Module 2 – Regression and Prediction

CASE STUDY ACTIVITY TUTORIAL

Case Study 5: The Effect of Gun Ownership on Homicide Rates

MIT PROFESSIONAL EDUCATION
Digital Programs

Regression 3.4: Inference using Modern Nonlinear Regression Methods. Case Study

V. Chernozhukov
MIT

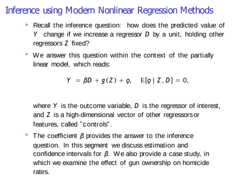
August 4, 2016

Inference using Modern Nonlinear Regression Methods

- Recall the inference question: how does the predicted value of Y change if we increase a regressor D by a unit, holding other regressors Z fixed?
- We answer this question within the context of the partially linear model, which reads:

$$Y = \beta D + g(Z) + \varrho, \quad \mathrm{E}[\varrho \mid Z, D] = 0,$$

  where Y is the outcome variable, D is the regressor of interest, and Z is a high-dimensional vector of other regressors or features, called "controls".
- The coefficient β provides the answer to the inference question. In this segment we discuss estimation and confidence intervals for β. We also provide a case study, in which we examine the effect of gun ownership on homicide rates.



In this segment we consider inference for the modern nonlinear regression. Recall the inference question: how does the predicted value of Y change if we increase a regressor D by a unit, holding other regressors Z fixed? Here we answer this question within the context of the partially linear model, which reads,

$$Y = \beta D + g(Z) + \varrho, \quad \mathrm{E}[\varrho \mid Z, D] = 0,$$

where Y is the outcome variable, D is the regressor of interest, and Z is a high-dimensional vector of other regressors or features, called "controls". The coefficient β provides the answer to the inference question. In this segment we discuss estimation and confidence intervals for β. We also provide a case study, in which we examine the effect of gun ownership on homicide rates.

- We can rewrite the model in the partialled-out form as:

$$\tilde{Y} = \beta\tilde{D} + \varrho, \quad \mathrm{E}(\varrho\tilde{D}) = 0, \tag{1}$$

where $\tilde{Y}$ and $\tilde{D}$ are the residuals left after predicting Y and D using Z, namely,

$$\tilde{Y} := Y - \ell(Z), \quad \tilde{D} := D - m(Z),$$

where $\ell(Z)$ and $m(Z)$ are defined as conditional expectations of Y and D given Z:

$$\ell(Z) := \mathrm{E}[Y \mid Z], \quad m(Z) := \mathrm{E}[D \mid Z].$$

Regression 3.4

- We can rewrite the model in the partialled-out form as:

$$\tilde{Y} = \beta\tilde{D} + \varrho, \quad \mathrm{E}(\varrho\tilde{D}) = 0, \tag{1}$$

where $\tilde{Y}$ and $\tilde{D}$ are the residuals left after predicting Y and D using Z, namely,

$$\tilde{Y} := Y - \ell(Z), \quad \tilde{D} := D - m(Z),$$

where $\ell(Z)$ and $m(Z)$ are defined as conditional expectations of Y and D given Z:

$$\ell(Z) := \mathrm{E}[Y \mid Z], \quad m(Z) := \mathrm{E}[D \mid Z].$$

We can rewrite the model in the partialled-out form as:

$$\tilde{Y} = \beta\tilde{D} + \varrho, \quad \mathrm{E}(\varrho\tilde{D}) = 0, \tag{2}$$

where $\tilde{Y}$ and $\tilde{D}$ are the residuals left after predicting Y and D using Z, namely,

$$\tilde{Y} := Y - \ell(Z), \quad \tilde{D} := D - m(Z),$$

where $\ell(Z)$ and $m(Z)$ are conditional expectations of Y and D given Z:

$$\ell(Z) := \mathrm{E}[Y \mid Z], \quad m(Z) := \mathrm{E}[D \mid Z].$$

The equation $\mathrm{E}\epsilon\tilde{D} = 0$ above is the Normal Equation for the population regression of $\tilde{Y}$ on $\tilde{D}$. This implies the following result:

Theorem (Frisch-Waugh-Lovell for Partially Linear Model)
The population regression coefficient β can be recovered from the population linear regression of $\tilde{Y}$ on $\tilde{D}$:

$$\beta = \arg\min_b \mathrm{E}(\tilde{Y} - b\tilde{D})^2 = (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}\tilde{D}\tilde{Y},$$

where β is uniquely defined if D can not be perfectly predicted by Z, i.e. $\mathrm{E}\tilde{D}^2 > 0$.

