008	-.0007	-.0007
	200	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0004
	400	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005
1(b)	25	-.0006	-.0008	-.0011	-.0005	-.0007	-.0008	-.0008	-.0008	-.0011	-.0008	-.0008	-.0007	-.0007
	50	-.0007	-.0007	-.0006	-.0006	-.0006	-.0006	-.0007	-.0007	-.0006	-.0006	-.0007	-.0007	-.0007
	100	.0001	.0001	-.0001	.0001	.0001	.0001	.0001	.0001	.0000	.0001	.0001	.0001	.0001
	200	.0005	.0003	.0003	.0002	.0002	.0002	.0002	.0003	.0003	.0003	.0003	.0002	.0002
	400	-.0007	-.0006	-.0006	-.0005	-.0005	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006
1(c)	25	-.0007	.0015	.0014	.0003	.0002	.0003	.0003	.0014	.0015	.0004	.0005	.0013	.0012
	50	.0052	.0028	.0027	.0027	.0030	.0032	.0034	.0028	.0027	.0027	.0031	.0032	.0033
	100	.0010	.0008	.0008	.0009	.0011	.0012	.0012	.0009	.0009	.0009	.0011	.0013	.0013
	200	.0008	.0007	.0007	.0007	.0007	.0007	.0007	.0005	.0005	.0005	.0006	.0005	.0005
	400	.0003	-.0005	-.0005	-.0005	-.0005	-.0005	-.0005	-.0006	-.0006	-.0006	-.0005	-.0006	-.0006
1(d)	25	.0004	-.0002	-.0002	.0014	.0020	.0011	.0023	.0012	.0012	.0010	.0023	.0022	.0036
	50	.0078	.0006	.0006	.0008	.0006	-.0005	.0001	.0000	.0000	-.0003	-.0002	-.0011	-.0009
	100	-.0040	-.0017	-.0017	-.0017	-.0014	-.0025	-.0026	-.0027	-.0027	-.0027	-.0027	-.0035	-.0036
	200	-.0051	.0013	.0013	.0013	.0010	.0002	.0002	.0028	.0028	.0028	.0022	.0022	.0022
	400	-.0036	.0017	.0017	.0017	.0016	.0019	.0018	.0020	.0020	.0020	.0018	.0023	.0023
2(a)	25	.0019	.0023	.0022	.0021	.0021	.0022	.0022	.0022	.0022	.0022	.0021	.0022	.0022
	50	.0002	.0001	.0001	.0001	.0002	.0003	.0003	.0000	.0000	.0001	.0001	.0003	.0003
	100	-.0005	-.0006	-.0006	-.0006	-.0006	-.0006	-.0006	-.0007	-.0007	-.0006	-.0006	-.0006	-.0006
	200	-.0003	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	.0000	.0000
	400	.0001	.0001	.0001	.0001	.0001	.0000	.0000	.0001	.0001	.0001	.0002	.0001	.0001
2(b)	25	-.0018	-.0009	-.0009	-.0009	-.0009	-.0009	-.0010	-.0008	-.0008	-.0009	-.0009	-.0010	-.0011
	50	.0010	.0009	.0009	.0009	.0009	.0011	.0011	.0009	.0009	.0009	.0009	.0011	.0010
	100	-.0004	.0000	.0000	.0000	.0000	-.0001	-.0001	.0001	.0001	.0001	.0001	.0000	-.0001
	200	.0009	.0007	.0007	.0007	.0007	.0006	.0006	.0008	.0008	.0008	.0008	.0007	.0006
	400	.0009	.0001	.0001	.0001	.0001	-.0001	-.0001	.0000	.0000	.0000	.0000	-.0002	-.0002
3(a)	25	-.0006	-.0001	-.0004	.0002	.0001	.0004	.0005	-.0002	-.0006	.0001	.0001	.0003	.0004
	50	.0027	.0017	.0017	.0021	.0020	.0015	.0016	.0014	.0014	.0016	.0018	.0012	.0013
	100	.0023	.0020	.0020	.0020	.0021	.0020	.0020	.0018	.0018	.0018	.0019	.0017	.0017
	200	-.0001	.0000	.0000	.0000	-.0001	-.0001	-.0001	.0000	.0000	.0000	.0000	-.0001	-.0001
	400	.0003	-.0002	-.0002	-.0002	-.0002	-.0003	-.0003	-.0003	-.0003	-.0003	-.0003	-.0003	-.0003
3(b)	25	-.0032	-.0007	-.0012	-.0009	-.0013	-.0006	-.0012	-.0005	-.0011	-.0010	-.0014	-.0010	-.0014
	50	.0030	.0021	.0023	.0022	.0020	.0019	.0020	.0020	.0021	.0022	.0019	.0017	.0018
	100	-.0018	-.0022	-.0022	-.0021	-.0021	-.0021	-.0021	-.0021	-.0020	-.0020	-.0020	-.0020	-.0020
	200	-.0020	-.0023	-.0023	-.0023	-.0023	-.0025	-.0025	-.0024	-.0024	-.0024	-.0024	-.0025	-.0025
	400	.0000	.0000	.0000	.0000	-.0001	-.0001	-.0001	.0000	.0000	.0000	.0000	.0000	.0000
4(a)	25	.0012	.0005	.0002	.0005	.0005	.0009	.0007	.0006	.0001	.0003	.0005	.0005	.0006
	50	-.0001	.0000	-.0002	-.0002	-.0002	.0000	.0000	.0000	-.0001	-.0002	-.0001	.0001	.0001
	100	.0004	.0005	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0007	.0006	.0004	.0004
	200	-.0003	-.0005	-.0005	-.0006	-.0005	-.0005	-.0005	-.0005	-.0005	-.0006	-.0005	-.0006	-.0006
	400	-.0008	-.0008	-.0008	-.0009	-.0009	-.0009	-.0009	-.0008	-.0008	-.0009	-.0009	-.0009	-.0009
4(b)	25	-.0010	.0004	.0000	.0004	.0002	.0008	.0005	.0003	-.0003	.0007	.0001	.0006	.0005
	50	-.0013	-.0010	-.0010	-.0009	-.0010	-.0013	-.0012	-.0008	-.0007	-.0009	-.0010	-.0012	-.0012
	100	.0003	.0009	.0010	.0010	.0010	.0008	.0008	.0008	.0009	.0007	.0009	.0008	.0008
	200	.0013	.0018	.0018	.0017	.0016	.0017	.0017	.0019	.0019	.0018	.0017	.0018	.0017
	400	.0009	.0014	.0014	.0014	.0013	.0013	.0013	.0013	.0013	.0013	.0012	.0012	.0012

Table 8: Estimated bias of estimators for $h(\beta) := \beta_2$.

DGP	n	Model 1 for $\omega^2(X;\gamma)$ OLS	WLS	ALS	MIN VAR	LIN COM	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ OLS	WLS	ALS	MIN VAR	LIN COM	MWLS -HC3
1(a)	25	.90	.96	.96	.93	.92	.98	.91	.96	.96	.94	.93	.99
	50	.93	.97	.97	.95	.95	.99	.94	.97	.97	.96	.96	.99
	100	.96	.99	.99	.98	.97	1.00	.97	.99	.99	.98	.98	1.00
	200	.98	.99	.99	.99	.99	1.00	.98	1.00	1.00	.99	.99	1.00
	400	.99	1.00	1.00	1.00	1.00	1.00	.99	1.00	1.00	1.00	1.00	1.00
1(b)	25	.99	.96	.98	.95	.94	.99	1.00	.97	.99	.96	.96	1.00
	50	1.06	.98	1.01	.99	.98	1.00	1.07	.98	1.01	.99	.98	1.00
	100	1.08	.99	1.00	1.00	.99	1.00	1.08	.99	1.00	1.00	1.00	1.00
	200	1.11	.99	1.00	1.00	1.00	1.00	1.11	1.00	1.00	1.00	1.00	1.00
	400	1.12	1.00	1.00	1.00	1.00	1.00	1.12	1.00	1.00	1.00	1.00	1.00
1(c)	25	1.28	.96	.99	.98	.97	1.03	1.30	.98	1.01	1.00	.99	1.03
	50	1.47	.98	1.00	1.00	.99	1.01	1.47	1.00	1.01	1.00	1.00	1.01
	100	1.52	.98	.98	.99	.99	1.00	1.51	.99	1.00	1.00	1.00	1.00
	200	1.55	.99	.99	.99	1.00	1.00	1.53	1.00	1.00	1.00	1.00	1.00
	400	1.57	1.00	1.00	1.00	1.00	1.00	1.55	1.01	1.01	1.01	1.01	1.00
1(d)	25	2.51	.88	.88	.90	.90	1.18	2.63	.95	.95	.95	.96	1.13
	50	3.61	.94	.94	.94	.94	1.08	3.64	.97	.97	.97	.98	1.06
	100	3.97	.94	.94	.94	.94	1.04	3.97	.99	.99	.99	.99	1.03
	200	4.24	.97	.97	.97	.97	1.01	4.13	1.00	1.00	1.00	1.00	1.01
	400	4.75	.98	.98	.98	.98	1.00	4.57	1.01	1.01	1.01	1.01	1.00
2(a)	25	2.20	1.04	1.04	1.04	1.05	1.22	2.02	1.07	1.07	1.06	1.06	1.14
	50	3.19	1.16	1.16	1.16	1.16	1.09	2.73	1.15	1.15	1.15	1.15	1.05
	100	3.59	1.24	1.24	1.24	1.24	1.05	2.88	1.20	1.20	1.20	1.20	1.02
	200	3.72	1.24	1.24	1.24	1.24	1.02	2.82	1.14	1.14	1.14	1.14	1.01
	400	4.07	1.29	1.29	1.29	1.29	1.01	3.02	1.17	1.17	1.17	1.17	1.00
2(b)	25	3.41	.92	.93	.92	.92	1.33	3.38	1.00	1.00	1.00	1.00	1.21
	50	5.47	1.04	1.04	1.04	1.04	1.13	5.05	1.06	1.06	1.06	1.06	1.08
	100	7.23	1.10	1.10	1.10	1.10	1.10	6.55	1.20	1.20	1.20	1.21	1.06
	200	9.46	1.29	1.29	1.29	1.30	1.06	7.96	1.37	1.37	1.37	1.37	1.03
	400	12.03	1.50	1.50	1.50	1.50	1.03	9.42	1.53	1.53	1.53	1.54	1.01
3(a)	25	1.05	.95	.99	.95	.95	.99	1.07	.96	1.00	.97	.96	1.00
	50	1.12	.98	1.02	1.00	.99	1.00	1.14	.99	1.02	1.00	.99	1.00
	100	1.15	.99	1.00	1.00	.99	1.00	1.18	.99	1.00	1.00	1.00	1.00
	200	1.16	.99	1.00	1.00	1.00	1.00	1.18	1.00	1.00	1.00	1.00	1.00
	400	1.20	1.00	1.00	1.00	1.00	1.00	1.22	1.00	1.00	1.00	1.00	1.00
3(b)	25	1.25	.94	.98	.96	.95	1.01	1.30	.95	.99	.98	.97	1.02
	50	1.42	.98	1.00	1.00	.99	1.01	1.46	.99	1.01	1.00	1.00	1.01
	100	1.45	.99	.99	1.00	.99	1.00	1.50	.99	.99	.99	1.00	1.00
	200	1.48	1.00	1.00	1.00	1.00	1.00	1.53	1.00	1.00	1.00	1.00	1.00
	400	1.55	1.00	1.00	1.00	1.00	1.00	1.60	1.00	1.00	1.00	1.00	1.00
4(a)	25	.99	.97	.98	.95	.94	.99	1.01	.98	.99	.97	.96	1.00
	50	1.04	.99	1.00	.99	.98	1.00	1.05	.99	1.00	1.00	.99	1.00
	100	1.07	1.00	1.00	1.01	1.00	1.00	1.08	1.00	1.00	1.01	1.00	1.00
	200	1.08	1.00	1.00	1.01	1.00	1.00	1.09	1.00	1.00	1.00	1.00	1.00
	400	1.10	1.00	1.00	1.01	1.00	1.00	1.11	1.00	1.00	1.00	1.00	1.00
4(b)	25	1.26	.97	1.04	.98	.96	1.01	1.31	.99	1.07	1.01	.99	1.02
	50	1.35	1.00	1.02	1.02	1.00	1.01	1.37	1.00	1.02	1.02	1.00	1.01
	100	1.39	1.02	1.02	1.03	1.00	1.00	1.42	1.00	1.01	1.01	1.00	1.00
	200	1.45	1.02	1.02	1.02	1.01	1.00	1.47	1.00	1.00	1.01	1.00	1.00
	400	1.47	1.02	1.02	1.02	1.01	1.00	1.50	1.01	1.01	1.01	1.00	1.00

Table 9: Ratio of estimated MSE of estimators for $h(\beta) := \beta_1$ with respect to estimated MSE of $h(\beta)$'s MWLS-HC1 estimator, i.e., MSE-Estimator/MSE-MWLS-HC1, in Model 1 and Model 2.

DGP	n	Model 1 for $\omega^2(X;\gamma)$ OLS	WLS	ALS	MIN VAR	LIN COM	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ OLS	WLS	ALS	MIN VAR	LIN COM	MWLS -HC3
1(a)	25	.91	.97	.97	.95	.94	.98	.93	.99	.99	.96	.96	1.00
	50	.94	.98	.98	.96	.96	.99	.95	.99	.99	.97	.97	1.00
	100	.97	.99	.99	.98	.98	1.00	.97	.99	.99	.98	.98	1.00
	200	.99	1.00	1.00	.99	.99	1.00	.99	1.00	1.00	.99	.99	1.00
	400	.99	1.00	1.00	.99	.99	1.00	.99	1.00	1.00	1.00	.99	1.00
1(b)	25	1.01	.98	1.00	.98	.97	1.00	1.02	.99	1.01	.99	.98	1.01
	50	1.05	.99	1.02	1.00	.99	1.00	1.05	1.00	1.02	1.00	.99	1.00
	100	1.07	1.00	1.01	1.01	1.00	1.00	1.07	1.00	1.01	1.01	1.00	1.00
	200	1.09	1.00	1.00	1.00	1.00	1.00	1.08	1.00	1.00	1.00	1.00	1.00
	400	1.10	1.00	1.00	1.00	1.00	1.00	1.10	1.00	1.00	1.00	1.00	1.00
1(c)	25	1.22	.99	1.01	1.01	1.00	1.02	1.21	.99	1.01	1.01	.99	1.02
	50	1.33	.99	1.00	1.00	1.00	1.01	1.31	.99	1.00	1.01	.99	1.00
	100	1.38	1.00	1.00	1.00	1.00	1.00	1.36	1.00	1.00	1.01	1.00	1.00
	200	1.41	1.00	1.00	1.00	1.00	1.00	1.39	1.00	1.00	1.00	1.00	1.00
	400	1.43	1.00	1.00	1.00	1.00	1.00	1.41	1.00	1.00	1.00	1.00	1.00
1(d)	25	1.75	.95	.95	.97	.97	1.08	1.77	.95	.95	.97	.97	1.05
	50	2.30	1.00	1.00	1.00	1.00	1.03	2.27	.98	.98	.98	.99	1.02
	100	2.67	.98	.98	.98	.98	1.02	2.63	.99	.99	.99	.99	1.01
	200	2.88	.99	.99	.99	.99	1.01	2.79	.99	.99	.99	1.00	1.00
	400	3.07	.99	.99	.99	1.00	1.00	2.94	1.00	1.00	1.00	1.00	1.00
2(a)	25	1.45	.98	.98	.99	.99	1.06	1.37	.98	.98	.99	.98	1.03
	50	1.76	1.02	1.02	1.02	1.03	1.02	1.61	1.00	1.00	1.00	1.00	1.01
	100	1.92	1.05	1.05	1.05	1.05	1.01	1.68	1.01	1.01	1.01	1.01	1.00
	200	1.96	1.04	1.04	1.04	1.04	1.00	1.71	1.00	1.00	1.00	1.00	1.00
	400	2.03	1.05	1.05	1.05	1.05	1.00	1.75	1.01	1.01	1.01	1.00	1.00
2(b)	25	1.89	.97	.97	.97	.98	1.13	1.78	.94	.95	.96	.97	1.06
	50	2.68	1.07	1.07	1.07	1.08	1.05	2.39	.99	.99	1.00	1.01	1.02
	100	3.53	1.12	1.12	1.12	1.12	1.05	3.04	1.06	1.06	1.06	1.07	1.02
	200	4.20	1.23	1.23	1.23	1.23	1.03	3.35	1.12	1.12	1.12	1.13	1.01
	400	4.75	1.27	1.27	1.27	1.27	1.01	3.67	1.16	1.16	1.16	1.16	1.00
3(a)	25	1.08	.98	1.02	.98	.98	1.00	1.10	.99	1.03	.99	.98	1.01
	50	1.15	.99	1.03	1.00	1.00	1.00	1.17	.99	1.03	1.01	1.00	1.00
	100	1.18	1.00	1.01	1.01	1.00	1.00	1.20	1.00	1.01	1.01	1.00	1.00
	200	1.18	1.00	1.00	1.00	1.00	1.00	1.20	1.00	1.00	1.00	1.00	1.00
	400	1.21	1.00	1.00	1.00	1.00	1.00	1.23	1.00	1.00	1.00	1.00	1.00
3(b)	25	1.27	.98	1.01	1.00	.99	1.01	1.31	.99	1.02	1.01	1.00	1.01
	50	1.41	1.00	1.01	1.01	1.01	1.01	1.45	1.00	1.02	1.01	1.01	1.00
	100	1.45	1.00	1.00	1.00	1.00	1.00	1.51	1.00	1.00	1.00	1.01	1.00
	200	1.48	1.00	1.00	1.00	1.00	1.00	1.54	1.00	1.00	1.00	1.01	1.00
	400	1.52	1.00	1.00	1.00	1.00	1.00	1.57	1.00	1.00	1.00	1.00	1.00
4(a)	25	.99	.98	.99	.97	.96	1.00	1.01	1.00	1.01	.99	.98	1.00
	50	1.04	1.00	1.00	1.00	.99	1.00	1.04	1.00	1.01	1.00	.99	1.00
	100	1.06	1.00	1.00	1.01	1.00	1.00	1.06	.99	.99	1.01	.99	1.00
	200	1.08	1.00	1.00	1.01	1.00	1.00	1.08	1.00	1.00	1.01	1.00	1.00
	400	1.08	1.00	1.00	1.01	1.00	1.00	1.08	1.00	1.00	1.01	1.00	1.00
4(b)	25	1.17	.99	1.03	1.01	.99	1.01	1.17	.99	1.03	1.01	.98	1.01
	50	1.24	.99	1.00	1.01	.99	1.00	1.23	.98	1.00	1.01	.98	1.00
	100	1.29	1.01	1.01	1.02	1.00	1.00	1.29	1.00	1.00	1.02	.99	1.00
	200	1.34	1.01	1.01	1.01	1.00	1.00	1.34	1.01	1.01	1.01	1.00	1.00
	400	1.36	1.01	1.01	1.01	1.00	1.00	1.36	1.01	1.01	1.01	1.00	1.00

Table 10: Ratio of estimated MSE of estimators for $h(\beta) := \beta_2$ with respect to estimated MSE of $h(\beta)$'s MWLS-HC1 estimator, i.e., MSE-Estimator/MSE-MWLS-HC1, in Model 1 and Model 2.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3
1(a)	25	.403	.432	.432	.418	.415	.449	.439	.426	.426	.415	.412	.443	.437
	50	.188	.196	.196	.193	.192	.203	.200	.194	.194	.192	.191	.200	.198
	100	.096	.099	.099	.098	.097	.100	.100	.098	.098	.097	.097	.099	.099
	200	.045	.045	.045	.045	.045	.045	.045	.045	.045	.045	.045	.045	.045
	400	.023	.023	.023	.023	.023	.023	.023	.023	.023	.023	.023	.023	.023
1(b)	25	.728	.708	.719	.698	.694	.735	.728	.706	.718	.697	.693	.726	.727
	50	.345	.318	.327	.320	.318	.324	.324	.318	.326	.319	.318	.323	.323
	100	.186	.171	.173	.172	.171	.172	.172	.171	.173	.172	.171	.172	.172
	200	.088	.079	.079	.080	.079	.080	.080	.080	.080	.080	.080	.080	.080
	400	.047	.041	.041	.041	.041	.042	.041	.042	.042	.042	.042	.042	.042
1(c)	25	1.603	1.199	1.237	1.232	1.218	1.251	1.289	1.213	1.248	1.229	1.218	1.234	1.272
	50	.784	.523	.530	.530	.528	.531	.539	.532	.539	.535	.535	.534	.539
	100	.453	.294	.294	.295	.296	.298	.300	.299	.299	.299	.300	.300	.301
	200	.217	.140	.140	.140	.140	.140	.141	.142	.142	.142	.142	.142	.142
	400	.116	.074	.074	.074	.074	.074	.074	.075	.075	.075	.075	.075	.075
1(d)	25	14.303	5.016	5.016	5.136	5.124	5.694	6.696	5.150	5.150	5.166	5.191	5.432	6.162
	50	7.346	1.907	1.907	1.907	1.919	2.037	2.196	1.961	1.961	1.961	1.970	2.016	2.129
	100	4.352	1.027	1.027	1.027	1.032	1.097	1.141	1.084	1.084	1.084	1.087	1.097	1.126
	200	2.147	.490	.490	.490	.491	.507	.513	.520	.520	.520	.520	.519	.523
	400	1.187	.246	.246	.246	.246	.250	.251	.263	.263	.263	.263	.260	.261
2(a)	25	.159	.075	.075	.075	.075	.072	.088	.084	.084	.084	.083	.079	.090
	50	.081	.029	.029	.029	.029	.025	.028	.034	.034	.034	.034	.030	.031
	100	.048	.017	.017	.017	.017	.013	.014	.020	.020	.020	.020	.017	.017
	200	.023	.008	.008	.008	.008	.006	.006	.009	.009	.009	.009	.008	.008
	400	.013	.004	.004	.004	.004	.003	.003	.005	.005	.005	.005	.004	.004
2(b)	25	.207	.056	.056	.056	.056	.061	.081	.061	.061	.061	.061	.061	.074
	50	.107	.020	.020	.020	.020	.020	.022	.022	.022	.022	.023	.021	.023
	100	.064	.010	.010	.010	.010	.009	.010	.012	.012	.012	.012	.010	.010
	200	.032	.004	.004	.004	.004	.003	.004	.006	.006	.006	.006	.004	.004
	400	.017	.002	.002	.002	.002	.001	.001	.003	.003	.003	.003	.002	.002
3(a)	25	.865	.789	.821	.789	.783	.827	.820	.780	.812	.782	.776	.809	.812
	50	.420	.368	.382	.374	.370	.374	.374	.364	.377	.368	.365	.369	.369
	100	.231	.198	.200	.201	.199	.200	.200	.194	.196	.196	.196	.196	.196
	200	.108	.093	.093	.093	.093	.093	.093	.092	.092	.092	.092	.092	.092
	400	.057	.047	.047	.047	.047	.048	.048	.047	.047	.047	.047	.047	.047
3(b)	25	1.521	1.138	1.190	1.171	1.156	1.216	1.224	1.113	1.161	1.140	1.129	1.168	1.186
	50	.756	.521	.534	.533	.528	.533	.536	.510	.520	.517	.514	.517	.520
	100	.417	.284	.285	.288	.286	.288	.288	.275	.276	.277	.277	.278	.278
	200	.203	.136	.136	.136	.137	.137	.136	.132	.132	.132	.133	.132	.132
	400	.110	.071	.071	.071	.071	.071	.071	.068	.068	.068	.068	.069	.069
4(a)	25	.547	.537	.543	.526	.523	.554	.547	.531	.538	.523	.520	.542	.542
	50	.266	.254	.256	.255	.251	.256	.255	.251	.254	.252	.250	.253	.253
	100	.143	.133	.133	.134	.133	.133	.133	.131	.132	.133	.132	.132	.132
	200	.068	.063	.063	.063	.063	.063	.063	.062	.062	.063	.062	.062	.062
	400	.037	.033	.033	.033	.033	.033	.033	.033	.033	.033	.033	.033	.033
4(b)	25	.957	.739	.796	.745	.734	.762	.768	.727	.781	.736	.725	.732	.746
	50	.480	.355	.364	.364	.356	.356	.358	.350	.359	.356	.350	.350	.353
	100	.273	.199	.200	.202	.197	.196	.196	.194	.194	.195	.192	.193	.193
	200	.136	.095	.095	.096	.094	.094	.094	.093	.093	.093	.092	.092	.092
	400	.071	.049	.049	.049	.049	.048	.048	.048	.048	.048	.048	.048	.048

