On the use of comparison regions in visualizing stochastic uncertainty in some two-parameter estimation problems

Maren Eckert¹ | Werner Vach²

¹Institute of Medical Biometry and Statistics, Section of Health Care Research and Rehabilitation Research, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
²Department of Orthopaedics and Traumatology, University Hospital Basel, Basel, Switzerland

Correspondence
Maren Eckert, Institute of Medical Biometry and Statistics, Section of Health Care Research and Rehabilitation Research, Faculty of Medicine and Medical Center, University of Freiburg, Hebelstr. 11, 79104 Freiburg, Germany.
Email: eckert@imbi.uni-freiburg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: VA 88/5-1

Abstract
When considering simultaneous inference for two parameters, it is very common to visualize stochastic uncertainty by plotting two-dimensional confidence regions. This allows us to test post hoc null hypotheses about a single point in a simple manner. However, in some applications the interest is not in rejecting hypotheses on single points, but in demonstrating evidence for the two parameters to be in a convex subset of the parameter space. The specific convex subset to be considered may vary from one post hoc analysis to another. Then it is of interest to have a visualization allowing to perform corresponding analyses. We suggest comparison regions as a simple tool for this task.

KEYWORDS
comparison region, confidence region, likelihood ratio test, two-parameter problem, Wald test

1 | INTRODUCTION

When considering simultaneous inference for a two-dimensional parameter vector \( \theta = (\theta_1, \theta_2)' \), it is very common to visualize the point estimate and its stochastic uncertainty by plotting the point estimate and a \((1 - \alpha)\)-confidence region in a two-dimensional plane. This allows a simple judgement whether for any given value \( \theta_0 \) the null hypothesis \( H_0 : \theta = \theta_0 \) can be rejected at level \( \alpha \) or not, just by inspecting the visualization and checking whether \( \theta_0 \) is outside or within the confidence region. This is in particular helpful in post hoc testing, that is, if there is an interest in testing a specific hypothesis generated after publication of the original analysis.

However, in some applications, it can be expected that the interest is not in rejecting certain values of \( \theta \), but in demonstrating evidence for \( \theta \) to be in a convex subset of the two-dimensional parameter space. A few examples should illustrate this point.

(i) Two-dimensional equivalence problems

In establishing new measurement or intervention procedures, it is often of interest to demonstrate that certain characteristics have not changed in comparison to established procedures. Such characteristics can often only be estimated, and equivalence is then established by demonstrating that the true difference is within a certain margin \( \pm \epsilon \). If two characteristics are considered simultaneously, we can consider two margins \( \epsilon_1 \) and \( \epsilon_2 \). Then we may aim to demonstrate that the point corresponding to the two true differences \( \theta_1 \) and \( \theta_2 \) is within the corresponding rectangle, that is, a convex set. It has been argued...
that it might be preferable to consider alternative types of convex sets, allowing some type of compensation between \( \theta_1 \) and \( \theta_2 \). Consequently, besides rectangles, ellipses also have been considered as the regions of interest in bivariate equivalence problems, cf. Hoffelder, Goessl, and Wellek (2015).

(ii) Joint evaluation of sensitivity and specificity

In evaluating the sensitivity \( \theta_{SE} \) and the specificity \( \theta_{SP} \) of a diagnostic test, it is often cumbersome to specify in advance the two necessary thresholds to perform one sided hypothesis tests for each parameter: If the test performs much better than expected for one parameter, one may be satisfied with a lower value for the other parameter. To circumvent this problem, one may specify a weight \( w \) and a threshold \( \xi \) and aim at demonstrating \( w\theta_{se} + (1-w)\theta_{sp} > \xi \), cf. Vach, Gerke, and Hoilund-Carlsen (2012). However, clinicians, patients, or policy makers may have different opinions on how to choose weights or thresholds. Consequently, it might be necessary to take several choices into account simultaneously, that is, to demonstrate \( w_i\theta_{se} + (1-w_i)\theta_{sp} > \xi_i \) for \( i = 1, \ldots, I \). Both with a single choice or a multiple choice the aim is to demonstrate that the true values of sensitivity and specificity are within a convex subset of the parameter space.

(iii) Cost-effectiveness analyses

In cost-effectiveness analyses, we compare the incremental average costs \( \theta_{\Delta(C)} \) of a new intervention with the gain in effectiveness \( \theta_{\Delta(E)} \) with respect to some patient relevant outcome, based on estimates of both. A crucial quantity is the cost effectiveness ratio \( \theta_{\Delta(C)}/\theta_{\Delta(E)} \). Given a specific value \( r \) representing the acceptable ratio, we can consider the net benefit \( b_r = \theta_{\Delta(E)} - r\theta_{\Delta(C)} \), expressing how much the effectiveness exceeds the necessary effectiveness \( r\theta_{\Delta(C)} \) according to the actual costs and the acceptable ratio (Stinnett & Mullaly, 1998). The aim is to demonstrate \( b_r \geq 0 \), or at the best \( b_r \geq b \) for some \( b \), that is, that the pair \( (\theta_{\Delta(C)}, \theta_{\Delta(E)}) \) is in a given convex subset of the the cost-effectiveness plane. Note that the net benefit approach is applicable also if \( \theta_{\Delta(C)} \), \( \theta_{\Delta(E)} \), or both are negative. A similar situation arises when there is a need to balance the benefit and the risk of medical interventions. Indeed, many approaches in benefit–risk assessment consider a benefit–risk plane similar to the cost-effectiveness plane (Guo et al., 2010; Mt-Isa et al., 2014). Gladstone and Vach (2015) suggested to evaluate noninferiority trials by considering an advantage-deficit plane.

It is a first common property of the three examples that they involve half spaces or other convex sets as regions of interest. Here we denote with region of interest a subset of the two-dimensional parameter space for which we are interested to demonstrate that the true parameter values are within this region of interest. They share also a second property: The concrete regions of interest may be defined after publication of the original analysis. Equivalence margins may change due to new regulatory guidelines, or a new type of application may require a stricter margin or may allow a more liberal margin. New stakeholders may have a different opinion on how to weight sensitivity and specificity and which threshold to be used. Policy makers may change their opinion about the acceptable cost-effectiveness ratio. Hence, there is a need to visualize the two parameter estimates and their uncertainty in a way, which allows us to check easily for any given (convex) region of interest, whether we can reject the null hypothesis of \( \theta \) not to be in this region, that is, to generate evidence for the true parameter value to be covered by the region of interest. So we have to support this way of post hoc testing.

