RESEARCH STATEMENT

ANDREAS HAGEMANN

The primary objective of my research is to develop new **econometric methodology** that addresses the complexities of **clustered data** to enhance the accuracy and reliability of empirical work in economics and related fields. Typical examples of clusters are firms, cities, or states. The central challenge is that units within clusters may influence one another or may be influenced by similar environmental factors in ways that cannot be observed. Empirical researchers know that neglecting to account for clusters can yield results where non-existent effects erroneously appear as highly significant. My research agenda develops new tools to address this issue in challenging and empirically relevant scenarios where (i) only few clusters are available (e.g., an intervention in an economic development context targeted only a small number of villages), (ii) only a single cluster received treatment (e.g., a state passed a new law), or (iii) the target parameter is an entire function (e.g., the response of an entire distribution to treatment is of interest instead of just the average response). A secondary aspect of my research agenda is **interdisciplinary collaboration**. Econometric theory is a small and technical field where contributions can be subtle and narrow. Collaborations with colleagues in **health and labor economics, marketing, strategy, and management** have helped me reach audiences outside of econometrics.

My work has had substantial impact in my field. I have seven peer-reviewed publications in highly regarded journals, two papers revised and resubmitted, and one paper under revision for resubmission. Seven of these papers are solo-authored. I am frequently invited to present my work at high-profile conferences, workshops, university seminars, and in the industry. Methodology I developed is now the standard option for clustering in the canonical implementation of quantile regression in the statistical programming language R. My work was prominently featured in two of the (so far) three chapters of the new *Journal of Econometrics* "How-To" papers, a widely read paper series on connecting econometric theory and empirical practice. I am frequently called upon to be a peer-reviewer in my field. For instance, I have reviewed 23 manuscripts for the *Journal of Econometrics* since 2014.

In the following, I give an overview of my work in econometrics, my interdisciplinary collaborations, and my currently ongoing research.

Research in econometrics

Clustered data can be roughly divided into two groups. Typically, there are either (i) a large number of small clusters (e.g., many classrooms) or (ii) a small number of large clusters (e.g., entire states). For inference about a finite-dimensional parameter such as a coefficient in a linear regression model, case (i) typically involves a simple adjustment to a standard inference procedure, which has led the literature to focus on this special case. In all other situations, inference becomes surprisingly difficult. These situations are the focus of my current research. My starting point was inference in case (i) with an infinite-dimensional parameter but my more recent work has shifted towards case (ii), which is very common in empirical practice. Small numbers of large clusters occur, for example, in the analysis of policy reforms, where entire states are treated with the passage of a new law, or in a development context, where the introduction of a new technology affects entire villages. Below is a summary of the five papers I have written on the broader topic. I then summarize my other papers.

Inference about an infinite-dimensional object such as the quantile regression function in the cluster context requires analysis of a nonstandard Gaussian process whose distributional properties cannot be tabulated. I am the first to point this out in **Cluster-Robust Bootstrap Inference in Quantile Regression Models** (Hagemann, 2017, *Journal of the American Statistical Association*), where I provide an exhaustive solution to this problem for quantile regression and develop the *wild gradient bootstrap* for cluster-robust inference in linear quantile regression (QR) models. Despite the fact that it involves resampling the QR gradient process, the wild gradient bootstrap is fast and easy to implement because it does not involve finding zeros of the gradient during each bootstrap iteration. I show that the wild gradient bootstrap allows for the construction of asymptotically valid bootstrap standard errors, hypothesis tests over ranges of quantiles, and confidence bands for the QR coefficient function.

A disadvantage of QR in comparison to least squares methods is that the asymptotic variance of the QR coefficient function is notoriously difficult to estimate due to its dependence on the unknown conditional density of the response variable. An analytical estimate of the asymptotic variance therefore requires a user-chosen kernel and bandwidth. Furthermore, a common concern in applied work is that analytical estimates of asymptotic variances perform poorly in the cluster context when the number of clusters is small or the within-cluster correlation is high. As a consequence, true null hypotheses are rejected far too often. Similar problems also occur when the cluster sizes differ substantially.