Table 11: Estimated variance (MCVar) of estimators for $h(\beta) := \beta_1$.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$						Model 2 for $\omega^2(X;\gamma)$					
			WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3	WLS	ALS	MIN VAR	LIN COM	MWLS -HC1	MWLS -HC3
1(a)	25	.043	.046	.046	.045	.044	.047	.047	.046	.046	.045	.044	.046	.046
	50	.023	.024	.024	.024	.024	.025	.024	.024	.024	.024	.024	.024	.024
	100	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013
	200	.006	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007
	400	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
1(b)	25	.100	.097	.099	.096	.095	.098	.099	.097	.099	.096	.095	.098	.099
	50	.055	.052	.053	.052	.052	.052	.052	.052	.053	.052	.052	.052	.052
	100	.031	.029	.029	.029	.029	.029	.029	.029	.029	.029	.029	.029	.029
	200	.016	.015	.015	.015	.015	.015	.015	.015	.015	.015	.015	.015	.015
	400	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008
1(c)	25	.289	.234	.239	.238	.235	.237	.242	.237	.241	.240	.236	.239	.242
	50	.160	.119	.120	.121	.120	.121	.121	.121	.122	.123	.121	.122	.122
	100	.094	.068	.068	.069	.069	.068	.068	.069	.069	.069	.069	.069	.069
	200	.050	.036	.036	.036	.036	.036	.036	.036	.036	.036	.036	.036	.036
	400	.027	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
1(d)	25	3.471	1.891	1.891	1.918	1.919	1.986	2.150	1.877	1.877	1.910	1.906	1.966	2.059
	50	1.939	.841	.841	.842	.847	.844	.873	.838	.838	.840	.846	.853	.870
	100	1.145	.421	.421	.421	.423	.430	.439	.430	.430	.430	.432	.436	.440
	200	.596	.204	.204	.204	.205	.207	.209	.212	.212	.212	.213	.214	.215
	400	.319	.103	.103	.103	.103	.104	.104	.108	.108	.108	.108	.108	.109
2(a)	25	.032	.022	.022	.022	.022	.022	.024	.023	.023	.023	.023	.023	.024
	50	.019	.011	.011	.011	.011	.011	.011	.011	.011	.012	.011	.012	.012
	100	.011	.006	.006	.006	.006	.006	.006	.007	.007	.007	.007	.007	.007
	200	.006	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
	400	.003	.002	.002	.002	.002	.001	.001	.002	.002	.002	.002	.002	.002
2(b)	25	.052	.027	.027	.027	.027	.028	.031	.028	.028	.028	.028	.029	.031
	50	.029	.012	.012	.012	.012	.011	.012	.012	.012	.012	.012	.012	.013
	100	.017	.006	.006	.006	.006	.005	.005	.006	.006	.006	.006	.006	.006
	200	.009	.003	.003	.003	.003	.002	.002	.003	.003	.003	.003	.003	.003
	400	.005	.001	.001	.001	.001	.001	.001	.002	.002	.002	.002	.001	.001
3(a)	25	.139	.126	.131	.126	.125	.128	.129	.125	.131	.126	.125	.127	.128
	50	.076	.065	.067	.066	.065	.065	.066	.064	.067	.065	.065	.065	.065
	100	.044	.037	.038	.037	.037	.037	.037	.037	.037	.037	.037	.037	.037
	200	.022	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	400	.012	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
3(b)	25	.293	.226	.234	.231	.229	.232	.235	.222	.229	.227	.224	.225	.228
	50	.160	.113	.116	.115	.115	.114	.115	.111	.113	.112	.112	.111	.111
	100	.094	.064	.064	.065	.065	.064	.065	.062	.062	.062	.063	.062	.062
	200	.048	.032	.032	.032	.032	.032	.032	.031	.031	.031	.031	.031	.031
	400	.025	.017	.017	.017	.017	.017	.017	.016	.016	.016	.016	.016	.016
4(a)	25	.077	.076	.077	.075	.075	.077	.077	.076	.077	.076	.075	.076	.077
	50	.044	.042	.042	.042	.042	.042	.042	.042	.042	.042	.042	.042	.042
	100	.025	.024	.024	.024	.024	.024	.024	.024	.024	.024	.024	.024	.024
	200	.013	.012	.012	.012	.012	.012	.012	.012	.012	.012	.012	.012	.012
	400	.007	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006
4(b)	25	.176	.148	.155	.151	.149	.150	.152	.149	.155	.152	.148	.150	.152
	50	.102	.081	.083	.084	.082	.082	.083	.082	.083	.084	.082	.083	.083
	100	.061	.048	.048	.049	.048	.048	.048	.048	.048	.048	.047	.048	.048
	200	.033	.025	.025	.025	.025	.024	.024	.024	.024	.025	.024	.024	.024
	400	.017	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013

Table 12: Estimated variance (MCVar) of estimators for $h(\beta) := \beta_2$.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	.558	.542	.542	.519	.507	.487	.546	.546	.527	.505	.501
	50	.217	.215	.215	.210	.202	.202	.215	.215	.211	.205	.205
	100	.102	.101	.101	.100	.098	.098	.101	.101	.100	.099	.099
	200	.047	.047	.047	.046	.046	.046	.047	.047	.046	.046	.046
	400	.024	.024	.024	.023	.023	.023	.024	.024	.023	.023	.023
1(b)	25	.993	.912	.912	.866	.922	.845	.918	.918	.884	.867	.851
	50	.396	.354	.354	.347	.330	.331	.356	.356	.351	.337	.337
	100	.198	.176	.176	.175	.169	.169	.177	.177	.176	.172	.172
	200	.091	.082	.082	.081	.080	.080	.082	.082	.082	.081	.081
	400	.047	.042	.042	.042	.042	.042	.042	.042	.042	.042	.042
1(c)	25	2.188	1.575	1.575	1.516	1.787	1.572	1.583	1.583	1.552	1.579	1.521
	50	.911	.589	.589	.585	.552	.555	.597	.597	.595	.565	.567
	100	.479	.306	.306	.306	.293	.294	.312	.312	.312	.300	.301
	200	.224	.144	.144	.144	.140	.141	.146	.146	.146	.143	.143
	400	.118	.074	.074	.074	.073	.073	.075	.075	.075	.074	.074
1(d)	25	19.015	6.378	6.378	6.327	7.544	7.250	6.312	6.312	6.303	7.411	6.842
	50	8.465	2.045	2.045	2.045	1.985	2.055	2.122	2.122	2.122	2.045	2.089
	100	4.645	1.045	1.045	1.045	1.025	1.043	1.106	1.106	1.106	1.073	1.083
	200	2.210	.492	.492	.492	.484	.486	.524	.524	.524	.513	.514
	400	1.192	.248	.248	.248	.245	.246	.265	.265	.265	.260	.261
2(a)	25	.219	.097	.097	.097	.105	.101	.106	.106	.106	.103	.102
	50	.093	.030	.030	.030	.025	.026	.035	.035	.035	.030	.030
	100	.051	.016	.016	.016	.013	.013	.020	.020	.020	.017	.017
	200	.024	.008	.008	.008	.006	.006	.010	.010	.010	.008	.008
	400	.013	.004	.004	.004	.003	.003	.005	.005	.005	.004	.004
2(b)	25	.275	.060	.060	.060	.064	.074	.067	.067	.067	.065	.070
	50	.124	.020	.020	.020	.017	.018	.023	.023	.023	.020	.021
	100	.069	.009	.009	.009	.008	.008	.012	.012	.012	.009	.010
	200	.033	.004	.004	.004	.003	.003	.006	.006	.006	.004	.004
	400	.018	.002	.002	.002	.001	.001	.003	.003	.003	.002	.002
3(a)	25	1.174	1.025	1.025	.967	1.075	.968	1.017	1.017	.980	.988	.959
	50	.484	.408	.408	.399	.382	.383	.404	.404	.399	.384	.384
	100	.243	.204	.204	.202	.195	.196	.201	.201	.200	.195	.195
	200	.113	.095	.095	.095	.093	.093	.094	.094	.093	.092	.092
	400	.059	.049	.049	.049	.048	.048	.048	.048	.048	.048	.048
3(b)	25	2.044	1.497	1.497	1.416	1.666	1.468	1.457	1.457	1.417	1.488	1.416
	50	.867	.579	.579	.572	.543	.545	.566	.566	.563	.538	.539
	100	.451	.297	.297	.296	.283	.284	.288	.288	.288	.279	.279
	200	.211	.139	.139	.139	.136	.136	.135	.135	.135	.133	.133
	400	.112	.072	.072	.072	.071	.071	.069	.069	.069	.069	.069
4(a)	25	.748	.693	.693	.654	.702	.644	.693	.693	.665	.658	.646
	50	.308	.283	.283	.274	.265	.265	.282	.282	.276	.268	.268
	100	.152	.139	.139	.137	.133	.133	.138	.138	.137	.134	.134
	200	.071	.065	.065	.064	.063	.063	.064	.064	.064	.063	.063
	400	.036	.033	.033	.033	.033	.033	.033	.033	.033	.032	.032
4(b)	25	1.286	.965	.965	.907	1.087	.955	.943	.943	.915	.959	.915
	50	.559	.399	.399	.390	.372	.373	.392	.392	.387	.373	.374
	100	.295	.209	.209	.207	.196	.196	.202	.202	.202	.195	.196
	200	.140	.098	.098	.098	.094	.094	.095	.095	.095	.093	.093
	400	.073	.051	.051	.050	.049	.049	.049	.049	.049	.048	.048

Table 13: Estimated variance (ASVar) of estimators for $h(\beta) := \beta_1$.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	.058	.057	.057	.055	.052	.052	.057	.057	.055	.053	.053
	50	.027	.026	.026	.026	.025	.025	.026	.026	.026	.025	.025
	100	.014	.014	.014	.013	.013	.013	.014	.014	.013	.013	.013
	200	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007	.007
	400	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
1(b)	25	.132	.122	.122	.117	.115	.114	.122	.122	.118	.115	.114
	50	.063	.057	.057	.056	.054	.054	.057	.057	.056	.054	.054
	100	.034	.030	.030	.030	.029	.029	.031	.031	.030	.030	.030
	200	.017	.015	.015	.015	.015	.015	.015	.015	.015	.015	.015
	400	.009	.008	.008	.008	.008	.008	.008	.008	.008	.008	.008
1(c)	25	.376	.291	.291	.284	.288	.281	.292	.292	.287	.277	.275
	50	.184	.132	.132	.131	.126	.126	.133	.133	.132	.127	.128
	100	.101	.072	.072	.071	.069	.069	.072	.072	.072	.070	.071
	200	.051	.036	.036	.036	.035	.035	.036	.036	.036	.036	.036
	400	.027	.018	.018	.018	.018	.018	.019	.019	.019	.019	.019
1(d)	25	4.468	2.268	2.268	2.240	2.291	2.304	2.246	2.246	2.231	2.221	2.193
	50	2.194	.871	.871	.870	.839	.853	.889	.889	.889	.856	.864
	100	1.217	.427	.427	.427	.418	.422	.444	.444	.444	.435	.437
	200	.612	.209	.209	.209	.206	.207	.219	.219	.219	.217	.217
	400	.322	.103	.103	.103	.102	.103	.109	.109	.109	.108	.108
2(a)	25	.042	.026	.026	.026	.026	.026	.028	.028	.028	.027	.027
	50	.021	.011	.011	.011	.011	.011	.012	.012	.012	.012	.012
	100	.012	.006	.006	.006	.006	.006	.007	.007	.007	.007	.007
	200	.006	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
	400	.003	.002	.002	.002	.001	.001	.002	.002	.002	.002	.002
2(b)	25	.067	.030	.030	.030	.029	.031	.032	.032	.032	.030	.031
	50	.033	.012	.012	.012	.010	.011	.013	.013	.013	.012	.012
	100	.019	.005	.005	.005	.005	.005	.006	.006	.006	.006	.006
	200	.009	.003	.003	.003	.002	.002	.003	.003	.003	.003	.003
	400	.005	.001	.001	.001	.001	.001	.002	.002	.002	.001	.001
3(a)	25	.181	.156	.156	.150	.150	.146	.154	.154	.150	.145	.145
	50	.086	.072	.072	.071	.068	.068	.071	.071	.070	.068	.068
	100	.047	.039	.039	.039	.037	.037	.038	.038	.038	.037	.037
	200	.023	.019	.019	.019	.019	.019	.019	.019	.019	.019	.019
	400	.012	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010
3(b)	25	.383	.282	.282	.274	.283	.273	.274	.274	.269	.263	.260
	50	.183	.124	.124	.124	.118	.118	.121	.121	.120	.115	.116
	100	.099	.067	.067	.067	.064	.064	.065	.065	.065	.063	.063
	200	.049	.033	.033	.033	.032	.032	.032	.032	.032	.032	.032
	400	.026	.017	.017	.017	.017	.017	.016	.016	.016	.016	.016
4(a)	25	.102	.095	.095	.091	.090	.088	.095	.095	.091	.089	.089
	50	.050	.046	.046	.045	.044	.044	.046	.046	.045	.044	.044
	100	.027	.025	.025	.024	.024	.024	.025	.025	.024	.024	.024
	200	.013	.012	.012	.012	.012	.012	.012	.012	.012	.012	.012
	400	.007	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006
4(b)	25	.231	.186	.186	.178	.185	.179	.185	.185	.180	.175	.174
	50	.118	.092	.092	.091	.087	.088	.092	.092	.091	.088	.088
	100	.066	.050	.050	.050	.048	.048	.050	.050	.050	.048	.048
	200	.033	.025	.025	.025	.025	.025	.025	.025	.025	.025	.025
	400	.017	.013	.013	.013	.013	.013	.013	.013	.013	.013	.013

Table 14: Estimated variance (ASVar) of estimators for $h(\beta) := \beta_2$.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	4.6	6.0	6.0	6.1	9.0	8.1	5.7	5.7	5.8	7.6	7.2
	50	4.5	5.3	5.3	5.4	6.9	6.5	5.1	5.1	5.2	6.2	6.0
	100	4.8	5.3	5.3	5.2	5.9	5.9	5.2	5.2	5.2	5.7	5.7
	200	4.7	4.8	4.8	4.8	5.0	5.0	4.7	4.7	4.8	4.9	4.9
	400	4.9	5.0	5.0	5.0	5.1	5.1	5.0	5.0	5.0	5.1	5.1
1(b)	25	3.6	5.9	6.1	6.1	8.4	7.7	5.5	5.6	5.6	7.0	6.8
	50	4.4	5.5	5.8	5.7	6.8	6.7	5.3	5.5	5.4	6.3	6.2
	100	4.5	5.3	5.4	5.4	6.0	5.9	5.1	5.2	5.2	5.8	5.7
	200	4.8	4.8	4.8	4.8	5.0	5.0	4.8	4.8	4.8	5.0	5.0
	400	4.7	4.9	4.9	4.9	5.1	5.1	4.9	4.9	4.9	5.0	5.0
1(c)	25	2.7	5.5	5.8	5.7	7.4	6.6	5.0	5.3	5.0	5.9	5.8
	50	3.8	5.2	5.3	5.3	6.3	6.3	4.8	5.0	4.9	5.8	5.7
	100	4.7	5.3	5.3	5.4	6.1	6.1	5.2	5.2	5.2	5.8	5.8
	200	4.8	5.2	5.2	5.2	5.5	5.5	5.1	5.1	5.1	5.2	5.3
	400	4.7	5.1	5.1	5.1	5.2	5.2	5.1	5.1	5.1	5.1	5.1
1(d)	25	2.3	4.3	4.3	4.3	5.2	4.9	3.7	3.7	3.8	4.6	4.6
	50	4.2	5.1	5.1	5.1	5.9	6.2	4.6	4.6	4.6	5.2	5.5
	100	4.8	5.5	5.5	5.5	6.4	6.4	5.2	5.2	5.2	5.6	5.8
	200	5.0	5.4	5.4	5.4	6.0	6.0	5.0	5.0	5.0	5.2	5.1
	400	5.3	5.1	5.1	5.1	5.4	5.4	5.0	5.0	5.0	5.0	5.0
2(a)	25	1.8	2.2	2.2	2.2	3.1	3.3	2.3	2.3	2.3	2.9	3.3
	50	3.6	4.2	4.2	4.2	4.8	5.1	4.1	4.1	4.1	4.5	4.7
	100	4.5	4.7	4.7	4.7	4.9	5.2	4.6	4.6	4.6	4.7	4.8
	200	4.9	4.8	4.8	4.8	4.8	4.9	4.9	4.9	4.9	4.7	4.8
	400	5.2	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.8	4.8
2(b)	25	2.1	2.7	2.8	2.7	3.5	3.6	3.2	3.2	3.2	3.8	4.0
	50	4.0	5.0	5.0	5.0	5.8	6.0	4.6	4.6	4.6	5.5	5.7
	100	4.9	5.2	5.2	5.2	5.9	6.2	4.8	4.8	4.8	5.2	5.3
	200	5.0	4.9	4.9	4.9	5.2	5.3	4.8	4.8	4.8	4.7	4.6
	400	4.7	5.0	5.0	5.0	4.8	4.9	5.0	5.0	5.0	4.5	4.5
3(a)	25	3.3	5.9	6.3	6.1	8.3	7.3	5.4	5.8	5.6	6.7	6.6
	50	4.3	5.6	5.9	5.8	6.9	6.6	5.4	5.7	5.5	6.4	6.3
	100	4.8	5.6	5.8	5.8	6.1	6.0	5.3	5.5	5.4	5.9	5.8
	200	4.8	5.3	5.4	5.5	5.6	5.6	5.2	5.2	5.3	5.5	5.5
	400	4.7	4.9	4.9	4.9	5.1	5.1	4.9	4.9	4.9	5.1	5.1
3(b)	25	2.9	5.8	6.3	6.0	7.8	6.9	5.4	5.8	5.5	6.7	6.3
	50	4.0	5.5	5.7	5.7	6.7	6.5	5.3	5.4	5.4	6.1	6.0
	100	4.5	5.5	5.5	5.6	6.2	6.1	5.4	5.4	5.4	5.8	5.8
	200	4.8	4.9	4.9	4.9	5.3	5.3	4.8	4.8	4.8	5.0	5.0
	400	5.0	5.1	5.1	5.1	5.2	5.2	5.0	5.0	5.0	5.2	5.2
4(a)	25	3.5	5.9	6.0	6.1	8.1	7.3	5.5	5.7	5.7	6.7	6.5
	50	4.2	5.3	5.4	5.5	6.5	6.3	5.1	5.1	5.2	6.0	5.8
	100	4.7	5.3	5.3	5.6	5.8	5.8	5.3	5.3	5.5	5.6	5.7
	200	4.9	5.0	5.0	5.1	5.1	5.1	4.9	4.9	5.0	5.1	5.1
	400	5.2	5.4	5.4	5.4	5.5	5.5	5.3	5.3	5.3	5.3	5.3
4(b)	25	2.8	5.7	6.5	5.9	7.0	6.3	5.1	5.9	5.2	5.9	5.7
	50	4.0	5.6	5.8	5.8	6.2	6.1	5.2	5.4	5.4	5.7	5.7
	100	4.6	5.6	5.6	5.7	6.0	5.9	5.4	5.4	5.4	5.6	5.5
	200	4.9	5.1	5.1	5.1	5.2	5.2	5.0	5.0	5.0	5.1	5.1
	400	4.8	4.9	4.9	4.9	5.1	5.1	4.9	4.9	4.9	5.0	5.0

Table 15: Empirical size (in %) of two-sided 5% Wald test for $h(\beta) := \beta_1$ using t_{n-2} critical value as in Romano and Wolf (2017).

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	4.0	5.0	5.0	5.1	6.4	6.0	4.8	4.8	4.8	5.6	5.6
	50	4.6	5.1	5.1	5.2	5.9	5.8	5.0	5.0	5.1	5.7	5.6
	100	4.6	5.0	5.0	5.0	5.5	5.5	5.0	5.0	5.0	5.4	5.3
	200	4.9	5.1	5.1	5.1	5.3	5.3	5.1	5.1	5.1	5.2	5.2
	400	5.0	5.1	5.1	5.1	5.2	5.2	5.1	5.1	5.1	5.2	5.2
1(b)	25	3.6	4.3	4.5	4.5	5.4	5.2	4.2	4.3	4.4	5.0	4.9
	50	4.4	5.0	5.2	5.3	5.7	5.7	4.9	5.1	5.1	5.5	5.5
	100	4.7	5.0	5.1	5.1	5.4	5.4	5.0	5.1	5.1	5.2	5.2
	200	5.3	5.2	5.2	5.3	5.5	5.5	5.1	5.1	5.3	5.5	5.5
	400	4.8	4.7	4.7	4.7	4.8	4.8	4.6	4.6	4.7	4.8	4.8
1(c)	25	3.6	4.2	4.5	4.5	5.0	5.1	4.1	4.3	4.3	4.8	4.8
	50	4.4	4.5	4.6	4.7	5.2	5.2	4.5	4.6	4.7	5.1	5.1
	100	5.0	4.9	4.9	5.0	5.3	5.3	4.9	4.9	4.9	5.1	5.1
	200	5.2	5.0	5.0	5.0	5.1	5.2	5.0	5.0	5.0	5.1	5.1
	400	5.3	5.4	5.4	5.4	5.5	5.5	5.4	5.4	5.4	5.5	5.5
1(d)	25	4.0	4.4	4.4	4.6	5.2	5.4	4.3	4.3	4.5	5.3	5.5
	50	5.0	5.2	5.2	5.2	5.6	5.8	4.9	4.9	5.0	5.7	5.7
	100	4.9	5.5	5.5	5.5	5.8	5.9	5.1	5.1	5.1	5.3	5.4
	200	5.2	4.9	4.9	4.9	5.2	5.3	4.8	4.8	4.8	4.9	5.0
	400	5.1	5.1	5.1	5.1	5.3	5.3	5.1	5.1	5.1	5.2	5.2
2(a)	25	3.6	3.7	3.7	3.8	4.3	4.6	3.8	3.8	3.9	4.5	4.6
	50	4.6	4.8	4.8	4.9	5.2	5.2	4.7	4.7	4.8	5.2	5.2
	100	4.7	4.8	4.8	4.8	5.0	5.1	4.8	4.8	4.8	5.0	5.0
	200	4.9	4.9	4.9	4.9	5.1	5.1	4.9	4.9	4.9	5.1	5.1
	400	5.0	5.0	5.0	5.0	5.0	5.1	5.0	5.0	5.0	5.0	5.0
2(b)	25	4.1	4.5	4.5	4.6	5.4	5.6	4.5	4.5	4.7	5.8	5.9
	50	4.8	5.1	5.1	5.1	5.9	6.0	5.0	5.0	5.1	6.0	6.1
	100	5.1	5.3	5.3	5.3	5.8	5.9	5.1	5.1	5.1	5.6	5.5
	200	5.4	5.4	5.4	5.4	5.4	5.4	5.3	5.3	5.3	5.3	5.3
	400	5.1	5.4	5.4	5.4	5.2	5.3	5.3	5.3	5.3	5.3	5.3
3(a)	25	3.7	4.7	5.0	4.9	5.8	5.8	4.6	4.9	4.7	5.2	5.3
	50	4.4	4.7	5.0	4.9	5.4	5.4	4.7	5.1	4.9	5.3	5.3
	100	4.9	5.1	5.2	5.2	5.6	5.6	5.0	5.1	5.1	5.4	5.4
	200	4.9	4.8	4.8	4.9	5.1	5.0	4.8	4.8	4.8	5.0	5.0
	400	5.1	5.2	5.2	5.2	5.3	5.3	5.2	5.2	5.2	5.2	5.3
3(b)	25	3.6	4.5	4.9	4.7	5.3	5.2	4.4	4.8	4.5	5.2	5.1
	50	4.6	5.0	5.2	5.2	5.5	5.5	4.9	5.0	5.1	5.3	5.3
	100	5.0	5.2	5.2	5.3	5.7	5.7	5.1	5.1	5.2	5.4	5.4
	200	5.1	5.1	5.1	5.1	5.5	5.4	5.0	5.0	5.0	5.2	5.2
	400	4.8	4.9	4.9	4.9	5.1	5.1	4.9	4.9	4.9	5.0	5.0
4(a)	25	3.6	4.5	4.6	4.7	5.5	5.3	4.5	4.5	4.6	5.0	5.0
	50	4.3	4.8	4.9	5.0	5.4	5.4	4.8	4.8	5.0	5.3	5.3
	100	4.7	5.0	4.9	5.2	5.4	5.3	4.9	4.9	5.1	5.2	5.2
	200	5.0	4.9	4.9	5.1	5.1	5.1	4.9	4.9	5.0	5.1	5.1
	400	5.0	5.0	5.0	5.1	5.1	5.1	4.9	4.9	5.1	5.0	5.0
4(b)	25	3.5	4.2	4.7	4.6	4.9	4.9	4.1	4.5	4.4	4.8	4.8
	50	4.3	4.4	4.6	4.8	5.1	5.1	4.3	4.5	4.7	5.0	5.0
	100	4.8	5.1	5.1	5.3	5.3	5.3	4.9	4.9	5.2	5.2	5.2
	200	5.1	5.1	5.1	5.1	5.4	5.3	4.9	4.9	4.9	5.1	5.2
	400	5.2	5.2	5.2	5.2	5.2	5.2	5.1	5.1	5.1	5.2	5.2

Table 16: Empirical size (in %) of two-sided 5% Wald test for $h(\beta) := \beta_2$ using t_{n-2} critical value as in Romano and Wolf (2017).

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	5.6	7.0	7.0	7.2	10.3	9.4	6.8	6.8	6.9	8.9	8.4
	50	5.0	5.8	5.8	5.9	7.4	7.1	5.6	5.6	5.7	6.8	6.7
	100	5.2	5.5	5.5	5.5	6.2	6.2	5.5	5.5	5.5	6.0	5.9
	200	4.8	4.9	4.9	5.0	5.2	5.1	4.9	4.9	4.9	5.1	5.1
	400	5.0	5.1	5.1	5.1	5.2	5.2	5.1	5.1	5.1	5.2	5.1
1(b)	25	4.7	7.0	7.2	7.3	9.7	8.9	6.6	6.8	6.8	8.2	8.0
	50	4.9	6.1	6.5	6.3	7.5	7.3	5.8	6.0	5.9	7.0	6.8
	100	4.8	5.5	5.6	5.7	6.4	6.2	5.5	5.6	5.6	6.1	6.1
	200	4.9	4.9	4.9	5.0	5.2	5.1	5.0	5.0	5.0	5.1	5.1
	400	4.8	5.0	5.0	5.0	5.2	5.2	4.9	4.9	4.9	5.1	5.1
1(c)	25	3.6	6.6	7.0	6.8	8.5	7.8	6.0	6.4	6.1	7.2	7.0
	50	4.3	5.8	5.9	5.9	6.9	6.8	5.4	5.5	5.5	6.2	6.3
	100	5.0	5.7	5.7	5.7	6.4	6.4	5.5	5.5	5.5	6.1	6.1
	200	5.0	5.3	5.3	5.3	5.7	5.7	5.2	5.2	5.2	5.4	5.4
	400	4.8	5.2	5.2	5.2	5.3	5.3	5.2	5.2	5.2	5.2	5.2
1(d)	25	3.1	5.3	5.3	5.4	6.4	6.2	4.8	4.8	4.9	5.7	5.9
	50	4.7	5.6	5.6	5.6	6.6	6.7	5.1	5.1	5.1	5.9	6.0
	100	5.1	5.8	5.8	5.8	6.7	6.8	5.5	5.5	5.5	5.9	6.2
	200	5.1	5.6	5.6	5.6	6.1	6.2	5.1	5.1	5.1	5.3	5.3
	400	5.3	5.2	5.2	5.2	5.4	5.5	5.0	5.0	5.0	5.1	5.1
2(a)	25	2.6	3.1	3.1	3.2	4.2	4.4	3.3	3.3	3.3	3.9	4.3
	50	4.1	4.8	4.8	4.8	5.4	5.7	4.6	4.6	4.6	5.1	5.4
	100	4.8	4.9	4.9	4.9	5.2	5.5	4.9	4.9	4.9	5.0	5.2
	200	5.1	4.9	4.9	4.9	5.0	5.1	5.0	5.0	5.0	4.9	4.9
	400	5.3	5.0	5.0	5.0	5.0	5.0	4.9	4.9	4.9	4.8	4.8
2(b)	25	2.9	3.8	3.8	3.8	4.6	4.9	4.3	4.3	4.3	4.9	5.2
	50	4.4	5.6	5.6	5.6	6.5	6.7	5.3	5.3	5.3	6.3	6.3
	100	5.2	5.5	5.5	5.5	6.2	6.5	5.1	5.1	5.1	5.5	5.7
	200	5.1	5.1	5.1	5.1	5.3	5.5	5.0	5.0	5.0	4.8	4.8
	400	4.8	5.1	5.1	5.1	4.9	5.0	5.0	5.0	5.0	4.6	4.5
3(a)	25	4.2	7.1	7.6	7.3	9.5	8.5	6.7	7.1	6.7	8.1	7.8
	50	4.7	6.3	6.6	6.5	7.4	7.3	6.0	6.3	6.1	7.0	6.9
	100	5.2	5.9	6.1	6.1	6.5	6.3	5.6	5.7	5.7	6.1	6.1
	200	5.0	5.5	5.5	5.6	5.8	5.8	5.4	5.4	5.4	5.7	5.7
	400	4.8	4.9	4.9	5.0	5.2	5.2	4.9	4.9	4.9	5.2	5.2
3(b)	25	3.8	6.9	7.4	7.2	9.2	8.3	6.4	6.9	6.6	7.9	7.5
	50	4.4	6.2	6.4	6.4	7.2	7.1	5.8	6.0	5.9	6.7	6.6
	100	4.8	5.8	5.8	5.9	6.5	6.4	5.7	5.7	5.7	6.1	6.1
	200	4.9	5.1	5.1	5.1	5.4	5.4	5.0	5.0	5.0	5.2	5.2
	400	5.0	5.2	5.2	5.2	5.3	5.3	5.1	5.1	5.1	5.3	5.3
4(a)	25	4.4	7.1	7.2	7.3	9.3	8.5	6.6	6.7	6.8	8.0	7.8
	50	4.7	6.0	6.0	6.2	7.1	6.9	5.7	5.7	5.9	6.5	6.4
	100	5.0	5.5	5.5	5.8	6.1	6.1	5.5	5.5	5.7	6.0	5.9
	200	5.0	5.1	5.1	5.3	5.2	5.3	5.0	5.0	5.1	5.2	5.2
	400	5.3	5.4	5.4	5.5	5.5	5.5	5.3	5.3	5.4	5.4	5.4
4(b)	25	3.7	6.8	7.6	7.0	8.2	7.5	6.2	7.0	6.4	7.1	6.9
	50	4.4	6.2	6.4	6.4	6.8	6.7	5.8	6.1	6.0	6.2	6.2
	100	4.8	5.9	5.9	6.0	6.3	6.1	5.7	5.7	5.7	5.8	5.8
	200	5.0	5.2	5.2	5.3	5.3	5.3	5.2	5.2	5.2	5.3	5.3
	400	4.9	5.0	5.0	5.0	5.1	5.1	5.0	5.0	5.0	5.1	5.1

Table 17: Empirical size (in %) of two-sided 5% Wald test for $h(\beta) := \beta_1$ using $N(0,1)$ critical value as in (20) of our paper.

