TOOL CHOICE MATTERS

Kavaler, Trockman, Vasilescu, Filkov
SOFTWARE DEVELOPMENT KEEPS CHANGING

➤ Waterfall
  ➤ OOP
    ➤ flexible off the shelf
      ➤ modular
        ➤ collaborative
          ➤ agile
            ➤ platform independence
              ➤ containers
                ➤ automation, independence

DevOps, CI, CD
SOFTWARE DEVELOPMENT KEEPS CHANGING

➤ Waterfall

➤ OOP

➤ flexible off the shelf

➤ modular

➤ collaborative

➤ agile

➤ platform independence

➤ containers

➤ automation, independence

➤ DevOps, CI, CD

Relies on Tools
More and more

time
NEW TECH INCREASES PRODUCTIVITY AND PREDICTABILITY

From Boehm and Valerdi, 2008
I GOT TOOLS FOR THIS, I GOT TOOLS FOR THAT

- Tools available for many tasks
  - QA: linters, package managers, coverage, testing, deployment

![Logos of various tools](image)
I GOT **MULTIPLE TOOLS FOR THIS**

➤ Many tools available for the same task

➤ E.g., dependency managers

➤ Projects adopt tools with features needed, presumably
WHICH ONES?

➤ But how are the tools chosen?
➤ What discussions precede the choices?
➤ Are any benefits seen/goals achieved after tool adoption?

### Tool Adoption Summary Statistics

<table>
<thead>
<tr>
<th>Tool</th>
<th>Task class</th>
<th># Adoption Events Across Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Per tool</td>
</tr>
<tr>
<td>david</td>
<td>Dependency Management</td>
<td>20,763</td>
</tr>
<tr>
<td>bithound</td>
<td></td>
<td>900</td>
</tr>
<tr>
<td>gemnasium</td>
<td></td>
<td>3,093</td>
</tr>
<tr>
<td>codecov</td>
<td>Coverage</td>
<td>2,785</td>
</tr>
<tr>
<td>codeclimate</td>
<td></td>
<td>2,328</td>
</tr>
<tr>
<td>coveralls</td>
<td></td>
<td>11,221</td>
</tr>
<tr>
<td>ESLint</td>
<td>Linter</td>
<td>7,095</td>
</tr>
<tr>
<td>JSHint</td>
<td></td>
<td>2,876</td>
</tr>
<tr>
<td>standardJS</td>
<td></td>
<td>3,435</td>
</tr>
</tbody>
</table>

*Note:*
- 54,440 total projects under study
- 38,948 projects which adopt tools under study
- 2,283 projects use different tools in the same task class
PROJECTS USE MULTIPLE TOOLS

- Number of projects: 0, 12,500, 25,000, 37,500, 50,000
- Number of tools: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+
Sometimes projects switch from one tool to another in the same task class.

Why do they switch? Is there a benefit?
We expected to find at least some discussions of the choices.
RESEARCH QUESTIONS

➤ RQ1: How often do projects change between tools within the same task class?

➤ RQ2: Are there measurable changes, in terms of monthly churn, pull requests, number of contributors, and issues, associated with adopting a tool? Are different tools within an equivalence class associated with different outcomes?

➤ RQ3: Are certain tool adoption sequences more associated with changes in our outcomes of interest than others?
STUDY DESIGN

➤ Research methodology
  ➤ Software Repository mining
  ➤ Quasi-experiments, modeling, hypothesis testing
  ➤ Case studies for triangulation, theory building

➤ Focus: 3 task classes (linters, dependency managers, code coverage)

➤ Data: 54,440 projects, 38,948 tool adoptions
RQ1: TOOL SWITCHING ALLUVIAL DIAGRAMS

Dependency Manager

Tool
- bithound
- david
- gemnasium
Most projects choose one tool within a task class and stick with it. When projects adopt additional tools within the same task class, they go with the most popular choice.
RQ2: EFFECTIVENESS BEFORE AND AFTER ADOPTION

Effectiveness variables: churn, #pull requests, #unique authors, #issues
INTERRUPTED TIME SERIES: REGRESSION DISCONTINUITY

\[
y_i = \alpha + \beta \cdot \text{time}_i + \gamma \cdot \text{intervention}_i + \delta \cdot \text{time_after_intervention}_i + \varepsilon_i
\]
SLOPE INCREASES OR DECREASES, AND DISCONTINUITY
### Dependency Manager Models

