A Appendix

A.1 Calculation of firm total factor productivity

Following Griffith, Redding and Simpson (2009), the level of LogTFP can be calculated as:

\[
\text{LogTFP}_{it} = \text{Log}(Y_{it}/\bar{Y}_j) - \sum_{z=1}^{Z} \sigma^z_i \text{Log}(x^z_{it}/\bar{x}^z_j),
\]

where \(i, j,\) and \(t\) are firm-, industry-, and time-specific subscripts. \(Y_{it}\) is the output of firm \(i\) in year \(t\) in the form of total sales, and \(\bar{Y}_j\) is the corresponding geometric mean in industry \(j\). \(x^z_{it}\) denotes the use of factor \(z\). We consider three factors of production, labor, capital, and material input costs. Labor input is measured by the total wage bill, capital by gross fixed assets, and material input costs by raw material expenditures. \(\bar{x}^z_j\) captures the industry-specific geometric mean of each factor. \(\sigma^z_i = (\alpha^z_i + \bar{\alpha}^z_j)/2\), where \(\alpha^z_i\) is the share of the factor \(z\) in output. \(\sigma^z_i\) captures the average of the factor share in each firm \(i\), and the geometric mean factor share of the corresponding industry \(j\). Similarly, total factor productivity growth \(\Delta \text{LogTFP}_{it}\) is given by

\[
\Delta \text{LogTFP}_{it} = \Delta \text{Log}Y_{it} - \sum_{z=1}^{Z} \tilde{\alpha}^z_{it} \Delta \text{Log}(x^z_{it}),
\]

where \(\tilde{\alpha}^z_{it} = (\alpha^z_{it} + \bar{\alpha}^z_{it-1})/2\). The superlative index number approach assumes constant returns to scale, which requires \(\sum_z \sigma^z_i = 1\) and \(\sum_z \tilde{\alpha}^z_{it} = 1\).

Sales, the wage bill, gross fixed assets, raw material expenses, and exports are in million rupees. Sales and the export volume are deflated by the Indian industry-specific wholesale price index, and all other monetary values are deflated by the Indian overall wholesale price index.
Table A.1: The impact of product sophistication on foreign investments

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dFDIpos</td>
<td>dFDIpos</td>
<td>dFDIneg</td>
<td>dFDIneg</td>
<td>ChangeFDI</td>
<td>ChangeFDI</td>
</tr>
<tr>
<td>LogEXS</td>
<td>0.0008</td>
<td>0.0009*</td>
<td>0.0325</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0005)</td>
<td>(0.0273)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>horizontal_EXS</td>
<td>0.0000</td>
<td>-0.0000</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forward_EXS</td>
<td>0.0000</td>
<td>-0.0000</td>
<td>0.0001**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>backward_EXS</td>
<td>-0.0000</td>
<td>-0.0000**</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogTFP</td>
<td>-0.0029</td>
<td>-0.0018</td>
<td>-0.0032*</td>
<td>-0.0067***</td>
<td>-0.0569</td>
<td>-0.0477</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0019)</td>
<td>(0.0017)</td>
<td>(0.0017)</td>
<td>(0.0920)</td>
<td>(0.0687)</td>
</tr>
<tr>
<td>LogAge</td>
<td>0.0070***</td>
<td>0.0062***</td>
<td>0.0054***</td>
<td>0.0014</td>
<td>0.3507</td>
<td>0.3463*</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0013)</td>
<td>(0.0013)</td>
<td>(0.0013)</td>
<td>(0.2167)</td>
<td>(0.2071)</td>
</tr>
<tr>
<td>HHio</td>
<td>0.0346</td>
<td>0.0174</td>
<td>-0.0547</td>
<td>0.0203</td>
<td>8.3733*</td>
<td>3.5602</td>
</tr>
<tr>
<td></td>
<td>(0.1023)</td>
<td>(0.1067)</td>
<td>(0.1098)</td>
<td>(0.1104)</td>
<td>(4.6762)</td>
<td>(3.0751)</td>
</tr>
<tr>
<td>L.LogEXS</td>
<td>0.0010</td>
<td>0.0006</td>
<td>0.0001</td>
<td></td>
<td>0.0140</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0005)</td>
<td>(0.0001)</td>
<td></td>
<td>(0.0169)</td>
<td></td>
</tr>
<tr>
<td>L.horizontal_EXS</td>
<td>-0.0000</td>
<td>-0.0000</td>
<td>0.0001*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L.forward_EXS</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L.backward_EXS</td>
<td>-0.0000</td>
<td>-0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>36,087</td>
<td>30,323</td>
<td>36,278</td>
<td>30,515</td>
<td>34,864</td>
<td>31,693</td>
</tr>
<tr>
<td>Sample</td>
<td>All firms</td>
<td>All firms</td>
<td>All firms</td>
<td>All firms</td>
<td>All firms</td>
<td>All firms</td>
</tr>
<tr>
<td>Time-FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>State-FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Industry-FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Firm-FE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Method</td>
<td>Probit</td>
<td>Probit</td>
<td>Probit</td>
<td>Probit</td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.0651</td>
<td>0.0479</td>
<td>0.0413</td>
<td>0.0325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. R²</td>
<td>0.000776</td>
<td>0.000217</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are clustered at the industry-year level. ***, **, and * denote significance at 0.01, 0.05, and 0.10 levels.

