latebalance — Check balance after lateffects estimation+

+This command is part of StataNow.

Description

The latebalance postestimation commands produce diagnostic statistics, test statistics, and diagnostic plots to assess whether a lateffects command balanced the covariates over treatment-assignment levels. latebalance is not available after lateffects balancing.

Syntax

latebalance *subcommand* ... [, *options*]

subcommand	Description
summarize	compare means and variances in raw and balanced data
overid	overidentification test
density	kernel density plots for raw and balanced data

Remarks and examples

This entry provides an overview of the commands in latebalance. We recommend that you read this entry before proceeding to [CAUSAL] **latebalance summarize**, [CAUSAL] **latebalance overid**, or [CAUSAL] **latebalance density** for command-specific syntax and details.

A covariate is said to be balanced when its distribution does not vary over treatment-assignment levels.

Covariates are balanced in experimental data with full compliance because treatment assignment is independent of the covariates because of the study design. In contrast, covariates must be balanced by weighting or matching in observational data if treatment assignment is related to the covariates that also affect the outcome of interest.

The estimators implemented in lateffects use a model to make the potential outcomes independent of the treatment assignment conditional on covariates. If the model is well specified, it should balance the covariates. Balance diagnostic techniques and tests check the specification of the conditioning method used by a lateffects estimator.

latebalance implements three methods to check for balance after lateffects: latebalance summarize, latebalance overid, and latebalance density. latebalance overid implements a formal test, while the other two methods are exploratory diagnostic techniques.

Austin (2009, 2011) and Guo and Fraser (2015, sec. 5.52) provide introductions to covariate balance. Imai and Ratkovic (2014) derived a test for balance implemented in latebalance overid. For an illustration, see example 2 of [CAUSAL] **lateffects**.

Methods and formulas

Methods and formulas are presented under the following headings:

Introduction

For covariate x, we observe values $\{x_1, x_2, \ldots, x_N\}$. Define a treatment-assignment indicator variable for treatment levels 0 and 1 as $t_i \in \{0,1\}$, for $i = 1, \ldots, N$, and frequency weights as $\{w_1, w_2, \ldots, w_N\}$. The sample mean and variance of x for level t are

$$\widehat{\mu}_x(t) = \frac{\sum_i^N I(t_i = t) w_i x_i}{N_t} \qquad \text{and}$$

$$\widehat{\sigma}^2_x(t) = \frac{\sum_i^N I(t_i = t) w_i \left\{x_i - \widehat{\mu}_x(t)\right\}^2}{N_t - 1}$$

where $N_t = \sum_i^N w_i I(t_i = t)$, and

$$I(t_i = t) = \begin{cases} 1 & \text{if } t_i = t \\ 0 & \text{otherwise} \end{cases}$$

As shown in Austin (2011), the standardized differences for covariate x between level t and the control t_0 are computed as

$$\delta_x(t) = \frac{\widehat{\mu}_x(t) - \widehat{\mu}_x(t_0)}{\sqrt{\frac{\widehat{\sigma}^2_x(t) + \widehat{\sigma}^2_x(t_0)}{2}}}$$

The variance ratio is $\rho_x(t) = \{\widehat{\sigma}^2_x(t)\}/\{\widehat{\sigma}^2_x(t_0)\}$.

The kernel density estimates from `latebalance density` are computed by `kdensity` using the normalized instrument inverse propensity scores for each treatment level.

Testing the propensity-score model specification

We estimate the probability of treatment conditioned on a set of covariates with a propensity-score model. Imai and Ratkovic (2014) derive a test for whether the estimated propensity score balances the covariates. The score equations for parameters of the propensity-score model define an exactly identified generalized method of moments (GMM) estimator. Imai and Ratkovic (2014) use the conditions imposed by mean balance as overidentifying conditions. A standard GMM test for the validity of the overidentifying conditions is then a test for covariate balance. See [R] **gmm** for a discussion of this overidentifying test, which is known as Hansen's J test in the econometrics literature.

Here are the details about the score equations and the overidentifying balance conditions. Recall from *Methods and formulas* of [CAUSAL] **teffects aipw** that we have the first-order condition of the treatment model

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{s}_{\mathrm{tm},i}(\mathbf{x}_i, \widehat{\gamma}) = 0$$

For a two-level treatment-effects model with conditional treatment $\tilde{t}$ and control t_0, the score is

$$\mathbf{s}_{\text{tm},i}(\mathbf{x}_i,\gamma) = \frac{I\left(t_i=\tilde{t}\right)}{p(\mathbf{x}_i,\tilde{t},\gamma)}\frac{\partial p(\mathbf{x}_i,t,\gamma)}{\partial\gamma'} - \left\{\frac{I(t_i=t_0)}{1-p\left(\mathbf{x}_i,\tilde{t},\gamma\right)}\right\}\frac{\partial p\left(\mathbf{x}_i,\tilde{t},\gamma\right)}{\partial\gamma'}\Bigg|_{\gamma=\widehat{\gamma}}$$