DGP	n	OLS	Model 1 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3	Model 2 for $\omega^2(X;\gamma)$ WLS	ALS	MIN VAR	MWLS -HC1	MWLS -HC3
1(a)	25	5.0	6.1	6.1	6.2	7.5	7.2	5.8	5.8	5.9	6.7	6.7
	50	5.0	5.6	5.6	5.7	6.5	6.4	5.5	5.5	5.6	6.3	6.2
	100	4.9	5.2	5.2	5.2	5.8	5.7	5.3	5.3	5.2	5.6	5.6
	200	5.0	5.3	5.3	5.2	5.4	5.4	5.3	5.3	5.2	5.4	5.4
	400	5.1	5.2	5.2	5.2	5.2	5.2	5.1	5.1	5.1	5.3	5.3
1(b)	25	4.5	5.5	5.6	5.7	6.7	6.5	5.4	5.4	5.5	6.2	6.2
	50	4.9	5.5	5.8	5.8	6.3	6.3	5.4	5.6	5.7	6.0	6.0
	100	4.9	5.3	5.4	5.4	5.7	5.7	5.2	5.3	5.3	5.5	5.5
	200	5.4	5.4	5.4	5.5	5.6	5.7	5.4	5.4	5.5	5.6	5.6
	400	4.9	4.7	4.7	4.8	4.9	4.9	4.7	4.7	4.7	4.8	4.8
1(c)	25	4.6	5.2	5.5	5.5	6.2	6.3	5.1	5.3	5.4	6.0	6.1
	50	5.0	5.1	5.2	5.3	5.8	5.8	5.0	5.1	5.2	5.6	5.6
	100	5.2	5.2	5.2	5.2	5.6	5.6	5.1	5.1	5.2	5.5	5.4
	200	5.3	5.2	5.2	5.2	5.3	5.3	5.1	5.1	5.1	5.3	5.2
	400	5.4	5.5	5.5	5.5	5.6	5.6	5.5	5.5	5.5	5.5	5.5
1(d)	25	5.0	5.6	5.6	5.8	6.5	6.7	5.4	5.4	5.6	6.7	6.8
	50	5.6	5.7	5.7	5.7	6.2	6.4	5.5	5.5	5.5	6.2	6.3
	100	5.2	5.8	5.8	5.8	6.1	6.1	5.3	5.3	5.3	5.6	5.7
	200	5.3	5.1	5.1	5.1	5.4	5.4	4.9	4.9	4.9	5.1	5.1
	400	5.1	5.2	5.2	5.2	5.3	5.4	5.1	5.1	5.1	5.2	5.3
2(a)	25	4.6	4.7	4.7	4.8	5.5	5.8	4.8	4.8	4.9	5.6	5.9
	50	5.1	5.3	5.3	5.3	5.8	5.9	5.3	5.3	5.4	5.8	5.9
	100	5.0	5.1	5.1	5.1	5.3	5.4	5.0	5.0	5.0	5.2	5.2
	200	5.0	5.1	5.1	5.1	5.3	5.3	5.0	5.0	5.1	5.2	5.3
	400	5.1	5.0	5.0	5.0	5.2	5.1	5.1	5.1	5.1	5.1	5.1
2(b)	25	5.2	5.6	5.6	5.7	6.6	6.9	5.7	5.7	5.9	7.2	7.2
	50	5.3	5.6	5.6	5.6	6.6	6.6	5.6	5.6	5.6	6.7	6.7
	100	5.3	5.6	5.6	5.6	6.1	6.2	5.4	5.4	5.4	5.8	5.8
	200	5.6	5.5	5.5	5.5	5.5	5.5	5.4	5.4	5.4	5.5	5.5
	400	5.2	5.4	5.4	5.4	5.2	5.3	5.3	5.3	5.3	5.4	5.4
3(a)	25	4.7	5.8	6.2	6.0	7.1	7.1	5.6	6.0	5.8	6.4	6.5
	50	4.9	5.3	5.7	5.5	6.0	6.0	5.2	5.6	5.5	5.9	5.9
	100	5.1	5.3	5.4	5.4	5.8	5.8	5.2	5.3	5.3	5.7	5.7
	200	5.0	5.0	5.0	5.0	5.2	5.2	4.9	4.9	5.0	5.1	5.1
	400	5.2	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3
3(b)	25	4.8	5.7	6.1	5.9	6.6	6.4	5.4	5.9	5.7	6.3	6.2
	50	5.2	5.6	5.8	5.8	6.1	6.1	5.4	5.6	5.6	5.9	5.9
	100	5.2	5.5	5.5	5.6	6.0	5.9	5.4	5.4	5.4	5.7	5.7
	200	5.2	5.2	5.2	5.3	5.6	5.6	5.1	5.1	5.1	5.4	5.4
	400	4.9	5.0	5.0	5.0	5.2	5.2	5.0	5.0	5.0	5.1	5.1
4(a)	25	4.5	5.6	5.6	5.8	6.6	6.5	5.5	5.6	5.7	6.1	6.1
	50	4.8	5.4	5.5	5.6	6.1	6.0	5.3	5.4	5.6	5.9	5.9
	100	5.0	5.2	5.2	5.5	5.7	5.7	5.2	5.2	5.5	5.5	5.6
	200	5.1	5.0	5.0	5.2	5.3	5.3	5.0	5.0	5.2	5.2	5.2
	400	5.1	5.0	5.0	5.1	5.2	5.2	5.0	5.0	5.2	5.1	5.1
4(b)	25	4.6	5.3	5.9	5.8	6.1	6.1	5.1	5.7	5.5	6.0	6.0
	50	4.8	4.9	5.1	5.3	5.6	5.7	4.8	5.0	5.2	5.6	5.6
	100	5.1	5.3	5.3	5.5	5.6	5.6	5.2	5.2	5.5	5.5	5.5
	200	5.3	5.2	5.2	5.3	5.5	5.5	5.1	5.1	5.1	5.3	5.3
	400	5.3	5.3	5.3	5.3	5.3	5.3	5.2	5.2	5.2	5.3	5.3

Table 18: Empirical size (in %) of two-sided 5% Wald test for $h(\beta) := \beta_2$ using $N(0,1)$ critical value as in (20) of our paper.