We start this paper by looking at the role of confidence intervals and confidence regions with respect to supporting post hoc testing. This way we will identify the need for a new type of region, which we call comparison regions. Then we will present a general construction principle for comparison regions based on one-sided tests for linear combinations of two parameters. We consider the special cases of using likelihood ratio (LR) tests or Wald tests, respectively, allowing to provide explicit representations of comparison regions and a comparison with the shape and size of confidence regions. We further present an illustrative example considering the case of two independent estimates of proportions and take a short look at finite sample properties. Our considerations will take into account that in some applications regions of interest are bounded convex sets, whereas in other applications regions of interest can be unbounded, but share the directions in which they are unbounded.

## 2 THE ROLE OF CONFIDENCE INTERVALS AND CONFIDENCE REGIONS IN POST HOC TESTING

In one-parameter problems, we are used to depict stochastic uncertainty of a parameter estimate \( \hat{\theta} \) by presenting a \( (1-\alpha) \)-confidence interval \( C \) and a \( p \)-value referring to a test for a prespecified null hypothesis. Most research questions require to demonstrate that the true parameter value is above a certain threshold \( \theta_t \), that is, the region of interest is given by \( R = \{ \theta \mid \theta > \theta_t \} \). This suggests to use one-sided tests for the null hypothesis \( H_0 : \theta \leq \theta_t \). However, there is a tradition to prefer two-sided tests
for the null hypothesis $H_0 : \theta = \theta_t$, mainly to avoid the temptation to reverse the region of interest post hoc. Analogously, we prefer to present two-sided confidence intervals to one-sided confidence intervals. If we explicitly use one-sided tests, we typically choose the level $\alpha/2$ to avoid an advantage compared to two-sided testing.

There is a well-known correspondence between confidence intervals and hypothesis tests. If a family $(\phi_{\theta_i})_{\theta_i \in R}$ of level-$\alpha$ tests for $H_0 : \theta = \theta_i$ is given, then $C := \{\theta_i | \phi_{\theta_i} = 0\}$ defines a $(1 - \alpha)$-confidence interval. And if a $(1 - \alpha)$-confidence interval $C$ is given, $\phi_{\theta_i} := 1(\theta_i \notin C)$ defines a level-$\alpha$ test. Consequently, confidence intervals play a role not only for illustrating the imprecision of $\hat{\theta}$, but also in supporting post hoc testing: Whenever after publication of a statistical analysis there is the interest to test the null hypothesis $H_0 : \theta = \theta_i$, we can perform a corresponding hypothesis test by checking whether $C$ is not covering $\theta_i$. This can also be extended to the situation where the region of interest is a half interval as considered above, if the two-sided $(1 - \alpha)$-confidence interval $C = [\theta_{low}, \theta_{up}]$ is constructed in a way such that both $[\theta_{low}, \infty]$ and $[\infty, \theta_{up}]$ are $(1 - \alpha/2)$-confidence intervals. Then the rule $1(C \subseteq R)$ corresponds to a one-sided level-$\alpha/2$ test of $H_0 : \theta \notin R$. This rule to check whether the confidence interval is completely within the region of interest is also probably what many readers of scientific publications actually do if they have to interpret a confidence interval. Figure 1 illustrates this point.

We are less used to two-parameter problems, as they occur rather rarely in biomedical research. In general, the same principles as in the one-parameter case should be applied: We are interested in depicting the stochastic uncertainty of the two-dimensional parameter estimates and the evidence we have against a prespecified null hypothesis, reflecting the research question of interest. The first can be approached by a two-dimensional confidence region, the latter by a $p$-value of a corresponding hypothesis test. We still have a correspondence between confidence regions and hypothesis tests, as long as we consider tests for a single point null hypothesis. So, confidence regions also support the post hoc testing of such null hypotheses.

However, as pointed out above, in two-parameter problems the regions of interest $R$ are often half spaces or convex sets, such that corresponding hypotheses tests for $H_0 : \theta \notin R$ are not testing single point null hypotheses. To support post hoc testing corresponding to such regions of interest, we are hence interested in alternative data-dependent regions $C$, such that $1(C \subseteq R)$ defines a level-$\alpha$ test for $H_0 : \theta \notin R$ for any convex set $R$. We will call such a region a level-$\alpha$ comparison region, and Figure 2 illustrates the use of such a comparison region in post hoc testing. And such comparison regions do exist. As already pointed out by Munk and Pflüger (1999)—also referring to previous work by Aitchison (1964) and Lehmann (1959)—any $(1 - \alpha)$-confidence region $C$ provides also a level-$\alpha$ comparison region: If $\theta_0 \notin R$, then $C \subseteq R$ implies $\theta_0 \notin C$, and hence $P(C \subseteq R) \leq P(\theta_0 \notin C) \leq \alpha$ by definition of a confidence region.

However, the comparison of $R$ with a $(1 - \alpha)$-confidence region is a very conservative approach. We will demonstrate in this paper that it is possible to construct level-$\alpha$ comparison regions, which are distinctly smaller than $(1 - \alpha)$-confidence regions, and hence provide a more powerful approach to post hoc testing. Consequently, if we want to support post hoc testing with respect to convex regions of interest, we should present not only confidence regions, but also comparison regions.

### 3 | A GENERAL CONSTRUCTION PRINCIPLE

If we have for any half space $H$ of $\mathbb{R}^2$ a test $\phi$ for $H_0 : \theta \in H$, then a comparison region $C$ has to fulfil $C \subseteq \bar{H}$ whenever $\phi_H = 1$, that is, whenever we can reject $H_0 : \theta \in H$. Consequently, a straightforward idea is to define $C$ as the intersection of the complement of all half spaces, which we can reject when considered as null hypothesis. In Appendix A, we show that this
idea indeed works under some regularity conditions on the family of tests considered. It is sufficient to consider half spaces, as any convex set can be represented as an intersection of half spaces, and the intersection-union principle allows to define a test based on testing all half spaces.

