

Does the Choice of Model or Benchmark Affect Inference in Measuring Mutual Fund Performance?*

Jeffrey L. Coles
Department of Finance
W. P. Carey School of
Business
Arizona State University
Tel: (480) 965-4475
Email:
jeffrey.coles@asu.edu
(corresponding author)

Naveen D. Daniel
Department of Finance
Krannert School of
Management
Purdue University
Tel: (765) 494-4461
Email: nav@purdue.edu

Federico Nardari
Department of Finance
W. P. Carey School of
Business
Arizona State University
Tel: (480) 965-7961
Email: fnardari@asu.edu

This Draft: January 31, 2006

JEL Classification: G12, G20, G23

Key words: Mutual Funds Performance, Market Timing, Model Misspecification, Bootstrap Analysis

Abstract: We address the practical question of whether investors and researchers are likely to make invalid inferences about fund manager performance when using the wrong model and/or benchmark. We consider three well-known models, those of Jensen (1968), Treynor and Mazuy (1966), and Henriksson and Merton (1981), and two commonly used timing benchmarks, the S&P 500 index and CRSP value-weighted index. Although prior studies recognize the possibility of model and benchmark misspecification, the existing literature does not explore empirically the existence, magnitude, and significance, if any, of potential inferential errors. Based on Monte Carlo simulations calibrated to real mutual fund data, we find that: (1) model misspecification results in severely biased measures of both selectivity and timing ability, especially for extreme (good and bad) performers; (2) *but* biases in measures of overall performance are economically insignificant; (3) benchmark misspecification results in qualitatively similar difficulties, with the addition that overall performance as well can be biased; and (4) model and benchmark misspecification do not appreciably alter the power to detect ability and distinguish a good fund from a bad fund. These results are robust to alternative asset pricing specifications, alternative simulation schemes, varying length of the return series, and periodicity of the simulated series. The use of daily fund returns amplifies our conclusions about the biases induced by model misspecifications. Moreover, the biases we identify appear to be difficult to correct by using standard model selection criteria and misspecification tests. If the benchmark is known but the timing model is not, investors should use measures of overall performance to evaluate funds and managers.

* We thank George Aragon, Hank Bessembinder, Rick Burdick, Chris Gadarowski, John Griffin, Dan Kapanak, Stuart Low, Cyndi McDonald, Erik Sirri, Kumar Venkatraman, Doug Witte, and especially Seung (Min) Ahn, Jason Greene, Mike Lemmon, Lalitha Naveen, and Russ Wermers for their helpful comments and discussion. We also thank seminar participants at Arizona State University, Babson College, Financial Management Association, and University of Missouri-Columbia for suggestions. We thank Jeff Busse for daily data on mutual fund returns and momentum returns.