In Hagemann (2017), I show that the wild gradient bootstrap performs well with as few as 20 clusters even when the within-cluster dependence is high and the cluster sizes are heterogenous. The wild gradient bootstrap consistently estimates the asymptotic distribution and covariance functions of the QR coefficients without relying on kernels, bandwidths, or other user-chosen parameters. Methods available prior to my paper did not allow for uniform inference across quantiles because the limiting QR process generally has an analytically intractable distribution. In contrast, the bootstrap approximations of the distribution and covariance functions developed in Hagemann (2017) can be combined to perform uniform Wald-type inference about the QR coefficient function. The wild gradient bootstrap is implemented as the 'cluster' option in the 'quantreg' package in R. Hagemann (2017) has been used in a wide variety of applications in fields such as economics, finance, criminology, biology, and agricultural science. Examples range from assessing the impact of the COVID-19 pandemic on labor market outcomes across clusters of industries to investigating the effect of microloans on clusters of villages to assessing the impact of urbanization on CO_2 emissions in China across clusters of provinces.

The results in Hagemann (2017) are designed for situations with 20 or more clusters. I improve on this limitation in **Inference on Quantile Processes with a Finite Number of Clusters** (Hagemann, 2023a, revised and resubmitted to the *Journal of Econometrics*). There, I develop a generic method for inference on the entire quantile or regression quantile process in the presence of five or more large and arbitrarily heterogeneous clusters. The method, which I refer to as *cluster-randomized Kolmogorov-Smirnov* (*CRK*) test, asymptotically controls size by generating Kolmogorov-Smirnov statistics that exhibit enough distributional symmetry at the cluster level such that randomization tests based on sign changes can be applied. The CRK test is not limited to the pure quantile regression setting and can be used in distributional difference-in-differences estimation and related situations where quantile treatment effects are identified by between-cluster comparisons.

The CRK test performs well even when the dependence varies from cluster to cluster and the cluster sizes are heterogenous. The reason for this robustness is that the CRK test does not rely on clustered covariance matrices to rescale the estimates. I instead use randomization inference to generate random critical values that automatically scale to the data. There are no kernels, bandwidths, or spatio-temporal orderings of the data to choose. The test achieves consistency with a finite number of large but heterogeneous clusters under interpretable high-level conditions. Randomization is performed with a fixed set of estimates and does not require repeated estimation to obtain its critical values. A practical issue with some alternative methods is that they require treated clusters to be matched ex-ante with an equal number of control clusters. Each match corresponds to a separate test and two researchers working with the same data could reach different conclusions based on which matches they choose. If there is not an equal number of treated and control clusters, then alternative tests require that some clusters have to be combined or dropped in an ad-hoc manner. The CRK test sidesteps these issues completely and explicitly merges all potential tests into a single, uniquely determined test decision. The CRK test is the only currently available test in the literature that can perform uniform inference on quantile processes with a finite number of heterogeneous clusters.

Moving away from the quantile setting, the main motivation for **Placebo Inference on Treatment Effects When the Number of Clusters Is Small** (Hagemann, 2019, *Journal of Econometrics*) is that the majority of analytical and bootstrap procedures perform poorly with a small to moderate number of clusters. In this paper, I introduce a testing framework based on permutation inference that allows for nearly exact inference about the lack of effect of a treatment when the size of the clusters is large relative to the number of clusters. The framework applies to situations where a binary treatment occurs in some but not all clusters and the treatment effect of interest is identified by between-cluster comparisons.

In a randomized trial, the average effect of a treatment is estimated by comparing the means of treatment and control groups. Computing this comparison of means for all possible ways in which individuals could have been assigned to the two groups generates "placebo" estimates. If treatment has no effect, the placebo estimates have the same distribution as the estimated treatment effect. A permutation or placebo test takes these observations as the null distribution to test the "sharp" hypothesis that there is no effect because the difference of treatment and control potential outcomes is zero for each individual. Such tests can be made exact under conventional assumptions. They are particularly attractive when only a small number of observations are available because the set of placebo estimates that have to be computed grows quickly with the sample size. More recently, placebo-type Monte Carlo experiments have been used in empirical economics as informal robustness exercises. I formalize and extend the notion of a placebo test to the cluster case by developing statistics that measure the size of a treatment effect of interest but are amenable to a placebo-like reassignment mechanism. Under simple and easily verifiable conditions, this placebo test leads to asymptotically valid, cluster-robust inference about conventional (non-sharp) null hypotheses in a very large class of empirically relevant models. The theoretical justification for the test comes from the fact that consistent permutation tests of certain hypotheses are possible even in situations where the joint distribution of the data is not invariant to permutations under the null hypothesis. Hagemann (2019) has been used in the literature to assess, for example, the effect of welfare reforms on intergenerational welfare participation, the behavior of local governments' reporting of air pollution, and the effects of illegal moneylending in Singapore.