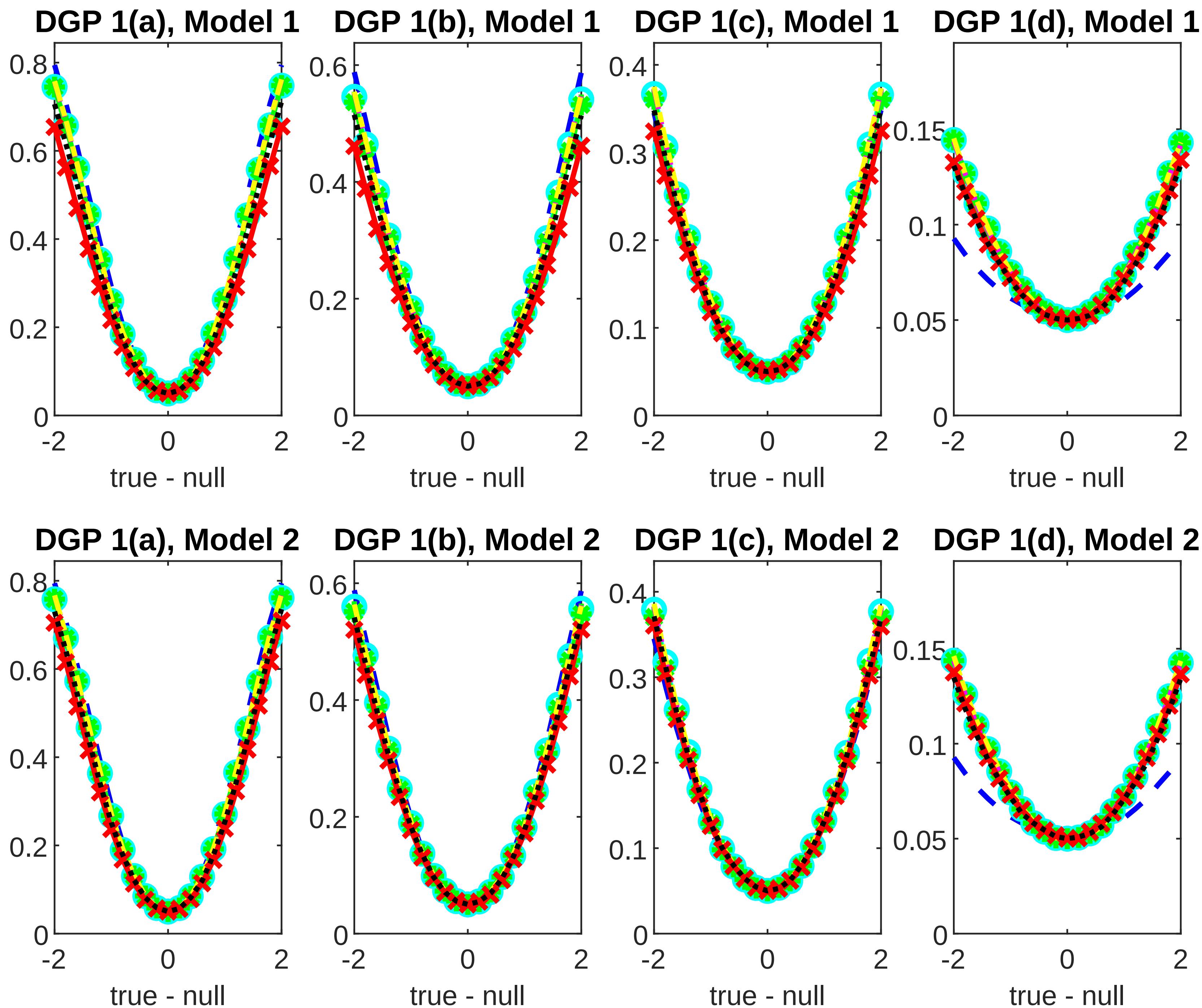

Figure 10: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

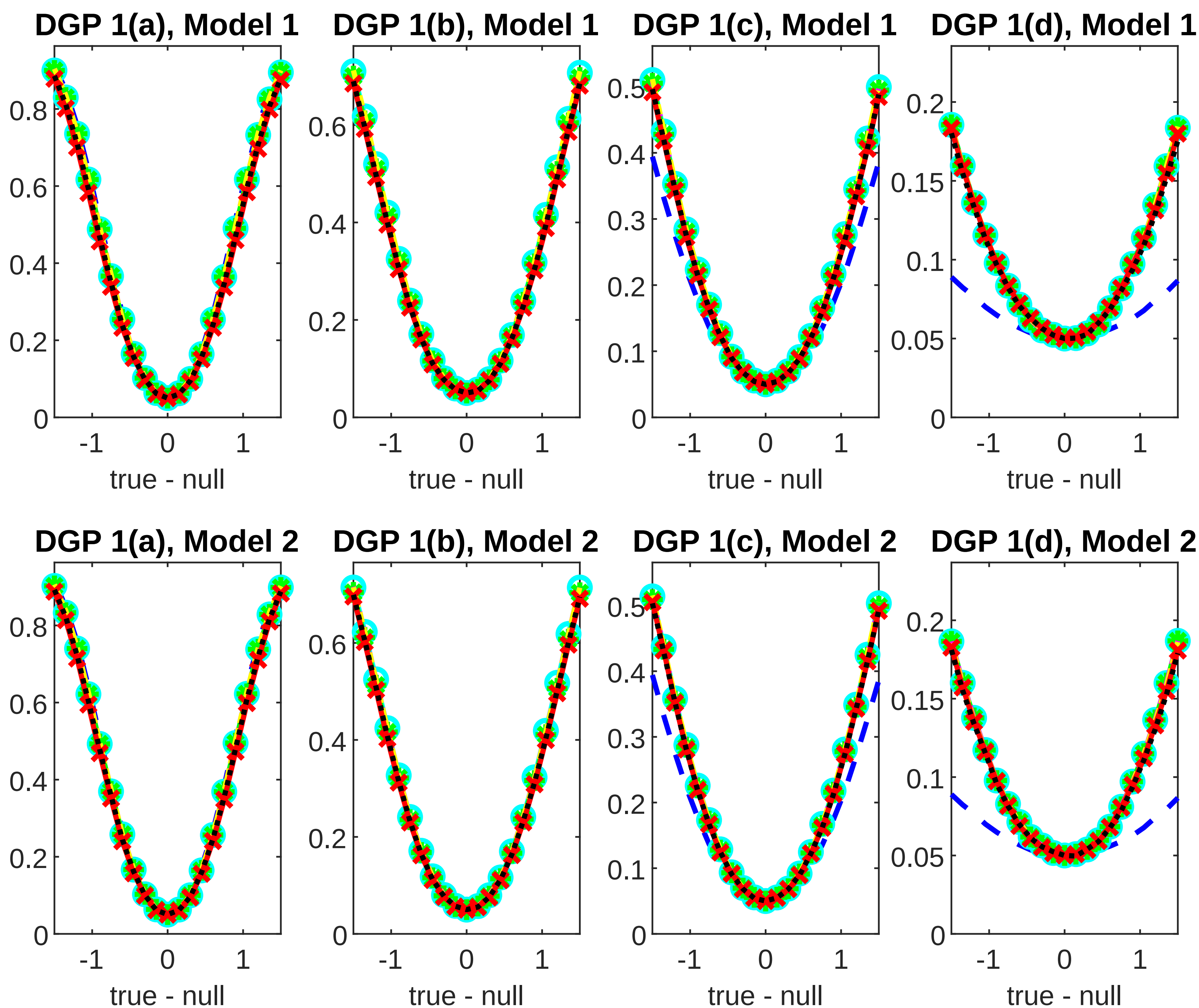

Figure 11: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

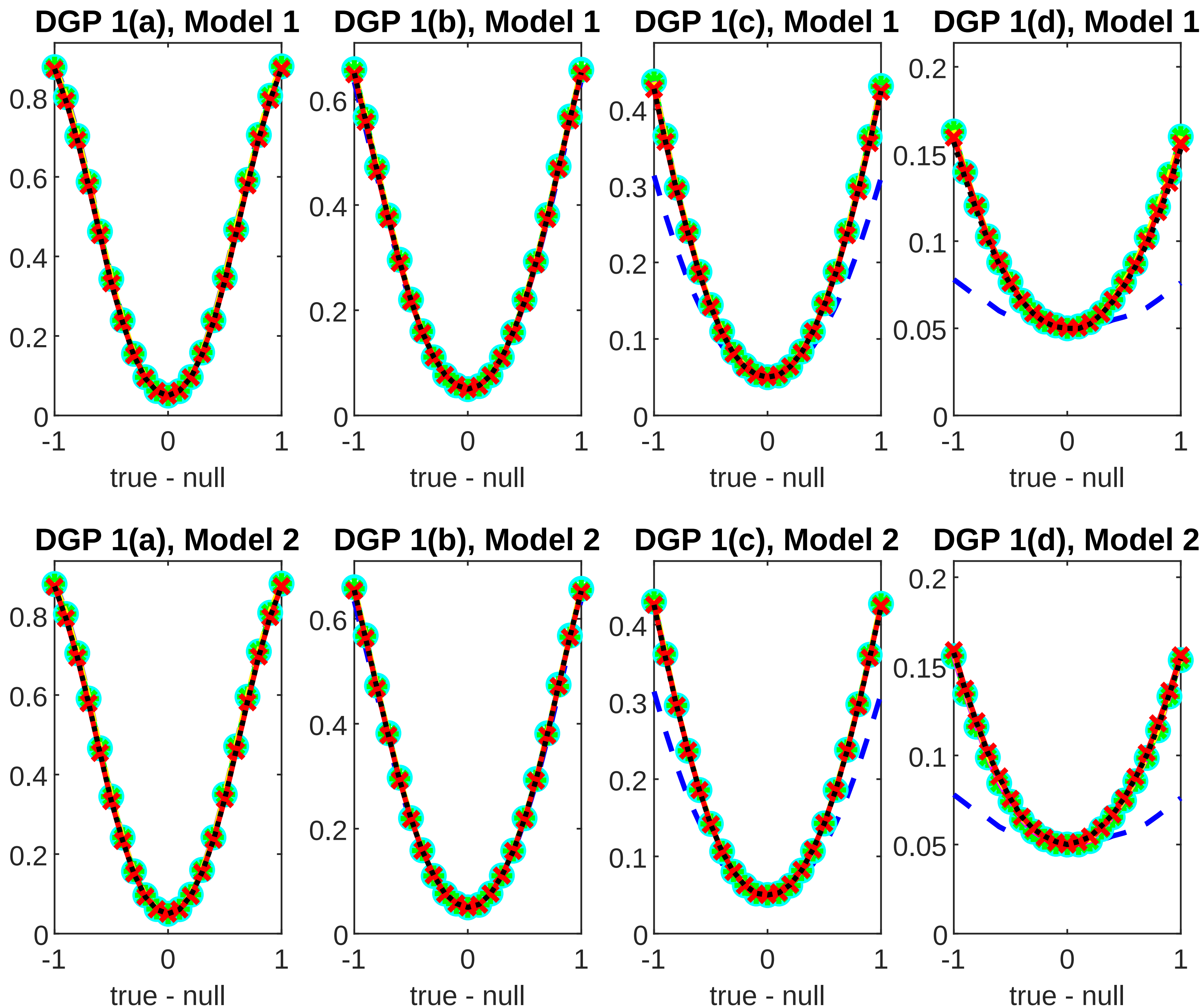

Figure 12: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

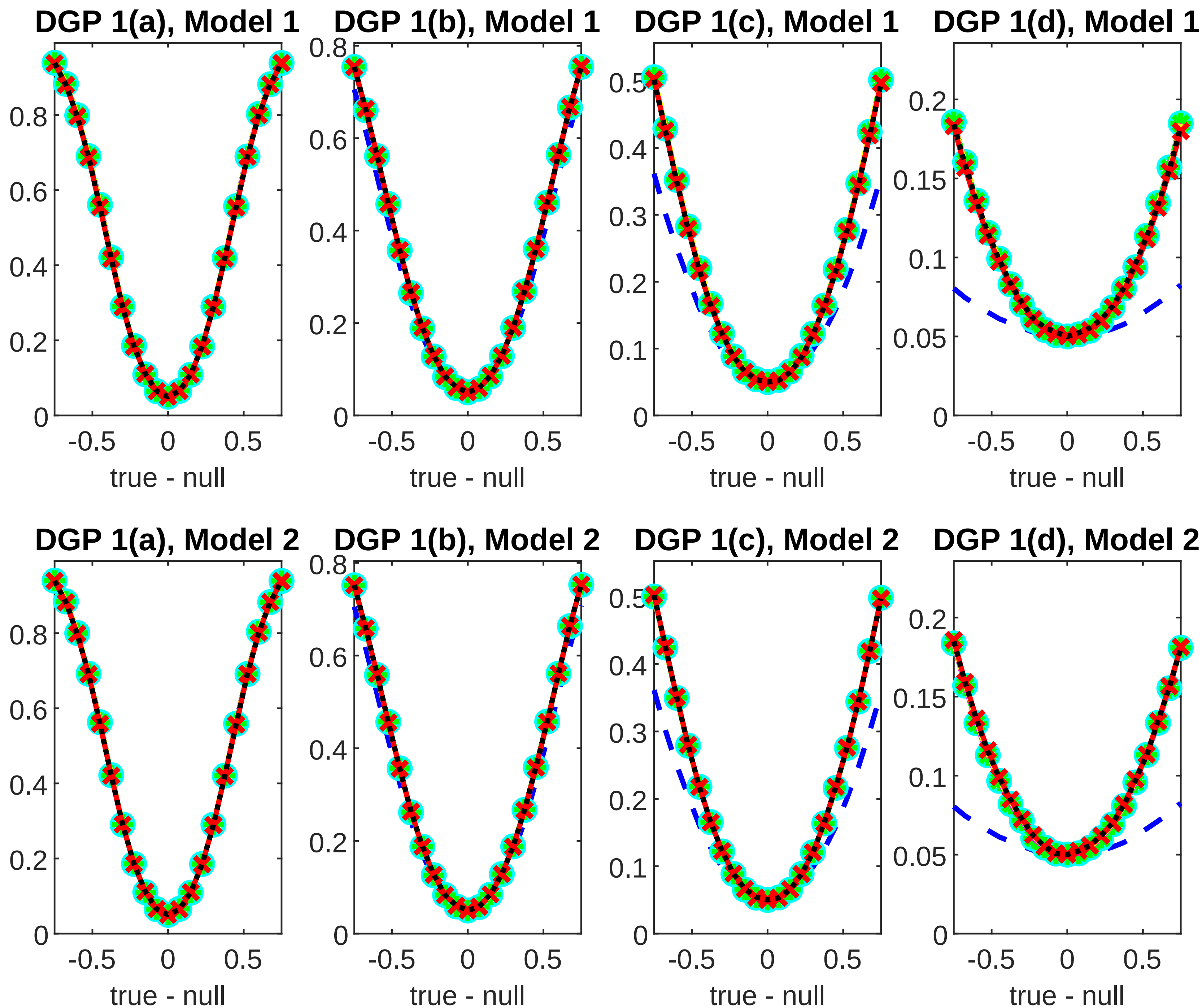

Figure 13: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

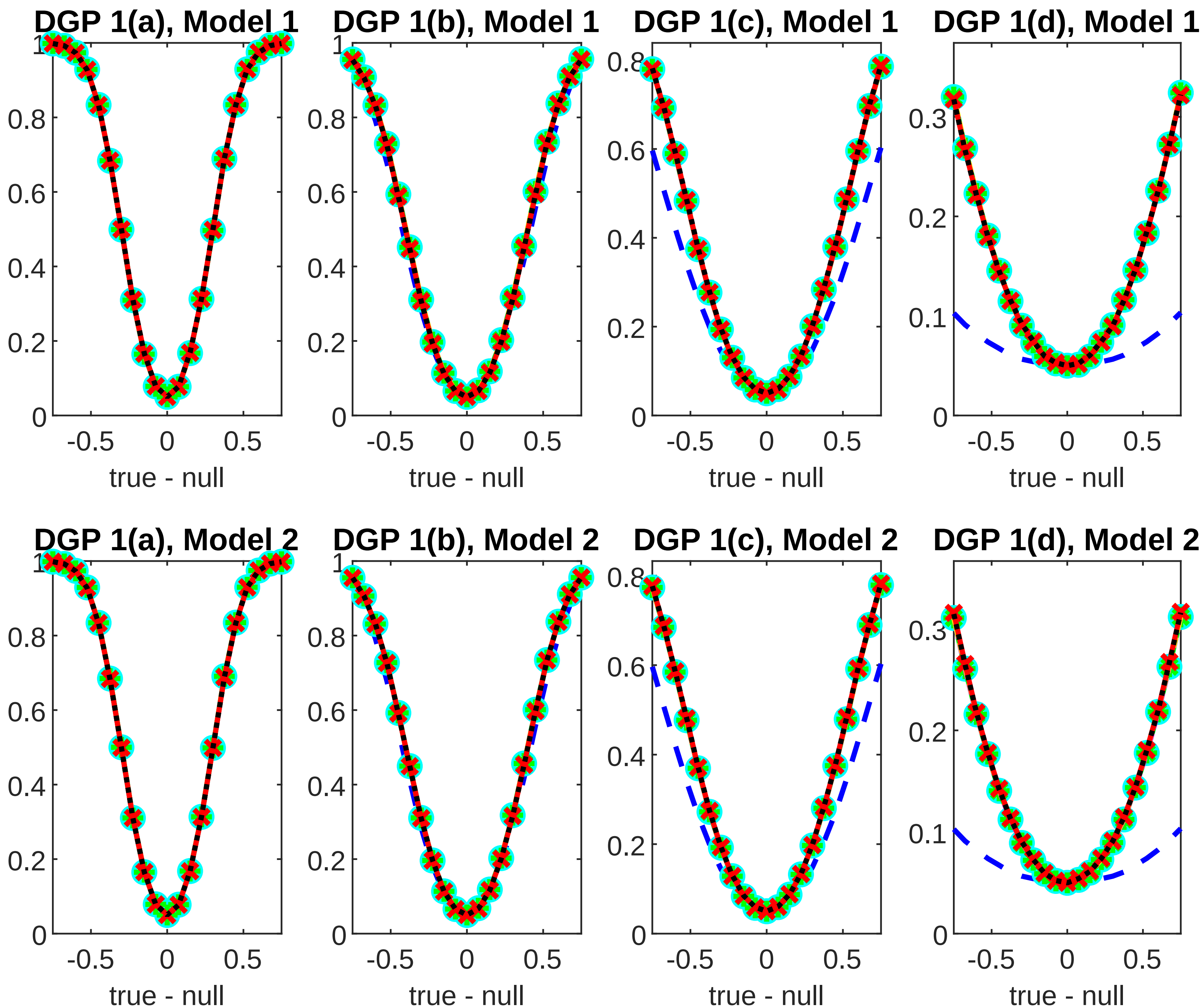

Figure 14: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

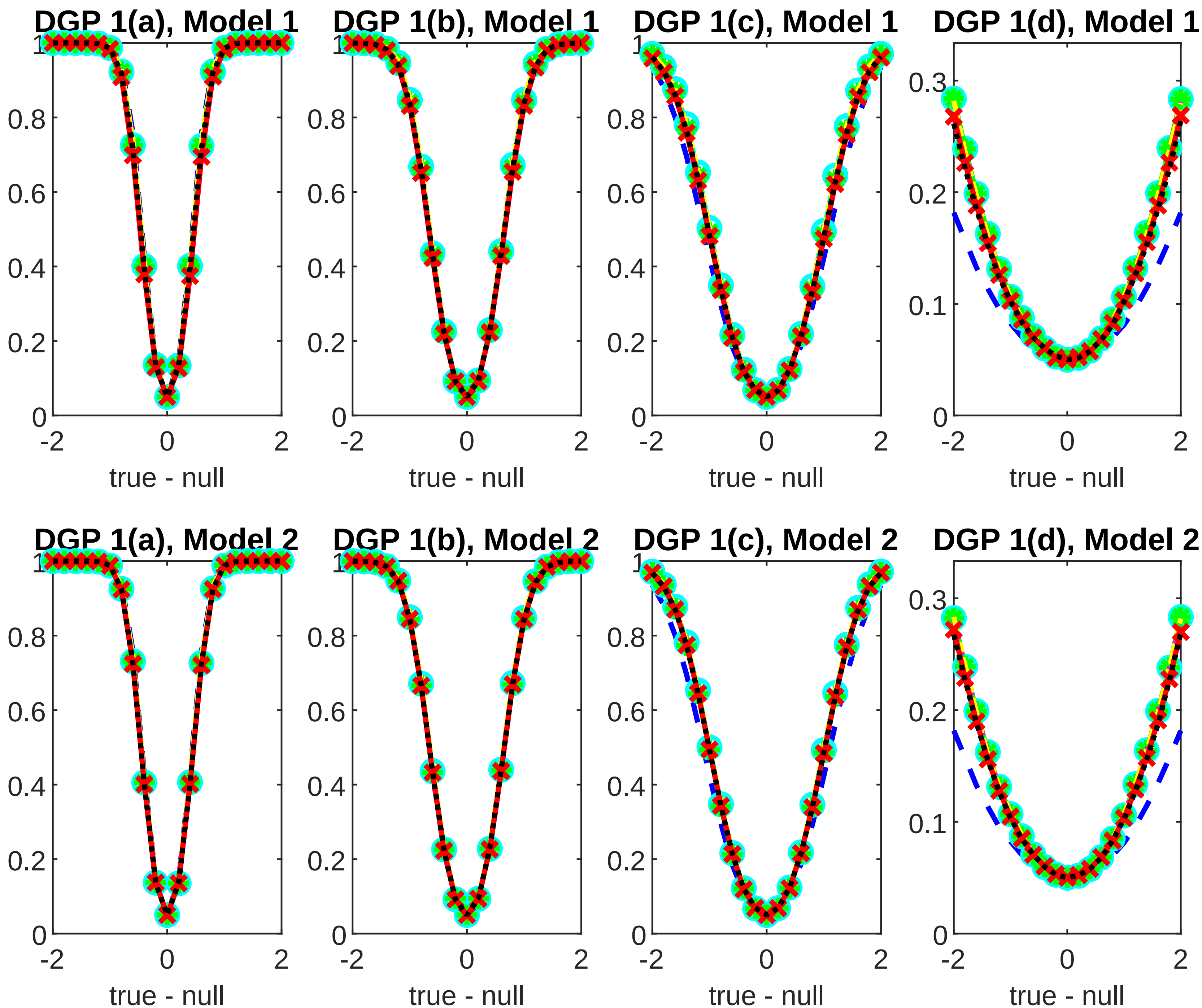

Figure 15: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

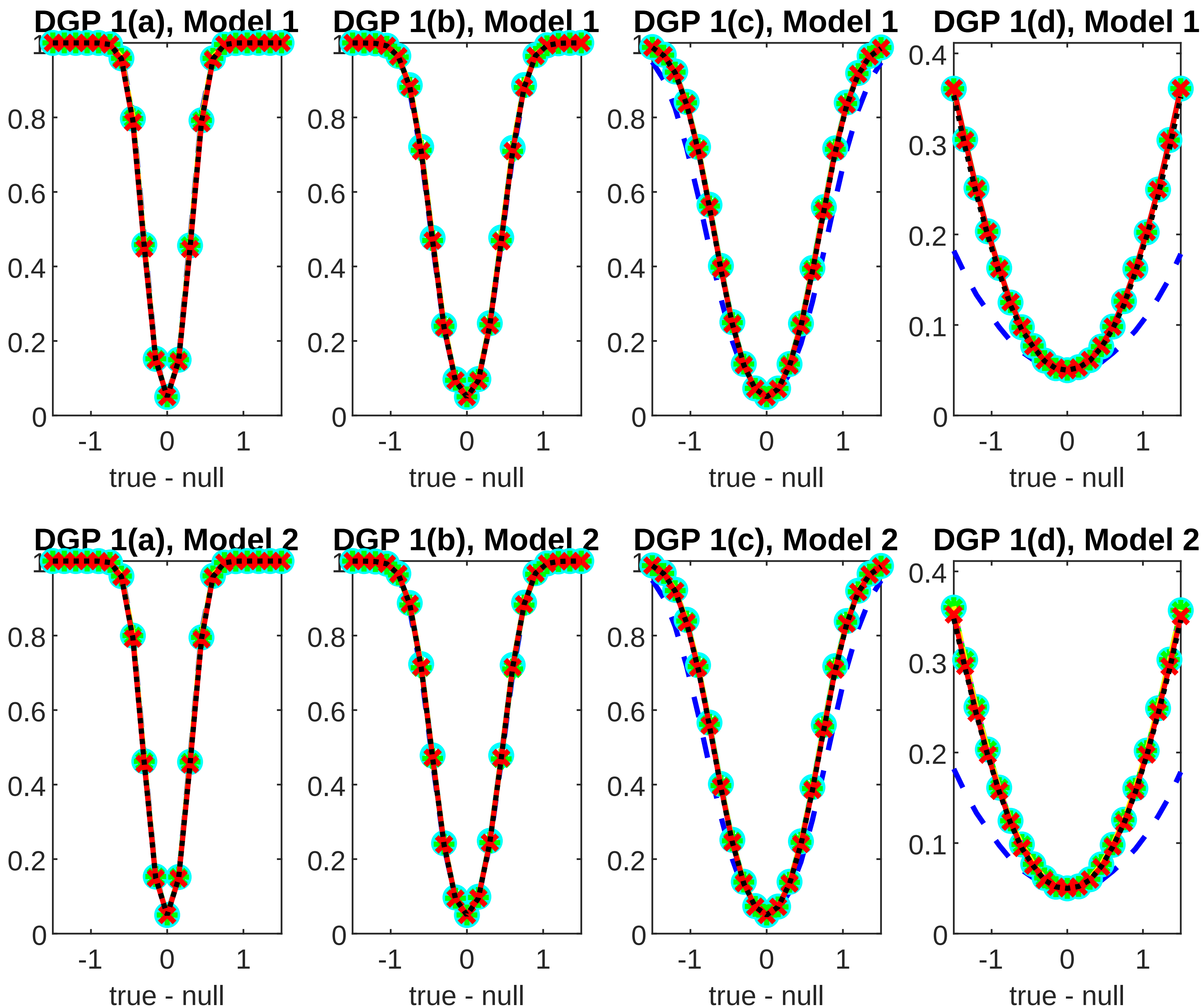

Figure 16: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

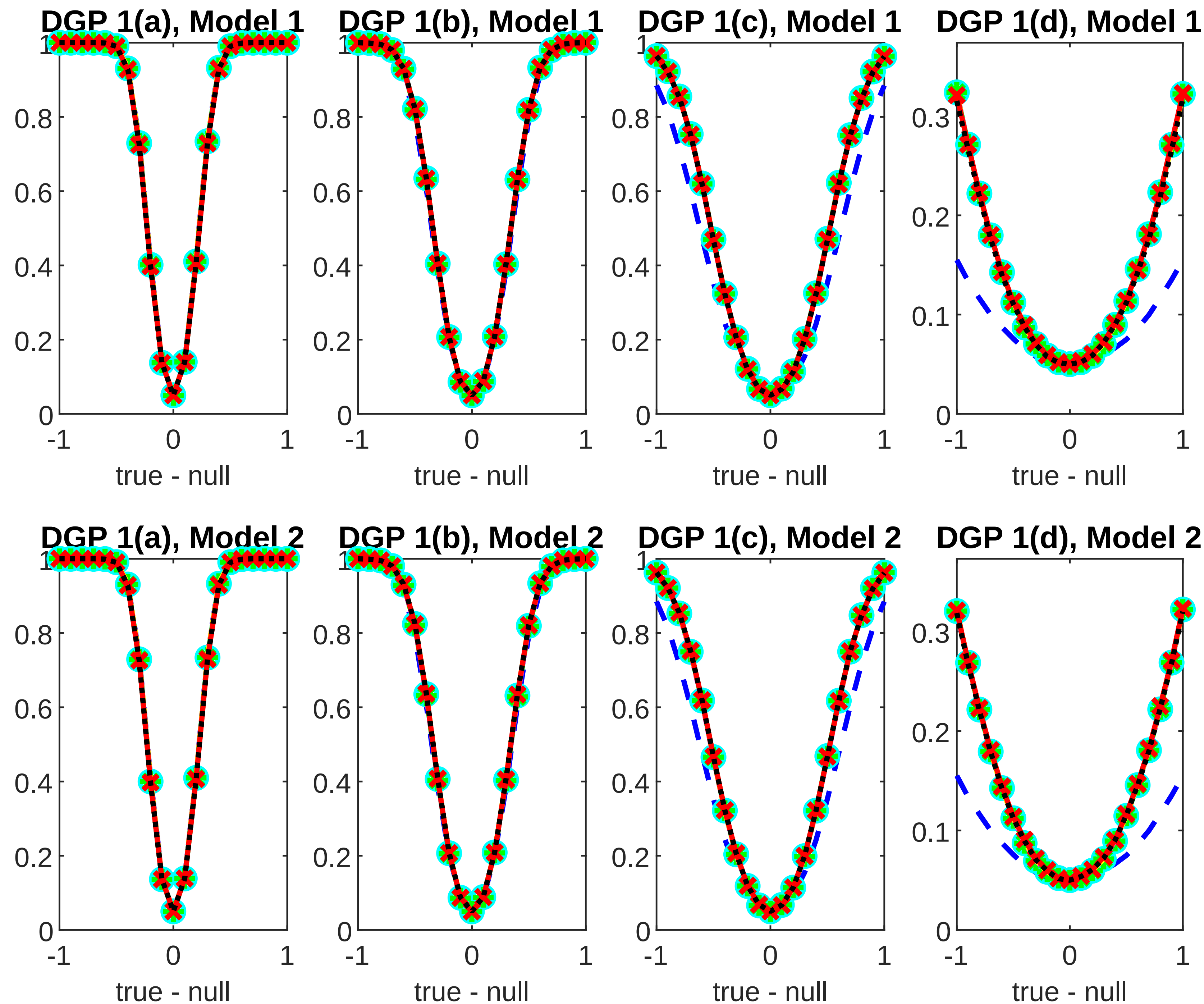

Figure 17: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

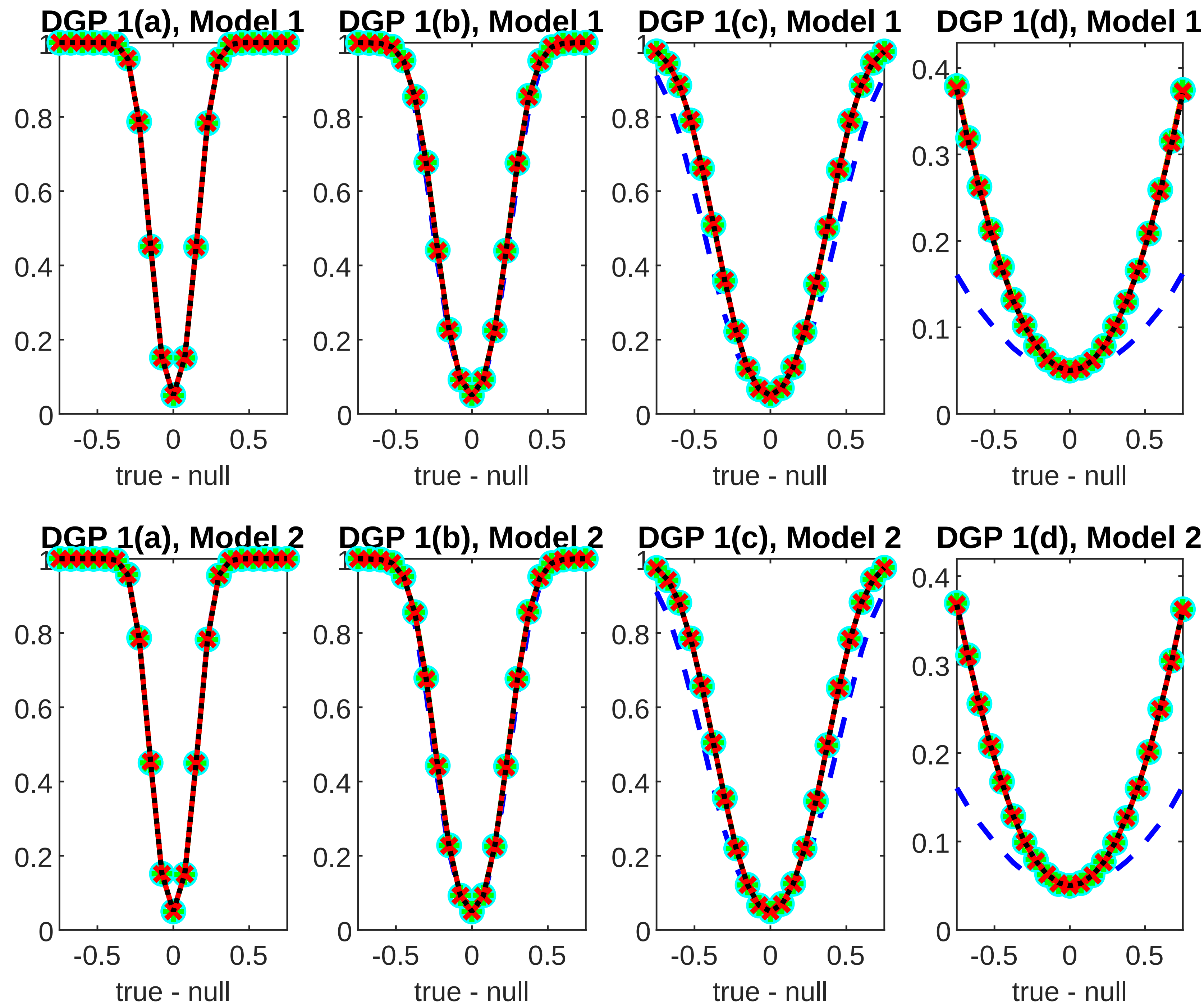

Figure 18: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

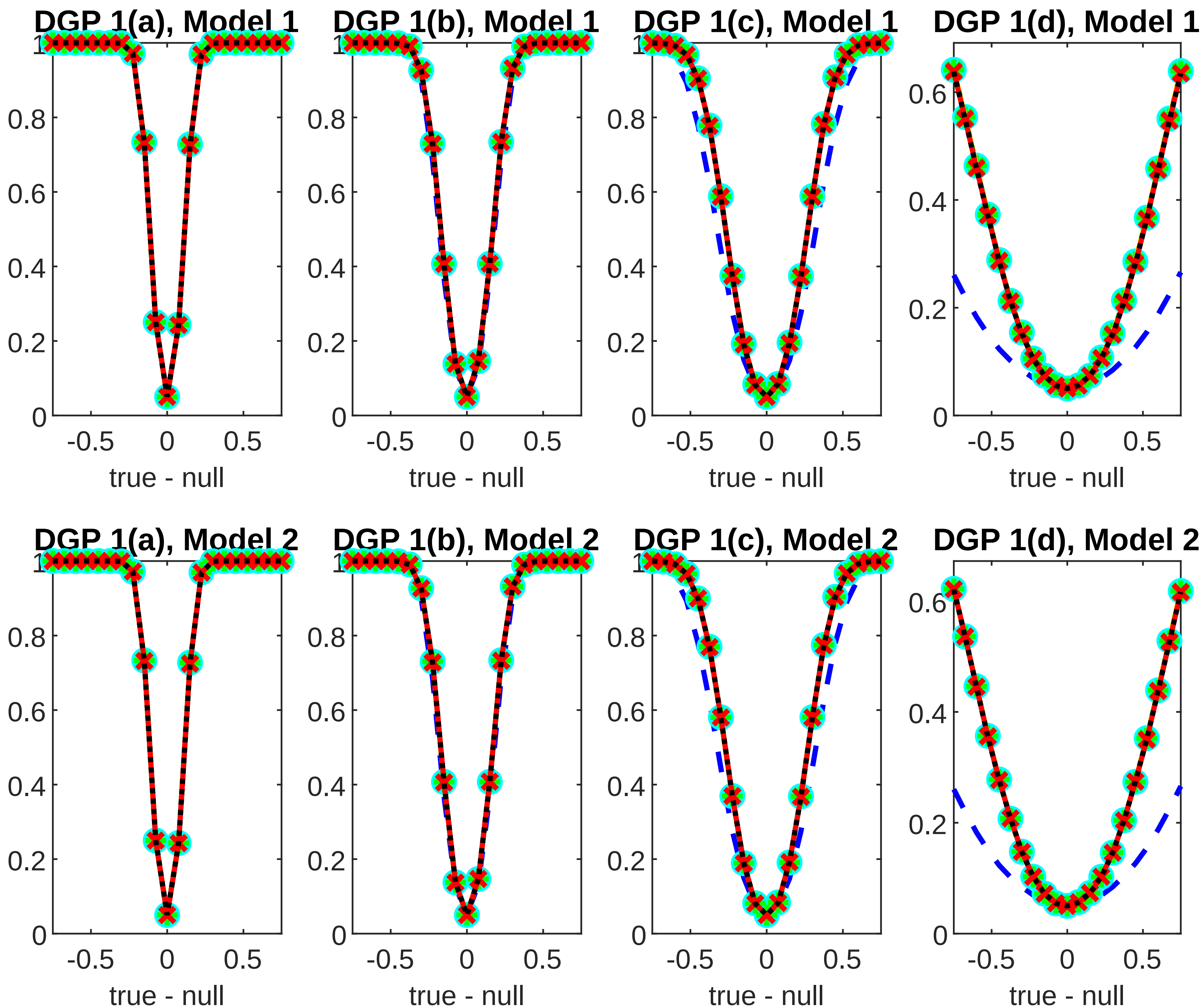