So β can be interpreted as a regression coefficient of *residualized* Y on *residualized* D, where the residuals are defined by taking-out the conditional expectation of Y and D given Z, from Y and D. Here we recall that the conditional expectations of Y and D given Z are best predictors of Y and D using Z.

Regression 3.4

The equation $\mathrm{E}\varrho\tilde{D} = 0$ above is the Normal Equation for the population regression of $\tilde{Y}$ on $\tilde{D}$. This implies the following result:

Theorem (Frisch-Waugh-Lovell for Partially Linear Model)
The population regression coefficient β can be recovered from the population linear regression of $\tilde{Y}$ on $\tilde{D}$:

$$\beta = \arg\min_b \mathrm{E}(\tilde{Y} - b\tilde{D})^2 = (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}\tilde{D}\tilde{Y},$$

where β is uniquely defined if D can not be perfectly predicted by Z, i.e. $\mathrm{E}\tilde{D}^2 > 0$.

So β can be interpreted as a regression coefficient of *residualized* Y on *residualized* D, where the residuals are defined by taking-out the conditional expectation of Y and D given Z, from Y and D. Here we recall that the conditional expectations of Y and D given Z are best predictors of Y and D using Z.

Estimation of β: The Procedure

- Our estimation procedure for β in the sample will mimic the partialling out procedure in the population.
- In order to avoid the possibility of overfitting we rely on sample splitting. We have data $(Y_i, D_i, Z_i)^n_{i=1}$. We randomly split the data into two halves: one half will serve as an auxilliary sample, which will be used to estimate the best predictors of Y and D, given Z, and then estimate the residualized Y and residualized D. Another half will serve as the main sample and will be used to estimate the regression coefficients.
- Let A denote the set of observation names in the auxiliary sample, and M the set of observations names in the main sample.

└─Estimation of β: The Procedure

Estimation of β: The Procedure

- Our estimation procedure for β in the sample will mimic the partialling out procedure in the population.
- In order to avoid the possibility of overfitting we rely on sample splitting. We have data $(Y_i, D_i, Z_i)_{i=1}^n$. We randomly split the data into two halves: one half will serve as an auxilliary sample, which will be used to estimate the best predictors of Y and D, given Z, and then estimate the residualized Y and residualized D. Another half will serve as the main sample and will be used to estimate the regression coefficients.
- Let A denote the set of observation names in the auxiliary sample, and M the set of observations names in the main sample.

Now we proceed to set up an estimation procedure for β. Our estimation procedure in the sample will mimic the partialling out procedure in the population.

In order to avoid the possibility of overfitting we rely on sample splitting. We have data $(Y_i, D_i, Z_i)^n_{i=1}$. We randomly split the data into two halves: one half will serve as an auxilliary sample, which will be used to estimate the best predictors of Y and D, given Z, and then estimate the residualized Y and residualized D. Another half will serve as the main sample and will be used to estimate the regression coefficients.

Let A denote the set of observation names in the auxiliary sample, and M the set of observations names in the main sample.

Estimation of β

Step 1: using auxiliary sample, we employ modern nonlinear regression methods to build estimators $\hat{\ell}(Z)$ and $\hat{m}(Z)$ of the best predictors $\ell(Z)$ and $m(Z)$. Then, using the main sample, we obtain the estimates of the residualized quantities:

$$\check{Y}_i = Y_i - \hat{\ell}(Z_i), \quad \check{D}_i = D_i - \hat{m}(Z_i), \quad \text{for each } i \in M,$$

and then using ordinary least squares of $\check{Y}_i$ on $\check{D}_i$ obtain the estimate of β, denoted by $\hat{\beta}^1$ and defined by the formula:

$$\hat{\beta}^1 = \arg\min_b \sum_{i \in M} (\check{Y}_i - b\check{D}_i)^2.$$

Step 2: we reverse the roles of the auxiliary and main samples, repeat Step 1, and obtain another estimate of β, denoted by $\hat{\beta}^2$.

