Using the lasso in Stata for inference in high-dimensional models

David M. Drukker

Executive Director of Econometrics
Stata

London Stata Conference
5-6 September 2019

Outline

1. What are high-dimensional models?
2. What is the lasso?
3. Using the lasso for inference

Using the lasso in applied statistics

- The least absolute shrinkage and selection operator (lasso) is a method
  - that produces point estimates for model coefficients and
  - can be used to select which covariates should be included in a model
- The lasso is used for problems of prediction and problems in statistical inference
  - I am going to focus on estimating and getting reliable inference for a parameter that has a causal interpretation

- Stata 16 has
  - `lasso` and `elasticnet` commands for prediction problems
  - Inferential lasso commands
    - `poregress`, `pologit`, `popoisson`, `poivregress`
    - `dsregress`, `dslogit`, `dspoisson`
    - `xporegress`, `xpologit`, `xpopoisson`, `xpoivregress`

Estimating the effect of no2_class

- I have an extract of the data Sunyer et al. (2017) used to estimate the effect air pollution on the response time of primary school children

$$\texttt{htime}_i = \texttt{no2\_class}_i\gamma + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$$

htime	measure of the response time on test of child *i* (hit time)
no2_class	measure of the pollution level in the school of child *i*
$\mathbf{x}_i$	vector of control variables that might need to be included

- I want to estimate the effect `no2_class` on `htime` and a confidence interval for the size of this effect
- There are 252 controls in $\mathbf{x}$, but I only have 1,036 observations
- This is a high-dimensional model
- I cannot reliably estimate γ if I include all 252 controls

Data

- Use extract of data from Sunyer et al. (2017)

```
. use breathe7, clear
. local ccontrols "sev_home sev_sch age ppt age_start_sch  oldsibl "
. local ccontrols "`ccontrols´ youngsibl no2_home ndvi_mn noise_sch"
.
. local fcontrols "grade sex lbweight lbfeed smokep "
. local fcontrols "`fcontrols´ feduc4 meduc4 overwt_who"
.
. local allcontrols "c.(`ccontrols´) i.(`fcontrols´) "
. local allcontrols "`allcontrols´ i.(`fcontrols´)#c.(`ccontrols´) "
```

Potential Controls II

```
. describe htime no2_class `fcontrols´ `ccontrols´
```

variable name	storage type	display format	value label	variable label
htime	double	%10.0g		ANT: mean hit reaction time (ms)
no2_class	float	%9.0g		Classroom NO2 levels (g/m3)
grade	byte	%9.0g	grade	Grade in school
sex	byte	%9.0g	sex	Sex
lbweight	float	%9.0g		1 if low birthweight
lbfeed	byte	%19.0f	bfeed	duration of breastfeeding
smokep	byte	%3.0f	noyes	1 if smoked during pregnancy
feduc4	byte	%17.0g	edu	Paternal education
meduc4	byte	%17.0g	edu	Maternal education
overwt_who	byte	%32.0g	over_wt	WHO/CDC-overweight 0:no/1:yes
sev_home	float	%9.0g		Home vulnerability index
sev_sch	float	%9.0g		School vulnerability index
age	float	%9.0g		Child´s age (in years)
ppt	double	%10.0g		Daily total precipitation
age_start_sch	double	%4.1f		Age started school
oldsibl	byte	%1.0f		Older siblings living in house
youngsibl	byte	%1.0f		Younger siblings living in house
no2_home	float	%9.0g		Residential NO2 levels (g/m3)
ndvi_mn	double	%10.0g		Home greenness (NDVI), 300m buffer
noise_sch	float	%9.0g		Measured school noise (in dB)

An estimate of the effect

```
. poregress htime no2_class, controls(`allcontrols´)
Estimating lasso for htime using plugin
Estimating lasso for no2_class using plugin
Partialing-out linear model                 Number of obs                 =      1,036
                                            Number of controls            =        252
                                            Number of selected controls   =         11
                                            Wald chi2(1)                  =      24.19
                                            Prob > chi2                   =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.354892	.4787494	4.92	0.000	1.416561	3.293224

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
```

Another microgram of NO2 per cubic meter increases the mean reaction time by 2.35 milliseconds.

