

The IPCC's GHG Attribution Methodology is Fundamentally Flawed

Post by Ross McKittrick
August 17, 2021

Introduction

One day after the IPCC released the AR6 I published [a paper in *Climate Dynamics*](#) showing that their “Optimal Fingerprinting” methodology on which they have long relied for attributing climate change to greenhouse gases is seriously flawed and its results are unreliable and largely meaningless. Some of the errors would be obvious to anyone trained in regression analysis, and the fact that they went unnoticed for 20 years despite the method being so heavily used does not reflect well on climatology as an empirical discipline.

My paper is a critique of “Checking for model consistency in optimal fingerprinting” by Myles Allen and Simon Tett, which was [published in *Climate Dynamics*](#) in 1999 and to which I refer as AT99. Their attribution methodology was instantly embraced and promoted by the IPCC in the 2001 Third Assessment Report (coincident with their embrace and promotion of the Mann hockey stick). The IPCC promotion continues today: see AR6 [Section 3.2.1](#). It has been used in dozens and possibly hundreds of studies over the years. Wherever you begin in the Optimal Fingerprinting literature ([example](#)), all paths lead back to AT99, often via [Allen and Stott \(2003\)](#). So its errors and deficiencies matter acutely.

The abstract of my paper reads as follows:

Allen and Tett (1999, herein AT99) introduced a Generalized Least Squares (GLS) regression methodology for decomposing patterns of climate change for attribution purposes and proposed the “Residual Consistency Test” (RCT) to check the GLS specification. Their methodology has been widely used and highly influential ever since, in part because subsequent authors have relied upon their claim that their GLS model satisfies the conditions of the Gauss-Markov (GM) Theorem, thereby yielding unbiased and efficient estimators. But AT99 stated the GM Theorem incorrectly, omitting a critical condition altogether, their GLS method cannot satisfy the GM conditions, and their variance estimator is inconsistent by construction. Additionally, they did not formally state the null hypothesis of the RCT nor identify which of the GM conditions it tests, nor did they prove its distribution and critical values, rendering it uninformative as a specification test. The continuing influence of AT99 two decades later means these issues should be corrected. I identify 6 conditions needing to be shown for the AT99 method to be valid.

The Allen and Tett paper had merit as an attempt to make operational some ideas emerging from an engineering (signal processing) paradigm for the purpose of analyzing climate data. The errors they made come from being experts in one thing but not another, and the review process in both climate journals and IPCC reports is notorious for not involving people with relevant statistical expertise (despite the reliance on statistical methods). If someone trained in econometrics had refereed their paper 20 years ago the problems would have immediately been spotted, the methodology would have been heavily modified or abandoned and a lot of papers since then would probably never have been published (or would have, but with different conclusions—I suspect most would have failed to report “attribution”).

Optimal Fingerprinting

AT99 made a number of contributions. They took note of previous proposals for estimating the greenhouse “signal” in observed climate data and showed that they were equivalent to a statistical technique called Generalized Least Squares (GLS). They then argued that, by construction, their GLS model satisfies the Gauss-Markov (GM) conditions, which according to an important theorem in statistics means it yields unbiased and efficient parameter estimates. (“Unbiased” means the expected value of an estimator equals the true value. “Efficient” means all the available sample information is used, so the estimator has the minimum variance possible.) If an estimator satisfies the GM conditions, it is said to be “BLUE”—the Best (minimum variance) Linear Unbiased Estimator; or the best option out of the entire class of estimators that can be expressed as a linear function of the dependent variable. AT99 claimed that their estimator satisfies the GM conditions and therefore is BLUE, a claim repeated and relied upon subsequently by other authors in the field. They also introduced a “Residual Consistency” (RC) test which they said could be used to assess the validity of the fingerprinting regression model.

Unfortunately these claims are untrue. Their method is not a conventional GLS model. It does not, and cannot, satisfy the GM conditions and in particular it violates an important condition for unbiasedness. And rejection or non-rejection of the RC test tells us nothing about whether the results of an optimal fingerprinting regression are valid.

AT99 and the IPCC

AT99 was heavily promoted in the 2001 IPCC Third Assessment Report (TAR Chapter 12, Box 12.1, Section 12.4.3 and Appendix 12.1) and has been referenced in every IPCC Assessment Report since. TAR Appendix 12.1 was headlined “Optimal Detection is Regression” and began

The detection technique that has been used in most “optimal detection” studies performed to date has several equivalent representations (Hegerl and North, 1997; Zwiers, 1999). It has recently been recognised that it can be cast as a multiple regression problem with respect to generalised least squares (Allen and Tett, 1999; see also Hasselmann, 1993, 1997)

The growing level of confidence regarding attribution of climate change to GHG’s expressed by the IPCC and others over the past two decades rests principally on the many studies that employ the AT99 method, including the RC test. The methodology is still in wide use, albeit with a couple of minor changes that don’t address the flaws identified in my critique. (Total Least Squares or TLS, for instance, introduces new biases and problems which I analyze elsewhere; and regularization methods to obtain a matrix inverse do not fix the underlying theoretical flaws). There have been a small number of attribution papers using other methods, including ones which the TAR mentioned. “Temporal” or time series analyses have their own flaws which I will address separately (put briefly, regressing $I(0)$ temperatures on $I(1)$ forcings creates obvious problems of interpretation).

