Zelig: Tools to expand the reach of R

Koskue Imai\textsuperscript{1}  Gary King\textsuperscript{2}  Olivia Lau\textsuperscript{2}

\textsuperscript{1}Department of Politics  Princeton University
\textsuperscript{2}Department of Government  Harvard University

Vienna, Austria  June 16, 2006

What does Zelig do?

For developers:
Makes it easier to translate user-inputs

For researchers, students, and instructors:
Computes quantities of substantive interest for every model:

\[
\begin{align*}
E(Y) = E(Y|X) = E(Y|X_0) \\
e \text{ a command or function call for every model}
\end{align*}
\]

3 commands \rightarrow Simplified \text{\textit{ui}}

For researchers, students, and instructors who don't use R:
An automatically-generated \textit{gui}

What does Zelig do?

For developers:
Makes it easier to translate user-inputs

For researchers, students, and instructors:
Computes quantities of substantive interest for every model:

\[
\begin{align*}
E(Y) = E(Y|X) = E(Y|X_0) \\
e \text{ a command or function call for every model}
\end{align*}
\]

3 commands \rightarrow Simplified \text{\textit{ui}}

For researchers, students, and instructors who don't use R:
An automatically-generated \textit{gui}
What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  Computes quantities of substantive interest for every model

\[
E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots
\]
- 3 commands → Simplified UI

- For researchers, students, and instructors who don’t use R:
  An automatically-generated GUI

Kosuke Imai, Gary King, Olivia Lau

Zelig: Everyone’s Statistical Software
What does Zelig do?

For developers:
- Makes it easier to translate user-inputs

For researchers, students, and instructors:
- Computes quantities of substantive interest for every model

\[ E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots \]

- 3 commands → Simplified UI

For researchers, students, and instructors who don’t use R:
An automatically-generated GUI

Goal: R world (with the help of Zelig)

Users only need to know 3 commands

\[ z <- \text{zelig}(\text{vote}\sim\text{race}+\text{educate}, \quad \text{Select vars} \]
\[ \text{data=turnout,} \quad \text{Select data set} \]
\[ \text{model="probit"}) \quad \text{Select model} \]

\[ x <- \text{setx}(z.\text{out}, \text{educate}=12) \quad \text{Select QIs} \]

\[ s <- \text{sim}(z, x=x) \quad \text{Calculate QIs} \]
z <- zelig(vote~race+educate, data=turnout, model="probit")

Select vars
Select data set
Select model

x <- setx(z.out, educate=12)

Select QIs

s <- sim(z, x=x)

Calculate QIs

Koskue Imai, Gary King, Olivia Lau
Zelig: Everyone's Statistical Software
<table>
<thead>
<tr>
<th>Built-in Zelig functionality</th>
<th>Built-in Zelig functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handle sets of multiply-imputed data frames</td>
<td>Handle sets of multiply-imputed data frames</td>
</tr>
<tr>
<td>Bootstrap quantities of interest</td>
<td>Bootstrap quantities of interest</td>
</tr>
<tr>
<td>Stratify data</td>
<td>Stratify data</td>
</tr>
<tr>
<td>Conditional prediction</td>
<td>Conditional prediction</td>
</tr>
</tbody>
</table>

Koskue Imai, Gary King, Olivia Lau

Zelig: Everyone's Statistical Software
No need to change existing packages
Developers only need to add a few functions

Estimate
zelig()
  (1) zelig2mymodel()
  (2) mymodel()

Interpret
sim()
  (3) param.myclass()
  (4) qi.myclass()

Koskue Imai, Gary King, Olivia Lau
Zelig: Everyone’s Statistical Software

R framework for interpreting user-inputs

User specifies a formula
- model.frame
- model.matrix
- model.extract (or model.response)
R framework for interpreting user-inputs

- User specifies a formula
- model.frame
- model.matrix
- model.extract (or model.response)

User specifies a formula
model.frame
model.matrix
model.extract (or model.response)

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks → different and diverse UIs
- Challenging for programmers → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API
But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks → different and diverse UIs
- Challenging for programmers → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Koskue Imai, Gary King, Olivia Lau
Zelig: Everyone’s Statistical Software
But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks → different and diverse UIs
- Challenging for programmers → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Represent arbitrarily complicated models

Use a list of equations:

\[
f <- \text{list}(\mu_1 = y_1 \sim x_1 + x_2 + x_3, \\
\quad \mu_2 = y_2 \sim x_1 + x_4 + x_5)
\]

\[
f <- \text{list}(\mu_1 = y_1 \sim x_1 + \text{tag}(x_2, \text{beta}2), \\
\quad \mu_2 = y_2 \sim x_3 + \text{tag}(x_4, \text{beta}2), \\
\quad \rho = \sim z_1 \sim 1)
\]

\[
f <- \text{list}(\text{cbind}(y_1, y_2) \sim x_1 + x_2)
\]

Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`
Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`

Koskue Imai, Gary King, Olivia Lau
Zelig: Everyone's Statistical Software

The Virtual Data Center GUI for Zelig

Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI "dynamically" extends to include new models!
- Developer doesn’t need to write a graphical module
Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI “dynamically” extends to include new models!
- Developer doesn’t need to write a graphical module

Everyone should use R... ...and can with Zelig!

- Partnership with the US Census Bureau (DataFerrett)
- Partnership with the Broad Institute (GenePattern)
- Your suggestions!
Everyone should use R... ...and can with Zelig!

- Partnership with the US Census Bureau (DataFerrett)
- Partnership with the Broad Institute (GenePattern)
- Your suggestions!

Visit Zelig on the web at

http://gking.harvard.edu/zelig/