Step 3: we take the average of the two estimates from Steps 1 and 2 obtaining the final estimate:

$$\hat{\beta} = \frac{1}{2}\hat{\beta}^1 + \frac{1}{2}\hat{\beta}^2.$$

Estimation of β

Step 1: using auxiliary sample, we employ modern nonlinear regression methods to build estimators $\hat{\ell}(Z)$ and $\hat{m}(Z)$ of the best predictors $\ell(Z)$ and $m(Z)$. Then, using the main sample, we obtain the estimates of the residualized quantities:

$$\check{Y}_i = Y_i - \hat{\ell}(Z_i), \quad \check{D}_i = D_i - \hat{m}(Z_i), \quad \text{for each } i \in M,$$

and then using ordinary least squares of $\check{Y}_i$ on $\check{D}_i$ obtain the estimate of β, denoted by $\hat{\beta}^1$ and defined by the formula:

$$\hat{\beta}^1 = \arg\min_b \sum_{i \in M} (\check{Y}_i - b\check{D}_i)^2.$$

Step 2: we reverse the roles of the auxiliary and main samples, repeat Step 1, and obtain another estimate of β, denoted by $\hat{\beta}^2$

Step 3: we take the average of the two estimates from Steps 1 and 2 obtaining the final estimate:

$$\hat{\beta} = \frac{1}{2}\hat{\beta}^1 + \frac{1}{2}\hat{\beta}^2.$$

Our algorithm proceeds in three steps.

In Step 1, using auxiliary sample, we employ modern nonlinear regression methods to build estimators $\hat{\ell}(Z)$ and $\hat{m}(Z)$ of the best predictors $\ell(Z)$ and $m(Z)$. Then, using the main sample, we obtain the estimates of the residualized quantities:

$$\check{Y}_i = Y_i - \hat{\ell}(Z_i), \quad \check{D}_i = D_i - \hat{m}(Z_i), \quad \text{for each } i \in M,$$

and then using ordinary least squares of $\check{Y}_i$ on $\check{D}_i$ obtain the estimate of β, denoted by $\hat{\beta}^1$ and defined by the formula:

$$\hat{\beta}^1 = \arg\min_b \sum_{i \in M} (\check{Y}_i - b\check{D}_i)^2.$$

In Step 2 we reverse the roles of the auxiliary and main samples, repeat Step 1, and obtain another estimate of β, denoted by $\hat{\beta}^2$.

In Step 3, we take the average of the two estimates from Steps 1 and 2 obtaining the final estimate: $\hat{\beta} = \frac{1}{2}\hat{\beta}^1 + \frac{1}{2}\hat{\beta}^2$.

Inference Result: Theory

Theorem (Inference)

If estimators $\hat{\ell}(Z)$ and $\hat{m}(Z)$ provide approximation to the best predictors $\ell(Z)$ and $m(Z)$ that is of sufficiently high quality, then the estimation error in $\check{D}_i$ and $\check{Y}_i$ has no first order effect on $\hat{\beta}$, and

$$\hat{\beta} \overset{a}{\sim} N(\beta, V/n)$$

where $V = (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}(\tilde{D}^2\varrho^2)(\mathrm{E}\tilde{D}^2)^{-1}$.

The above statement means that $\hat{\beta}$ concentrates in a $\sqrt{V/n}$-neighborhood of β, with deviations controlled by the normal law.

Precise definition of the term "sufficiently high quality" is given in the supplementary course materials.

Inference Result: Theory

Theorem (Inference)
If estimators $\hat\ell(Z)$ and $\hat m(Z)$ provide approximation to the best predictors $\ell(Z)$ and $m(Z)$ that is of sufficiently high quality, then the estimation error in $\check D_i$ and $\check Y_i$ has no first order effect on $\hat\beta$, and

$$\hat\beta \overset{a}{\sim} N(\beta, V/n)$$

where $V = (\mathrm{E}\tilde D^2)^{-1}\mathrm{E}(\tilde D^2\epsilon^2)(\mathrm{E}\tilde D^2)^{-1}$.

The above statement means that $\hat\beta$ concentrates in a $\sqrt{V/n}$-neighborhood of β, with deviations controlled by the normal law.

Precise definition of the term "sufficiently high quality" is given in the supplementary course materials.