The Gauss-Markov (GM) Theorem

As with regression methods generally, everything in this discussion centres on the GM Theorem. There are two GM conditions that a regression model needs to satisfy to be BLUE. The first, called homoskedasticity, is that the error variances must be constant across the sample. The second, called conditional independence, is that the expected values of the error terms must be independent of the explanatory variables. If homoskedasticity fails, least squares coefficients will still be unbiased but their variance estimates will be biased. If conditional independence fails, least squares coefficients and their variances will be biased and inconsistent, and the regression model

output is unreliable. (“Inconsistent” means the coefficient distribution does not converge on the right answer even as the sample size goes to infinite.)

I teach the GM theorem every year in introductory econometrics. (As an aside, that means I am aware of the ways I have oversimplified the presentation, but you can refer to the paper and its sources for the formal version). It comes up near the beginning of an *introductory* course in regression analysis. It is not an obscure or advanced concept, it is the foundation of regression modeling techniques. Much of econometrics consists of testing for and remedying violations of the GM conditions.

The AT99 Method

(It is not essential to understand this paragraph, but it helps for what follows.) Optimal Fingerprinting works by regressing observed climate data onto simulated analogues from climate models which are constructed to include or omit specific forcings. The regression coefficients thus provide the basis for causal inference regarding the forcing, and estimation of the magnitude of each factor’s influence. Authors prior to AT99 argued that failure of the homoskedasticity condition might thwart signal detection, so they proposed transforming the observations by premultiplying them by a matrix \mathbf{P} which is constructed as the matrix root of the inverse of a “climate noise” matrix \mathbf{C} , itself computed using the covariances from preindustrial control runs of climate models. But because \mathbf{C} is not of full rank its inverse does not exist, so \mathbf{P} can instead be computed using a Moore-Penrose pseudo inverse, selecting a rank which in practice is far smaller than the number of observations in the regression model itself.

The Main Error in AT99

AT99 asserted that the signal detection regression model applying the \mathbf{P} matrix weights is homoscedastic by construction, therefore it satisfies the GM conditions, therefore its estimates are unbiased and efficient (BLUE). Even if their model yields homoscedastic errors (which is not guaranteed) their statement is obviously incorrect: they left out the conditional independence assumption. Neither AT99 nor—as far as I have seen—anyone in the climate detection field has ever mentioned the conditional independence assumption nor discussed how to test it nor the consequences should it fail.

And fail it does—routinely in regression modeling; and when it fails the results can be spectacularly wrong, including wrong signs and meaningless magnitudes. But you won’t know that unless you test for specific violations. In the first version of my paper (written in summer 2019) I criticized the AT99 derivation and then ran a suite of AT99-style optimal fingerprinting regressions using 9 different climate models and showed they routinely fail standard conditional independence tests. And when I implemented some standard remedies, the greenhouse gas signal was no longer detectable. I sent that draft to Allen and Tett in late summer 2019 and asked for their comments, which they undertook to provide. But hearing none after several months I submitted it to the *Journal of Climate*, requesting Allen and Tett be asked to review it. Tett provided a constructive (signed) review, as did two other anonymous reviewers, one of whom was clearly an econometrician (another might have been Allen but it was anonymous so I don’t know). After several rounds the paper was rejected. Although Tett and the econometrician supported publication the other reviewer and the editor did not like my proposed alternative methodology. But none of the reviewers disputed my critique of AT99’s handling of the GM theorem. So I carved that part out and sent it in winter 2021 to *Climate Dynamics*, which accepted it after 3 rounds of review.

Other Problems

In my paper I list five assumptions which are necessary for the AT99 model to yield BLUE coefficients, not all of which AT99 stated. All 5 fail by construction. I also list 6 conditions that need to be proven for the AT99 method to be valid. In the absence of such proofs there is no basis for claiming the results of the AT99 method are unbiased or consistent, and the results of the AT99 method (including use of the RC test) should not be considered reliable as regards the effect of GHG's on the climate.

One point I make is that the assumption that an estimator of \mathbf{C} provides a valid estimate of the error covariances means the AT99 method cannot be used to test a null hypothesis that greenhouse gases have no effect on the climate. Why not? Because an elementary principle of hypothesis testing is that the distribution of a test statistic under the assumption that the null hypothesis is true cannot be conditional on the null hypothesis being false. The use of a climate model to generate the homoscedasticity weights requires the researcher to assume the weights are a true representation of climate processes and dynamics. The climate model embeds the assumption that greenhouse gases have a significant climate impact. Or, equivalently, that natural processes alone cannot generate a large class of observed events in the climate, whereas greenhouse gases can. It is therefore not possible to use the climate model-generated weights to construct a test of the assumption that natural processes alone could generate the class of observed events in the climate.

