



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Constructing a summary index using the standardized inverse-covariance weighted average of indicators

Benjamin Schwab
Kansas State University
Department of Agricultural Economics
Manhattan, KS
benschwab@ksu.edu

Sarah Janzen
University of Illinois at Urbana Champaign
Department of Agricultural and Consumer Economics
Urbana, IL
sjanzen@illinois.edu

Nicholas P. Magnan
University of Georgia
Department of Agricultural and Applied Economics
Athens, GA
nmagnan@uga.edu

William M. Thompson
IDInsight
New Delhi, India
will.thompson@idinsight.org

Abstract. Researchers often want to examine the relationship between a variable of interest and multiple related outcomes. To avoid problems of inference that arise from testing multiple hypotheses, one can create a summary index of the outcomes. Summary indices facilitate generalizing findings and can be more powerful than individual tests. In this article, we introduce a command, `swindex`, that implements the generalized least-squares method of index construction proposed by Anderson (2008, *Journal of the American Statistical Association* 103: 1481–1495). We describe the command and its options and provide an example based on Blattman, Fiala, and Martinez’s (2014, *Quarterly Journal of Economics* 129: 697–752) evaluation of a cash transfer program in Uganda.

Keywords: `st0622`, `swindex`, index construction, GLS

1 Introduction

In statistical analysis, one is often interested in looking at the causal relationship between a variable of interest and multiple outcomes. Intuitively, the more hypotheses one tests, the more likely one is to erroneously reject a null hypothesis. Various corrections exist, including the familywise error rate, the false discovery rate, or the particularly conservative Bonferroni correction. Using a summary index as an outcome is an alter-

native or complementary solution. A summary index combines multiple indicators into a single index, allowing one to test the single hypothesis of whether the index is affected by a causal variable of interest, as opposed to testing multiple related hypotheses individually. This accommodates a statistical test for whether a program has a general effect on a set of outcomes as opposed to a series of related tests.

For example, a study might be interested in estimating the impact of a program on empowerment. Empowerment is a multidimensional concept, and a variety of indicators could be used to measure it. Such indicators might include decision-making power regarding spending over a particular item or items, ownership of a particular asset or assets, freedom to travel to a specific location or locations, and values or behaviors regarding intimate partner violence. Rather than test for the relationship between each of these indicators and the program of interest, which suffers from the multiple comparison problem, a researcher may wish to simply say something about the effect of a program on empowerment, broadly speaking. A summary index compiles information from each of these indicators into a single measure of empowerment, accommodating a single hypothesis test.

Aggregating multiple related outcomes into a summary index of a broader outcome can also facilitate generalizing findings, particularly when it comes to making broad conclusions about an intervention's effectiveness. For instance, a researcher could find a program increases women's control over spending decisions and improves a woman's freedom to travel, while having no significant impact on asset ownership or reported intimate partner violence. With only the estimates of impact for multiple indicators over a range of related outcomes, coming to a succinct conclusion as to whether the program improves women's empowerment can be difficult. Suppose instead the researcher creates a summary index of empowerment, which reveals a positive impact on the summary index. The researcher can then say the program, broadly speaking, increases women's empowerment.

One obvious criticism of this generalized conclusion is that it fails to really explain or unpack what is going on. In some sense, the summary index is a "black box". In response to this criticism, the researcher could then unpack the index by estimating the same regression using each of the components of the index as individual outcomes, as we will show in the example presented later in this article. In this way, using a summary index does not preclude researchers from going into greater detail about the mechanisms behind the generalized findings. Reporting and discussing individual indicator impacts remains common practice and can help unpack the "black box" index-based findings—with the caveat that doing so retains the multiple comparison problem and is therefore often considered exploratory.

Summary index-based tests can also be more statistically powerful than individual-level tests. As described in Anderson (2008), multiple outcomes that approach significance may aggregate into a single index that attains statistical significance. Each indicator is measured with some error and may exhibit (pretreatment) imbalance in finite samples. When one aggregates variables into an index, random errors that are uncorrelated across indicators are more likely to cancel each other out as the number

of indicators increases. Thus, summary indices as outcomes can be less noisy than individual variables.

