

# Scaling Model Size and Robustness to Covariate Shift in Tabular Data Models

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the impact of scaling model size on the robustness of tabular models to covariate shift, as evaluated by accuracy degradation on out-of-distribution datasets from the TableShift benchmark. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Benchmarking Distribution Shift in Tabular Data with TableShift. Research question: What is the impact of scaling model size on the robustness of tabular models to covariate shift, as evaluated by accuracy degradation on out-of-distribution datasets from the TableShift benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

## 3 Results

14 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 3.5/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

Claim	Verified	Confidence
Food insecurity affects more than 10% of households (13.5 million) across the United States in 2021.	×	0.00
The American Community Survey (ACS) is used as the data source for the Food Stamps task.	×	0.03
The Food Stamps task filters data for low-income adults aged 18-62 in households with at least one child.	×	0.02
An income threshold of \$30,000 is used based on the U.S. poverty threshold for a family with one child.	×	0.03
The ACS includes 10 regions: Puerto Rico; New England (Northeast region); Middle Atlantic (Northeast region); East North	×	0.01
East South Central (South region) is used as the holdout domain for the Food Stamps task.	×	0.02
The relationships between other forms of shift showed much weaker correlation, roughly $\rho \approx -0.2$ .	×	0.03
TableShift does not host or distribute the data; each data source is publicly available.	×	0.04
The datasets in TableShift do not contain personally identifiable information, offensive content, or proprietary informa	×	0.02
For data collected from human subjects, the conditions of collection and the ethics approval under which the data were c	×	0.03

## References

- <http://arxiv.org/abs/2306.11113v2>

- <http://arxiv.org/abs/2206.02435v2>
- <http://arxiv.org/abs/2312.07577v3>