Another less-obvious problem is the assumption that use of the Moore-Penrose pseudo inverse has no implications for claiming the result satisfies the GM conditions. But the reduction of rank of the resulting covariance matrix estimator means it is biased and inconsistent and the GM conditions automatically fail. As I explain in the paper, there is a simple and well-known alternative to using \mathbf{P} matrix weights—use of White's (1980) heteroskedasticity-consistent covariance matrix estimator, which has long been known to yield consistent variance estimates. It was already 20 years old and in use everywhere (other than climatology apparently) by the time of AT99, yet they opted instead for a method that is much harder to use and yields biased and inconsistent results.

The RC Test

AT99 claimed that a test statistic formed using the signal detection regression residuals and the \mathbf{C} matrix from an independent climate model follows a centered chi-squared distribution, and if such a test score is small relative to the 95% chi-squared critical value, the model is validated. More specifically, the null hypothesis is not rejected.

But what is the null hypothesis? Astonishingly it was never written out mathematically in the paper. All AT99 provided was a vague group of statements about noise patterns, ending with a far-reaching claim that if the test doesn't reject, "then we have no explicit reason to distrust uncertainty estimates based on our analysis." As a result, researchers have treated the RC test as encompassing every possible specification error, including ones that have no rational connection to it, erroneously treating non-rejection as comprehensive validation of the signal detection regression model specification.

This is incomprehensible to me. If in 1999 someone had submitted a paper to even a low-rank economics journal proposing a specification test in the way that AT99 did, it would have been annihilated at review. They didn't state the null hypothesis mathematically or list the assumptions necessary to prove its distribution (even asymptotically, let alone exactly), they provided no analysis of its power against alternatives nor did they state any alternative hypotheses in any form so readers have no idea what rejection or non-rejection implies. Specifically, they established no link between the RC test and the GM conditions. I provide in the paper a simple description of a case

in which the AT99 model might be biased and inconsistent by construction, yet the RC test would never reject. And supposing that the RC test does reject, which GM condition therefore fails? Nothing in their paper explains that. It's the only specification test used in the fingerprinting literature and it is utterly meaningless.

The Review Process

When I submitted my paper to CD I asked that Allen and Tett be given a chance to provide a reply which would be reviewed along with it. As far as I know this did not happen, instead my paper was reviewed in isolation. When I was notified of its acceptance in late July I sent them a copy with an offer to delay publication until they had a chance to prepare a response, if they wished to do so. I did not hear back from either of them so I proceeded to edit and approve the proofs. I then wrote them again, offering to delay further if they wanted to produce a reply. This time Tett wrote back with some supportive comments about my earlier paper and he encouraged me just to go ahead and publish my comment. I hope they will provide a response at some point, but in the meantime my critique has passed peer review and is unchallenged.

Guessing at Potential Objections

1. *Yes but* look at all the papers over the years that have successfully applied the AT99 method and detected a role for GHGs. Answer: the fact that a flawed methodology is used hundreds of times does not make the methodology reliable, it just means a lot of flawed results have been published. And the failure to spot the problems means that the people working in the signal detection/Optimal Fingerprinting literature aren't well-trained in GLS methods. People have assumed, falsely, that the AT99 method yields "BLUE" – i.e. unbiased and efficient – estimates. Maybe some of the past results were correct. The problem is that the basis on which people said so is invalid, so no one knows.

2. *Yes but* people have used other methods that also detect a causal role for greenhouse gases. Answer: I know. But in past IPCC reports they have acknowledged those methods are weaker as regards proving causality, and they rely even more explicitly on the assumption that climate models are perfect. And the methods based on time series analysis have not adequately grappled with the problem of mismatched integration orders between forcings and observed temperatures. I have some new coauthored work on this in process.

3. *Yes but* this is just theoretical nitpicking, and I haven't proven the previously-published results are false. Answer: What I have proven is that the basis for confidence in them is non-existent. AT99 correctly highlighted the importance of the GM theorem but messed up its application. In other work (which will appear in due course) I have found that common signal detection results, even in recent data sets, don't survive remedying the failures of the GM conditions. If anyone thinks my arguments are mere nitpicking and believes the AT99 method is fundamentally sound, I have listed the six conditions needing to be proven to support such a claim. Good luck.

I am aware that AT99 was followed by [Allen and Stott \(2003\)](#) which proposed TLS for handling errors-in-variables. This doesn't alleviate any of the problems I have raised herein. And in a separate paper I argue that TLS over-corrects, imparting an upward bias as well as causing severe inefficiency. I am presenting a paper at this year's climate econometrics conference discussing these results.

Implications

The AR6 [Summary](#) paragraph A.1 upgrades IPCC confidence in attribution to "Unequivocal" and the [press release](#) boasts of "major advances in the science of attribution." In reality, for the past 20 years, the climatology profession has been oblivious to the errors in AT99, and untroubled by the

complete absence of specification testing in the subsequent fingerprinting literature. These problems mean there is no basis for treating past attribution results based on the AT99 method as robust or valid. The conclusions might by chance have been correct, or totally inaccurate; but without correcting the methodology and applying standard tests for failures of the GM conditions it is mere conjecture to say more than that.