I push the results of Hagemann (2019) further in **Permutation Inference With a Finite Number of Heterogeneous Clusters** (Hagemann, 2022, *Review of Economics and Statistics*). There, I introduce an adjusted permutation procedure that is able to asymptotically control the size of tests about the effect of a binary treatment in the presence of finitely many large and heterogeneous clusters. Hagemann (2019) still required the number of clusters to grow, albeit very slowly. The procedure applies to difference-in-differences estimation and other situations where treatment occurs in some but not all clusters and the treatment effect of interest is identified by between-cluster comparisons.

The main theoretical insight of this paper is that classical permutation inference can be adjusted to test the null hypothesis of equality of means of two finite samples of mutually independent but arbitrarily heterogeneous normal variables. This runs counter to classical permutation testing, where the data under the null are presumed to be exchangeable. The adjustment corrects the significance level of the test downwards to account for heterogeneity. I prove that this is possible for empirically relevant levels of significance if both samples consist of more than three observations. I also show that if a random vector of interest converges weakly to multivariate normal with diagonal covariance matrix, then permutation inference remains approximately valid for that vector. To exploit this result in a cluster context, I construct asymptotically normal statistics from each cluster and then apply adjusted permutation inference to the collection of these statistics.

I tackle the important limiting case where only one cluster received treatment in **Inference With a Single Treated Cluster** (Hagemann, 2023b, revised and resubmitted to *Review of Economic Studies*). This paper is motivated by the fact that prominent journals routinely publish studies where a single treated group is compared to multiple control groups. Statistical inference in this context is challenging and the results of some studies have been questioned specifically because they only have a single treated group or cluster. With one treated cluster, currently available inferential procedures assume identically distributed clusters or other undesirable homogeneity conditions that are unlikely to hold in empirical practice. In an attempt to avoid statistical issues stemming from having a single treated cluster, researchers therefore routinely resort to splitting large groups into smaller clusters that are presumed to be independent.

In Hagemann (2023b), I introduce an asymptotically valid method for inference with a single treated cluster that allows for heterogeneity of unknown form. The number of observations within each cluster is presumed to be large but the total number of clusters is fixed. The method, which I refer to as a *rearrangement test*, applies to standard difference-in-differences estimation and other settings where treatment occurs in a single cluster and the treatment effect is identified by between-cluster comparisons. The key theoretical insight for the rearrangement test is that a mild restriction on some but not all of the heterogeneity in two samples of independent normal variables allows testing the equality of their means even if one sample consists of only a single observation. I prove that this is possible for empirically relevant levels of significance if the other sample consists of at least twenty observations. The test is feasible with even fewer observations if other restrictions are strengthened. The rearrangement test compares the data to a reordered version of itself after attaching a special weight to the sample with a single observation. The weights can be tabulated and are easy to compute. I also show that the weights remain approximately valid if the two samples of independent heterogeneous normal variables arise as a distributional limit. This test is implemented in R and Stata. Hagemann (2023b) has been used in the literature to assess, for example, the effects of disclosure and enforcement on payday lending in Texas, the impact of medical malpractice reforms in North Carolina, and the effect of cash transfers on voter turnout in Alaska.