If estimators $\hat\ell(Z)$ and $\hat m(Z)$ provide approximation to the best predictors $\ell(Z)$ and $m(Z)$ that is of sufficiently high quality, then the estimation error in $\check D_i$ and $\check Y_i$ has no first order effect on $\hat\beta$, and then the estimation error in the estimated residualized quantities have a negligible effect on $\hat\beta$, and $\hat\beta$ is approximately distributed as Normal variable with mean β and variance V/n, where the expression of the variance appears on the slide...
In words, we can say that the estimator $\hat\beta$ concentrates in a $\sqrt{V/n}$-neighborhood of β, with deviations controlled by the normal law...
We can now define the standard error of $\hat\beta$ as $\sqrt{\hat V/n}$, where $\hat V$ is an estimator of V.

Confidence Interval for $\hat{\beta}$

- The standard error of $\hat{\beta}$ is $\sqrt{\hat{V}/n}$, where $\hat{V}$ is an estimator of V.
- The result implies that the confidence interval

$$[\hat{\beta} - 2\sqrt{\hat{V}/n}, \hat{\beta} + 2\sqrt{\hat{V}/n}]$$

  covers β for most realizations of the sample, more precisely, approximately 95% of the realizations of the sample. Of course, if we replace 2 by other constants, we get other coverage probabilities.

- In other words, if our sample is not atypical, the interval covers the truth.

Confidence Interval for $\hat{\beta}$

- The standard error of $\hat{\beta}$ is $\sqrt{\hat{V}/n}$, where $\hat{V}$ is an estimator of V.
- The result implies that the confidence interval

$$[\hat{\beta} - 2\sqrt{\hat{V}/n}, \hat{\beta} + 2\sqrt{\hat{V}/n}]$$

  covers β for most realizations of the sample, more precisely, approximately 95% of the realizations of the sample. Of course, if we replace 2 by other constants, we get other coverage probabilities.
- In other words, if our sample is not atypical, the interval covers the truth.

The result implies that the confidence interval, given by the estimate plus/minus two standard errors, covers the true value β for most realizations of the data sample... more precisely, approximately 95% of realizations of the data sample. If we replace 2 by other constants, we get other coverage probabilities.
In other words, if our data sample is not extremely unusual, the interval covers the truth.

Selecting the Best Methods for Estimating Best Predictors $\ell(Z)$ and $m(Z)$

- In the above construction we used auxiliary sample A to estimate predictive models, using modern nonlinear regression methods. We can in principle use the main sample M as the validation/test sample to choose the best model for predicting Y and the best model for predicting D, following the procedures explained in the previous segment.
- We can also use the main sample M to aggregate the predictive models for Y and aggregate the predictive models for D, using least squares or lasso, following the procedures explained in the previous segment.

Given that we can use a wide variety of method for estimation of $\ell(Z)$ and $m(Z)$, it is natural to try to choose the best using sample splitting. In the above construction we used auxiliary sample A to estimate predictive models, using modern nonlinear regression methods. We can in principle use the main sample M as the validation/test sample to choose the best model for predicting Y and the best model for predicting D, following the procedures explained in the previous segment.

We can also use the main sample M to aggregate the predictive models for Y and aggregate the predictive models for D, using least squares or lasso, following the procedures explained in the previous segment.

Selecting the Best Methods for Estimating Best Predictors $\ell(Z)$ and $m(Z)$

- In the above construction we used auxiliary sample A to estimate predictive models, using modern nonlinear regression methods. We can in principle use the main sample M as the validation/test sample to choose the best model for predicting Y and the best model for predicting D, following the procedures explained in the previous segment.
- We can also use the main sample M to aggregate the predictive models for Y and aggregate the predictive models for D, using least squares or lasso, following the procedures explained in the previous segment.

Inference with Selecting the Best Methods for Estimating Best Predictors $\ell(Z)$ and $m(Z)$

Corollary

The previous inferential result continues to hold if the best or aggregated prediction rules are used as estimators $\hat{m}(Z)$ *and* $\hat{\ell}(Z)$ *of* $m(Z)$ *and* $\ell(Z)$ *in the algorithm we presented above. The required condition for this is that the number of rules we aggregate over or choose from is not "too large" relative to the overall sample size.*

A precise statement of the required condition is given in the supplementary course materials.

