This file explains the data and code that can be used for the replication of the paper: "Returns to Scale from Labor Specialization: Evidence from Asset Management Mergers," forthcoming in the *Review of Corporate Finance Studies*.

Authors: Mancy Luo, Alberto Manconi and David Schumacher
Date: August 2023

Introduction

Before going through the replication package, please read the manuscript and the Internet Appendix. They contain many details on the sample construction and variable definitions.

This replication package is made of three folders: data, code, and output. The "code" folder contains the Stata do-files necessary to generate all figures and tables in the main paper. The do-files will load the final datasets contained in the "data" folder and save all outputs in the "output" folder. Below we describe in detail how we construct the final datasets and how to run the code.

Data Availability Statement

The paper uses licensed data from the SDC Platinum, Zephyr Mergers and Acquisitions databases, FactSet Ownership database, the Morningstar Direct database, and Thomson Datastream / WorldScope. In line with *RCFS* guidelines, we share pseudo datasets that illustrate the structure of our final datasets but set all variables constructed based on licensed data to missing.

Sample Construction

Below we describe in detail the construction of all main datasets contained in the "data" folders. A variable dictionary can be obtained by loading the datasets and using Stata's "describe" command.

Matching Funds:

1. *matchingsample_fundlevel.dta* (in the "data" folder) contains treated funds and matching funds. The dataset is constructed as follows. First, we obtain information on mergers between asset managers that are completed between 2001 and 2013 from SDC Platinum and Zephyr Mergers and Acquisitions databases. We require that the buyer controls less than 50% of the target's shares before and more than 50% after the deal. Then, we merge the M&A deals with the FactSet ownership database by manually screening buyer and target names and keep all open-ended mutual funds from the buyer and the target.

   Next, we match funds in FactSet to the Morningstar Direct mutual fund database to obtain fund characteristics (e.g., investment objectives, returns, fees) and individual fund manager information. We obtain the names of the individual managers managing a given fund, together with the start and end date for each manager. In addition, Morningstar provides a

separate file that contains unique manager identifiers linked to the manager names which ensures the accuracy of our manager-fund mappings. We add the list of managers for a given fund at a given point of time in the variables *managerid**.

Third, we restrict to actively managed equity mutual funds (Morningstar variables *Broad Category Group* = "Equity" and *Index Fund* = "No", or the average percentage of fund TNA invested in equity is above 80% over the sample if *Broad Category Group* is missing). In addition, we require deals to have portfolio holdings information available for both the buyer and the target at least one year prior to the merger completion date.

Finally, we select a matching fund for each treated fund out of all active equity funds that are neither involved in a merger nor managed by an affiliated firm of our sample firms throughout the sample period. Out of those, we select the one that shares the same investment objective (Morningstar variable *Global Category*), is managed in the same country, and closest in terms of portfolio holdings, firm size, and fund size one year prior to the merger. We add the sample of matching funds together with the sample of treated funds and use the variable *treat* to separate them.

2. *matchingsample_managerlevel.dta* (in the "data" folder) changes the unit from fund level to individual manager level based on *matchingsample_fundlevel.dta* data.

Withdrawn Deals:

3. *withdrawndeals.dta* (in the "data" folder) contains our sample mergers and withdrawn deals. The dataset is constructed as follows. First, we obtain information on mergers between asset managers that are ultimately withdrawn between 2001 and 2013 from SDC Platinum and Zephyr Mergers and Acquisitions databases. Then we match them to FactSet to obtain all relevant open-ended funds and apply the same filters as selecting our sample.

   Second, for each withdrawn deal, we select a completed deal to serve as its treatment counterpart. From the completed deals, we select those that (1) are completed within the one-year window centered around the withdrawal date of the withdrawn deals; and (2) are closest in terms of the relative size ratio (i.e., target firm's total assets divided by buyer's total assets). Finally, we combine treated and withdrawn deals together and use the variable *treat* to separate them.

Firm-Level Data:

4. *statistics_firmlevel.dta* (in the "data" folder) contains firm-level statistics for mutual fund firms in our sample and in FactSet database. The dataset is constructed as follows. First, we construct relevant characteristics using treated funds (i.e., variable *treat* = 1) for each buyer- or target- firm based on *matchingsample_fundlevel.dta* data. The unit of observations is at deal × firm (buyer or target) × time level.

   Second, we construct similar firm-level characteristics for all asset management firms from the entire FactSet database. Specifically, we use all active equity open-ended mutual funds with available individual manager information for such construction. The unit of

observations is at firm × time level. Finally, we merge the two samples together and use the variable *factsetsample* to separate them.

5. *acquiror_target_matching.dta* (in the "data" folder) contains deal-level sample with matching targets. The dataset is constructed as follows. First, for every actual target in a given deal, we construct a pool of pseudo targets by selecting all asset management firms in FactSet that 1) are neither involved in our sample mergers nor affiliated with our real targets under the same parent firms and 2) share the same conglomerate status as the given real target one year prior to the merger. Second, we choose 25 closest hypothetical targets in terms of the pre-merger AUM and use the variable *closeness* as an indicator for how close the hypothetical target is to the actual target. Finally, we combine the samples with the actual targets and the corresponding 25 hypothetical targets together and set the variable *closeness* to zero for actual targets.

Aggregate Data:

6. *statistics_dealinfo.dta* (in the "data" folder) contains aggregate statistics on the number of deals and the total AUM of our overall sample by deal completion year. To calculate the deal volume, we first calculate the average TNA for each fund in each deal, and then aggregate the TNA by deal year.

Stata do-files

The do-files that replicate all tables and figures in the main paper are located inside the "code" folder. In order for the code to run correctly, you need the following packages installed in Stata:

- _reghdfe: Regression with multiple levels of fixed effects (install the package using command "ssc install reghdfe") .
- _estout: Output Stata tables in tex format. (Install the package using command "ssc install estout").
- _coefplot: Plots results from estimation commands or Stata matrices. (Install the package using command "ssc install coefplot").

The only do-file that needs to be run by the user is the *master.do* file. It will set the paths and call other do-files that generate all figures and tables in the main paper. Please note that, because the variables constructed based on licensed data are set to missing, most of the do-files will not generate the same tables and figures as the ones in the published paper. At the time of download, the output folder will be empty, but after running *master.do*, it will contain the tables and figures created by the do-files.

In order to run the *master.do* file, you need to update the folder path (global path in row 16 of *master.do*). Notice that the folders are set such that the starting point is the location of the read-me file. All paths are built around this starting point.

Conclusion

We hope this read-me file provides sufficient clarification regarding the replication folder. Thank you for your interest in our paper.

Mancy, Alberto and David