A.2 Cherry picking of MNEs

To look more deeply into reverse causality due to cherry picking of MNEs with regard to industry investment, we analyze the relationship between MNEs’ investment in a firm and industry sophistication in the same industry, in up- and downstream industries. In doing so, we define a measure that indicates the extent of product sophistication within each firm’s industry and within upstream and downstream industries of each firm. Specifically, we recalculate Horizontal, Backward, and Forward, taking the extent of product sophistication (EXS) instead of the foreign share of each firm. We then link the measures of upstream and downstream and intra-industry sophistication to the change in foreign direct investment in each firm. As a dependent variable we use a dummy variable dFDIpos_{it} (dFDIneg_{it}) equal to one if a firm increases (decreases) its foreign equity share in t and 0 otherwise. Additionally, we use the change in foreign equity share in percentage points ChangeFDI_{it} as a dependent variable. As explaining variables, we employ (the lags of) LogEXS and our recalculated spillover variables of product sophistication within and across linked industries. If we find a significant effect of LogEXS or our spillover measures on the change of foreign ownership, we take this as the first supportive evidence of reverse causality and cherry picking of MNEs.

The results in Table A.1 indicate that a firm’s own product sophistication and the presence of firms with products that are already sophisticated in vertically or horizontally linked industries largely does not alter the probability of investment in a certain firm. Only two out of 48 different estimates exhibit a significant, though very small influence.
on the change in foreign direct investment. In other words, the probability of an MNE investing in a firm does not seem to be systematically influenced by the extent of product sophistication within the same firm, within the same industry (horizontal), or by the extent of product sophistication in upstream or downstream industries. The same holds true for the continuous change in the foreign equity share. The results point toward endogeneity due to cherry picking being less severe than expected. However, we are of course aware that the reverse regressions cannot be taken as an ultimate test of endogeneity, and we therefore consider the estimates of our main results as indicative of the true effects.

A.3 Description of the product classification

Table A.2: Example of reclassification from CMIE codes to SITC categories

<table>
<thead>
<tr>
<th>CMIE product code</th>
<th>Name of products</th>
<th>3-digit</th>
<th>SITC description</th>
</tr>
</thead>
<tbody>
<tr>
<td>140405010000000000000</td>
<td>Conveyor systems</td>
<td>744</td>
<td>Mechanical handling equipment, and parts thereof, n.e.s.</td>
</tr>
<tr>
<td></td>
<td>Conveyors</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crusher Feed Conveyor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discharge Conveyor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>699070801000000000000</td>
<td>Fishing net</td>
<td>657</td>
<td>Special yarns, special textile fabrics and related products</td>
</tr>
<tr>
<td></td>
<td>Fish net</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fish Knitted Fabrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fishnet Fabrics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the Prowess database, product names as reported by the firms are assigned a 20-digit code based on an internal classification system by CMIE. In fact, one product code is usually linked to several different product names in the database. We first standardize product names according to their internal code. Since we are only interested in the products a firm actually manufactures, we delete product codes that refer to retail trading activities, rental income and other services performed. In doing so, we eliminate 316 different products. We then allocate each product code to the corresponding SITC three-digit category in order to determine the sophistication level of a product. This task was performed manually by a research assistant. We double-checked the reclassification and sorted out inconsistencies. Table A.2 provides an example of the concordance between the 20-digit internal code and the SITC Rev. 3 classification. Product names often differ in spelling (Fishing net vs. Fish net) or are more or less precise (Conveyors vs. Discharge Conveyor). We manage to classify 82 percent of all firm-product-year observations in our subsample at the three-digit level. These account for 88 percent of the total product output. For the remaining share of 12 percent of total output, we cannot determine the corresponding concordance because there is not sufficient information available on the type of the product. Assigning products to the four- or five-digit level would certainly be more satisfactory and reflect single products better than a more aggregate classification. However, given that we only observe the often rather uninformative names of the products, this is not feasible without sacrificing the precision of our concordance.
## A.4 Industry correspondence