The score reduces to

$$\mathbf{s}_{\text{tm},i}\left(\mathbf{x}_i,\widehat{\gamma}\right) = \left[\frac{I\left(t_i=\tilde{t}\right)-p\left(\mathbf{x}_i,\tilde{t},\gamma\right)}{p\left(\mathbf{x}_i,\tilde{t},\gamma\right)\left\{1-p\left(\mathbf{x}_i,\tilde{t},\gamma\right)\right\}}\right]\frac{\partial p\left(\mathbf{x}_i,\tilde{t},\gamma\right)}{\partial\gamma'}\Bigg|_{\gamma=\widehat{\gamma}}$$

The corresponding covariate balancing moment conditions are

$$\mathbf{w}_{\text{tm},i}(\mathbf{x}_i,\gamma) = \left[\frac{I\left(t_i=\tilde{t}\right)-p\left(\mathbf{x}_i,\tilde{t},\gamma\right)}{p\left(\mathbf{x}_i,\tilde{t},\gamma\right)\left\{1-p\left(\mathbf{x}_i,\tilde{t},\gamma\right)\right\}}\right]\mathbf{x}_i$$

We stack the moment conditions

$$\begin{aligned}\mathbf{g}_{\text{tm}}(\mathbf{X},\gamma) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\begin{array}{c}\mathbf{s}_{\text{tm},i}(\mathbf{x}_i,\gamma)\\ \mathbf{w}_{\text{tm},i}(\mathbf{x}_i,\gamma)\end{array}\right\}\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_{\text{tm},i}(\mathbf{x}_i,\gamma)\end{aligned}$$

The overidentified GMM estimator is then

$$\widetilde{\gamma} = \text{argmin}_{\gamma}\; N\,\mathbf{g}_{\text{tm}}(\mathbf{X},\gamma)'\,\mathbf{W}_{\text{tm}}(\mathbf{X},\gamma)^{-1}\,\mathbf{g}_{\text{tm}}(\mathbf{X},\gamma) \tag{1}$$

where

$$\mathbf{W}_{\text{tm}}(\mathbf{X},\gamma) = \frac{1}{N}\sum_{i=1}^{N}E_T\left\{\mathbf{g}_{\text{tm},i}(\mathbf{x},\gamma)\,\mathbf{g}_{\text{tm},i}(\mathbf{x},\gamma)'\right\}$$

and the expectation is taken with respect to treatment distribution. The weight matrix $\mathbf{W}_{\text{tm}}(\mathbf{X},\gamma)$ is computed explicitly (Imai and Ratkovic 2014), and (1), written as a maximization problem, is solved using `ml`.

Finally, Hansen's J statistic is evaluated at its minimum,

$$J = N\mathbf{g}_{\text{tm}}(\mathbf{X},\widetilde{\gamma})'\,\mathbf{W}_{\text{tm}}(\mathbf{X},\widetilde{\gamma})^{-1}\,\mathbf{g}_{\text{tm}}(\mathbf{X},\widetilde{\gamma})$$

and is asymptotically distributed as a χ^2 with degrees of freedom d,

$$d = \text{rank}\left\{\mathbf{W}_{\text{tm}}\left(\mathbf{X},\widetilde{\gamma}\right)\right\} - \text{rank}\left[\frac{1}{N}\sum_{i=1}^{N}E_T\left\{\mathbf{s}_{\text{tm},i}(\mathbf{x}_i,\widetilde{\gamma})\,\mathbf{s}_{\text{tm},i}\left(\mathbf{x}_i,\widetilde{\gamma}\right)'\right\}\right]$$

References

Austin, P. C. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine* 28: 3083–3107. https://doi.org/10.1002/sim.3697.

———. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46: 399–424. https://doi.org/10.1080/00273171.2011.568786.

Guo, S., and M. W. Fraser. 2015. *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: Sage.

Imai, K., and M. Ratkovic. 2014. Covariate balancing propensity score. *Journal of the Royal Statistical Society*, B ser., 76: 243–263. https://doi.org/10.1111/rssb.12027.

Jann, B. 2021. Relative distribution analysis in Stata. *Stata Journal* 21: 885–951.

Also see

[CAUSAL] **latebalance density** — Covariate balance density+

[CAUSAL] **latebalance overid** — Test for covariate balance+

[CAUSAL] **latebalance summarize** — Covariate-balance summary statistics+

[CAUSAL] **lateffects** — Local average treatment-effect estimation+