However, performing the above construction principle explicitly for a given family of tests is rather cumbersome. Fortunately, our main result also provides a simple sufficient condition for comparison regions. This way we can easily check whether a given definition of a convex set defines indeed a comparison region.

To present this main result—also derived in Appendix A—we need to introduce the concept of the direction of a half space. Any half space of $\mathbb{R}^2$ can be expressed as $H_{t,r} := \{(\theta_1, \theta_2)' \mid \theta_1 \cos t + \theta_2 \sin t \geq r\}$ with $t \in [0, 2\pi)$ and $r \in \mathbb{R}$. The parameter $t$ represents the direction of the half space, as it corresponds to the first polar coordinate of all points on a line orthogonal to the boundary line of the half space and hitting the origin (cf. Figure 3). For any direction $t$, it is reasonable to assume that $H_{t,r}$ can be rejected if $r$ is large enough, and by regularity conditions we can ensure that there is a minimal $r$, for which $H_{t,r}$ can be rejected. And we can expect that this half space plays a role in characterizing a comparison reason. Indeed, they will coincide with the outer tangent halfspace $H_{t,C}$ of $C$ in direction $t$. This outer tangent half space is the half space with direction $t$ with a boundary line equal to a tangent of $C$ and disjoint to $C$.

Our main result takes into account that in many applications not all possible directions $t \in [0, 2\pi)$ are of relevance. If all regions of interest are unbounded in some direction, the comparison region is allowed to be unbounded in this direction, too, and there is no outer tangent half space in this direction. We can now present our main result as a theorem after introducing two regularity conditions:

The family of tests should be monotone

$$H_1 \subseteq H_2 \Rightarrow \varphi_{H_1} \geq \varphi_{H_2}$$

and it should be right-continuous, that is, if for any sequence $(r_n)_{n \in \mathbb{N}_0} \in \mathbb{R}$ with $r_n \downarrow r_0$, we have

$$\varphi_{H_{t,r_n}} \rightarrow \varphi_{H_{t,0}} \quad \forall t \in [0, 2\pi).$$

**Theorem 3.1.** Let $(\varphi_H)_{H \in \mathcal{H}}$ be a family of level-$\alpha$-tests for $H_0: \theta \in H$, satisfying conditions (1) and (2). Then any open, convex set $C$ with $T_C \supseteq T$ and

$$H_{t,C} = H_{t,\partial} \quad \forall t \in T$$

defines a level-$\alpha$-comparison region for all $R$ with $T_R \subseteq T$.

It should be emphasized that the theorem specifies a sufficient, but not a necessary condition. In particular, any superset of a comparison region $C$ is by definition a comparison region, too. We will make use of this fact in the sequel, where we first construct bounded comparison regions. In a second step, we enlarge them to unbounded comparison regions, restricting boundedness to the directions of interest.
4 | LR TEST AND WALD TEST BASED COMPARISON REGIONS

In many two-parameter estimation problems, we can define hypothesis tests for any half space using the LR test principle or the Wald test principle. Hence, it is of interest to know how comparison regions based on these principles look like. This is considered in detail in Appendices B and C. We present here only the main results as two lemmata.

4.1 | LR test–based comparison regions

Lemma 4.1. Let $l(\theta)$ be a continuous and strictly concave log-likelihood function of a two-parameter problem and $l^*$ the value at the maximum likelihood estimator. Then the contour set

$$C_\alpha = \{ \theta \mid l(\theta) - l^* > -\frac{1}{2} \chi^2_{1,1-2\alpha} \}$$

is a level-$\alpha$-comparison region, for any choice of $T$ and $\alpha < 0.5$.

The interesting point here is that a level-$\alpha$-comparison region coincides with a $(1-\alpha')$-confidence region, but with $\alpha' > \alpha$: The LR test–based $(1-\alpha)$-confidence region is given by

$$C^*_\alpha = \{ \theta \mid l(\theta) - l^* > -\frac{1}{2} \chi^2_{2,1-\alpha} \}.$$ 

As $\chi^2_{1,0.9} = 2.71 = \chi^2_{2,0.742}$, a 5% comparison region coincides with a 74.2% confidence region.

4.2 | Wald test–based comparison regions

The level-$\alpha$-Wald test for $H_0 : \theta \in H_{t,r}$ is given by

$$\varphi_W^{t,r} = \begin{cases} 1 & \text{if } \sqrt{\tilde{\Sigma}} (\hat{\theta} - \theta) < u_\alpha \\ 0 & \text{otherwise} \end{cases},$$

where $u_\alpha$ is the $\alpha$-percentile of the standard normal distribution.

Theorem 3.1 allows to derive the following result:

Lemma 4.2. If $\varphi_W^{t,r}$ as defined above defines a level-$\alpha$-test, then

$$C^W_\alpha = \{ \theta \mid (\hat{\theta} - \theta)' \tilde{\Sigma}^{-1} (\hat{\theta} - \theta) < \chi^2_{1,1-2\alpha} \}$$

defines a level-$\alpha$-comparison region.

If we use Wald tests for a single point null hypothesis to construct $(1-\alpha)$-confidence regions for $\theta$, we arrive at

$$C^*_\alpha = \{ \theta \mid (\hat{\theta} - \theta)' \tilde{\Sigma}^{-1} (\hat{\theta} - \theta) < \chi^2_{2,1-\alpha} \}$$

and we have the same correspondence between comparison regions and confidence regions as above.