<table>
<thead>
<tr>
<th></th>
<th>log(Churn + 1)</th>
<th>PRs</th>
<th>Unique Authors</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeffs (Err.)</td>
<td>Coeffs (Err.)</td>
<td>Coeffs (Err.)</td>
<td>Coeffs (Err.)</td>
</tr>
<tr>
<td>Authors</td>
<td>0.456*** (0.007)</td>
<td>0.383*** (0.004)</td>
<td>0.009*** (0.000)</td>
<td>0.052*** (0.004)</td>
</tr>
<tr>
<td>Commits</td>
<td>0.052*** (0.001)</td>
<td>0.014*** (0.000)</td>
<td></td>
<td>0.015*** (0.000)</td>
</tr>
<tr>
<td>PRs</td>
<td>−0.078*** (0.003)</td>
<td></td>
<td>0.068*** (0.001)</td>
<td></td>
</tr>
<tr>
<td>Churn time</td>
<td>−0.049*** (0.000)</td>
<td>−0.040*** (0.001)</td>
<td></td>
<td>0.009*** (0.001)</td>
</tr>
<tr>
<td>gemnasium_int</td>
<td>−0.168*** (0.043)</td>
<td>−0.022 (0.025)</td>
<td>0.048*** (0.010)</td>
<td>0.061 (0.030)</td>
</tr>
<tr>
<td>gemnasium_after</td>
<td>−0.003 (0.003)</td>
<td>0.009*** (0.002)</td>
<td>0.001* (0.001)</td>
<td>−0.001 (0.002)</td>
</tr>
<tr>
<td>david_int</td>
<td>−0.182*** (0.021)</td>
<td>0.199*** (0.012)</td>
<td>0.082*** (0.005)</td>
<td>0.138*** (0.015)</td>
</tr>
<tr>
<td>david_after</td>
<td>−0.009*** (0.002)</td>
<td>0.005*** (0.001)</td>
<td>−0.001 (0.000)</td>
<td>−0.011*** (0.001)</td>
</tr>
<tr>
<td>bithound_int</td>
<td>0.002 (0.107)</td>
<td>0.418*** (0.061)</td>
<td>0.088*** (0.026)</td>
<td>0.078 (0.073)</td>
</tr>
<tr>
<td>bithound_after</td>
<td>−0.028* (0.014)</td>
<td>0.037*** (0.008)</td>
<td>0.000 (0.003)</td>
<td>−0.014 (0.009)</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.567*** (0.014)</td>
<td>0.289*** (0.009)</td>
<td>0.998*** (0.003)</td>
<td>0.415*** (0.011)</td>
</tr>
</tbody>
</table>

Marginal $R^2$ | 21.6% | 7.1% | 12.5% | 2.4%
Conditional $R^2$ | 54.1% | 58.9% | 44.8% | 56.8%

*Note: $p<0.05$; **$p<0.01$; ***$p<0.001$
### Dependency Manager Models

<table>
<thead>
<tr>
<th></th>
<th>log(Churn + 1) Coefs (Err.)</th>
<th>PRs Coefs (Err.)</th>
<th>Unique Authors Coefs (Err.)</th>
<th>Issues Coefs (Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>0.456*** (0.007)</td>
<td>0.383*** (0.004)</td>
<td></td>
<td>0.052*** (0.004)</td>
</tr>
<tr>
<td>Commits</td>
<td>0.052*** (0.001)</td>
<td>0.014*** (0.000)</td>
<td>0.009*** (0.000)</td>
<td>0.015*** (0.000)</td>
</tr>
<tr>
<td>PRs</td>
<td>-0.078*** (0.003)</td>
<td>-0.040*** (0.001)</td>
<td>0.068*** (0.001)</td>
<td></td>
</tr>
<tr>
<td>Churn time</td>
<td>-0.049*** (0.000)</td>
<td>0.011*** (0.000)</td>
<td>-0.005*** (0.000)</td>
<td>0.009*** (0.001)</td>
</tr>
<tr>
<td>gemnasium_int</td>
<td>-0.168*** (0.043)</td>
<td></td>
<td></td>
<td>0.048*** (0.010)</td>
</tr>
<tr>
<td>gemnasium_after</td>
<td>-0.003 (0.003)</td>
<td></td>
<td></td>
<td>0.001* (0.001)</td>
</tr>
<tr>
<td>david_int</td>
<td>-0.182*** (0.021)</td>
<td></td>
<td>0.082*** (0.005)</td>
<td>0.138*** (0.015)</td>
</tr>
<tr>
<td>david_after</td>
<td>-0.009*** (0.002)</td>
<td></td>
<td></td>
<td>0.001 (0.000)</td>
</tr>
<tr>
<td>bithound_int</td>
<td>0.002 (0.107)</td>
<td></td>
<td></td>
<td>0.088*** (0.026)</td>
</tr>
<tr>
<td>bithound_after</td>
<td>-0.028* (0.014)</td>
<td>0.00 (0.008)</td>
<td></td>
<td>0.078 (0.073)</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.567*** (0.014)</td>
<td>0.289*** (0.009)</td>
<td>0.998*** (0.003)</td>
<td>0.415*** (0.011)</td>
</tr>
</tbody>
</table>

**Marginal $R^2$** 21.6% 7.1% 12.5% 2.4%

**Conditional $R^2$** 54.1% 58.9% 44.8% 56.8%

*Note:* *p<0.05; **p<0.01; ***p<0.001
RQ3: ON ADOPTION ORDER
RQ3: SOME RESULTS
CONCLUSION AND FUTURE

➤ Tool choice matters but it is not discussed much
➤ Projects can benefit from adopting the right tool
➤ The order in which tools are adopted matters

➤ Future goal: bespoke tool pipelines, depending on project context
THANKS!

➤ NSF
➤ DECAL @ UCD
➤ Strudel @ CMU