In some cases, an established summary index exists for some broadly defined outcome. For example, Alkire et al. (2013) propose the women’s empowerment in agriculture index, which was developed through collaboration among the United States Agency for International Development, the International Food Policy Research Institute, and Oxford Poverty and Human Development Initiative and has been implemented as part of the United States Agency for International Development’s Feed the Future initiative to monitor changes in empowerment in many countries and across multiple continents. Established indices like the women’s empowerment in agriculture index can be helpful in many contexts, but by definition they are not very flexible. A given dataset may not always have the requisite indicators, such indices often rely on arbitrary weights, and missing observations can be problematic. Moreover, it is often the case that an established summary index does not exist, and it is up to the researcher to develop one from scratch. Such ad hoc indices can lead to p-hacking, especially if a researcher constructs the index *ex post*.

The new `swindex` command (standardized weighted index) presented here constructs a summary index using the standardized inverse-covariance weighted average of indicators proposed by Anderson (2008) and implemented recently by Haushofer and Shapiro (2016), Janzen et al. (2018), and others.

2 Constructing a summary index

Anderson (2008) proposes constructing a summary index using a generalized least-squares (GLS) weighting procedure, which has two primary advantages. First, it increases efficiency by ensuring highly correlated indicators receive less weight than uncorrelated indicators. Intuitively, uncorrelated indicators, which represent “new” information, receive more weight. Second, the procedure uses all available data but ascribes lower weight to indicators with missing values, which allows for the calculation of the summary index even for observations with missing indicators.

The index requires $k \geq 2$ indicators related to the outcome the researcher seeks to summarize. All indicators should be related to a singular theme (for example, an index of “freedom and prosperity” combines two separate concepts rather than a singular theme). Indices can be constructed using a mix of continuous and binary indicators, and combinations of the two are commonly used (see Haushofer and Shapiro [2016] and Janzen et al. [2018] for examples). The researcher is simply required to assign each variable with a direction (that is, polarity)—positive or negative. In the empowerment example, an indicator that takes a higher value for greater control over assets could enter into an empowerment index without any adjustment, and higher values for the indicator would also lead to a higher summary index. An indicator that takes a higher value for experiencing more or worse intimate partner violence decreases empowerment, so the indicator must enter the index negatively.

Categorical or ordinal variables that might otherwise be analyzed as outcomes using a multinomial or ordered logit regression could be incorporated by creating a binary variable for each response if we classify some responses as contributing positively to the index and others as contributing negatively. For an example of how a categorical variable can be incorporated into a summary index, suppose we want to create a “green behavior” index. Suppose we have binary variables for “walk”, “bicycle”, “public transportation”, and “private vehicle”. The first three indicators could enter positively into the index, while a binary variable for “private vehicle” could enter negatively into the index. Alternatively, a single binary variable could be created for “green transport” that takes on a value of 1 for “walk”, “bicycle”, or “public transport” and 0 otherwise, and this variable could enter positively into the index. With an ordinal variable, we could either treat it as cardinal or dichotomize it and choose a reference level above which it should enter positively into the index and below which it should enter negatively. As an example, suppose we wanted to include the response to a question about attitudes toward climate change posed using a Likert scale as an indicator to be included in a “green attitudes” index. Suppose the respondent is asked to state whether he or she agrees with the statement “Combatting climate change should be a top policy priority.” Response options are 1 = disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree. We could either allow the variable to enter as a continuous variable with values ranging from 1 to 5 or create a binary variable for a negative response (1 or 2) that enters negatively into the index and a binary variable for a positive response (4 or 5) that enters positively into the index.