Inference with Selecting the Best Methods for Estimating Best Predictors $\ell(Z)$ and $m(Z)$

Corollary

The previous inferential result continues to hold if the best or aggregated prediction rules are used as estimators $\hat{m}(Z)$ and $\hat{\ell}(Z)$ of $m(Z)$ and $\ell(Z)$ in the algorithm we presented above. The required condition for this is that the number of rules we aggregate over or choose from is not "too large" relative to the overall sample size.

A precise statement of the required condition is given in the supplementary course materials.

The previous inferential result continues to hold if the best or aggregated prediction rules are used as estimators $\hat{m}(Z)$ and $\hat{\ell}(Z)$ of $m(Z)$ and $\ell(Z)$ in the algorithm we presented above. The required condition for this is that the number of rules we aggregate over or choose from is not "too large" relative to the overall sample size.

We provide a precise statement of the required condition in the supplementary course materials.

A Case Study: The Effect of Gun Ownership on Gun-Homicide Rates

- We consider the problem of estimating the effect of gun ownership on the homicide rate.
- For this purpose, we estimate the partially linear model:

$$Y_{j,t} = \beta D_{j,(t-1)} + g(Z_{j,t}) + \epsilon_{j,t}.$$

- $Y_{j,t}$ is log homicide rate in county j at time t, $D_{j,t-1}$ is log fraction of suicides committed with a firearm in county j at time $t-1$, which we use as a proxy for gun ownership, and $Z_{j,t}$ is a set of demographic and economic characteristics of county j at time t.
- The parameter β is the effect of gun ownership on the homicide rates, controlling for county-level demographic and economic characteristics.

A Case Study: The Effect of Gun Ownership on Gun-Homicide Rates

- We consider the problem of estimating the effect of gun ownership on the homicide rate.
- For this purpose, we estimate the partially linear model:

$$Y_{j,t} = \beta D_{j,(t-1)} + g(Z_{j,t}) + \epsilon_{j,t}.$$

- $Y_{j,t}$ is log homicide rate in county j at time t, $D_{j,t-1}$ is log fraction of suicides committed with a firearm in county j at time $t-1$, which we use as a proxy for gun ownership, and $Z_{j,t}$ is a set of demographic and economic characteristics of county j at time t.
- The parameter β is the effect of gun ownership on the homicide rates, controlling for county-level demographic and economic characteristics.

We next consider the case study, where we consider the problem of estimating the effect of gun ownership on the homicide rate. For this purpose, we estimate the partially linear model:

$$Y_{j,t} = \beta D_{j,(t-1)} + g(Z_{j,t}) + \epsilon_{j,t}.$$

Here $Y_{j,t}$ is log homicide rate in county j at time t, $D_{j,t-1}$ is log fraction of suicides committed with a firearm in county j at time $t-1$, which we use as a proxy for gun ownership, and $Z_{j,t}$ is a set of demographic and economic characteristics of county j at time t.
The parameter β is the effect of gun ownership on the homicide rates, controlling for county-level demographic and economic characteristics.

A Case Study: The Effect of Gun Ownership on Homicide Rates

- To account for heterogeneity across counties and time trends in all variables, we have removed from them county-specific and time-specific effects.
- The sample covers 195 large United States counties between the years 1980 through 1999, giving us 3900 observations.
- Control variables $Z_{j,t}$ are from the U.S. Census Bureau and contain demographic and economic characteristics of the counties such as the age distribution, the income distribution, crime rates, federal spending, home ownership rates, house prices, educational attainment, voting paterns, employment statistics, and migration rates.

A Case Study: The Effect of Gun Ownership on Homicide Rates

To account for heterogeneity across counties and time trends in all variables, we have removed from them county-specific and time-specific effects.
The sample covers 195 large United States counties between the years 1980 through 1999, giving us 3900 observations.
Control variables $Z_{j,t}$ are from the U.S. Census Bureau and contain demographic and economic characteristics of the counties such as the age distribution, the income distribution, crime rates, federal spending, home ownership rates, house prices, educational attainment, voting paterns, employment statistics, and migration rates.

A Case Study: The Effect of Gun Ownership on Homicide Rates

- To account for heterogeneity across counties and time trends in all variables, we have removed from them county-specific and time-specific effects.
- The sample covers 195 large United States counties between the years 1980 through 1999, giving us 3900 observations.
- Control variables $Z_{j,t}$ are from the U.S. Census Bureau and contain demographic and economic characteristics of the counties such as the age distribution, the income distribution, crime rates, federal spending, home ownership rates, house prices, educational attainment, voting paterns, employment statistics, and migration rates.