5 | EXAMPLE: JOINT EVALUATION OF SENSITIVITY AND SPECIFICITY

In accuracy studies on a single diagnostic test, we observe $n_1$ diseased patients with $k_1$ having a positive test result, and $n_0$ disease-free patients with $k_0$ having a negative test result. The two estimates of interest are $\hat{\theta}_{SE} = k_1/n_1$ and $\hat{\theta}_{SP} = k_0/n_0$, that is, two independent proportions. For the construction of comparison regions, we can consider the LR test based on

$$l(\theta_{SE}, \theta_{SP}) = \log \left( n_1 \choose k_1 \right) \theta_{SE}^{k_1} (1-\theta_{SE})^{n_1-k_1} + \log \left( n_0 \choose k_0 \right) \theta_{SP}^{k_0} (1-\theta_{SP})^{n_0-k_0}$$
FIGURE 4 Comparison and confidence regions for sensitivity and specificity based on the observed values in the study of Xu et al. (2017):

\( n_1 = 112, \ k_1 = 106, \ n_0 = 216, \ k_0 = 198 \)

or the Wald test based on \( \hat{\Sigma} = \begin{pmatrix} \hat{\sigma}^2_{SE} & 0 \\ 0 & \hat{\sigma}^2_{SP} \end{pmatrix} \) with

\[ \hat{\sigma}^2_{SE} = \frac{\hat{\theta}_{SE}(1 - \hat{\theta}_{SE})}{n_1} \text{ and } \hat{\sigma}^2_{SP} = \frac{\hat{\theta}_{SP}(1 - \hat{\theta}_{SP})}{n_0}. \]

We illustrate the use of comparison regions in analyzing diagnostic accuracy using a multicenter study in 308 patients investigating the use of angiography-based quantitative flow ratio (QFR) measurements for online assessment of coronary stenosis (Xu et al., 2017). The study explores the sensitivity and specificity of QFR when using fractional flow reserve as reference standard. It results in estimates for sensitivity of 94.6% (95% CI [88.7, 98.0]) and for specificity of 91.7% (95% CI [87.1, 95.0]).

In Figure 4, 5% comparison regions and 95% confidence regions following both principles are shown for the data of this example. In this presentation, we take into account that in considering a weighted average of sensitivity and specificity, we are only interested in nonnegative weights. So all regions of interest represented by half spaces are unbounded for any \( t \in (0, \pi/2) \), and this holds also for intersections of such regions. So when starting with the full comparison region presented in the previous section, we can enlarge the comparison region in these directions and arrive at the presentation given in Figure 4.

The Figure illustrates nicely that in post hoc testing the use of comparison regions represents a more powerful approach than the use of confidence regions. For example, when we are interested to demonstrate that the average of sensitivity and specificity is above 0.9, we would be unable to reject the null hypothesis \( 0.5(\theta_{SE} + \theta_{SP}) < 0.9 \) when we use the comparison with a confidence region, but we would be able to reject this null hypothesis when we use the comparison with the comparison region.

6 | A SMALL INVESTIGATION OF FINITE SAMPLE PROPERTIES

We can observe in Figure 4 that LR test–based comparison regions are distinctly wider than Wald test–based comparison regions, mimicking the behavior of confidence regions. It is well known that Wald test–based confidence intervals for proportions behave poorly in small samples, in particular if the true probabilities are close to 0 or 1 (Agresti & Coull, 1998). We may expect such a behaviour also for Wald test–based comparison regions.

We investigate this question in a small investigation of finite sample properties. We consider a true sensitivity of 0.95 and and true specificity of 0.9—motivated by our example—and select the sample size accordingly to our example, but consider a doubled sample size also. We evaluate the exact level of post hoc tests based on a comparison of the region of interest with the comparison region. As region of interest, we consider five different half spaces and three convex sets built by intersection of
can easily be generalized to the two-dimensional case, whereas our approach allows to use the distinctly smaller region, as testing more than one single point hypothesis requires to adjust for multiple testing. We demonstrated in our paper that the level of confidence regions has to be increased to allow an interpretation as a comparison region. In that paper, it was argued that the level of confidence regions has to be increased to allow an interpretation as a comparison region.

A first systematic investigation about interpreting confidence regions as comparison regions has been given by Aitchison (1964). This also holds for LR test–based 5% comparison regions, but to a distinctly lower degree, in particular if the sample size is limited. This is not surprising, as our construction principle aimed at hitting with the tangents of the comparison regions exactly the boundary line of the maximal half space, which we can support as region of interest by hypothesis testing. For intersections of half spaces, we rely on the intersection-union principle, which is known to be conservative.

Two observations in Table 1 are noteworthy: First, if the regions of interest are half spaces, Wald test–based 5% comparison regions define a test with an actual level distinctly larger than the nominal level of 5%, in particular if the sample size is limited. This also holds for LR test–based 5% comparison regions, but to a distinctly lower degree, in particular if the borderlines of half spaces are not parallel to the coordinate axes. Second, when the regions of interest are intersections of half spaces, comparison regions turn out to be rather conservative. This is not surprising, as our construction principle aimed at hitting with the tangents of the comparison regions exactly the boundary line of the maximal half space, which we can support as region of interest by hypothesis testing. For intersections of half spaces, we rely on the intersection-union principle, which is known to be conservative.

### 7 | RELATION TO PREVIOUS WORK

A first systematic investigation about interpreting confidence regions as comparison regions has been given by Aitchison (1964). In that paper, it was argued that the level of confidence regions has to be increased to allow an interpretation as a comparison region, as testing more than one single point hypothesis requires to adjust for multiple testing. We demonstrated in our paper that the opposite is true.

Our results can be seen as a generalization of the confidence interval inclusion principle in one-dimensional equivalence testing, cf. chapter 3.1 in Wellek (2010): If a \((1 - 2\alpha)\)-confidence interval \(C\) for \(\theta \in \mathbb{R}\) is given, then for any bounded interval \(R\) the rule to reject \(H_0 : \theta \notin R\) if \(C \subseteq R\) defines a level-\(\alpha\)-test. Moreover, \(C\) can be expressed as the intersection of two \((1 - \alpha)\)-confidence intervals, cf. Schurmann (1987), which corresponds to our general construction principle.

Munk and Pfüger (1999) considered already a generalization of the interval inclusion principle to the higher dimensional case by demonstrating that in general \((1 - 2\alpha)\)-confidence regions are \(\alpha\)-comparison regions under certain regularity conditions. However, to obtain a 5% comparison region in the two-dimensional case, their approach suggests to use a 90% confidence region as comparison region, whereas our approach allows to use the distinctly smaller 74.2% confidence region. Note that our approach can easily be generalized to the \(p\)-dimensional case.