Figure 19: DGP 1 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

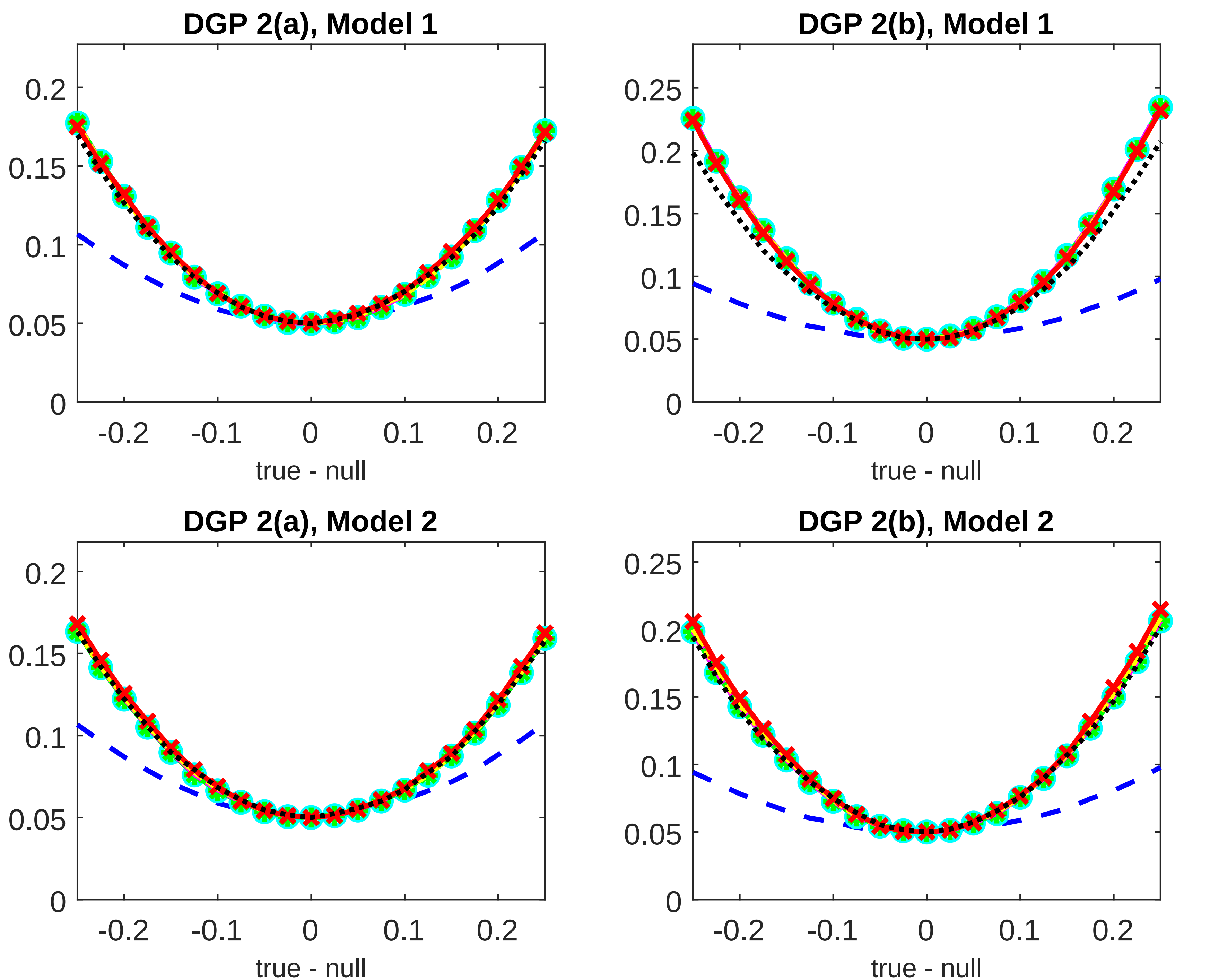

Figure 20: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

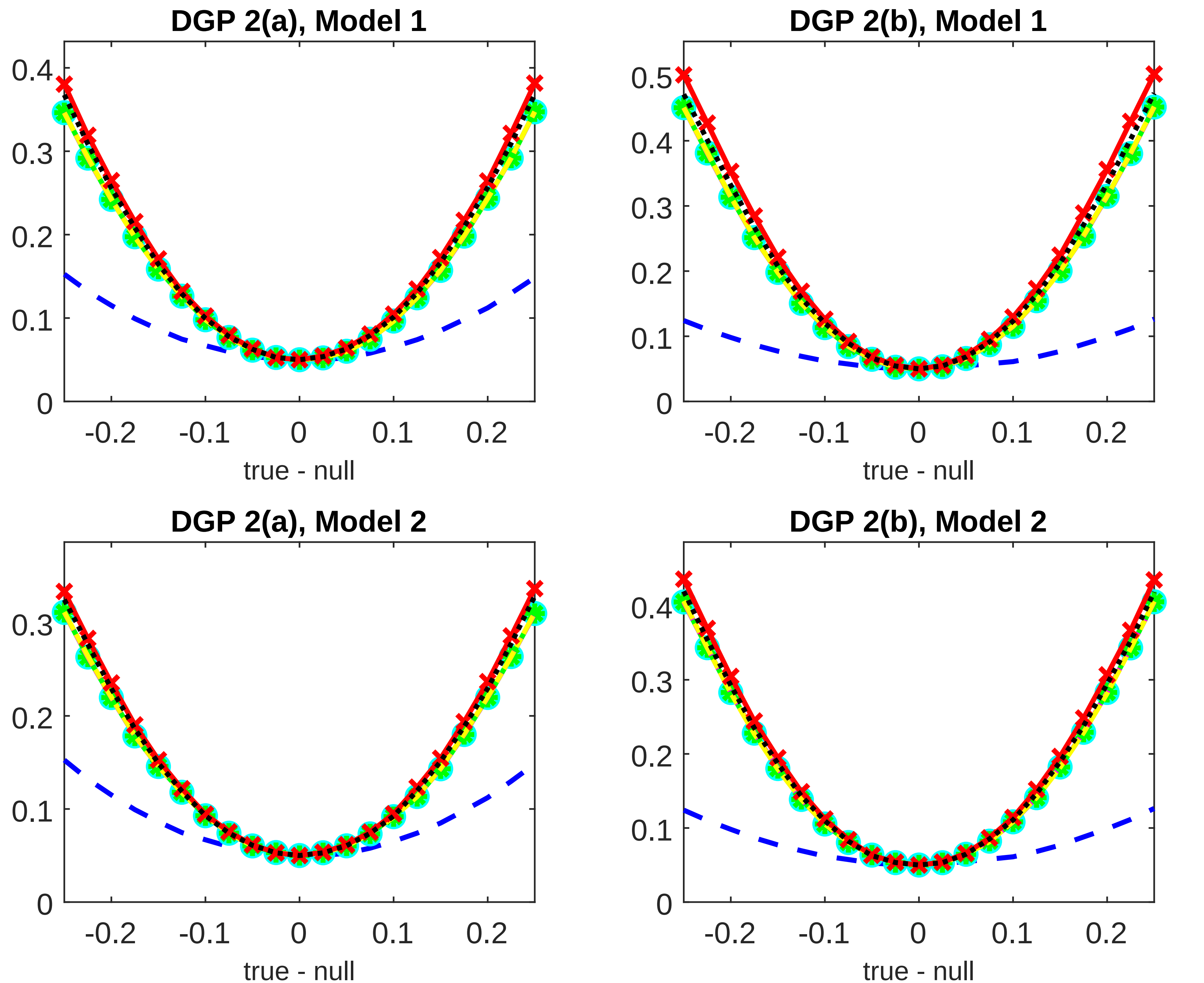

Figure 21: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

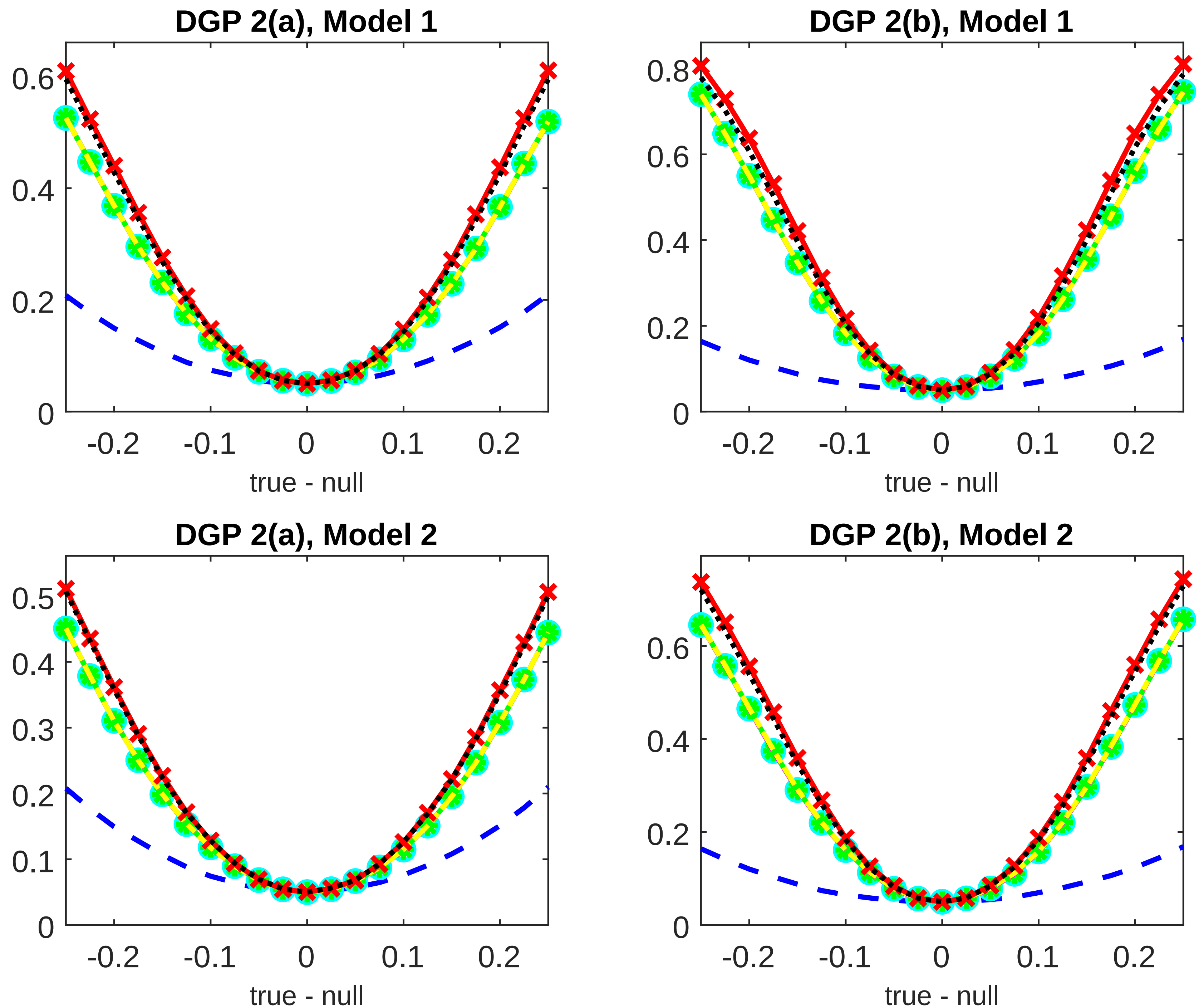

Figure 22: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

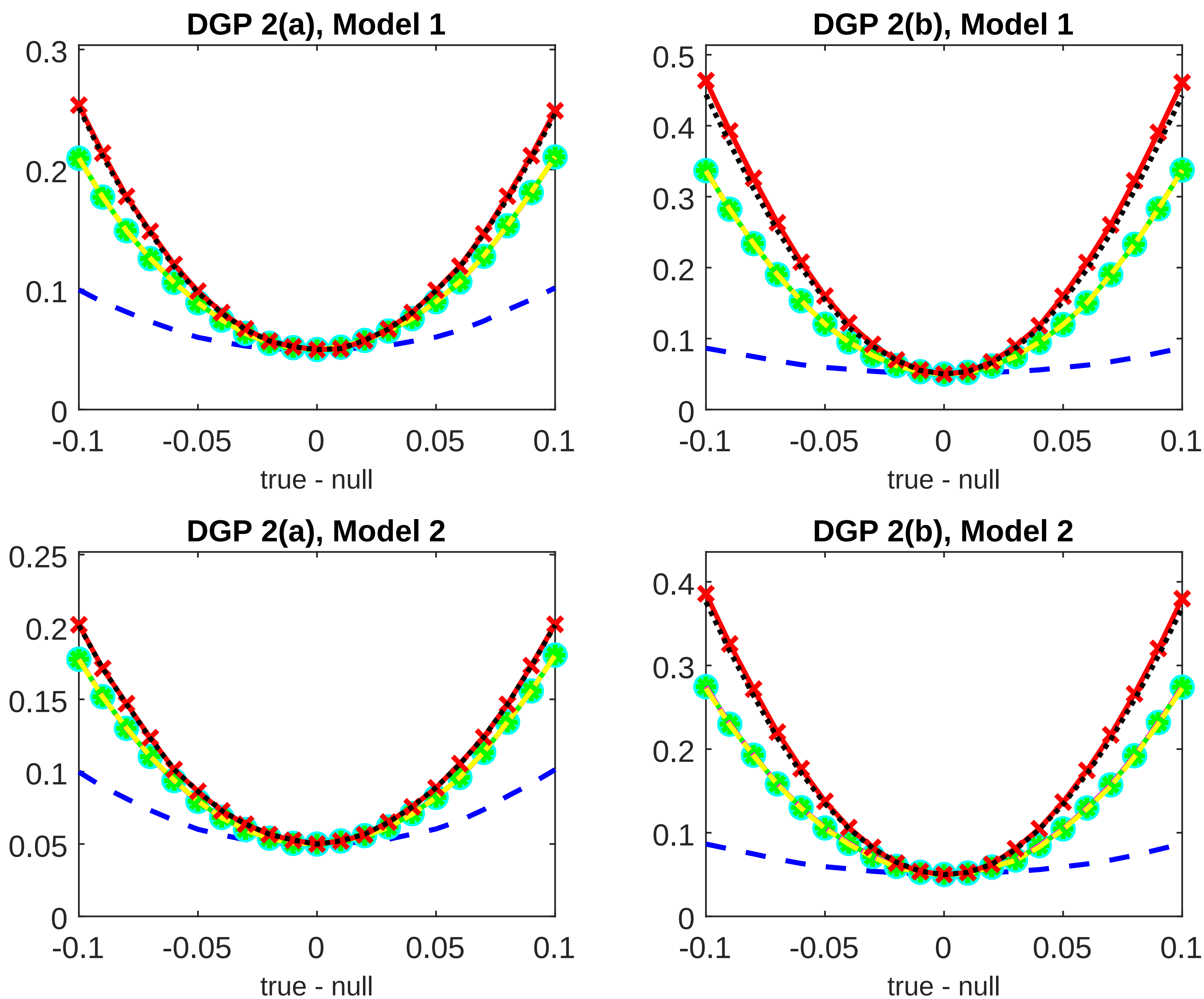

Figure 23: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

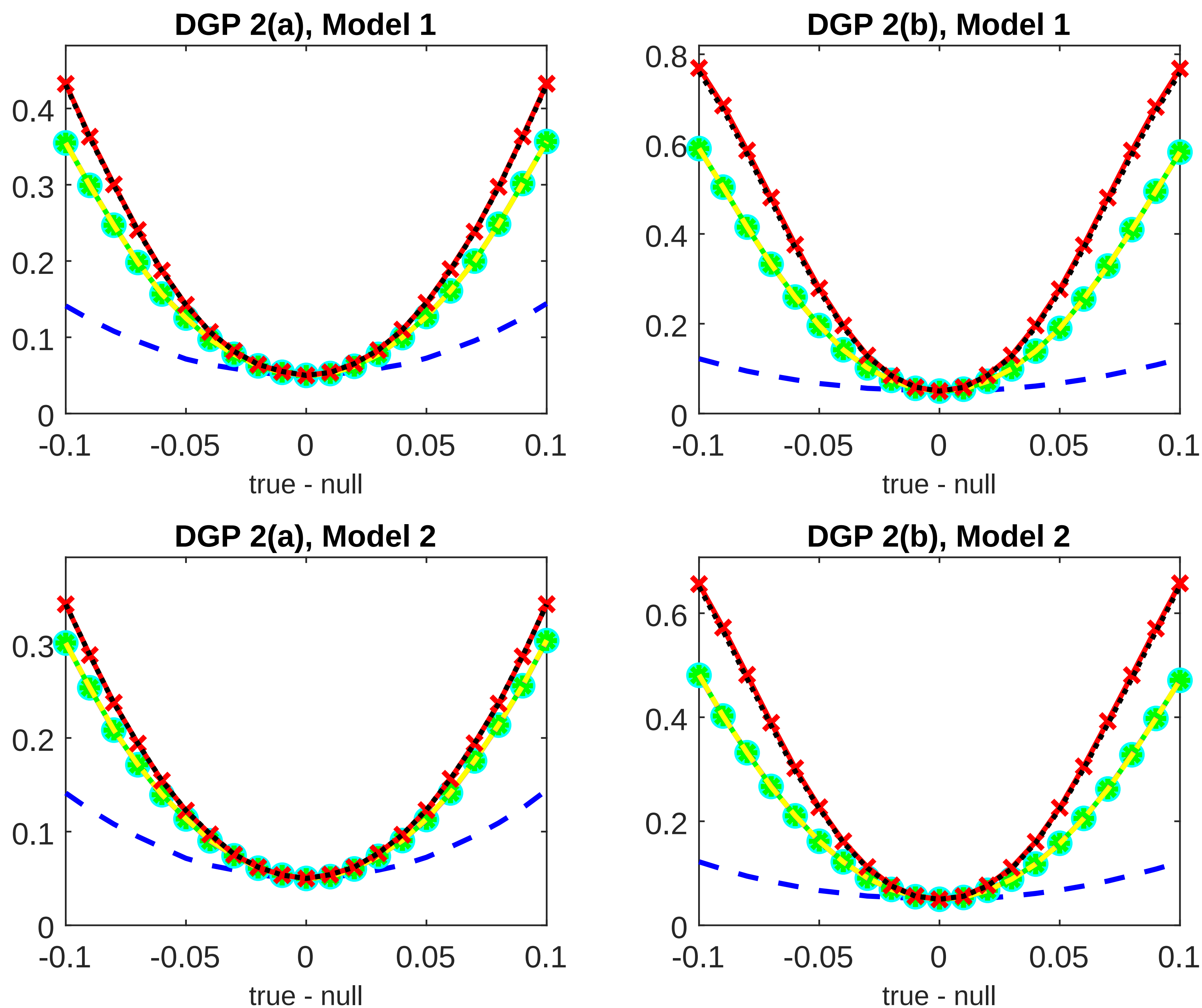

Figure 24: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

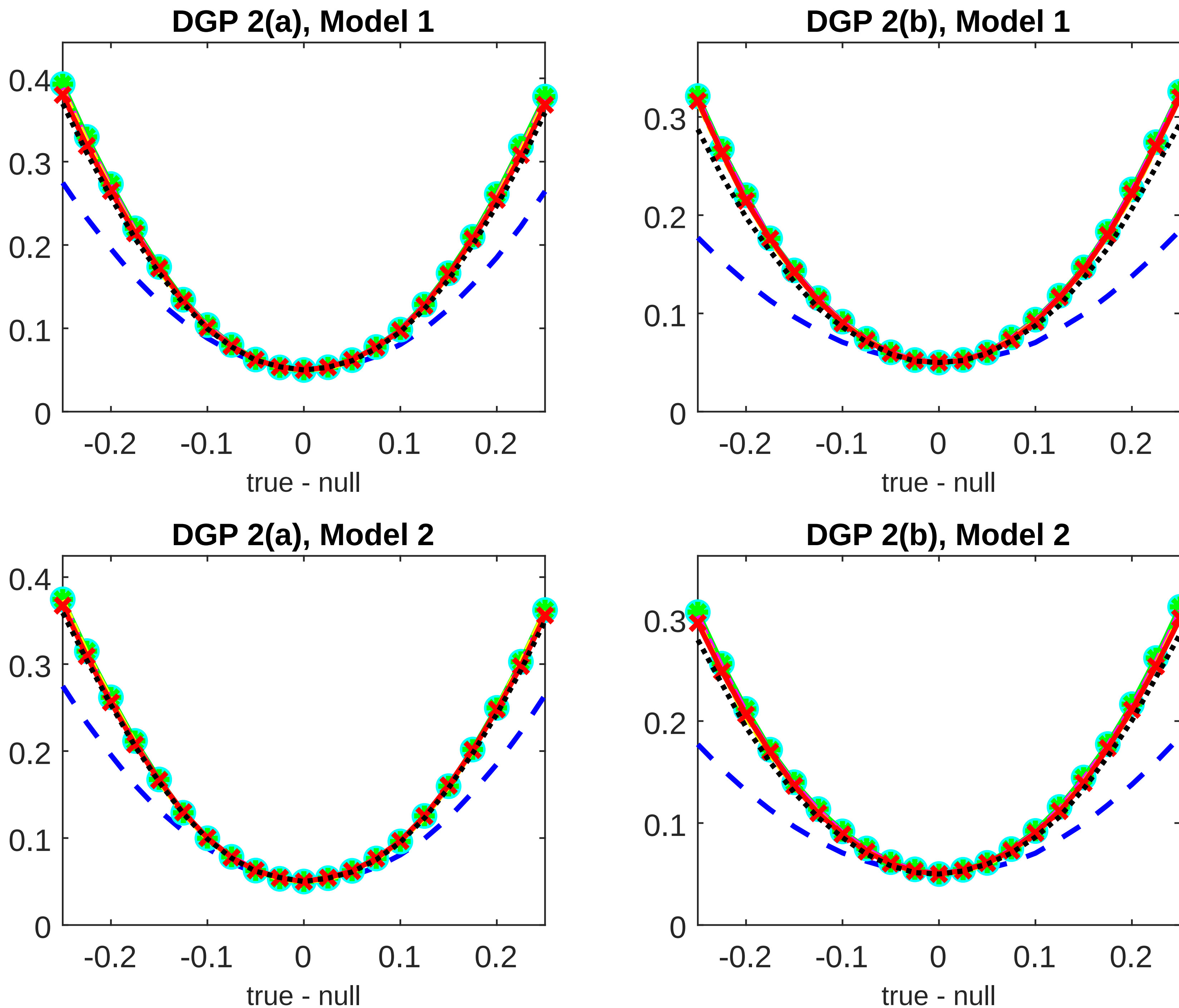

Figure 25: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

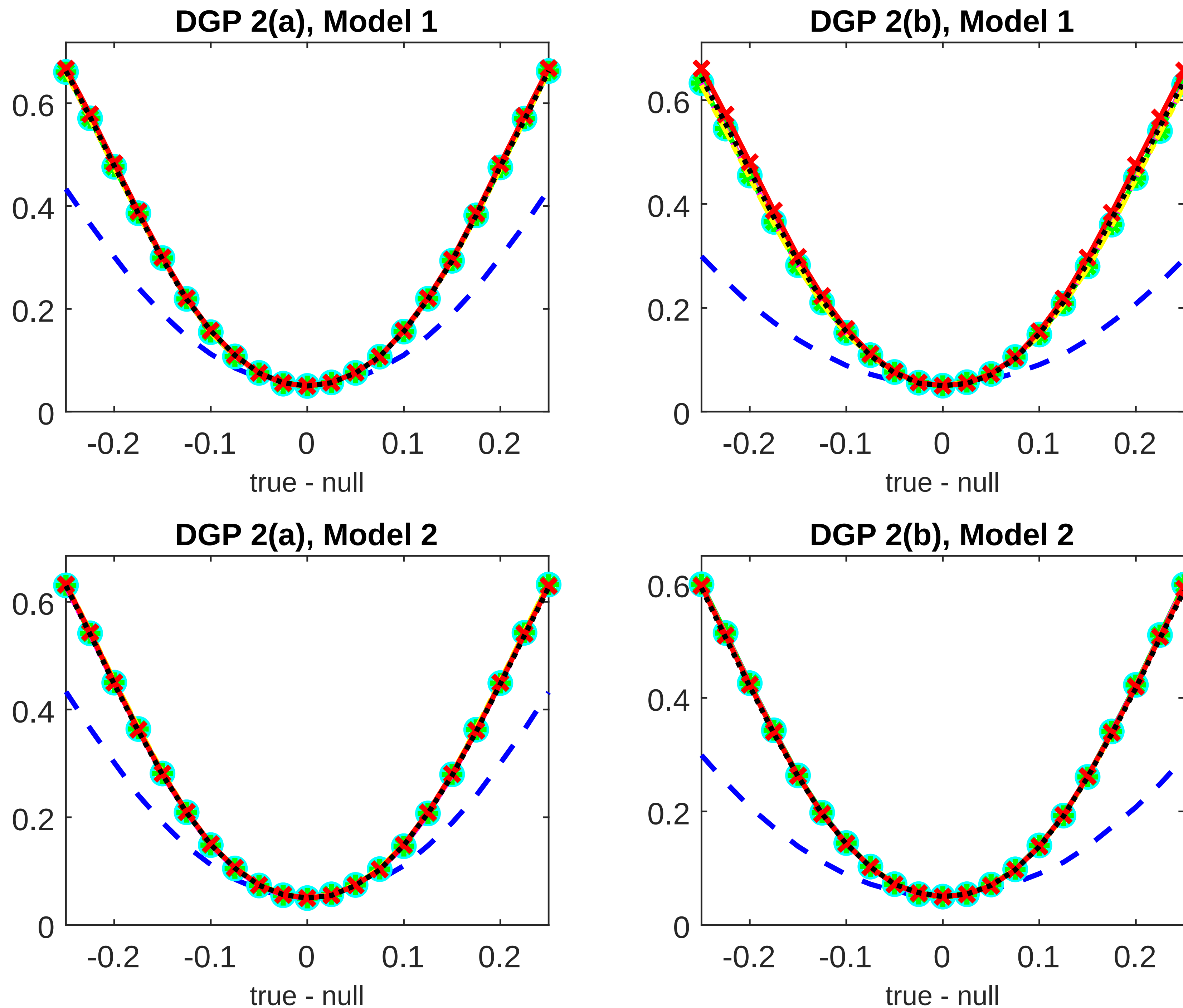

Figure 26: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

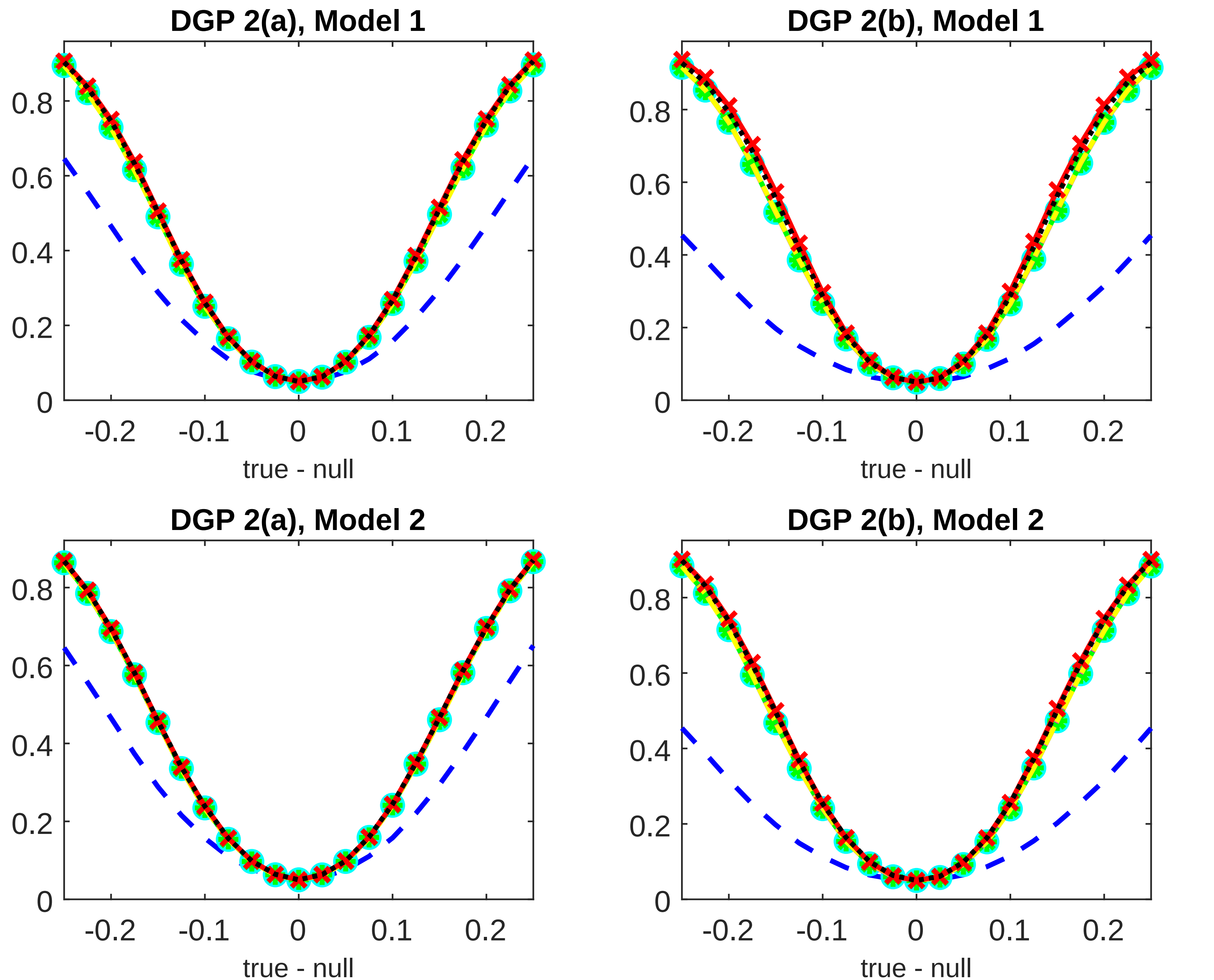

Figure 27: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

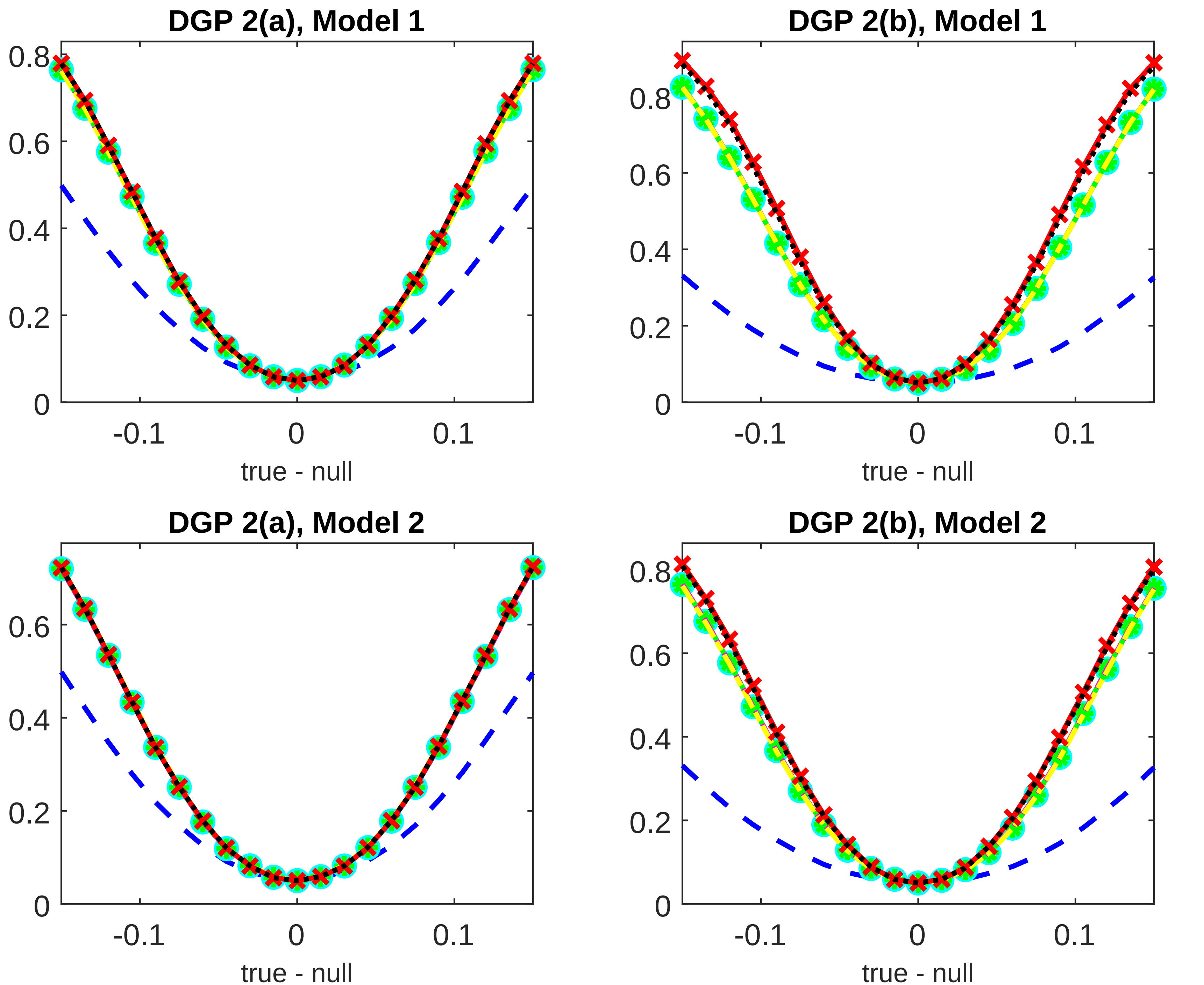