The Effect of Gun Ownership on Homicide Rates

- As a summary statistic we first look at simple regression of $Y_{j,t}$ on $D_{j,t-1}$ without controls. The point estimate is 0.282 with the confidence interval ranging from 0.17 to 0.39. This suggests that increases in gun ownership rates are related to gun homicide rates – if gun ownership increases by 1% relative to a trend then the predicted gun homicide rate goes up by .28%, without controlling for counties' characteristics.
- Since our goal is to estimate the effect of gun ownership after controlling for a rich set county characteristics we next include the controls and estimate the model by an array of the modern regression methods that we've learned.

The Effect of Gun Ownership on Homicide Rates

- As a summary statistic we first look at simple regression of $Y_{j,t}$ on $D_{j,t-1}$ without controls. The point estimate is 0.282 with the confidence interval ranging from 0.17 to 0.39. This suggests that increases in gun ownership rates are related to gun homicide rates – if gun ownership increases by 1% relative to a trend then the predicted gun homicide rate goes up by .28%, without controlling for counties' characteristics.
- Since our goal is to estimate the effect of gun ownership after controlling for a rich set county characteristics we next include the controls and estimate the model by an array of the modern regression methods that we've learned.

As a summary statistic we first look at simple regression of $Y_{j,t}$ on $D_{j,t-1}$ without controls. The point estimate is 0.282 with the confidence interval ranging from 0.17 to 0.39. This suggests that increases in gun ownership rates are related to gun homicide rates – if gun ownership increases by 1% relative to a trend then the predicted gun homicide rate goes up by .28%, without controlling for counties' characteristics.

Since our goal is to estimate the effect of gun ownership after controlling for a rich set county characteristics we next include the controls and estimate the model by an array of the modern regression methods that we've learned.

The Effect of Gun Ownership on Homicide Rates : Results

	Estimate	% 95 Confidence Interval
Least Squares (no controls)	0.282	[0.170 0.394]
Least Squares	0.227	[0.115 0.339]
Lasso	0.242	[0.124 0.360]
Post-Lasso	0.249	[0.133 0.365]
CV Lasso	0.189	[0.073 0.305]
CV Ridge	0.211	[0.099 0.323]
CV Elnet	0.197	[0.081 0.313]
Random Forest	0.252	[0.078 0.426]
Boosted Trees	0.190	[0.057 0.323]
Pruned Tree	0.152	[-0.013 0.317]
Neural Network	0.291	[0.081 0.501]
Best	0.244	[0.130 0.358]

Regression 3.4

The Effect of Gun Ownership on Homicide Rates : Results

The Effect of Gun Ownership on Homicide Rates : Results

	Estimate	% 95 Confidence Interval
Least Squares (no controls)	0.282	[0.170 0.394]
Least Squares	0.227	[0.115 0.339]
Lasso	0.242	[0.124 0.360]
Post-Lasso	0.249	[0.133 0.365]
CV Lasso	0.189	[0.073 0.305]
CV Ridge	0.211	[0.099 0.323]
CV Elnet	0.197	[0.081 0.313]
Random Forest	0.252	[0.078 0.426]
Boosted Trees	0.190	[0.057 0.323]
Pruned Tree	0.152	[-0.013 0.317]
Neural Network	0.291	[0.081 0.501]
Best	0.244	[0.130 0.358]

We present the results in a table, where the first column shows the method we used to estimate $m(Z)$ and $\ell(Z)$. The second column shows the estimated effect. And the final column shows the 95% confidence interval for the effect.

The Effect of Gun Ownership on Homicide Rates: Discussion

- The table shows the estimated effects of lagged gun ownership rate on the gun homicide rate as well as the 95% confidence bands for these effects.
- We first focus on the Lasso method: the estimated effect is about .25. This means that a 1% increase in gun ownership rate (as measured by the proxy) leads to a predicted quarter percent increase in gun homicide rates. The 95% confidence interval for the effect ranges from .12 to .36.
- Similar point estimates and confidence intervals are obtained by Least Squares Method. Random Forest also gives similar estimates to Lasso, however, confidence bands for this method are somewhat wider, covering the range from .07 to .42.