### Table 1

| Region of interest | LR test based | Wald test based |
|--------------------|---------------|-----------------|
| \(n_1 = 112, n_0 = 216\) |               |                 |
| \(R_1 \equiv (0.95, 1] \times [0, 1]\) | 0.0771         | 0.0771          |
| \(R_2 \equiv [0, 1] \times (0.9, 1]\) | 0.0506         | 0.0782          |
| \(R_3 \equiv R_1 \cap R_2 = (0.95, 1] \times (0.9, 1]\) | 0.0067         | 0.0060          |
| \(R_4 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.50 \theta_{SE} + 0.50 \theta_{SP} > 0.925\) | 0.0538         | 0.0803          |
| \(R_5 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.75 \theta_{SE} + 0.25 \theta_{SP} > 0.9375\) | 0.0594         | 0.0969          |
| \(R_6 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.25 \theta_{SE} + 0.75 \theta_{SP} > 0.9125\) | 0.0553         | 0.0699          |
| \(R_7 \equiv R_3 \cap R_4 \cap R_6\) | 0.0183         | 0.0278          |
| \(R_8 \equiv R_3 \cap R_7\) | 0.0067         | 0.0060          |

| Region of interest | LR test based | Wald test based |
|--------------------|---------------|-----------------|
| \(n_1 = 224, n_0 = 432\) |               |                 |
| \(R_1 \equiv (0.95, 1] \times [0, 1]\) | 0.0659         | 0.0659          |
| \(R_2 \equiv [0, 1] \times (0.9, 1]\) | 0.0559         | 0.0559          |
| \(R_3 \equiv R_1 \cap R_2 = (0.95, 1] \times (0.9, 1]\) | 0.0037         | 0.0037          |
| \(R_4 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.50 \theta_{SE} + 0.50 \theta_{SP} > 0.925\) | 0.0539         | 0.0692          |
| \(R_5 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.75 \theta_{SE} + 0.25 \theta_{SP} > 0.9375\) | 0.0558         | 0.0814          |
| \(R_6 \equiv (\theta_{SE}, \theta_{SP}) \mid 0.25 \theta_{SE} + 0.75 \theta_{SP} > 0.9125\) | 0.0530         | 0.0651          |
| \(R_7 \equiv R_3 \cap R_4 \cap R_6\) | 0.0169         | 0.0238          |
| \(R_8 \equiv R_3 \cap R_7\) | 0.0037         | 0.0037          |
8 | DISCUSSION

In this paper, we have introduced the concept of comparison regions to support post hoc testing in two-parameter estimation problems. Comparison regions allow to check visually whether we have evidence for the true parameter values to be in a given convex subset of the parameter space. In several areas of statistical analyses, we are confronted with the task to prepare such post hoc testing in publishing the results of a study.

We have suggested a first construction principle for comparison regions, but we cannot claim any optimality of our approach. The question of optimality should be a topic of future research. Our results for LR test and Wald test based comparison regions also raise the question: whether any convex 74.2% confidence region may define a 5% comparison region. We would like to point out already one important difference to confidence regions: Although confidence regions keep their defining property under monotone transformations of the single parameters, this does not hold for comparison regions. This is a simple consequence of the fact that already the convexity of subsets can be destroyed by such transformations.

One simple difference between confidence regions and comparison regions is that the latter are just smaller. This result may be a little bit counterintuitive, as when considering a half-space as a null hypothesis, we test many single point null hypotheses (cf. Aitchison, 1964). However, the property of being a convex set poses a structure onto this testing problem, such that it can be approached by a one-dimensional testing problem requiring only one degree of freedom. This stands in contrast to two-dimensional confidence regions related to tests requiring to use two degrees of freedom.

It might also be argued that in two-parameter estimation problems there is some danger to reverse regions of interest post hoc. This would suggest to use an $\alpha/2$-level comparison region, such that post hoc tests of half spaces are automatically performed at level $\alpha/2$. Indeed, hypotheses tests of half-spaces are logically one-sided tests (cf. the structure of the Wald test considered in Section 4.2). Whether there is a risk to reverse regions of interest in two-parameter estimation problems is a question of debate, which can only be clarified in the context of a concrete application. Moreover, sometimes we can limit this risk by a careful choice of how to present a comparison region. For example, the presentation chosen in our example does only support to check region of interests based on nonnegative weights and a lower threshold. If we consider an upper threshold, we can never reach $C \subseteq R$. In any case, it should be noted that $\alpha/2$-level comparison regions are still smaller than $(1 - \alpha)$-confidence regions. For example, 2.5% Wald test and LR test based comparison regions coincide with 85% confidence regions.

As long as post hoc testing is performed with respect to half spaces, the use of comparison regions seems to provide an efficient approach. This is not the case for convex regions that can be represented by an intersection of half spaces with varying directions. In this situation, it might be desirable to perform explicitly a test taking into account the shape of the region of interest. However, this requires that sufficient information is given in the publication. An example of this type arises in accuracy studies, when a stakeholder is not willing to consider a weighted average of sensitivity and specificity, but requires to demonstrate that both sensitivity and specificity are above certain thresholds. Then we can require that the corresponding tests are both significant at level $1 - \sqrt{1 - \alpha}$. We can perform these tests if both nominator and denominator of the sensitivity and specificity estimates are given, which is typically the case. This approach is more powerful than using the comparison with the comparison region.

We would like to recommend the use of comparison regions in publications, whenever it is expected that convex subsets of a two-dimensional parameter space have to be checked later for compatibility with the data. However, it can never be excluded that later single point hypotheses also need to be checked. Then it would be very dangerous if comparison regions are confused with confidence regions, as the actual coverage probability of an $\alpha$-level comparison region is much less than $1 - \alpha$. We hence recommend to present both confidence regions and comparison regions, and guiding the user in the correct application by using a dotted line for the boundary of confidence regions—reminding on the comparison with a single point—and using a solid line for the boundary of comparison regions—reminding on the comparison with a convex set.

ACKNOWLEDGMENT

This work was supported by the German Research Foundation (DFG) [VA 88/5-1]. The article processing charge was funded by the University of Freiburg in the funding programme Open Access Publishing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.
REFERENCES

Agresti, A., & Coull, B. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician, 52*, 119–126.

Aitchison, J. (1964). Confidence-region tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26, 462–476.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics, 25*, 573–578.

Friendly, M., Monette, G., & Fox, J. (2013). Elliptical insights: Understanding statistical methods through elliptical geometry. *Statistical Science, 28*, 1–39.