Using the Anderson (2008) approach, we can calculate the standardized weighted index \tilde{s} for each observation i as follows:

1. Select k indicators relevant for outcome j .
2. Adjust sign: For all k indicators, ensure the positive direction always indicates a “better outcome”.
3. Normalize indicators: Demean all k indicators by subtracting the mean of the indicator in the reference group (the full sample is the default reference group). Then, convert them to effect sizes, \tilde{y}_k , by dividing each indicator by its reference group standard deviation.¹
4. Construct weights: Create weights using Σ^{-1} , the inverse of the covariance matrix of the normalized indicators.² Specifically, set the weight \tilde{w}_k on each indicator equal to the sum of its row entries in Σ^{-1} . With this rule, highly correlated indicators are assigned small or offsetting weights, while less correlated outcomes receive larger weights.
5. Construct index: Calculate the weighted average of \tilde{y}_k for observation i . Formally, the weighted average \bar{s}_i is calculated using $\bar{s}_i = (\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})^{-1}(\mathbf{1}'\hat{\Sigma}^{-1}\tilde{\mathbf{y}}_i)$, where $\mathbf{1}$

1. In some cases, particularly when all subindicators are of comparable scales, the normalization may not be necessary. The `swindex` command accommodates this special case using the `nostd` option.

2. This is equivalent to using the correlation matrix in the special case that the normalization procedure described in step 3 standardizes against the full sample using the default option.

is a column vector of 1s and $\tilde{\mathbf{y}}_i$ is a column vector of all outcomes for observation i . This is an efficient GLS estimator.

6. Normalize index: Demean index \bar{s}_i by subtracting the mean of the index in the reference group, and convert it to effect sizes by dividing it by its reference group standard deviation.³ This normalization results in an index distributed with mean zero and standard deviation one in the reference group.

3 The swindex command

3.1 Syntax

The syntax of `swindex` is as follows:

```
swindex varlist [if] [in], generate(varname) [replace flip(varlist)
    normby(varname) nostd fullrescale norescale numvars(newvar) displayw]
```

3.2 Options

`generate(varname)` creates a new variable that takes the value of the constructed composite index. *varname* in `generate()` does not have to be a new variable name. Use the option `replace` if *varname* exists and is to be replaced. `generate()` is required.

`replace` changes the contents of an existing variable. Because `replace` alters data, the command cannot be abbreviated.

`flip(varlist)` alters the sign of variables in *varlist* to calculate the index. This command should be used when the variables in *varlist* move in the opposite direction as the summary index (for example, where increases indicate “worse” outcomes). The variables in *varlist* provided in `flip()` must be a subset of the variables in *varlist*. Invoking this option will not result in changes to the variables in *varlist* in the permanent dataset.

`normby(varname)` specifies the reference group over which to standardize the variables in *varlist* during the GLS weighting procedure. When one estimates impacts from a randomized intervention, `normby()` will typically be an indicator for being in the control group. The specified *varname* must be a binary variable. With `normby()`, the specified *varname* is also used to rescale the calculated index unless the `fullrescale` or `norescale` option is specified. This option cannot be combined with `nostd`.

3. This step is implicitly done by Anderson (2008), although it is not formally discussed. This step is optional using the `swindex` command with the `norescale` option. The default uses the reference group from step 3, but the full sample mean and standard deviation can alternatively be used to normalize the index with the `fullrescale` option.

nostd specifies that no standardization should take place during the GLS weighting procedure. When **nostd** is specified, the calculated index is also not rescaled unless **fullrescale** is specified. This option cannot be combined with **normby()**.

fullrescale specifies the full sample be used to rescale the calculated index. This is the default option when **normby()** and **nostd** are not specified.

nore scale specifies that the calculated index not be rescaled. This is the default option when **nostd** is specified.

numvars(*newvar*) stores the number of nonmissing variables in *varlist* for each observation.

displayw interactively displays the proportional weights used in the index calculation. The displayed weight matrix is also stored in the return list as **r(pw)**, while the raw weights are stored as **r(wt)**.