The Effect of Gun Ownership on Homicide Rates: Discussion

- The table shows the estimated effects of lagged gun ownership rate on the gun homicide rate as well as the 95% confidence bands for these effects.
- We first focus on the Lasso method: the estimated effect is about .25. This means that a 1% increase in gun ownership rate (as measured by the proxy) leads to a predicted quarter percent increase in gun homicide rates. The 95% confidence interval for the effect ranges from .12 to .36.
- Similar point estimates and confidence intervals are obtained by Least Squares Method. Random Forest also gives similar estimates to Lasso, however, confidence bands for this method are somewhat wider, covering the range from .07 to .42.

The table shows the estimated effects of lagged gun ownership rate on the gun homicide rate as well as the 95% confidence bands for these effects.

We first focus on the Lasso method: the estimated effect is about .25. This means that a 1% increase in gun ownership rate (as measured by the proxy) leads to a predicted quarter percent increase in gun homicide rates. The 95% confidence interval for the effect ranges from .12 to .36.

Similar point estimates and confidence intervals are obtained by Least Squares Method. Random Forest also gives similar estimates to Lasso, however, the confidence interval for this method is somewhat wider, covering the range from .07 to .42.

The Effect of Gun Ownership on Homicide Rates: Best Predictors

- The last row of the table provides the "best" estimates.
- To obtain "best" estimates we evaluate the performance of predictors $\hat{\ell}(Z)$ and $\hat{m}(Z)$ estimated by different methods on auxiliary samples using the main sample. Then we pick the methods giving the lowest MSE.
- In our case ridge regression and random forest give the best performances in predicting $Y_{j,t}$ and $D_{j,t-1}$, respectively. We then use the best methods as predictors in estimation procedure described above.
- The resulting estimate of the gun ownership effect is .24 and is similar to that of Lasso, and the confidence interval is somewhat tighter, now ranging from .13 to .35.

To obtain “best” estimates we evaluate the performance of predictors $\hat{\ell}(Z)$ and $\hat{m}(Z)$ estimated by different methods on auxiliary samples using the main sample. Then we pick the methods giving the lowest MSE.

In our case ridge regression and random forest give the best performances in predicting $Y_{j,t}$ and $D_{j,t-1}$, respectively. We then use the best methods as predictors in estimation procedure described above.

The resulting estimate of the gun ownership effect is .24 and is similar to that of Lasso, and the confidence interval is somewhat tighter, now ranging from .13 to .35.

The Effect of Gun Ownership on Homicide Rates: Best Predictors

- The last row of the table provides the “best” estimates.
- To obtain “best” estimates we evaluate the performance of predictors $\hat{\ell}(Z)$ and $\hat{m}(Z)$ estimated by different methods on auxiliary samples using the main sample. Then we pick the methods giving the lowest MSE.
- In our case ridge regression and random forest give the best performances in predicting $Y_{j,t}$ and $D_{j,t-1}$, respectively. We then use the best methods as predictors in estimation procedure described above.
- The resulting estimate of the gun ownership effect is .24 and is similar to that of Lasso, and the confidence interval is somewhat tighter, now ranging from .13 to .35.

Summary

- In this segment we discussed the use of modern nonlinear regression method for inference.
- The procedure relied on sample splitting in order to avoid overfitting, which may he hard to control theoretically.
- We applied these inference methods to the case study, where we estimated the effect of gun ownership rates on the homicide rates in the U.S., controlling for the counties' demographic and economic characteristics.

Summary

- In this segment we discussed the use of modern nonlinear regression method for inference.
- The procedure relied on sample splitting in order to avoid overfitting, which may he hard to control theoretically.
- We applied these inference methods to the case study, where we estimated the effect of gun ownership rates on the homicide rates in the U.S., controlling for the counties' demographic and economic characteristics.

In this segment we discussed the use of modern nonlinear regression method for inference.
The procedure relied on sample splitting in order to avoid overfitting, which may he hard to control theoretically.
We applied these inference methods to the case study, where we estimated the effect of gun ownership rates on the homicide rates in the U.S., controlling for the counties' demographic and economic characteristics.

Module 2 – Regression and Prediction

THANK YOU

Case Study 5: The Effect of Gun Ownership on Homicide Rates

PROFESSIONAL EDUCATION
Digital Programs