Table A.3: Correspondence between ISIC Rev.4 and ISIC Rev.3 for manufacturing industries

<table>
<thead>
<tr>
<th>ISIC Rev.4 (2-digit) Manufacture of</th>
<th>ISIC Rev.3 (2-digit) Name</th>
<th>I-O cat.</th>
<th>% firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Food products</td>
<td>15 Food products, beverages</td>
<td>C15,16</td>
<td>12.3</td>
</tr>
<tr>
<td>11 Beverages</td>
<td>15 Food products, beverages</td>
<td>Tobacco</td>
<td></td>
</tr>
<tr>
<td>12 Tobacco products</td>
<td>16 Tobacco</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Textiles</td>
<td>17 Textiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Wearing apparel</td>
<td>18 Textile products</td>
<td>C17,18,19</td>
<td>13.6</td>
</tr>
<tr>
<td>15 Leather and related products</td>
<td>19 Leather and footwear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Wood, wood and cork products, excl. furniture</td>
<td>20 Wood and products of wood and cork</td>
<td>C20</td>
<td>0</td>
</tr>
<tr>
<td>17 Paper and paper products</td>
<td>21 Pulp, paper, paper products</td>
<td>C21,22</td>
<td>0</td>
</tr>
<tr>
<td>18 Printing and reproduction of recorded media</td>
<td>22 Printing and publishing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 Coke and refined petroleum products</td>
<td>23 Coke, refined petroleum, nuclear fuel</td>
<td>C23</td>
<td>1.0</td>
</tr>
<tr>
<td>20 Chemicals and chemical products</td>
<td>24 Chemicals and chemical products</td>
<td>C24</td>
<td>21.6</td>
</tr>
<tr>
<td>21 Pharmaceuticals, medic., chem., and botan. products</td>
<td>24 Chemicals and chemical products</td>
<td>C24</td>
<td>21.6</td>
</tr>
<tr>
<td>22 Rubber and plastics products</td>
<td>25 Rubber and plastics products</td>
<td>C25</td>
<td>6.9</td>
</tr>
<tr>
<td>23 Other non-metallic mineral products</td>
<td>26 Other non-metallic mineral products</td>
<td>C26</td>
<td>4.4</td>
</tr>
<tr>
<td>24 Basic metals</td>
<td>27 Basic metals</td>
<td>C27</td>
<td>11.3</td>
</tr>
<tr>
<td>25 Fabricated metal products, excl. machinery and equipment</td>
<td>28 Fabricated metal products excl. machinery and equipment</td>
<td>C28</td>
<td>3.7</td>
</tr>
<tr>
<td>26 Computer, electronic and optical products</td>
<td>30 Office, accounting and computing machinery</td>
<td>C30</td>
<td></td>
</tr>
<tr>
<td>27(excl.2570) Electrical equipment</td>
<td>31 Electrical machinery and apparatus n.e.c</td>
<td>C31</td>
<td>4.5</td>
</tr>
<tr>
<td>2750 Domestic appliances</td>
<td>29 Machinery and equipment n.e.c</td>
<td>C29</td>
<td>7.4</td>
</tr>
<tr>
<td>28 Machinery and equipment n.e.c.</td>
<td>29 Machinery and equipment n.e.c</td>
<td>C29</td>
<td>7.4</td>
</tr>
<tr>
<td>29 Motor vehicles, trailers and semi-trailers</td>
<td>34 Motor vehicles, trailers and semi-trailers</td>
<td>C34</td>
<td>0.3</td>
</tr>
<tr>
<td>30 Other transport equipment</td>
<td>35 Other transport equipment</td>
<td>C35</td>
<td>7.2</td>
</tr>
<tr>
<td>31 Furniture</td>
<td>36 Manufacturing n.e.c</td>
<td>C36,37</td>
<td>1.9</td>
</tr>
<tr>
<td>32 Other manufacturing</td>
<td>36 Manufacturing n.e.c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33 Repair and installation of machinery and equipment</td>
<td>37 Recycling</td>
<td>C36,37</td>
<td>1.9</td>
</tr>
</tbody>
</table>

The industries captured by the OECD input-output tables are based on 23 2-digit ISIC Rev. 3 categories, but have already been aggregated to 18 industries. From the 18 industries defined in the OECD input-output tables, we combined industries C30, C32, and C33 into one industry, since these three industries correspond to one ISIC Rev.4 industry (code 26).

### References