I solve the central theoretical challenges in Hagemann (2017) with the help of empirical process and stochastic equicontinuity methods, which are the focus of my paper **Stochastic Equicontinuity in Nonlinear Time Series Models** (Hagemann, 2014, *Econometrics Journal*). Stochastic equicontinuity typically captures the key difficulty in weak convergence proofs of estimators with non-differentiable objective functions. Precise and elegant methods have been found to deal with cases where the data dependence structure can be described by mixing conditions. Mixing assumptions are convenient in this context because they measure how events generated by time series observations—rather than the observations themselves—relate to one another and therefore also measure dependence of functions of such time series. The downside to these assumptions is that they can be hard to verify for a given application.

In Hagemann (2014) I give simple and easily verifiable conditions under which objective functions of econometric estimators are stochastically equicontinuous when the underlying process is a stationary time series that can be described by a nonlinear system. The stochastic equicontinuity problem does not have to be parametric and no continuity conditions are needed. The nonlinear system theory developed in my paper allows for the construction of dependence measures that are directly related to the stochastic process and includes a large number of commonly-used stationary time series models.

Many of my papers revolve around resampling methods such as the bootstrap, randomization, and permutation inference. I also use the bootstrap to solve a central problem of my paper **A Simple Test for Regression Specification with Non-Nested Alternatives** (Hagemann, 2012, *Journal of Econometrics*). Models are non-nested if they are not special cases of one another. Hence, non-nested testing problems typically do not have a natural null hypothesis. For example, it is a priori not clear what should be the null hypothesis when testing whether a specific covariate enters the regression equation in level or in log form. The literature therefore usually suggests a sequence of tests where each possible null hypothesis is considered. In Hagemann (2012), I introduce a simple test for the presence of the correct model among several non-nested specifications that avoids sequential testing. The test, which I refer to as the MJ (minimum joint) test, is an extension of the classical J test and bases its decision on the model with the least significant J statistic.

Standard non-nested hypothesis tests rely heavily on the assumption that one of the models under consideration is correct, and therefore all other non-nested specifications must be wrong. However, it may well happen that a non-nested hypothesis test does not reject a model in the presence of an alternative model, but also does not reject the alternative in the presence of the original model when the hypotheses are reversed. This leaves the researcher in the unfortunate situation of having to conclude that both specifications "explain the data equally well" even though at most one of them can be correct. Similar problems arise when all models are rejected. A further issue is that the sequential testing is typically conducted without regard to overall size, and thus two researchers working with the same data can arrive at different specifications simply because they used different levels of significance. Non-nested testing procedures have been subject to substantial criticism because of these features. I show that the *MJ* test does not require the correct model to be among the considered specifications and avoids ambiguous test outcomes. Hagemann (2012) has been used widely in regional science and urban economics, where choosing an appropriate model for spacial dependence is of central concern.

Interdisciplinary collaboration

I have three completed manuscripts from interdisciplinary collaborations with colleagues in health economics, marketing, strategy, and management. My coauthors on these projects are Kasey Buckles (Notre Dame), Justin Frake (Ross), Tong Guo (Duke Fuqua), Ofer Malamud (Northwestern), Melinda Morill (NC State), Yeşim Orhun (Ross), Jose Uribe (Indiana Kelley), and Abigail Wozniak (Minneapolis Fed).

In **The Effect of College Education on Mortality** (Buckles, Hagemann, Malamud, Morrill, and Wozniak, 2016, *Journal of Health Economics*), we exploit exogenous variation in college completion induced by draft-avoidance behavior during the Vietnam War to examine the impact of college completion on adult mortality. The existence of state level variation allows us to decompose national induction risk into two constituent parts: induction risk faced by a young man's own state cohort and induction risk faced by young men of that cohort in the rest of the country. Our decomposition yields two instruments, which we use to identify the impact of the two endogenous variables—education and veteran status—in our empirical framework.

My contribution to this paper was twofold: First, I constructed a structural model that allowed us to formally state under which conditions our main parameters are identified. Second, I developed a cluster-robust test for instrument relevance that can reject the hypothesis that at most one instrument is relevant against the alternative that both instruments are relevant. Such a test was not available in the literature at this point.