3.3 Stored results

swindex stores the following in **r()**:

r(pw)	matrix of proportional weights used to construct index variable
r(wt)	matrix of the raw weights used to construct index variable

3.4 Summary of **swindex** command

swindex calculates a standardized weighted index from the variables in *varlist*. The procedure follows a GLS weighting procedure, as described in Anderson (2008).

The command stores the inverse-covariance weights for each variable in *varlist* calculated by **swindex**. The **r(wt)** matrix contains a vector of the raw weights, while the **r(pw)** matrix contains a vector of weights that have been scaled to sum to one (that is, proportions). Invoking the option **displayw** causes **swindex** to output the latter. The weights can be helpful for diagnostic purposes, though the user should be cautious about their interpretation. Because the procedure more heavily weights variables that provide new information not provided from the other variables in the index, highly correlated indicators are assigned small or offsetting weights. Thus, some weights may even be negative. For example, if variables y_i and y_j are highly correlated, then y_i may be assigned a large positive weight, while y_j is assigned a negative weight. The sum of the weights in this example will be similar to a weight for y_i if y_j were omitted from the index, and vice versa. Theoretically, a variable may even have a weight of exactly zero. Suppose y_c is correlated exactly 50% with y_a and 50% with y_b (that is, $y_c = 0.5y_a + 0.5y_b$). Including all three variables will result in a weight of zero for y_c ; y_c adds no new information, so it is effectively ignored.

Several options are provided to allow the user to customize the calculation. Variables included in the index should work in the same direction (for example, increases in the variables all indicate better outcomes). **swindex** allows users to include variables that

move in the opposite direction (for example, where increases indicate worse outcomes) by specifying them in *varlist* provided in the option `flip()`. The signs of variables included in `flip()` will change for the purposes of the calculation, but no changes are made to the dataset in memory.

The recommended method standardizes the indicator variables in *varlist* prior to constructing the inverted covariance matrix used in the GLS weighting procedure. This standardization can use a subsample (for example, the control group) as a reference group, or otherwise use the full sample as a reference group, for calculating the mean and standard deviation used for normalization. By default, the program normalizes against the full sample mean and standard deviation, which is equivalent to obtaining the weights by inverting the correlation matrix. If the user wishes to use the mean and standard deviation from a subsample, the user may specify the subsample varname using the `normby()` option. For example, in a randomized trial, the `normby()` option can be used to standardize based on the control group subsample. The user can also opt not to standardize by invoking `nostd`, though this is not recommended for most applications.

By default, the program rescales the calculated index to the mean and standard deviation of the sample used for the standardization in the GLS weighting procedure. This rescaling results in an “effect size” interpretation where the index is normally distributed with mean zero and standard deviation one for the sample used. The `full-rescale` option allows the user to rescale the calculated index using the full sample, even if `normby()` has been invoked for the GLS weighting procedure. Further, the user can opt not to rescale at all by specifying `norescale`.

The procedure accommodates construction of the index even when data on indicators are missing. It does so by setting missing indicator values to zero, which is the mean of the reference group following normalization. The `numvars()` option saves the number of variables in *varlist* missing for each observation.

4 Example

To illustrate how `swindex` works, we apply it to Uganda’s Youth Opportunities Program (YOP) evaluation by Blattman, Fiala, and Martinez (2014). The YOP was a government program in northern Uganda designed to help poor and unemployed adults become self-employed artisans. Young adults were invited to form groups for the preparation of grant proposals. Funding for vocational training and development of small start-up businesses was randomly assigned among 535 screened, eligible applicant groups. Successful group proposals received one-time unsupervised grants worth \$7,500 on average, or about \$382 per group member, roughly the member’s average annual income.

Blattman, Fiala, and Martinez (2014) measure intent-to-treat (ITT) estimates of the YOP impact across a range of individual outcomes. In each regression, they control for a set of individual baseline characteristics and district fixed effects. Errors are clustered at the group level (the level of treatment), and observations are weighted by the inverse

probability of selection into endline tracking. The original analysis groups the estimation into six key outcomes. At the four-year endline, more than one indicator is used for four of these main outcomes—business formality, income, employment, and migration and urbanization.