Figure 28: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

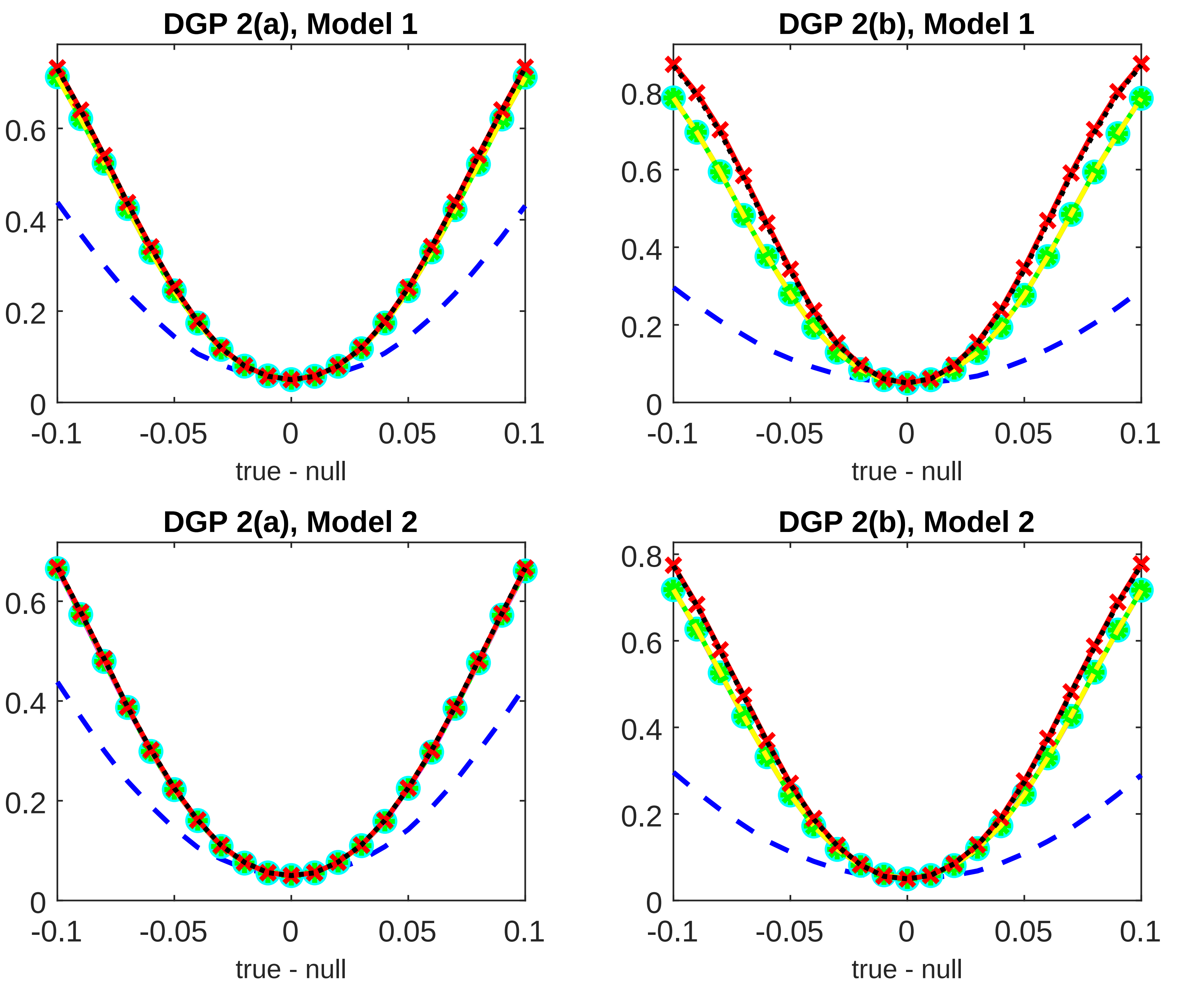

Figure 29: DGP 2 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

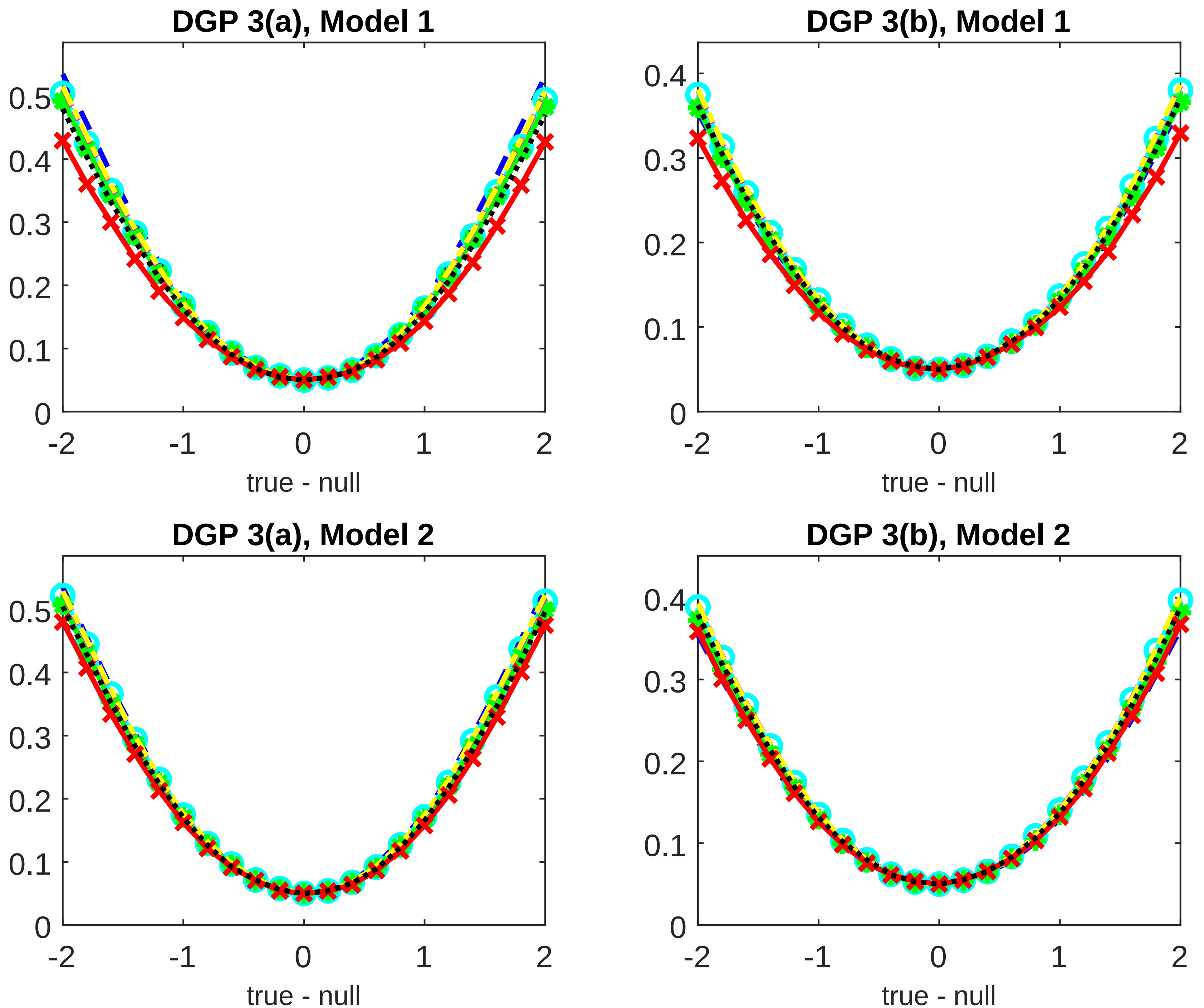

Figure 30: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

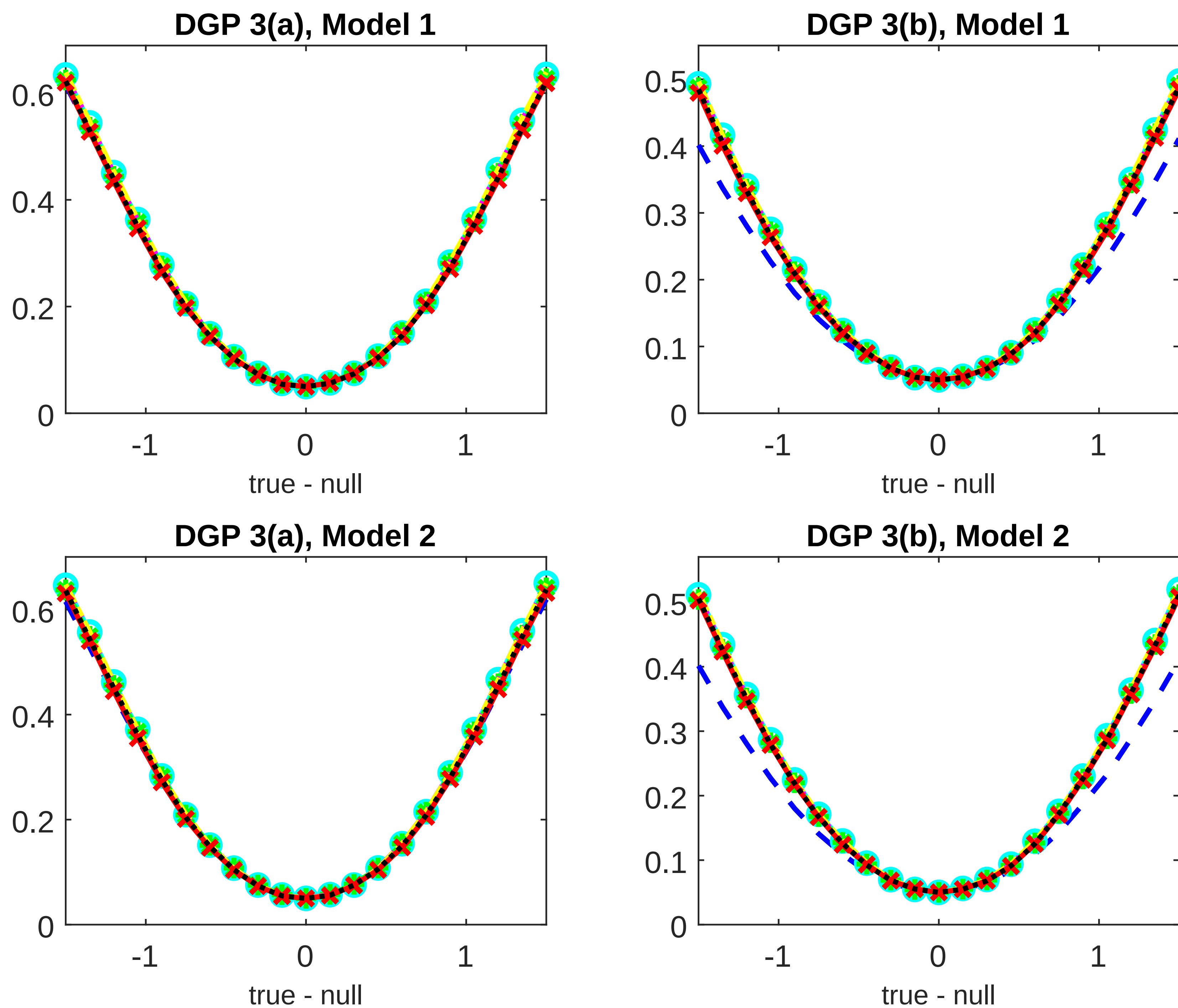

Figure 31: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

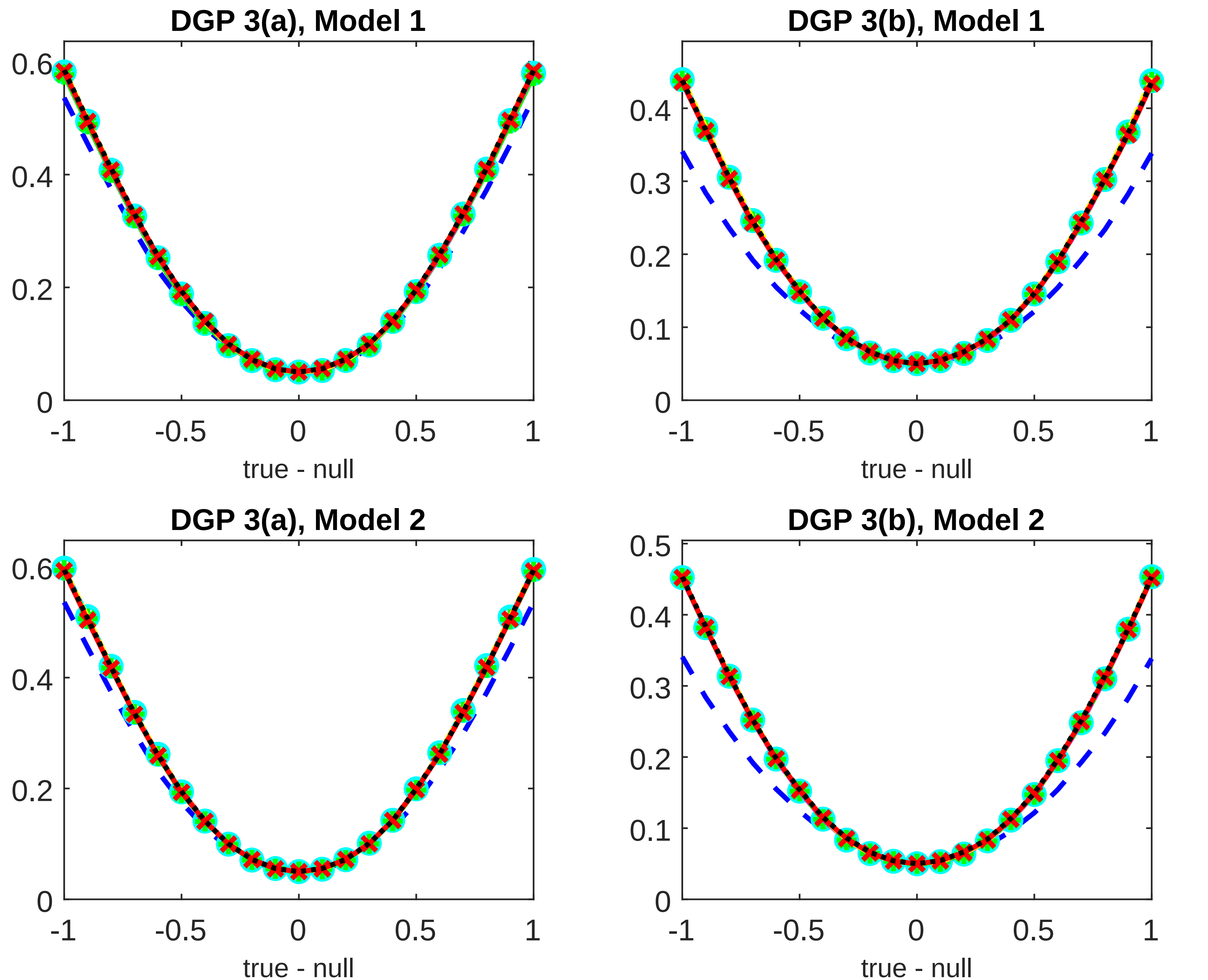

Figure 32: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

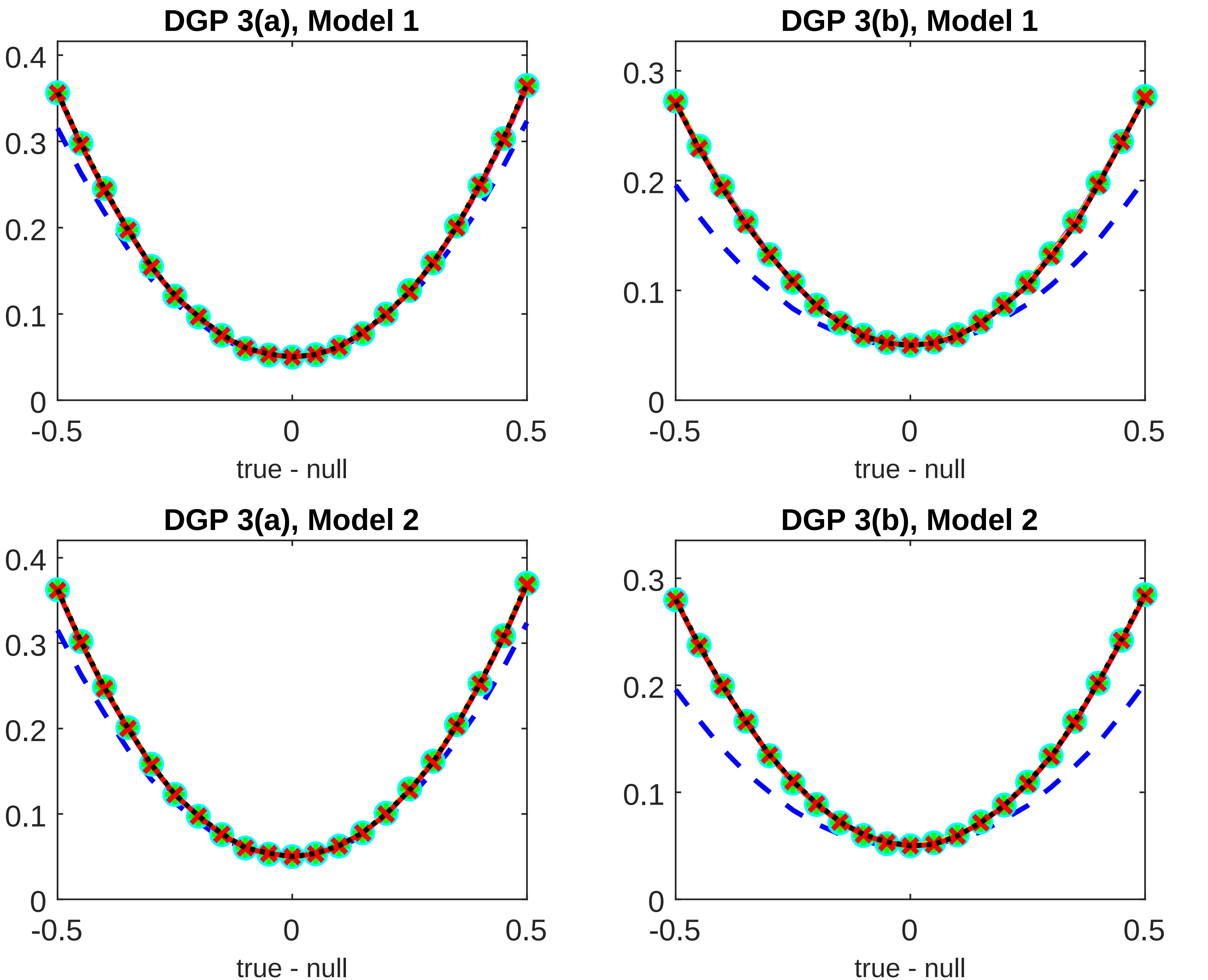

Figure 33: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

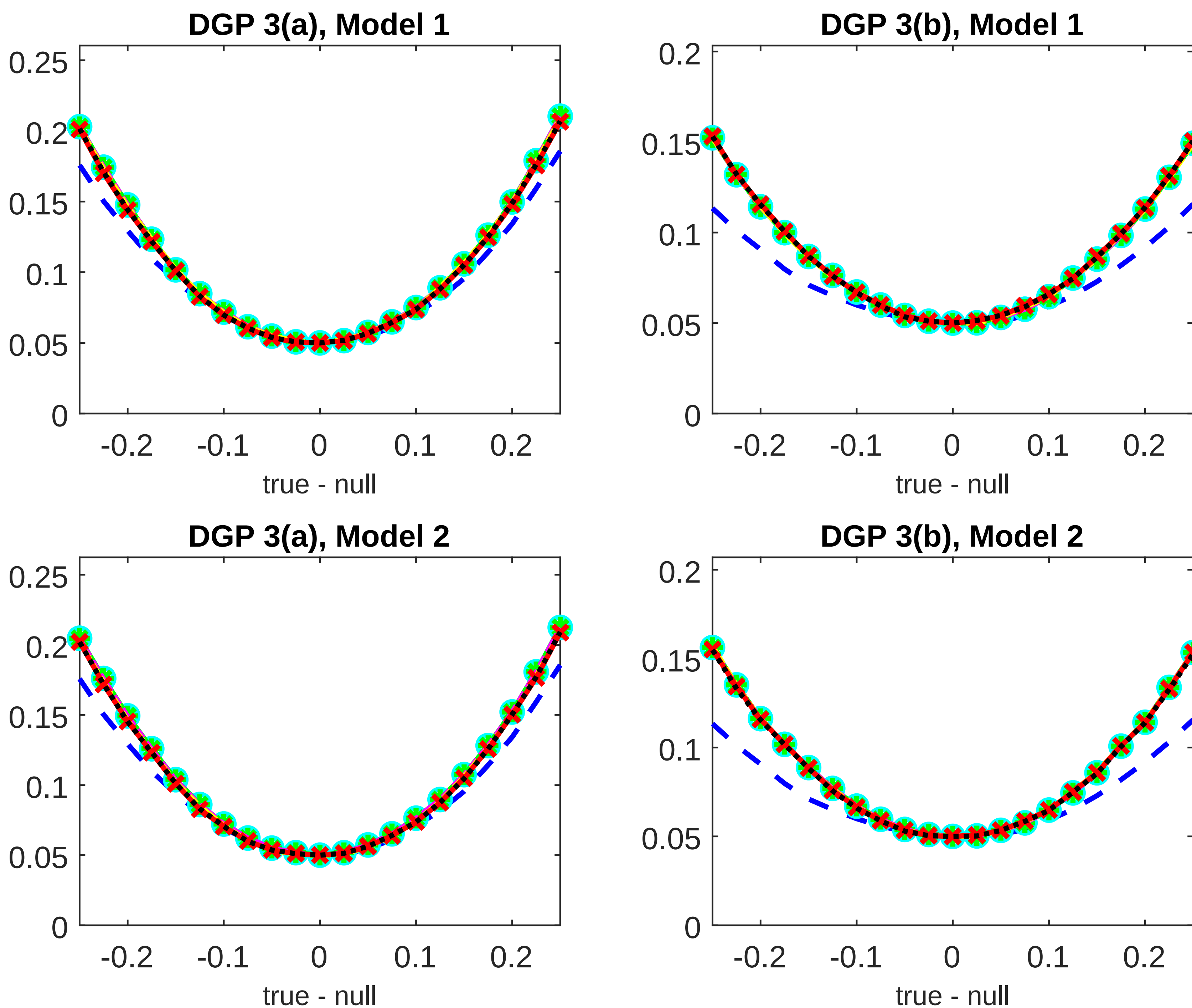

Figure 34: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

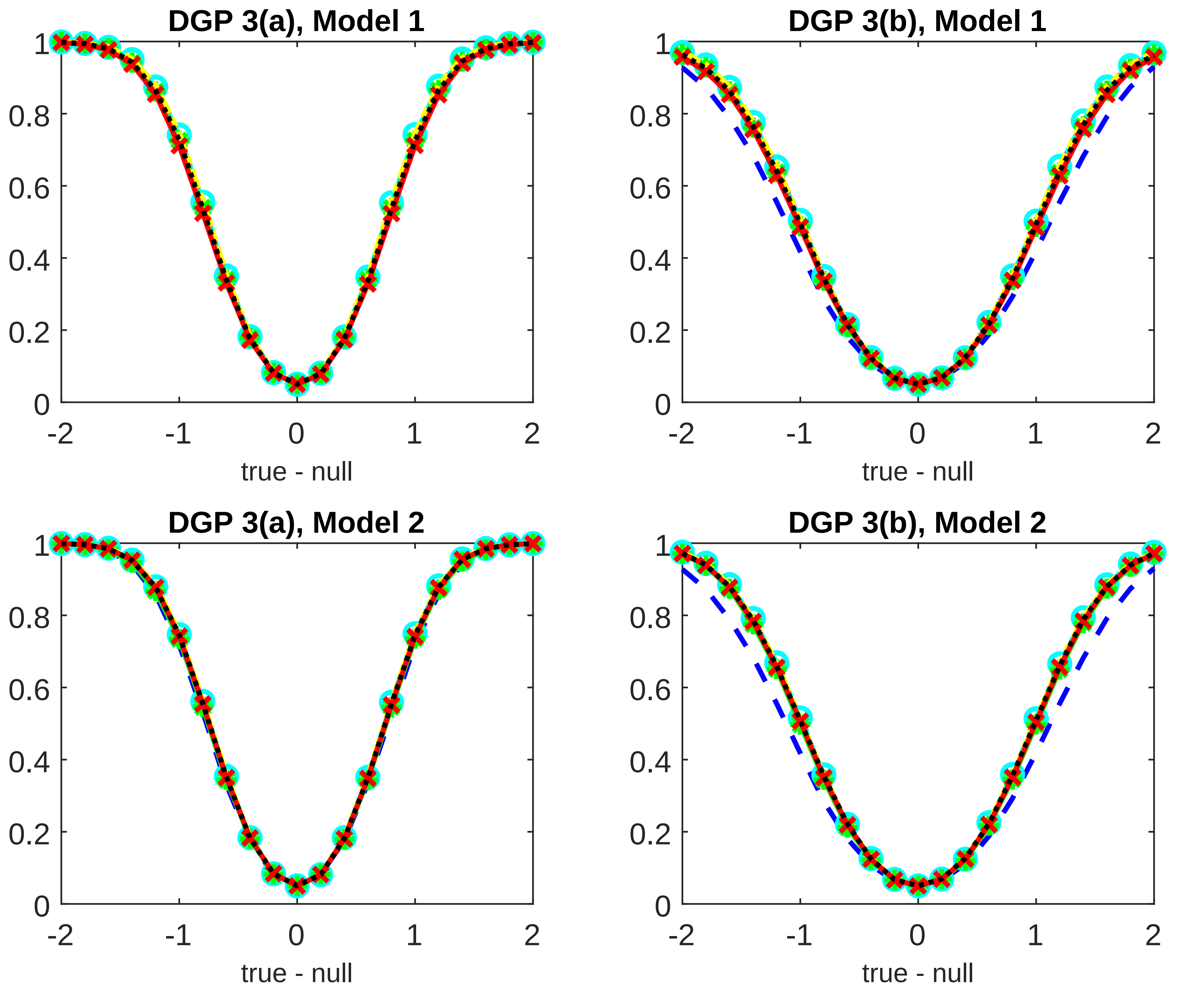

Figure 35: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

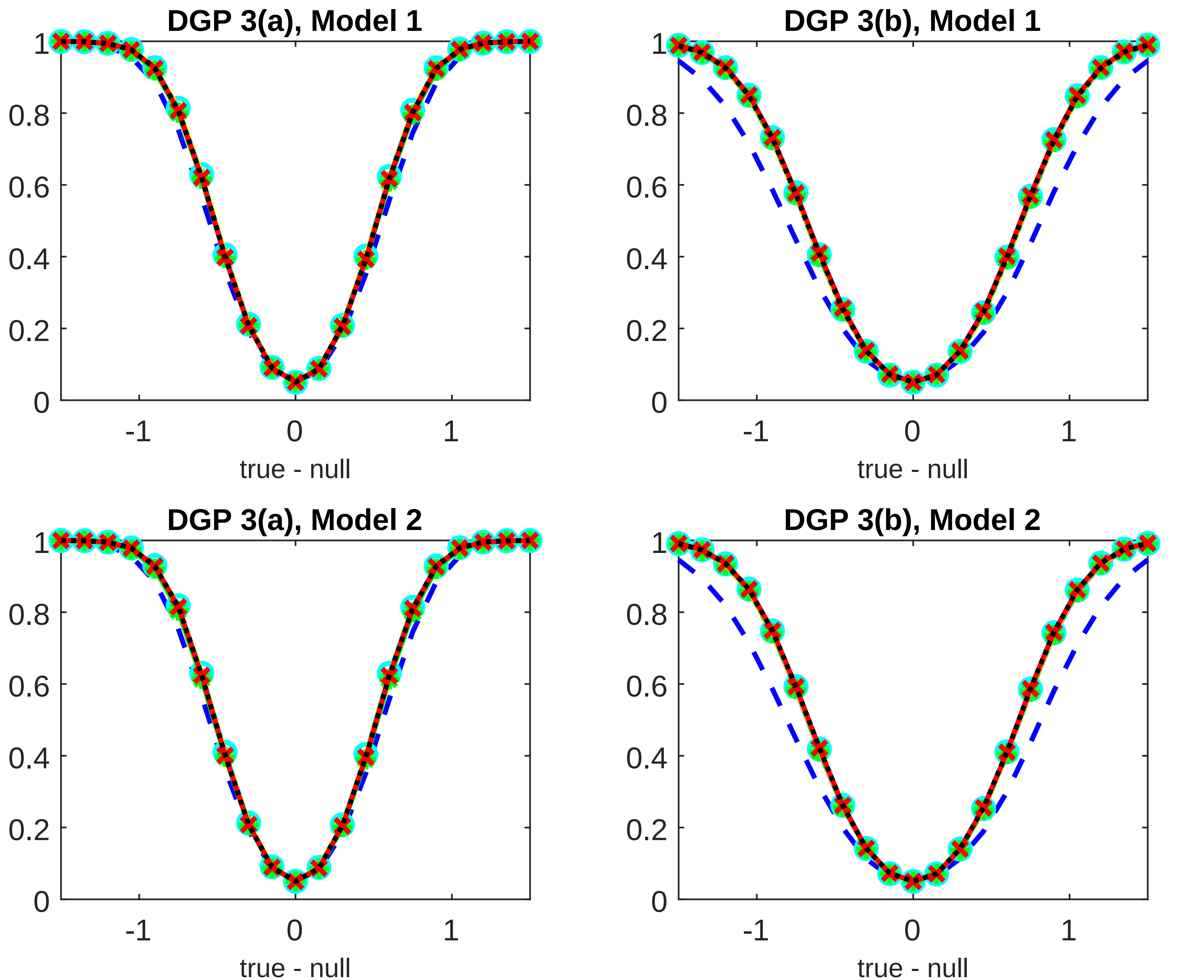

Figure 36: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

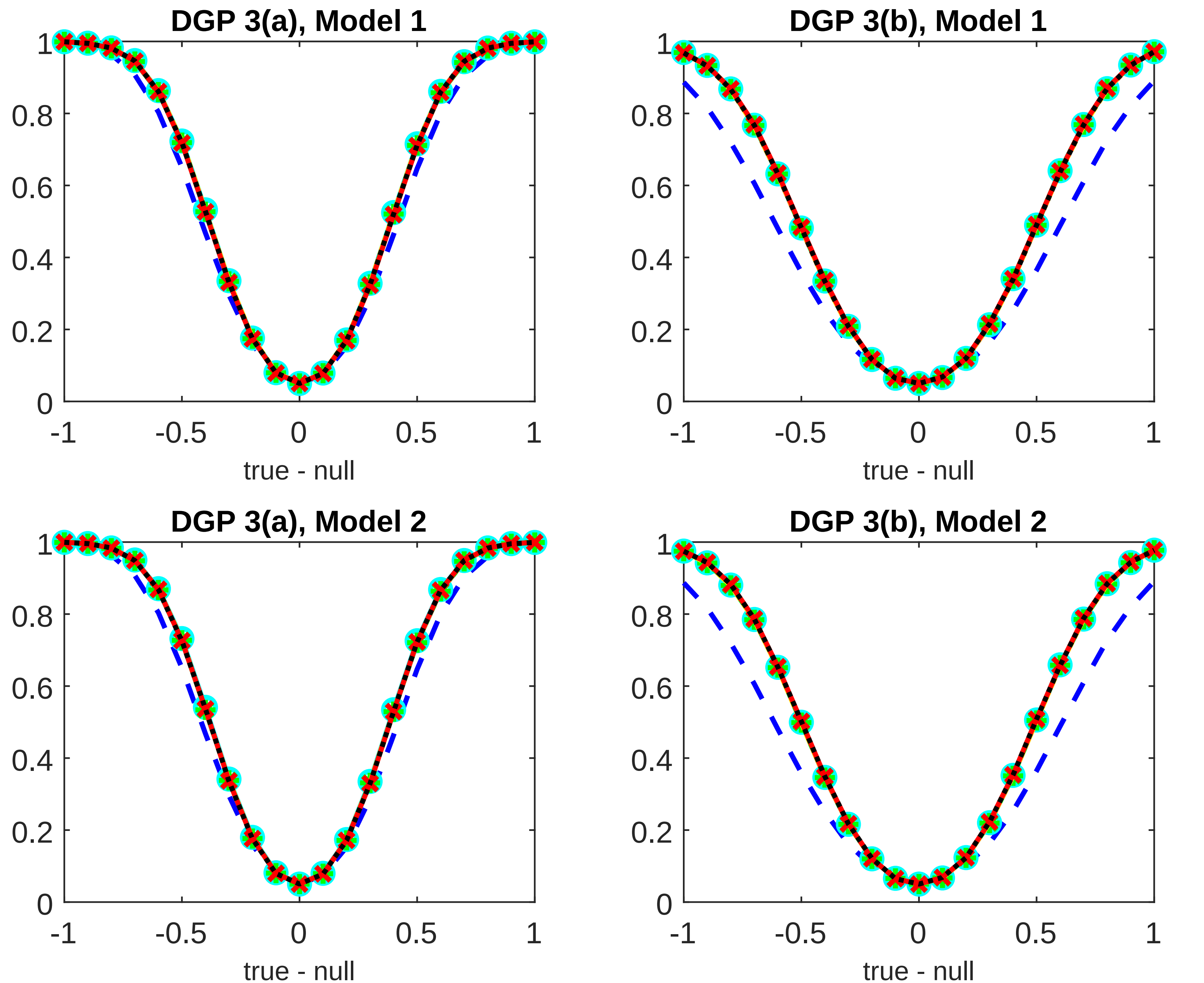