In **Reaching for Gold: Frequent-Flyer Status Incentives and Moral Hazard** (Orhun, Guo, and Hagemann, 2022, *Marketing Science*), we document systematic changes in the behavior of frequent flyers as they progress towards elite status. Using data from a leading U.S. airline, we show evidence for increased switching costs as the consumer approaches the target pace of point accumulation required to attain status. These switching costs reflect changes in booking behavior with the airline: Travelers become more likely to choose the airline even when it is less appealing than its competitors and pay higher prices than they otherwise would. These responses are reduced when travelers accumulate points at a rate substantially ahead of the target pace. The increase in switching costs is more pronounced for consumers at a hub of the airline and for business travelers. Moreover, we document a stronger willingness-to-pay response when consumers are less likely to shoulder the ticket costs themselves because they are traveling for business. This response suggests that asymmetric incentives induced by business travel explains much of the heterogeneity between business and leisure travelers, and moral hazard may be responsible for a large part of the profitability of frequent-flyer status incentives.

For this paper, I formalized a novel identification strategy and provided proofs that support the identification of the parameters of the model. In addition, I provided econometric guidance in how to interpret the empirical results in this non-standard setting.

In **Colliders in the Boardroom?** (Frake, Hagemann, and Uribe, 2023, revise and resubmit at the *Strategic Management Journal*), we show that estimates of effects of women in leadership positions on other women at the same organization can have incorrect signs and sizes because of endogenous selection bias. Similar effects can be produced for any underrepresented group. In line with published research, we estimate models suggesting that women and minority CEOs can reduce the compensation and representation of other women and minorities, respectively, on a company's top management team. However, we argue that these correlations are likely due to collider bias, an endogenous selection bias that has not received systematic attention in strategy and management. We use Monte Carlo simulations to illustrate conditions that reduce or amplify the problem and provide generalizable approaches to mitigate the risk of collider bias in applied research. In doing so, we find no evidence that women and minority CEOs damage the career outcomes of other women and minorities in their organizations and highlight the practical threat that collider bias can pose to empirical research.

My main contribution to this paper was a rigorous analysis of the endogenous selection bias in an econometric model of promotion decisions, which allowed us to identify the main drivers of the bias.

Ongoing Projects and Future Research

My immediate goals are to publish my resubmitted papers Hagemann (2023a,b) and to resubmit Frake et al. (2023). I plan to continue my research agenda on clustering with several projects that connect the recently emerging literature on design-based inference with the literature on inference with few clusters. I am also excited to add to my research agenda with projects on identification in event study designs. This is a widely used research design where standard least squares estimation can make unexpected and undesired assumptions about identification.

More broadly, I am excited about several interdisciplinary projects with colleagues in Ross and elsewhere. For instance, in joint work with Tom Buchmueller (Ross), Leontine Goldzahl (IESEG), and Sarah Miller (Ross), we are in the process of analyzing the effects of breast cancer screenings on later life outcomes with a view towards recent results on two-way fixed effects models. Other conversations with Ross colleagues in Strategy and Marketing are currently ongoing, and I am looking forward to contributing to the academic environment in Ross with my expertise in econometrics.

References

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2016). The effect of college education on mortality. *Journal of Health Economics 50*, 99–114.

Frake, J., A. Hagemann, and J. Uribe (2023). Colliders in the boardroom? Revise and resubmit at the *Strategic Management Journal.*

Hagemann, A. (2012). A simple test for regression specification with non-nested alternatives. *Journal of Econometrics 166*, 247–254.

Hagemann, A. (2014). Stochastic equicontinuity in nonlinear time series models. *Econometrics Journal 17*, 188–196.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association 112*, 446–456.

Hagemann, A. (2019). Placebo inference on treatment effects when the number of clusters is small. *Journal of Econometrics 213*, 190–209.

Hagemann, A. (2022). Permutation inference with a finite number of heterogeneous clusters. *Review of Economics and Statistics*, forthcoming.

Hagemann, A. (2023a). Inference on quantile processes with a finite number of clusters. Revise and resubmit at the *Journal of Econometrics*.

Hagemann, A. (2023b). Inference with a single treated cluster. Revise and resubmit at the *Review of Economic Studies*.

Orhun, Y., T. Guo, and A. Hagemann (2022). Reaching for gold: Frequent-flyer status incentives and moral hazard. *Marketing Science 43*, 548–574.