In this example, we will replicate the estimated impact of YOP after four years for each indicator related to business formality, income, employment, and migration and urbanization. This replicates a subset of the results originally presented in table 3 of Blattman, Fiala, and Martinez (2014). We then calculate a summary index of the indicators for each outcome using the `swindex` command. In a final step, we estimate the impact of the YOP using each summary index as the dependent variable. The results are presented in table 1.

Table 1. Descriptive statistics and ITT estimates of program impacts on key outcomes

	Control mean	Index weights (raw)	Obs	Coeff.	Std. error
<i>Business formality</i>					
Business formality summary index (IND.biz)	0.0	—	1,868	0.292	[0.054]***
Maintains formal records (bizlog_e)	0.260	0.57	1,868	0.124	[0.023]***
Enterprise formally registered (bizregister_e)	0.110	0.45	1,868	0.062	[0.019]***
Pays business taxes (biztaxes_e)	0.220	0.49	1,868	0.085	[0.023]***
<i>Income</i>					
Income summary index (IND_inc)	0.0	—	1,868	0.249	[0.051]***
Monthly cash earnings (profits4w_real_p99_e)	47.8	0.37	1,868	18.19	[4.898]***
Durable assets z score (wealthindex_e)	0.150	0.54	1,853	0.181	[0.055]***
Consumption z score (consumption_real_p99_z_e)	-0.011	0.57	1,862	0.180	[0.051]***
<i>Employment</i>					
Employment summary index (IND_emp)	0.0	—	1,868	0.355	[0.046]***
Avg. employment hours/week (totalhrs7da_zero)	32.2	0.19	1,864	5.5	[1.284]***
Nonagricultural hours (nonaghours7da_zero_e)	13.5	0.43	1,864	5.1	[0.998]***
Skilled trade hours (skilledtrade7da_zero_e)	2.8	-0.60	1,864	3.8	[0.548]***
No employment hours past month (zero_hours_e)	0.05	0.59	1,868	-0.022	[0.009]***
Engaged in any skilled trade (trade_dummy_e)	0.22	0.69	1,868	0.261	[0.026]***
30 hours/week skilled trade (skilledtrade7da_30_e)	0.03	0.89	1,868	0.039	[0.013]***
<i>Migration and urbanization</i>					
Migration and urbanization summary index (IND_mig_urb)	0.0	—	2,046	-0.115	[0.052]**
Has changed parish since baseline (migrate_e)	0.350	1.78	2,029	-0.077	[0.026]***
Lives in large town or city (urban_e)	0.170	0.98	1,859	0.01	[0.019]

NOTES: Column 1 reports the control group mean at the four-year endline, weighted by the inverse probability of selection into the endline sample. Columns 4–5 report the ITT estimate and standard error of program assignment at endline. Standard errors are heteroskedastic robust and clustered by group. We calculate the ITT via a weighted least-squares regression of the dependent variable on a program assignment indicator, 13 district (randomization stratum) fixed effects, and a vector of control variables that includes all the baseline covariates reported in table II of Blattman, Fiala, and Martinez (2014). Variable names preserved from the original replication code of Blattman, Fiala, and Martinez (2014).

*** $p < .01$, ** $p < .05$, * $p < .1$.

First, we consider business formality, which contains three dummy variables as indicators: maintenance of formal records (`bizlog_e`), formal enterprise registration (`bizregister_e`), and payment of business taxes (`biztaxes_e`). Blattman, Fiala, and Martinez (2014) show the YOP program has a statistically significant impact on each of these individual indicators. We aggregate these indicators into a standardized weighted average index using the `swindex` command. Our newly generated summary index is `IND_biz`. We normalize indicators and rescale the calculated index against the control using the `normby()` option. The binary variable, `control`, equals 1 for untreated units, so we specify the `normby()` option as follows:

```
. swindex bizlog_e bizregister_e biztaxes_e, g(IND_biz) normby(control) displayw
Weights
      bizlog_e      .37662597
    bizregister_e    .29655532
      biztaxes_e    .32681871
```