Potential solutions

$$htime_i = no2\_class_i\gamma + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$$

- Suppose that $\tilde{\mathbf{x}}$ contains the subset of $\mathbf{x}$ that must be included to get a good estimate of γ for the sample size that I have
- If I knew $\tilde{\mathbf{x}}$, I could use the model

$$htime_i = no2\_class_i\gamma + \tilde{\mathbf{x}}_i\tilde{\boldsymbol{\beta}} + \epsilon_i$$

  - I am willing to assume the number of variables in $\tilde{\mathbf{x}}_i$ is small relative to the sample size
    - This is a sparsity assumption
  - The problem is that I don't know which variables belong in $\tilde{\mathbf{x}}$ and which do not

Potential solutions

- I don't need to assume that the model

$$htime_i = no2\_class_i\gamma + \tilde{\mathbf{x}}_i\tilde{\boldsymbol{\beta}} + \epsilon_i \tag{1}$$

  is exactly the "true" process that generated the data
- I only need to assume that the model (1) is sufficiently close to the model that generated the data
  - Approximate sparsity assumption

Covariate-selection problem

- Now I have a covariate-selection problem
  - Which of the 252 potential controls in $\mathbf{x}$ belong in $\tilde{\mathbf{x}}$?

Theory-based model selection

- The traditional approach would be to use theory to determine which covariates should be included
- Theory tells us to include controls $\check{\mathbf{x}}$
  - The selected controls do not vary in repeated samples
- Regress `htime` on `no2_class` and controls $\check{\mathbf{x}}$

$$htime_i = no2\_class_i\gamma + \check{\mathbf{x}}_i\tilde{\boldsymbol{\beta}} + \epsilon_i$$

  - Bad news:
    Estimate $\widehat{\gamma}$ can have large-sample bias, because theory picked the wrong controls
  - Good news:
    The standard error for $\widehat{\gamma}$ is reliable, because the covariates do not vary in repeated samples

lasso to the rescue

- Many researchers want to use data-based methods like the lasso or other machine-learning methods to perform the covariate selection
    - These methods should be able to remove the bias (possibly) arising from non-data-based selection of $\tilde{\mathbf{x}}$
- Some post-covariate-selection estimators provide reliable inference for the few parameters of interest

  Some do not

What's a lasso?

- The linear lasso solves

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ 1/n \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i \boldsymbol{\beta}'\right)^2 + \lambda \sum_{j=1}^{p} \omega_j |\boldsymbol{\beta}_j| \right\}$$

  where
  - $\lambda > 0$ is the lasso penalty parameter
  - $\mathbf{x}$ contains the p potential covariates
  - the ω_j are parameter-level weights known as penalty loadings
  - λ and the ω_j are called the lasso tuning parameters

What's a lasso?

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ 1/n \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i \boldsymbol{\beta}'\right)^2 + \lambda \sum_{j=1}^{p} \omega_j |\boldsymbol{\beta}_j| \right\}$$

- You obtain the (unpenalized) OLS estimates at $\lambda = 0$, when $p < n$
- As λ grows, the coefficient estimates get "shrunk" towards zero
- The kink in the absolute value function causes some of the elements of $\widehat{\boldsymbol{\beta}}$ to be zero at the solution for some values of λ
- There is a finite value of $\lambda = \lambda_{max}$ for which all the estimated coefficients are zero

What's a lasso?