Figure 37: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

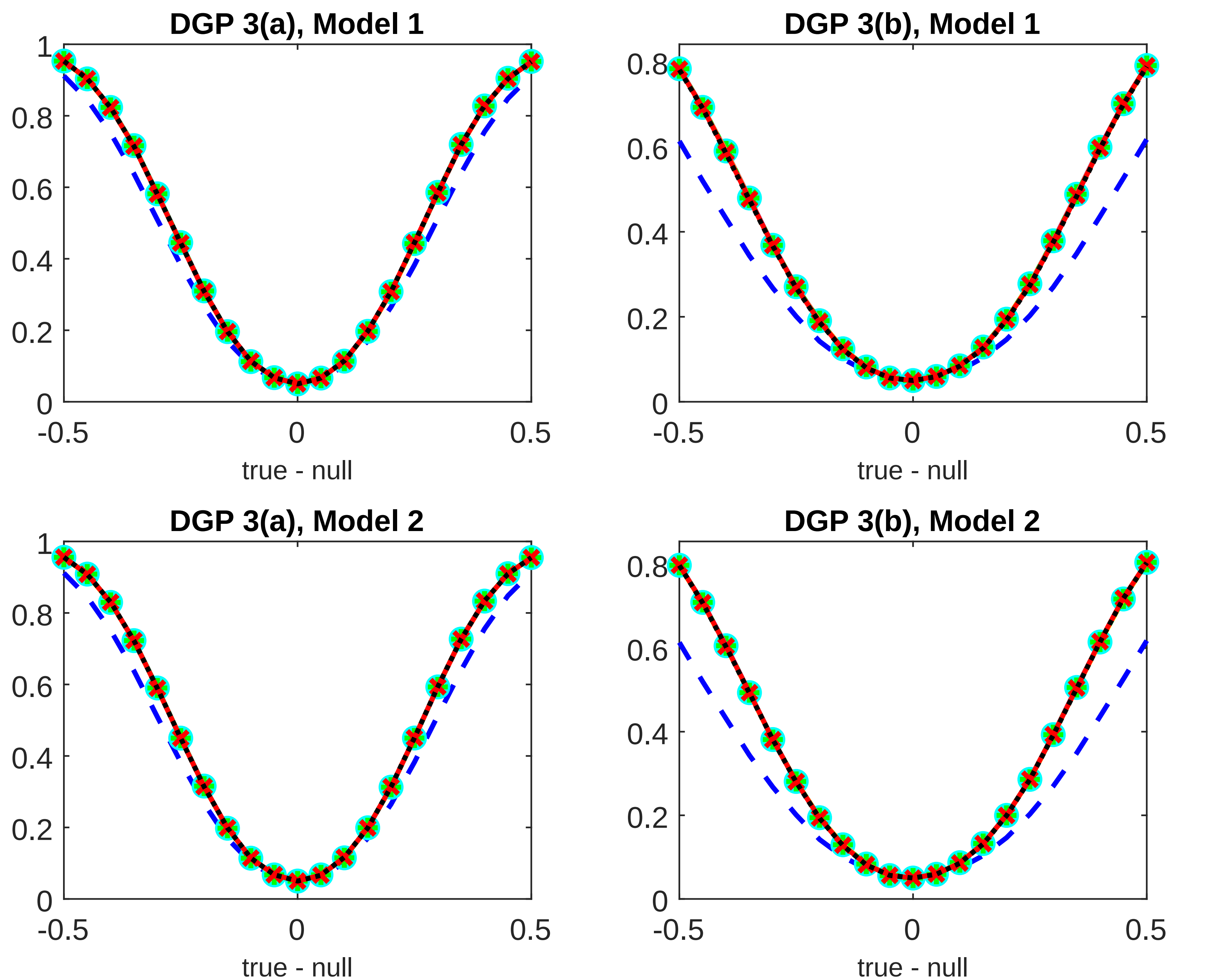

Figure 38: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

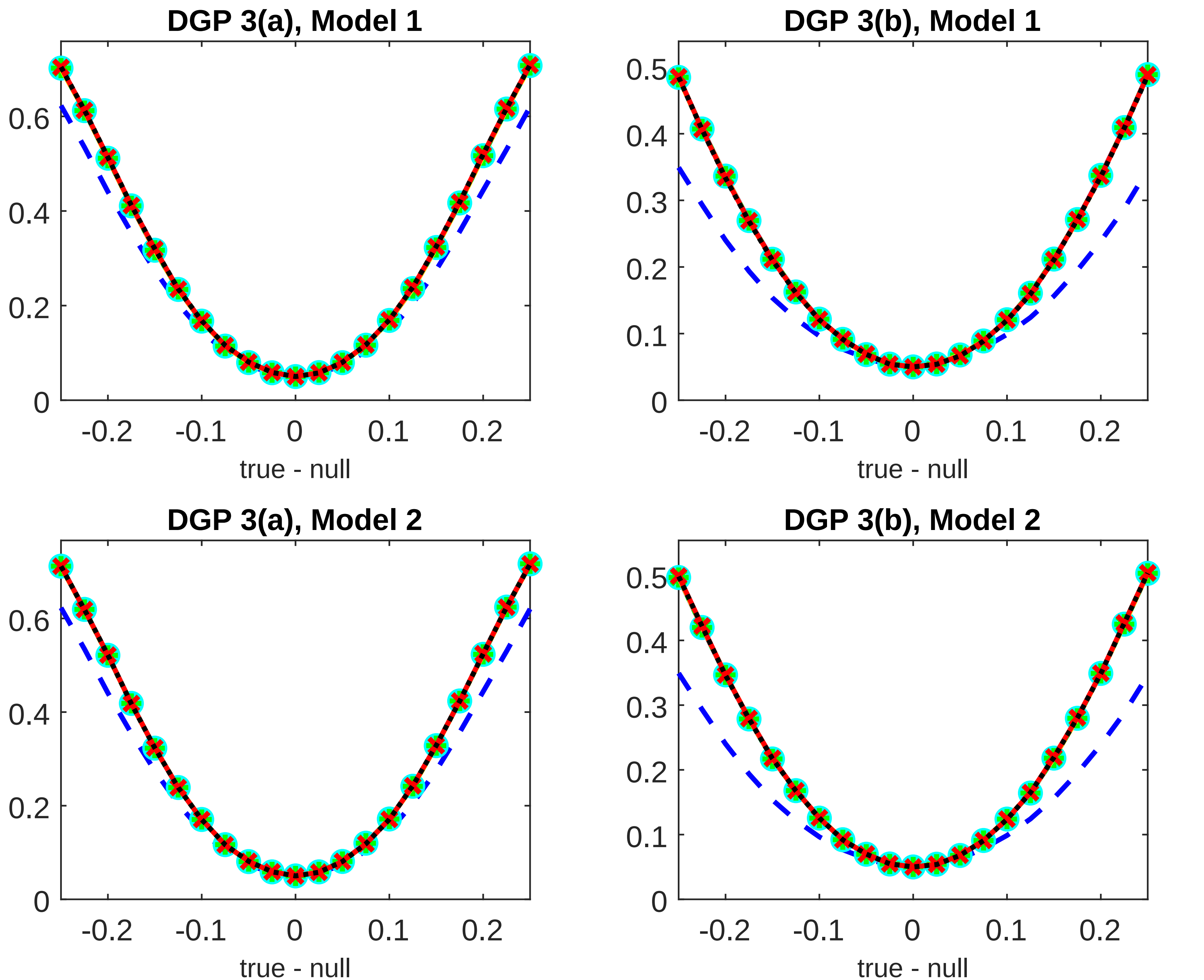

Figure 39: DGP 3 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

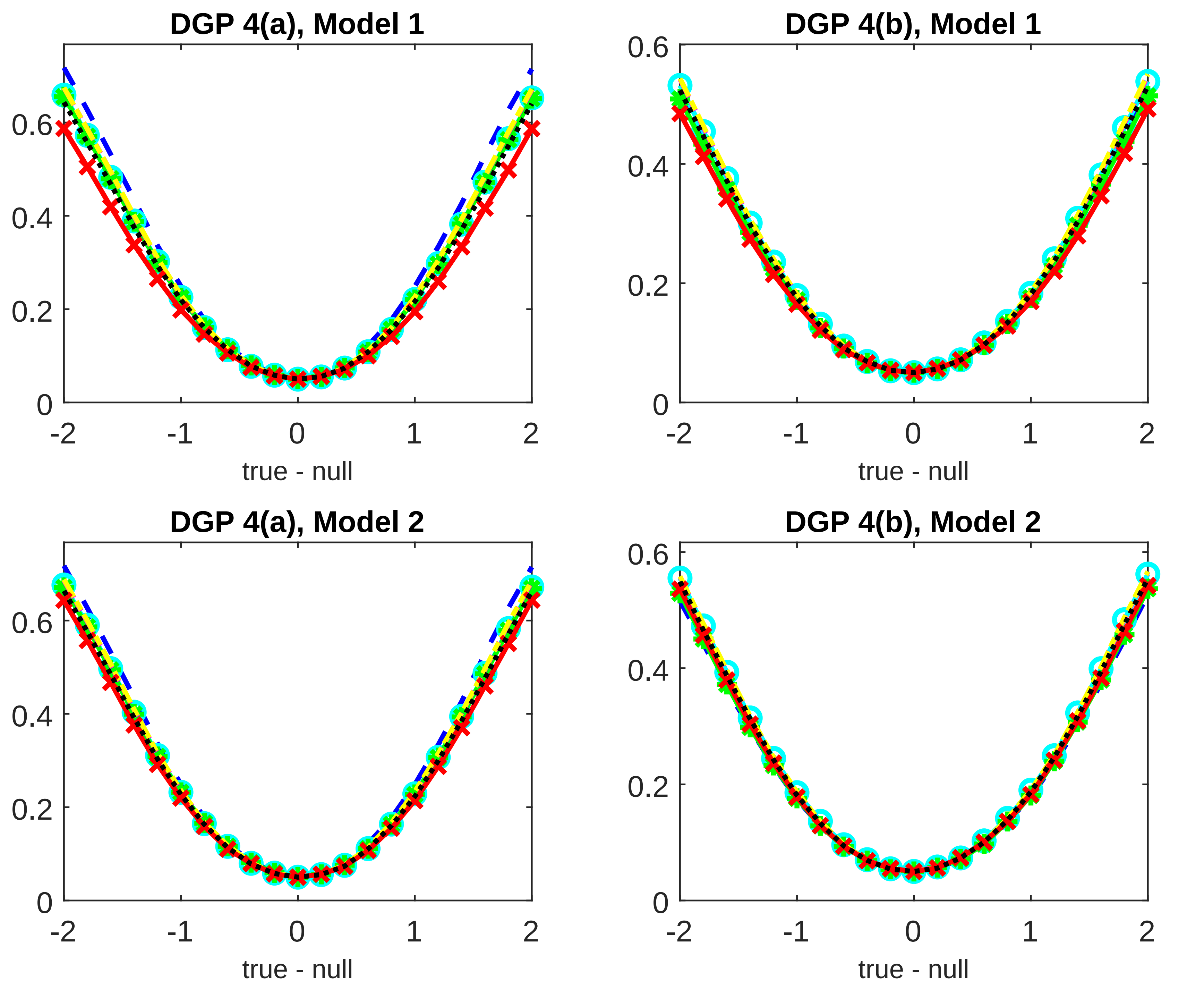

Figure 40: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

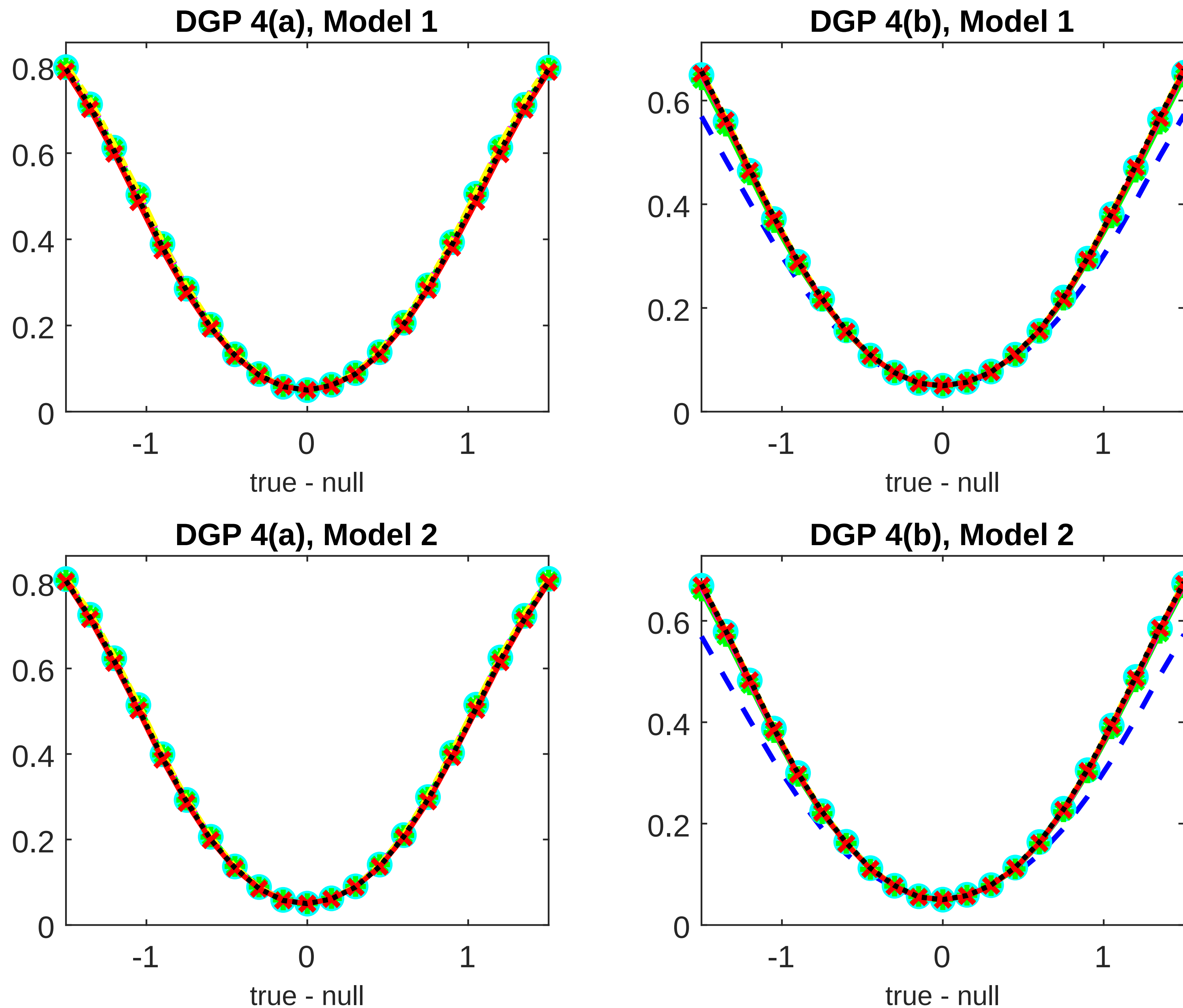

Figure 41: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

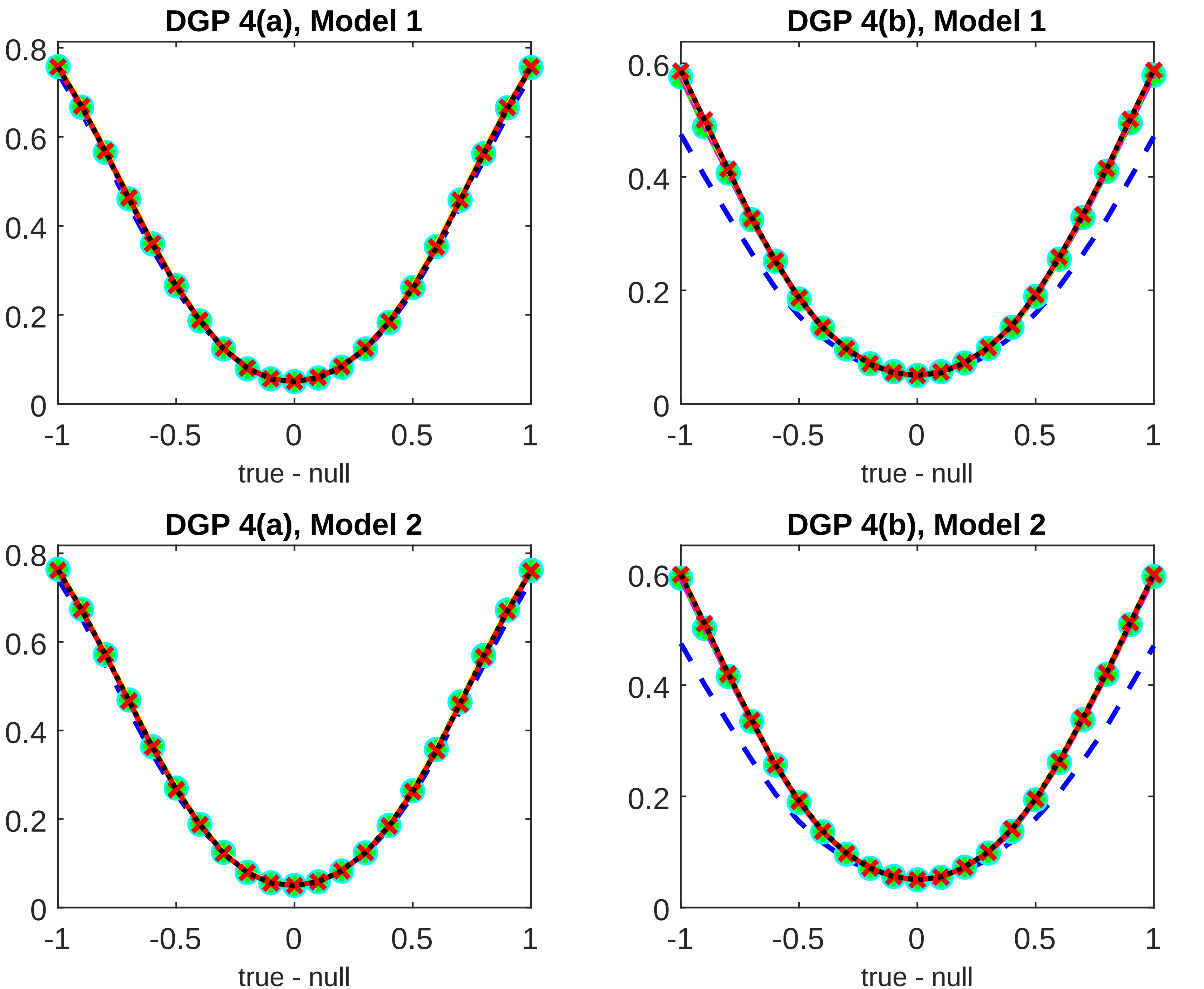

Figure 42: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

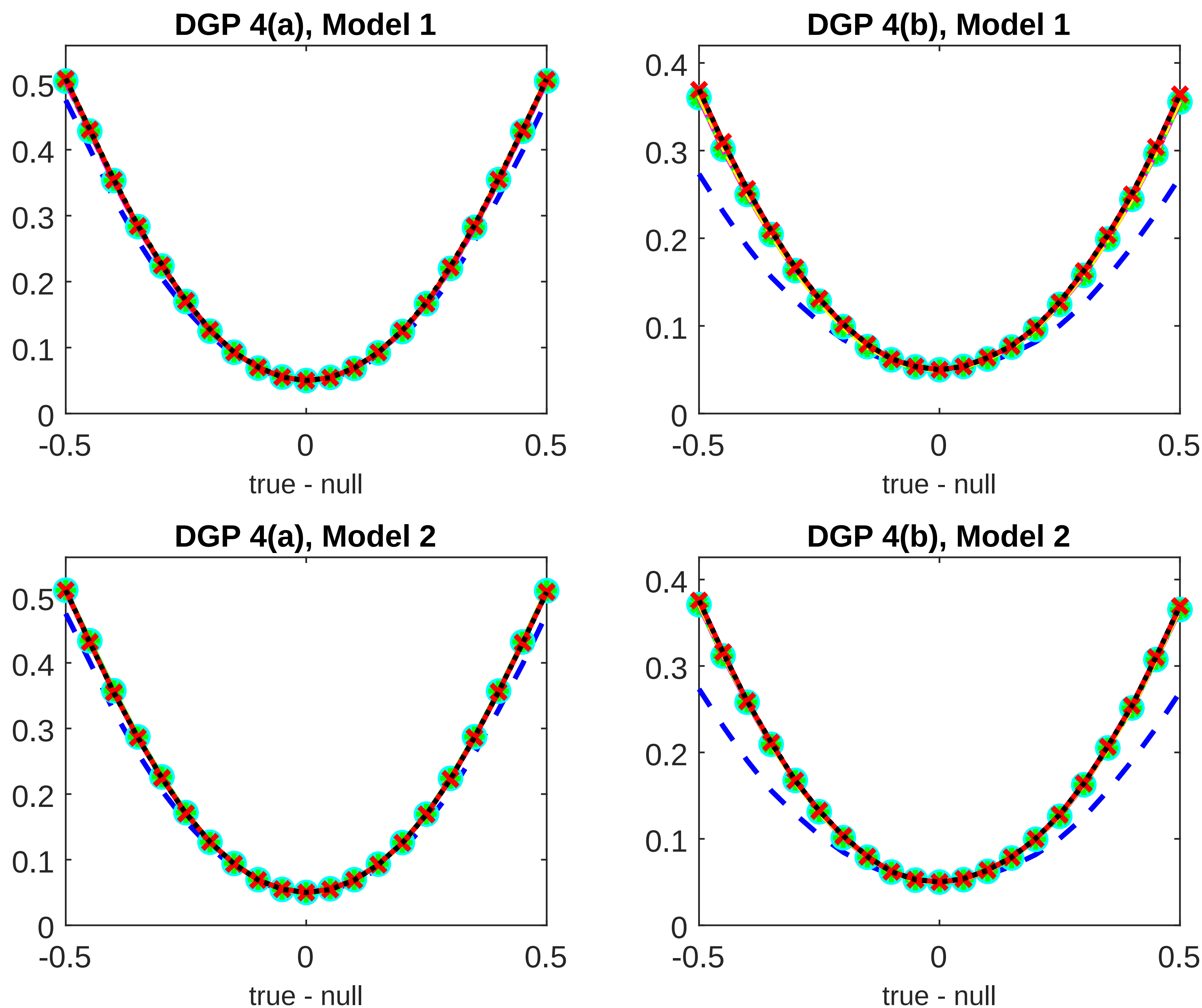

Figure 43: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

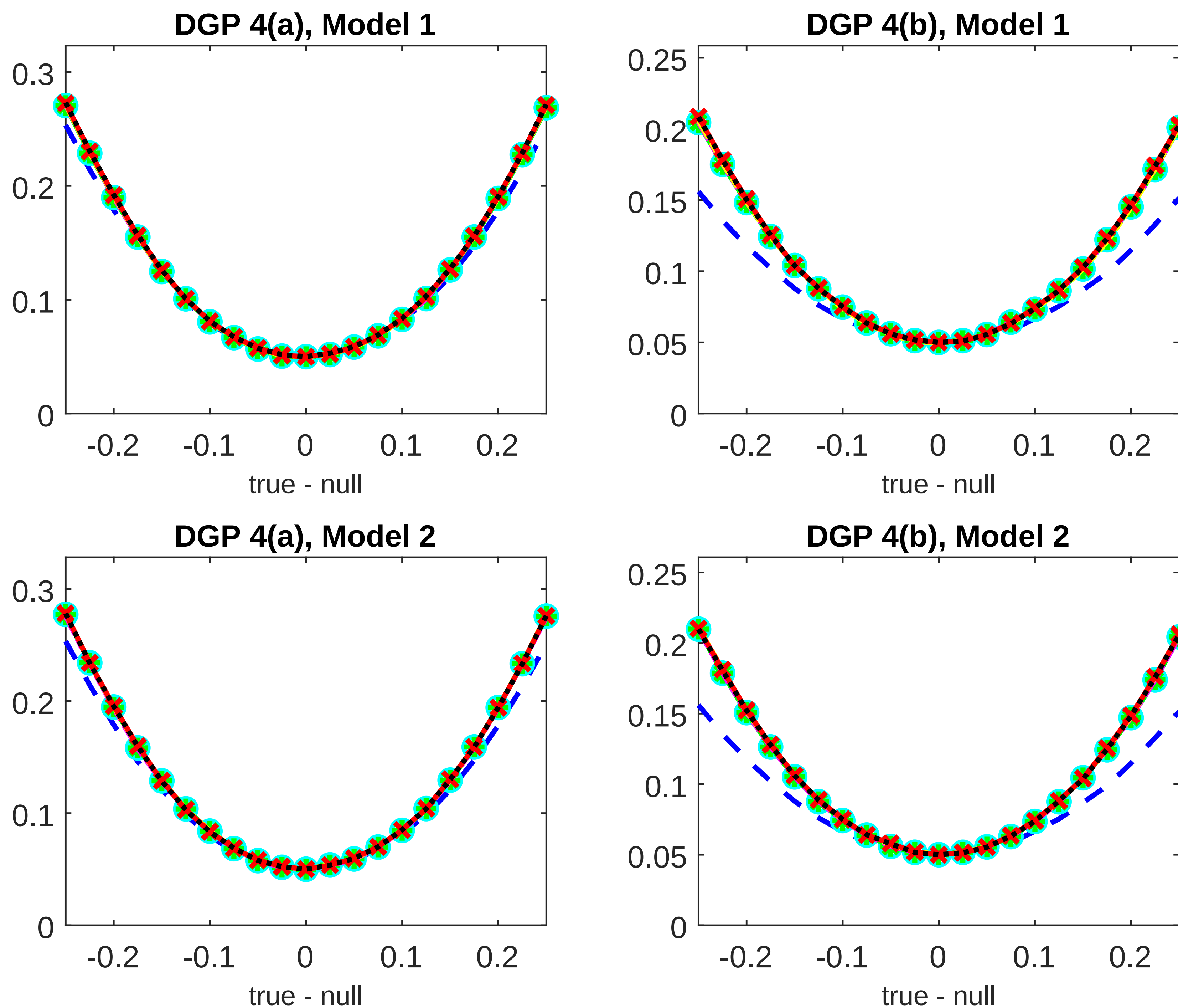

Figure 44: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_1$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

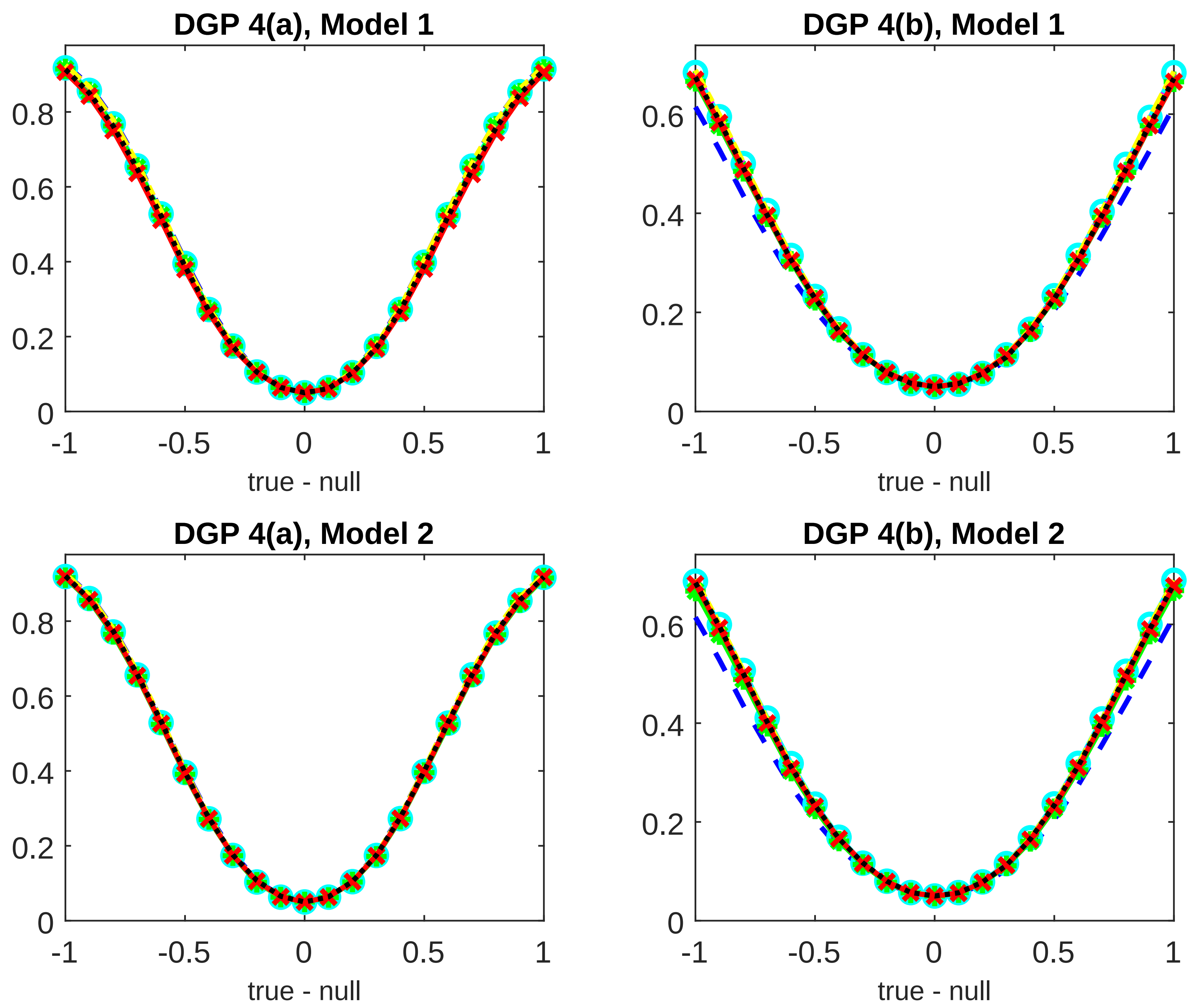

Figure 45: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 25$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