Gladstone, B. P., & Vach, W. (2015). Analyzing noninferiority trials: It is time for advantage deficit assessment – an observational study of published noninferiority trials. *Open Access Journal of Clinical Trials, 7*, 11–21.

Guo, J. J., Pandey, S., Doyle, J., Bian, B., Lis, Y., & Raisch, D. W. (2010). A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value in Health, 13*, 657–666.

Hoffelder, T., Goessl, R., & Wellek, S. (2015). Multivariate equivalence tests for use in pharmaceutical development. *Journal of Biopharmaceutical Statistics, 25*, 417–437.

Jaeger, A. (2016). Computation of two- and three-dimensional confidence regions with the likelihood ratio. *The American Statistician, 70*, 395–398.

Lehmann, E. L. (1959). *Testing statistical hypotheses*. Hoboken, NJ: Wiley.

Mt-Isa, S., Hallgreen, C. E., Wang, N., Callreus, T., Genov, G., Hirsch, I., … IMI-PROTECT benefit-risk participants. (2014). Balancing benefit and risk of medicines: A systematic review and classification of available methodologies. *Pharmacoepidemiology and Drug Safety, 23*, 667–678.

Munk, A., & Pflüger, R. (1999). 1 − 𝛼 equivariant confidence rules for convex alternatives are 𝛼/2-level tests—With applications to the multivariate assessment of bioequivalence. *Journal of the American Statistical Association, 94*, 1311–1319.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, 15*, 657–680.

Stinnett, A. A., & Mullahy, J. (1998). Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making, 18*, S68–S80.

Vach, W., Gerke, O., & Hoilund-Carlsen, P. F. (2012). Three principles to define the success of a diagnostic study could be identified. *Journal of Clinical Epidemiology, 65*, 293–300.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* (2nd ed.). Boca Raton, FL: CRC Press.

Xu, B., Tu, S., Qiao, S., Qu, X., Chen, Y., Yang, J., … Hu, S. (2017). Diagnostic accuracy of angiography-based quantitative flow ratio measurements for online assessment of coronary stenosis. *Journal of the American College of Cardiology, 70*, 3077–3087.

SUPPORTING INFORMATION

Additional supporting information including source code to reproduce the results may be found online in the Supporting Information section at the end of the article.

How to cite this article: Eckert M, Vach W. On the use of comparison regions in visualizing stochastic uncertainty in some two-parameter estimation problems. *Biometrical Journal*. 2020;62:598–609. https://doi.org/10.1002/bimj.201800232

APPENDIX A: DERIVATION OF THE MAIN RESULT

Any closed half-space can be expressed as $H_{t,r} := \{(\theta_1, \theta_2)' \mid \theta_1 \cos t + \theta_2 \sin t \geq r\}$ with $t \in [0, 2\pi)$ and $r \in \mathbb{R}$. Furthermore, $(H_{t,r})_{t \in [0, 2\pi), r \in \mathbb{R}}$ provides a parametrization of the set $H$ of all closed half spaces of $\mathbb{R}^2$.

It is well known that any open convex subset $A$ of $\mathbb{R}^2$ can be expressed as the intersection of the complements of its tangent half spaces, that is,

$$A = \bigcap_{t \in T_A} H_{t,A}$$
with $H_{t,A}$ denoting the maximal element in $H_{t,A} := \{ H_t, r \in \mathbb{R}, A \cap H_t, r = \emptyset \}$ and $T_A := \{ t \in [0, 2\pi) \mid H_{t,A} \neq \emptyset \}$. For two open convex subsets $A, A' \subseteq \mathbb{R}^2$ with $T_A \subseteq T_{A'}$, we have

$$A \supseteq A' \Leftrightarrow H_{t,A} \subseteq H_{t,A'} \forall t \in T_A.$$  \hfill (A.1)

Let us assume that all regions of interest are open convex subsets of $\mathbb{R}^2$. For any such region $R$, we have

$$R = \bigcap_{t \in T_R} H_{t,R}$$

and consequently

$$\overline{R} = \bigcup_{t \in T_R} H_{t,R}.$$  

If we have a family $(\varphi_H)_{H \in H}$ of level-$\alpha$-tests for $H_0 : \theta \in H$ versus $H_1 : \theta \in \overline{H}$, then the intersection-union principle provides us with the level-$\alpha$-test

$$\varphi_R^{IU} = \min_{t \in T_R} \varphi_{H_{t,R}}$$

for the null hypothesis $H_0 : \theta \in \overline{R}$. It is now a straightforward idea to consider the minimal region of interest for which $\varphi_R^{IU}$ is still equal to 1 as a candidate for a comparison region, and to try to construct this as

$$C = \bigcap_{t \in T} H_{t,\varphi}$$

with $H_{t,\varphi} := \bigcup_{r \in \mathbb{R}, \varphi_{H_t,r} = 1} H_{t,r}$ and $T$ chosen such that $T_R \subseteq T$ for all regions of interest.

This idea does indeed hold, if $\varphi_H$ is monotone, that is,

$$H_1 \subseteq H_2 \Rightarrow \varphi_{H_1} \geq \varphi_{H_2} \tag{1}$$

and if it is right-continuous, that is, if for any sequence $(r_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}$ with $r_n \downarrow r_0$, we have

$$\varphi_{H_{t,r_n}} \rightarrow \varphi_{H_{t,r_0}} \forall t \in [0, 2\pi) \tag{2}$$

and if one additional regularity condition holds, which is provided by the following theorem.

**Theorem 3.1.** Let $(\varphi_H)_{H \in H}$ be a family of level-$\alpha$-tests for $H_0 : \theta \in H$, satisfying conditions (1) and (2). Then any open, convex set $C$ with $T_C \supseteq T$ and

$$H_{t,C} = H_{t,\varphi} \forall t \in T$$  \hfill (3)

defines a level-$\alpha$-comparison region for all $R$ with $T_R \subseteq T$.