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ 1/n \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i \boldsymbol{\beta}'\right)^2 + \lambda \sum_{j=1}^{p} \omega_j |\boldsymbol{\beta}_j| \right\}$$

- For $\lambda \in (0, \lambda_{max})$ some of the estimated coefficients are exactly zero and some of them are not zero.
  - This is how the lasso works as a covariate-selection method
    - Covariates with estimated coefficients of zero are excluded
    - Covariates with estimated coefficients that not zero are included

Tuning parameters

- λ and the ω_j are called "tuning" parameters
  - They specify the weight that should be applied to the penalty term
- The tuning parameters must be selected before using the lasso for prediction or model selection
- Plug-in methods, cross validation, and the adaptive lasso are used to select the tuning parameters
- Plug-in methods are the default methods for the inferential lasso commands

A naive lasso-based approach

- Now consider using lasso to solve the covariate selection problem in our high-dimensional model

$$\texttt{htime}_i = \texttt{no2\_class}_i\gamma + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$$

- A "naive" solution is :
  1. Always include the covariates of interest
  2. Use covariate-selection to obtain an estimate of which covariates are in $\tilde{\mathbf{x}}$
     Denote estimate by `xhat`
  3. Use estimate `xhat` as if it contained the covariates in $\tilde{\mathbf{x}}$
     `regress htime no2_class xhat`

Why naive approach fails

- Unfortunately, naive estimators that use the selected covariates as if they were $\tilde{\mathbf{x}}$ provide unreliable inference in repeated samples
  - Covariate-selection methods make too many mistakes in estimating $\tilde{\mathbf{x}}$ when some of the coefficients are small in magnitude
- If your model only approximates the functional form of the true model, there are approximation terms
  - The coefficients on some of the approximating terms are most likely small

Why the naive estimator performs poorly

- The random inclusion or exclusion of the covariates with small coefficients causes
  - the distribution of the naive post-selection estimator to be not normal
  - the usual large-sample theory approximation to be invalid in theory and unreliable in finite samples
- Long literature about problems with naive estimators
  - See Leeb and Pötscher (2005); Leeb and Pötscher (2006); Leeb and Pötscher (2008); and Pötscher and Leeb (2009)
  - See Belloni, Chernozhukov, and Hansen (2014a) and Belloni, Chernozhukov, and Hansen (2014b)

Partialing-out estimators

$$htime_i = no2\_class_i\gamma + \tilde{\mathbf{x}}_i\tilde{\boldsymbol{\beta}} + \epsilon_i$$

- A series of seminal papers

  Belloni, Chen, Chernozhukov, and Hansen (2012);
  Belloni, Chernozhukov, and Hansen (2014b);
  Belloni, Chernozhukov, and Wei (2016); and
  Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018)

  derived partialing-out estimators that provide reliable inference for γ after using covariate selection to determine which covariates belong in $\tilde{\mathbf{x}}$

  - The cost of using covariate-selection methods is that these partialing-out estimators do not produce estimates for $\tilde{\boldsymbol{\beta}}$

An estimate of the effect

```
. poregress htime no2_class, controls(`allcontrols´)
Estimating lasso for htime using plugin
Estimating lasso for no2_class using plugin
Partialing-out linear model                 Number of obs                 =      1,036
                                            Number of controls            =        252
                                            Number of selected controls   =         11
                                            Wald chi2(1)                  =      24.19
                                            Prob > chi2                   =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.354892	.4787494	4.92	0.000	1.416561	3.293224

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
```

Another microgram of NO2 per cubic meter increases the mean reaction time by 2.35 milliseconds.