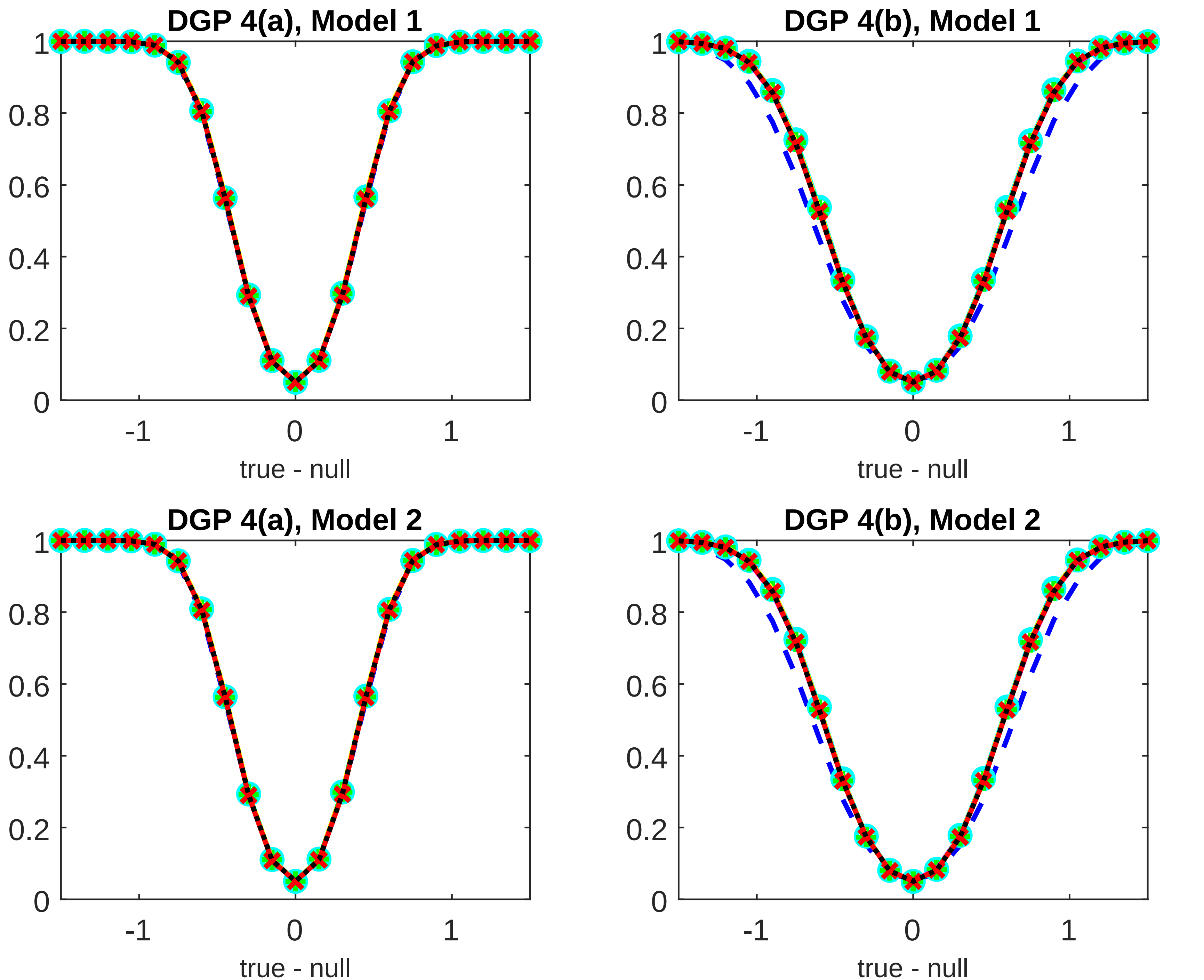

Figure 46: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 50$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

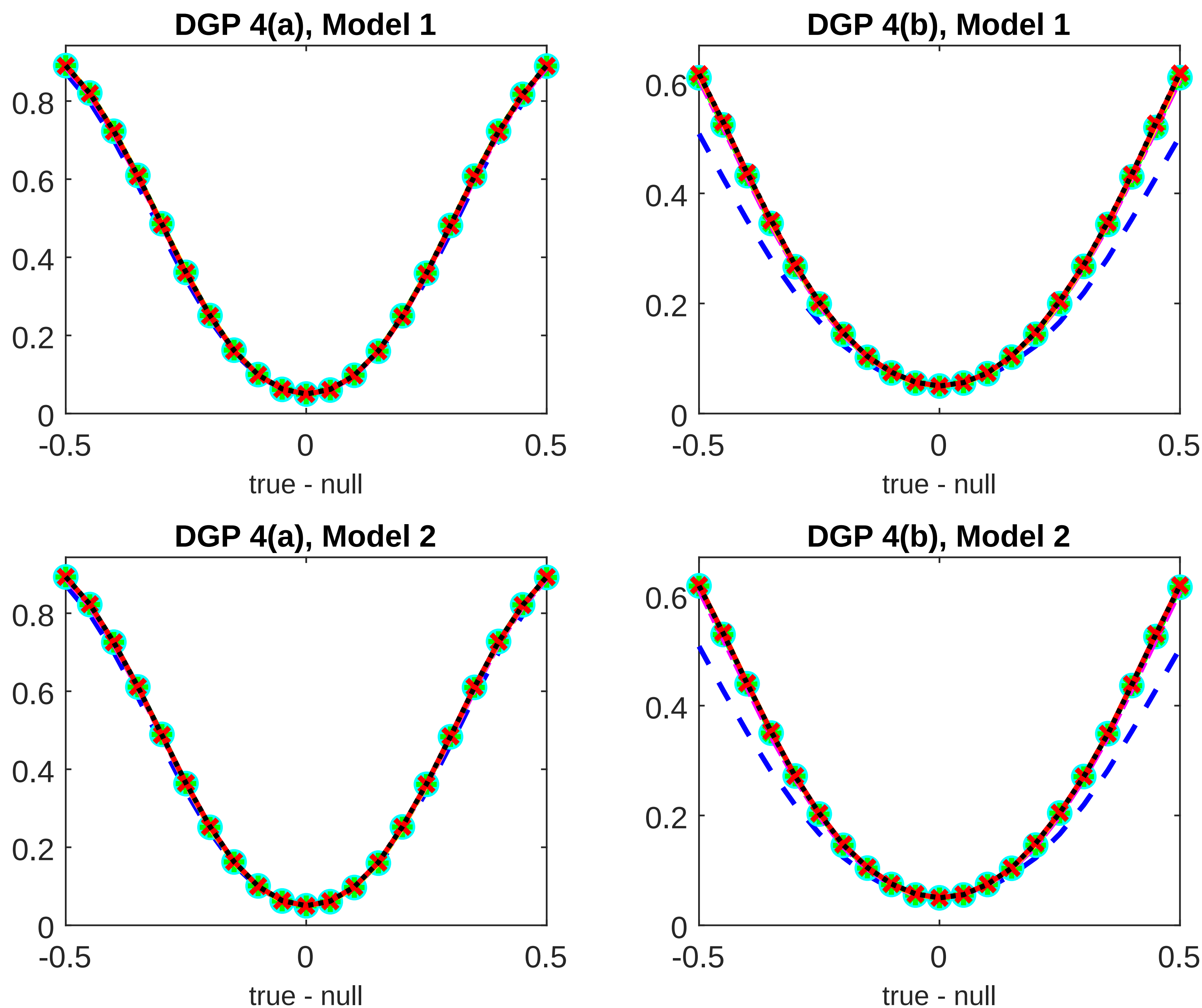

Figure 47: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 100$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

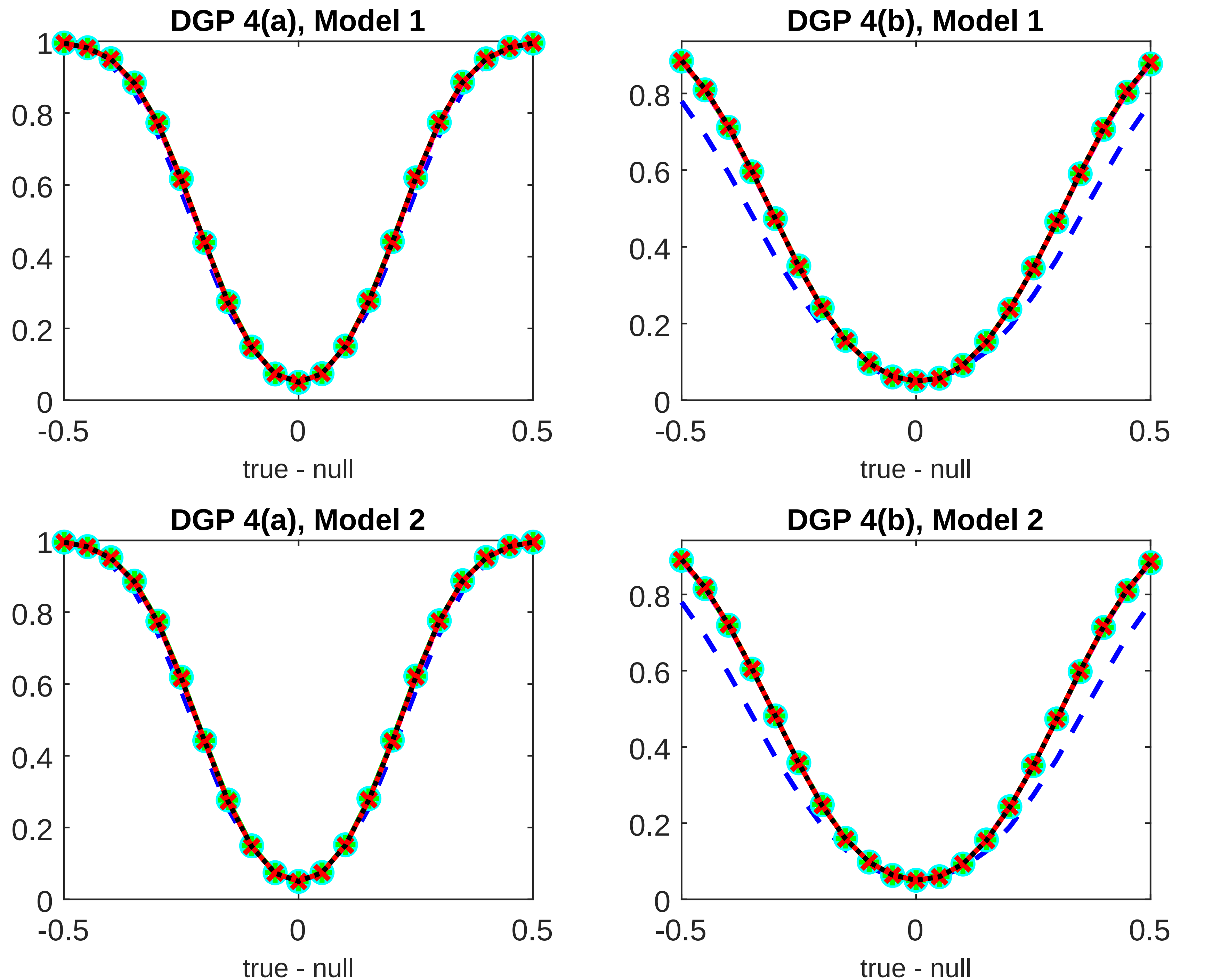

Figure 48: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 200$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.

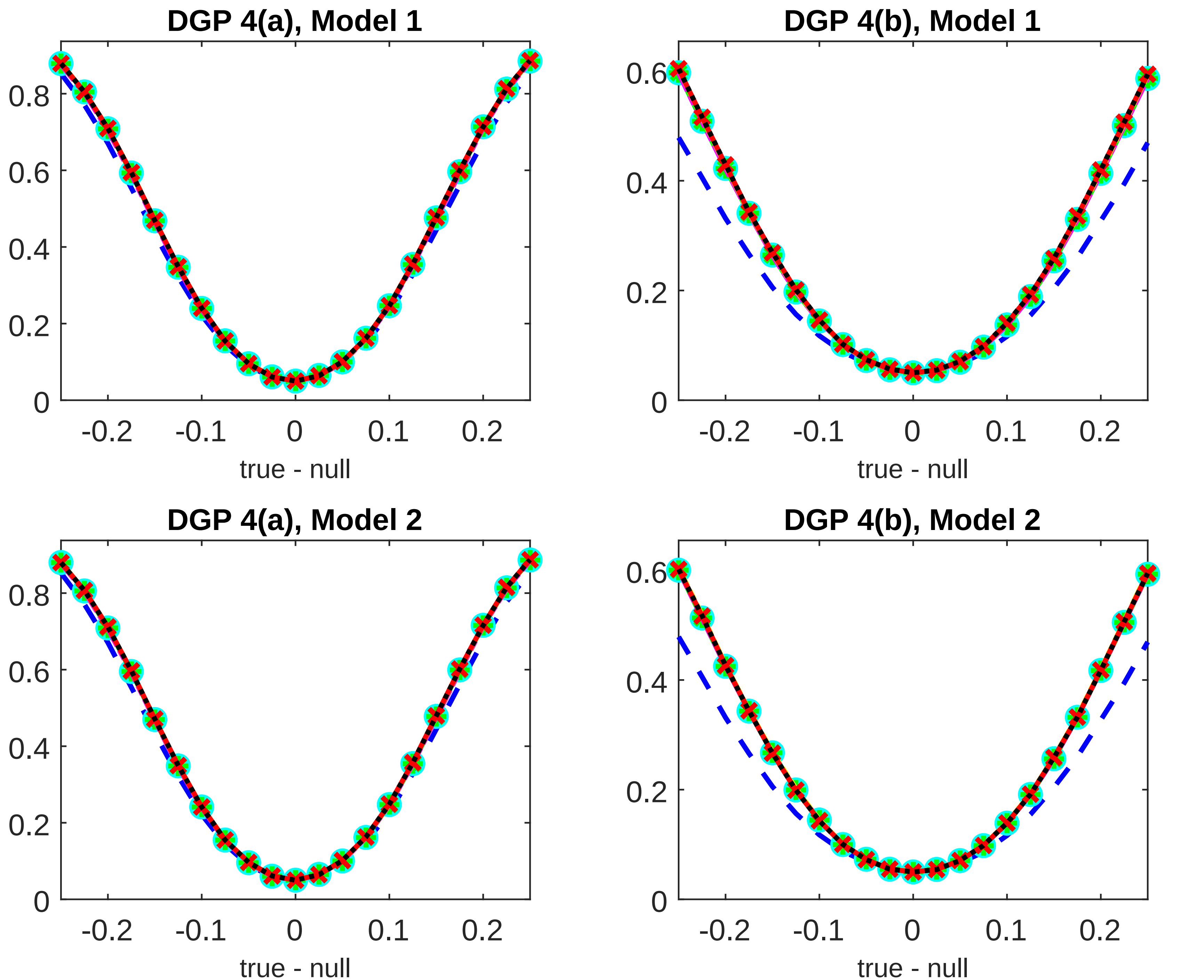

Figure 49: DGP 4 for $\omega_0^2(X)$: Empirical size-corrected power of two-sided 5% Wald test for $h(\beta) := \beta_2$ plotted against the deviation of the null from the truth. **Sample size is** $n = 400$. The Wald tests are based on the following estimators. OLS: blue (- -) line. WLS: cyan (o-) line. ALS: green (*-) line. MINVAR: magenta (-.) line. LINCOM: yellow (-.) line. MWLS-HC1: red (x-) line. MWLS-HC3: black (..) line.