Note that (3) is indeed a regularity condition, as not for any family $(\varphi_H)_{H \in H}$ the set $C = \bigcap_{t \in T} \overline{H_{t,\varphi}}$ satisfies (3). Examples can easily be constructed starting with a family and a set $C$ satisfying (3) and setting $\overline{H_{t,r}} := \varphi_{H_{t,r}}$ for $t \neq t_0$ and $\overline{H_{t,r_0}} = 0$ for all $r \in \mathbb{R}$ for one direction $t_0 \in T$. Then (1) and (2) hold for $\varphi$ and we have still $C = \bigcap_{t \in T} \overline{H_{t,\varphi}}$, but $H_{t_0,\varphi} = \emptyset$. Moreover, the theorem provides a simple tool to check the validity of construction methods for comparison regions without explicitly proving that they are given by $\bigcap_{t \in T} \overline{H_{t,\varphi}}$. We will make use of this in the following two sections.

**Proof.** We will show for any open, convex subset $R$ of $\mathbb{R}^2$ with $T_R \subseteq T$ that $C \subseteq R \Leftrightarrow \varphi_R^{IU} = 1$.

$\Leftarrow$ Let $t \in T_R$. $\varphi_R^{IU} = 1$ implies $\varphi_{H_{t,R}} = 1$. By definition of $H_{t,\varphi}$, we have $H_{t,\varphi} \supseteq H_{t,R}$.

(3) implies $H_{t,C} \supseteq H_{t,R}$. Consequently, $H_{t,C} \supseteq H_{t,R} \forall t \in T_R$ and $T_R \subseteq T \subseteq T_C$ implies $C \subseteq R$ by (A.1).

$\Rightarrow$ Let $t \in T_R$. The right-continuity of $\varphi_H$ (2) ensures that $\varphi_{H_{t,\varphi}} = 1$. $C \subseteq R$ and $T_R \subseteq T_C$ implies $H_{t,C} \supseteq H_{t,R}$ by (A.1), and therefore $H_{t,\varphi} \supseteq H_{t,R}$ by (3), and $\varphi_{H_{t,\varphi}} \leq \varphi_{H_{t,R}}$ by (1).

Consequently, $\varphi_{H_{t,R}} = 1 \forall t \in T_R$, and hence $\varphi_R^{IU} = 1$. \hfill $\square$
We would like to note that all considerations presented so far can also be applied if the parameter space is restricted to a closed, convex subset $\Theta$ of $\mathbb{R}^2$. This requires only to consider $H_0 := \{ H \cap \Theta | H \in \mathcal{H}, H \cap \Theta \neq \emptyset \}$ instead of $\mathcal{H}$. As $\Theta$ is closed and regions of interest are open, the concept of tangent sets is still applicable in $H_0$.

**APPENDIX B: LR TEST-BASED COMPARISON REGIONS**

Let $L(D;\theta)$ denote the likelihood for the observed data $D$ depending on a two-dimensional parameter vector $\theta$. Let $l(\theta) := \log L(D;\theta)$ denote the corresponding log-likelihood. Let $\hat{\theta}_{ML}$ denote the maximum likelihood estimator of $\theta$, $L^* = L(\hat{\theta}_{ML})$ the maximal likelihood, and $l^* = l(\hat{\theta}_{ML})$ the maximal log-likelihood. For testing $H_0: \theta \in H_{t,r}$, we can consider the LR test based on the test statistic $T_{t,r} = -2 \log(\max_{\theta \in H_{t,r}} l(\theta) / L^*) = -2(\max_{\theta \in H_{t,r}} l(\theta) - l^*)$. We assume in the following that the asymptotic distribution of $T_{t,r}$ on the boundary line of $H_{t,r}$ is $\frac{1}{2} \chi^2_0 + \frac{1}{2} \chi^2_1$ (Chernoff, 1954). Now for any $x \geq 0$, $P(\frac{1}{2} \chi^2_0 + \frac{1}{2} \chi^2_1 > x) = \frac{1}{2} + \frac{1}{2} P(\chi^2_1 > x)$, and $\chi^2_0$ is a distribution with mass 1 in the point 0. So if $P(\chi^2_1 > x) = 1 - 2\alpha$, we have $P(\frac{1}{2} \chi^2_0 + \frac{1}{2} \chi^2_1 > x) = \frac{1}{2} + \frac{1}{2}(1 - 2\alpha) = 1 - \alpha$. Consequently, $\phi_{t,r}^{LR} = 1(\Gamma_{t,r} \geq c_a)$ with $c_a = \chi^2_{1,1-2\alpha}$ defines a level-$\alpha$-test for $\alpha < 0.5$.

By definition, $\phi_{t,r}^{LR}$ is monotone and right-continuous if $l$ is continuous. The following lemma shows that the contour sets of $l(\theta)$ provide us with comparison regions under a mild regularity condition:

**Lemma 4.1.** Let $l(\theta)$ be a continuous and strictly concave log-likelihood function. Then the contour set

\[ C_\alpha = \left\{ \theta \mid l(\theta) - l^* > -\frac{1}{2} c_a = -\frac{1}{2} \chi^2_{1,1-2\alpha} \right\} \]

is a level-$\alpha$-comparison region, for any choice of $T$ and $\alpha < 0.5$.

**Proof.** By the strict concavity of $l$, $C_\alpha$ is an open and convex set. Let $t \in T$ and $H := H_{t,C_a}$. Because $C_a \subseteq \overline{H}$, we have $l(\theta) - l^* \leq -\frac{1}{2} c_a$ for all $\theta \in H$ by the definition of $C_a$. For $r$ with $H = H_{t,r}$, we hence have $T_{t,r} = -2(\max_{\theta \in H_{t,r}} l(\theta) - l^*) \geq c_a$ and hence $\phi_{t,r}^{LR} = 1$. As $l(\hat{\theta}_{ML}) - l^* = 0$, and $c_a > 0$, we know that $\hat{\theta}_{ML} \notin H$. Furthermore, the continuity of $l$ implies that there is a $\theta'$ on the boundary line of $H$ with $l(\theta') - l^* = -\frac{1}{2} c_a$. The strict concavity of $l$ implies that for any $\theta''$ on the connecting line between $\theta'$ and $\hat{\theta}_{ML}$, excluding $\theta'$, $l(\theta'') - l^* > -\frac{1}{2} c_a$ holds. For $r' < r$, the intersection between $H_{t,r'}$ and this line is not empty. Consequently, $\max_{\theta \in H_{t,r'}} l(\theta) - l^* > -\frac{1}{2} c_a$, hence $T_{t,r'} < c_a$, and hence $\phi_{t,r'}^{LR} = 0$. Consequently, $H = H_{t,\phi_{t,r}}$. The lemma follows then by Theorem 3.1.