Partialing-out estimator for linear model

- Consider model

$$y = d\gamma + \mathbf{x}\boldsymbol{\beta} + \epsilon$$

- For simplicity, d is a single variable, all methods handle multiple variables
- I discuss a linear model
  - Nonlinear models have similar methods that involve more details

PO estimator for linear model (I)

$$y = d\gamma + \mathbf{x}\boldsymbol{\beta} + \epsilon$$

1. Use a lasso of y on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_y$ that predict y
2. Regress y on $\tilde{\mathbf{x}}_y$ and let $\tilde{y}$ be residuals from this regression
3. Use a lasso of d on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_d$ that predict d
4. Regress d on $\tilde{\mathbf{x}}_d$ and let $\tilde{d}$ be residuals from this regression
5. Regress $\tilde{y}$ on $\tilde{d}$ to get estimate and standard error for γ

- Only the coefficient on d is estimated
- Not estimating $\boldsymbol{\beta}$ can be viewed as the cost of getting reliable estimates of γ that are robust to the mistakes that model-selection techniques make

PO estimator for linear model (II)

$$y = d\gamma + \mathbf{x}\boldsymbol{\beta} + \epsilon$$

1. Use a lasso of y on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_y$ that predict y
2. Regress y on $\tilde{\mathbf{x}}_y$ and let $\tilde{y}$ be residuals from this regression
3. Use a lasso of d on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_d$ that predict d
4. Regress d on $\tilde{\mathbf{x}}_d$ and let $\tilde{d}$ be residuals from this regression
5. Regress $\tilde{y}$ on $\tilde{d}$ to get estimate and standard error for γ

- This is an extension of the partialing-out method for obtaining the ordinary least squares (OLS) estimate for the coefficient and standard error on d (Also known as the result of the Frisch-Waugh-Lovell theorem)

$$y = d\gamma + \mathbf{x}\boldsymbol{\beta} + \epsilon$$

1. Use a lasso of y on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_y$ that predict y
2. Regress y on $\tilde{\mathbf{x}}_y$ and let $\tilde{y}$ be residuals from this regression
3. Use a lasso of d on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_d$ that predict d
4. Regress d on $\tilde{\mathbf{x}}_d$ and let $\tilde{d}$ be residuals from this regression
5. Regress $\tilde{y}$ on $\tilde{d}$ to get estimate and standard error for γ

- Heuristically, the moment conditions used in step 5 are unrelated to the selected covariates
- Formally, the moments conditions used in step 5 have been orthogonalized, or "immunized" to small mistakes in covariate selection
    - Chernozhukov, Hansen, and Spindler (2015a); and Chernozhukov, Hansen, and Spindler (2015b)

Double-selection estimators

$$y = d\gamma + \mathbf{x}\boldsymbol{\beta} + \epsilon$$

- Double-selection estimators extend the PO approach

1. Use a lasso of y on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_y$ that predict y
2. Use a lasso of d on $\mathbf{x}$ to select covariates $\tilde{\mathbf{x}}_d$ that predict d
3. Let $\tilde{\mathbf{x}}_u$ be the union of the covariates in $\tilde{\mathbf{x}}_y$ and $\tilde{\mathbf{x}}_d$
4. Regress y on d and $\tilde{\mathbf{x}}_u$
   The estimation results for the coefficient on d are the estimation results for γ

Double-selection estimators

- DS estimators include the extra control covariates that make the estimator robust to the mistakes that the lasso makes in selecting covariates that affect the outcome
- The DS estimator has two chances to find the relevant controls.
- Belloni et al. (2016) report that the DS estimator performed a little better than the PO in their simulations
- PO and DS have the same large-sample properties

```
. dsregress htime no2_class, controls(`allcontrols´)
Estimating lasso for htime using plugin
Estimating lasso for no2_class using plugin
Double-selection linear model          Number of obs                =      1,036
                                       Number of controls           =        252
                                       Number of selected controls  =         11
                                       Wald chi2(1)                 =      23.71
                                       Prob > chi2                  =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.370022	.4867462	4.87	0.000	1.416017	3.324027

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
. estimates store dsplugin
```

Another microgram of NO2 per cubic meter increases the mean reaction time by 2.37 milliseconds.