The proportional weights assigned to each indicator are presented as part of the output because the `displayw` option was specified. For this index, formal record keeping was assigned the largest weight (`bizlog_e` weight equals 0.38), followed by payment of business taxes (`biztaxes_e` weight equals 0.33), and formal enterprise registration (`bizregister_e` weight equals 0.30). Each indicator was assigned a positive weight and increases the value of the index. The largest weight is 27% larger than the smallest weight. This difference reflects that the indicator of formal record keeping offers relatively more distinct information than the indicator for formal enterprise registration. Nonetheless, these weights are not wildly different, and their relative similarity reflects the fact that each indicator provides additional information.

We are now ready to estimate the impact of the program on each indicator related to business formality (replicating the results of Blattman, Fiala, and Martinez [2014]), as well as our newly generated summary index (`IND_biz`). The estimation follows the replication files of Blattman, Fiala, and Martinez (2014) as closely as possible and estimates regressions of the form

```
. quietly: svy: regress IND_biz assigned `controls' if e2==1
```

The local `control` contains a vector of control variables, and the treatment indicator is `assigned`.⁴ Because we normalized indicators and rescaled the calculated index against the control, the control group has an index value with mean zero and standard deviation one. The estimated coefficient can thus be interpreted as an effect size: the intervention increases “business formality” by 0.292 standard deviations. For context, Cohen (1988) interprets a 0.20 standard deviation as a “small effect”, a 0.5 standard deviation as a “medium” effect, and a 0.9 standard deviation as a “large” effect.

We then analyze the impact of the intervention on income, where Blattman, Fiala, and Martinez (2014) also report impacts using three indicators: monthly cash earnings (`profits4w_real_p99_e`) reported in 1,000s of the 2008 UGX, durable assets (`wealthin-`

4. The full dataset `yop_data.dta` and estimation code `swindex.example.do` can be found on the *Stata Journal* website.

`dex.e`), and nondurable consumption (`consumption_real_p99_z.e`). Each of these indicators is included in our income summary index, `IND_inc`: the discrepancy in `IND_inc` weights is slightly larger than the differences across the `IND_biz` weights. While the z scores for durable assets and nondurable consumption assigned raw weights are 0.54 and 0.57, respectively, these similar weights are both approximately 50% greater than the weight assigned to monthly cash earnings (0.37). Regression results using summary index `IND_inc` as an outcome reveals a 0.249 standard-deviation impact on the YOP program.

The third outcome, employment, uses a greater number of indicators and requires a little more thought to construct and understand the index. The indicators included in our summary index are the following:

- average employment hours per week (`totalhrs7da_zero`),
- nonworking hours (`nonaghours7da_zero_e`),
- skilled trade working hours (`skilledtrade7da_zero_e`),
- a dummy variable for working zero hours in the past month that enters negatively (`zero_hours_e`),
- a dummy variable for being engaged in any skilled trade (`trade_dummy_e`), and
- a dummy variable if the individual worked at least 30 hours a week in a skilled trade (`skilledtrade7da_30_e`).

We omit one of the original indicators reported in Blattman, Fiala, and Martinez (2014)—the number of hours spent working on agricultural activities. We do this because the intervention targeted business development rather than agriculture, and it is not immediately obvious if agricultural hours should enter the index positively or negatively. (Notably, there is no discernible impact of the YOP intervention on the number of hours worked.) We use the `flip()` option to ensure that the dummy variable for working zero hours in the past month enters into the index negatively (if the variable equals one, it should reduce the employment index). We create the employment index `IND_emp` and estimate the impact on the summary index:

```
. local work totalhrs7da_zero nonaghours7da_zero_e skilledtrade7da_zero_e
> zero_hours_e trade_dummy_e skilledtrade7da_30_e
. swindex `work`, g(IND_work) normby(control) displayw flip(zero_hours_e)
Weights
totalhrs7da_z-e      .08643637
nonaghours7da-e      .19461683
skilledtrad-o_e      -.27265913
zero_hours_e         .2688965
trade_dummy_e        .3137599
skilledtrad-0_e      .40894953
```