Note that the shape of comparison regions is similar to the shape of LR test-based confidence regions. The $(1 - \alpha)$-confidence region is given by

\[ C_\alpha^* = \left\{ \theta \mid l(\theta) - l^* > -\frac{1}{2} c_a \right\} \]

with $c_a = \chi^2_{2,1-\alpha}$. As $\chi^2_{1,0.9} = 2.71 = \chi^2_{2,0.742}$, a 5% comparison region coincides with a 74.2% confidence region. To draw the comparison region, we can choose directions $t_j = \frac{j-1}{J-1} 2\pi$ for $j = 1, \ldots, J$, and solve for each $t_j$ the equation

\[ l(\hat{\theta}_{ML} + \gamma_j \xi_j) - l^* = -\frac{1}{2} c_a \]

with $\xi_j = (\cos t_j, \sin t_j)'$ for $\gamma_j > 0$. Because $\gamma_j \mapsto l(\hat{\theta}_{ML} + \gamma_j \xi_j)$ is a strictly monotone decreasing function with maximum 0, the solution $\gamma_j$ is unique and can easily be determined by numerical methods. We can then plot all points $(\hat{\theta}_{ML} + \gamma_j \xi_j)_{j=1,\ldots,J}$.

The same approach has been recently recommended by Jaeger (2016) for plotting LR test-based confidence regions. $J$ can simply be chosen large enough such that the boundary line of the comparison region appears as a smooth line in the graphical visualization.
APPENDIX C: WALD TEST-BASED COMPARISON REGIONS

If \( \hat{\theta} \) is asymptotically Gaussian distributed with mean \( \theta \) and variance \( \Sigma \), and a consistent estimate \( \hat{\Sigma} \) for \( \Sigma \) is available, the level-\( \alpha \)-Wald test for \( H_0 : \theta \in H_{t,r} \) is given by

\[
\phi_{t,r}^W = \begin{cases} 1 \left( \frac{\xi_t^\prime \hat{\theta} - r}{\sqrt{\xi_t^\prime \hat{\Sigma} \xi_t}} < u_{\alpha} \right) ，
\end{cases}
\]

where \( u_{\alpha} \) is the \( \alpha \)-percentile of the standard normal distribution.

**Lemma 4.2.** If \( \phi_{t,r}^W \) as defined above defines a level-\( \alpha \)-test, then

\[
C_W^{\alpha} = \left\{ \theta \mid (\theta - \hat{\theta})' \hat{\Sigma}^{-1} (\theta - \hat{\theta}) < \chi^2_{1,1-\alpha} \right\}
\]

defines a level-\( \alpha \)-comparison region.

**Proof.** Let \( t \in [0, 2\pi) \). \( H_{t,\phi}^W \) is given by \( H_{t,\phi}^r \) with \( r_t = \xi_t^\prime \hat{\theta} - u_{\alpha} \sqrt{\xi_t^\prime \hat{\Sigma} \xi_t} = \xi_t^\prime \hat{\theta} + u_{1-\alpha} \sqrt{\xi_t^\prime \hat{\Sigma} \xi_t} \). As shown in Appendix D, \( H_{t,\phi}^W \) is given by \( H_{t,\phi}^r \) with \( r_t' = \xi_t^\prime \hat{\theta} + u_{1-\alpha} \sqrt{\xi_t^\prime \hat{\Sigma} \xi_t} \). Hence, \( H_{t,\phi}^W = H_{t,\phi}^r \). The lemma follows then by Theorem 3.1. \( \square \)

\( C_\alpha \) can easily be drawn by considering the points \( \theta_t \) in which \( H_{t,\phi} \) touches \( C_W^{\alpha} \) on a grid \( t_j = \frac{j-1}{J-1} 2\pi \) for \( j = 1, \ldots, J \). As shown in Appendix D, these points are given by \( \theta_t = \hat{\theta} + \hat{\Sigma} \xi_t u_{1-\alpha} \sqrt{\xi_t^\prime \hat{\Sigma} \xi_t} \).

The Wald test can also be used to construct \((1-\alpha)\)-confidence regions for \( \theta \). They are given by

\[
C^*_\alpha = \left\{ \theta \mid (\theta - \hat{\theta})' \hat{\Sigma}^{-1} (\theta - \hat{\theta}) < \chi^2_{2,1-\alpha} \right\}
\]

Again, we can observe that comparison and confidence regions have similar shapes. As \( u_{1-\alpha}^2 = \chi^2_{1,1-2\alpha} \), we have again that a 5% comparison region corresponds to a 74.2% confidence region.

APPENDIX D

Let \( A \) be a positive definite symmetric matrix such that \( E = \{ \theta \mid \theta' A \theta \leq u^2 \} \) is a strictly convex ellipse for some \( u > 0 \). It is well known that for any \( \theta_B \) on the boundary of \( E \) the tangent through \( \theta_B \) is given by \( T = \{ \theta \mid \theta' B A \theta = u^2 \} \) (cf. Appendix A.2 in Friendly, Monette, & Fox, 2013). If this tangent is the boundary of a half-space \( H_{t,r} \), we also have the representation \( T = \{ \theta \mid \xi_t' \theta = r \} \) with \( r > 0 \). This implies \( \theta_B' A / u^2 = \xi_t' / r \), and hence \( \theta_B = A^{-1} \xi_t u^2 / r \). Now, \( u^2 = \theta_B' B A \theta_B = \xi_t' A^{-1} A A^{-1} \xi_t u^2 / r^2 \), and hence with \( r > 0 \) we have \( r = u \sqrt{\xi_t' A^{-1} \xi_t} \) and \( \theta_B = A^{-1} \xi_t u^2 / \sqrt{\xi_t' A^{-1} \xi_t} \). If \( E \) is centered around \( \theta_0 \), that is, \( E = \{ \theta \mid (\theta - \theta_0)' A (\theta - \theta_0) \} \), we have \( r = \xi_t' \theta_0 + u \sqrt{\xi_t' A^{-1} \xi_t} \) and \( \theta_B = \theta_0 + A^{-1} \xi_t u^2 / \sqrt{\xi_t' A^{-1} \xi_t} \).