We see much greater variation in the weights assigned to the indicators. Recall the index ascribes greater weight to an indicator if it contributes “new” information—that is,

if it is not highly correlated with other indicators. In this example, several indicators are highly correlated for obvious reasons. For example, a dummy variable for being engaged in any skilled trade and a dummy variable for being engaged in at least 30 hours of skilled trade are directly derived from the number of skilled trade hours worked; hence, we observe skilled trade hours entering the index with a partially offsetting negative weight. When estimating the impact of the treatment on the employment summary index, we observe a 0.355 standard-deviation impact.

The final outcome we consider is an index of migration and urbanization. For this index, we have only two indicators, one a dummy variable indicating recent migration to a different parish (`migrate_e`) and the other a dummy variable for residing in an urban area (`urban_e`). We aggregate them here to demonstrate the concept, but it is not clear that these two variables should be aggregated into a single index. To see why, notice that when we use `swindex`, we observe a statistically significant 0.115 standard-deviation reduction in the `IND_mig_urb` index. But this is somewhat misleading because the disaggregated results demonstrate very clearly that while migration is decreasing, there is no evidence that individuals are moving to or from urban areas. In this case, the indicators are capturing two different things—migration and urbanization—rather than contributing information toward a single idea. An aggregate index in this case may not be appropriate.

As a final word of caution, we note that `swindex`-derived weights are themselves estimates. Thus, standard errors based on estimates of constructed summary indices may be underestimated. A potential solution is to bootstrap any treatment effect estimation, while including the index estimation procedure within the bootstrap loop. We are not aware of literature directly addressing the validity of this approach and suggest this as a possible area for future research.

5 Conclusion

Using a summary index accommodates testing a single hypothesis of whether the index is affected by a causal variable of interest, as opposed to testing multiple related hypotheses individually. This article presented the `swindex` command, which calculates the standardized weighted index using the GLS weighting procedure described in Anderson (2008). The procedure increases efficiency by ensuring that highly correlated outcomes receive less weight than outcomes that are uncorrelated and uses all available data by ascribing lower weight to indicators with missing values. The command provides an easy way to generate a summary index that is increasingly being used by researchers globally.

6 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-4
. net install st0622      (to install program files, if available)
. net get st0622          (to install ancillary files, if available)
```

7 References

- Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymour, and A. Vaz. 2013. The women's empowerment in agriculture index. *World Development* 52: 71–91. <https://doi.org/10.1016/j.worlddev.2013.06.007>.
- Anderson, M. L. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103: 1481–1495. <https://doi.org/10.1198/016214508000000841>.
- Blattman, C., N. Fiala, and S. Martinez. 2014. Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics* 129: 697–752. <https://doi.org/10.1093/qje/qjt057>.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Haushofer, J., and J. Shapiro. 2016. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *Quarterly Journal of Economics* 131: 1973–2042. <https://doi.org/10.1093/qje/qjw025>.
- Janzen, S. A., N. P. Magnan, S. Sharma, and W. M. Thompson. 2018. Short-term impacts of a pay-it-forward livestock transfer and training program in Nepal. In *Proceedings of the 2018 ASSA Annual Meeting*, 422–425. Philadelphia: American Economic Association.

About the authors

Benjamin Schwab is an associate professor of agricultural economics at Kansas State University.

Sarah Janzen is an assistant professor of agricultural and consumer economics at the University of Illinois at Urbana Champaign.

Nicholas P. Magnan is an associate professor of agricultural and applied economics at the University of Georgia.

William M. Thompson is an economist at ID Insight.