About the same as poregress estimate

Cross-fitting / Double-machine-learning PO

- Cross-fitting is also known as double machine learning (DML)
- It uses split-sample techniques on PO estimators
  - to weaken the sparsity condition
  - to get better finite sample performance
- Split-sample techniques further reduce the impact of covariate selection on the estimator for γ
- It's the combination of a sample-splitting technique with a PO estimator that gives cross-fit PO estimators their reliability
- These cross-fit PO (XPO) estimators are recommended over DS estimators and PO estimators
- Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) discusses

```
. xporegress htime no2_class, controls(`allcontrols´)
Cross-fit fold 1 of 10 ...
Estimating lasso for htime using plugin
Estimating lasso for no2_class using plugin
[Output Omitted]
Cross-fit partialing-out                Number of obs                 =      1,036
linear model                            Number of controls            =        252
                                        Number of selected controls   =         16
                                        Number of folds in cross-fit  =         10
                                        Number of resamples           =          1
                                        Wald chi2(1)                  =      27.31
                                        Prob > chi2                   =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.533651	.48482	5.23	0.000	1.583421	3.483881

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
```

Another microgram of NO2 per cubic meter increases the mean reaction time by 2.53 milliseconds.

About the same as poregress estimate

Choosing λ

- Recall that we must choose the tuning parameters λ and ω_j before using the lasso for model selection
- The value of the tuning parameters determines which covariates will be included and which will be excluded

Choosing λ

- Plug-in estimators find the value of the λ that is large enough to dominate the estimation noise
  - Plug-in-based lasso tends to include the important covariates and it is really good at not including covariates that do not belong in the model
- Cross validation (CV) selects the λ value that minimizes the out-of-sample mean squared error (MSE) of the predictions
  - CV is excellent at including the important covariates and but it tends to include many extra covariates that do not belong in the model
- The adaptive lasso is a multistep version of CV
  - The adaptive lasso is excellent at including the important covariates and but it tends to include some extra covariates that do not belong in the model

Choosing λ

- Including too many extra covariates can cause out $\{$PO,DS,XPO$\}$ estimator to perform poorly
  - (Including too many extra covariates slows the convergence rate of the $\{$PO,DS,XPO$\}$ estimator)

```
. dsregress htime no2_class, controls(`allcontrols´) selection(cv)      ///
>                 rseed(12345)
Estimating lasso for htime using cv
Estimating lasso for no2_class using cv
Double-selection linear model                 Number of obs               =      1,036
                                              Number of controls          =        252
                                              Number of selected controls =         36
                                              Wald chi2(1)                =      24.72
                                              Prob > chi2                 =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.523082	.5074363	4.97	0.000	1.528525	3.517639

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
. estimates store dscv
```

CV included 36 controls, while plug-in included 11 controls

```
. dsregress htime no2_class, controls(`allcontrols´) selection(adaptive) ///
>                 rseed(12345)
Estimating lasso for htime using adaptive
Estimating lasso for no2_class using adaptive
Double-selection linear model                 Number of obs               =      1,036
                                              Number of controls          =        252
                                              Number of selected controls =         26
                                              Wald chi2(1)                =      23.92
                                              Prob > chi2                 =     0.0000
```

htime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
no2_class	2.476892	.5064696	4.89	0.000	1.48423	3.469554

```
Note: Chi-squared test is a Wald test of the coefficients of the variables
      of interest jointly equal to zero. Lassos select controls for model
      estimation. Type lassoinfo to see number of selected variables in each
      lasso.
. estimates store dsadaptive
```

Adaptive included 26 controls, while plug-in included 11 controls, and CV included 36 controls

```
. lassoinfo dsplugin dscv dsadaptive
    Estimate:  dsplugin
     Command:  dsregress
```

Variable	Model	Selection method	lambda	No. of selected variables
htime	linear	plugin	.1375306	5
no2_class	linear	plugin	.1375306	6

```
    Estimate:  dscv
     Command:  dsregress
```

Variable	Model	Selection method	Selection criterion	lambda	No. of selected variables
htime	linear	cv	CV min.	9.129345	12
no2_class	linear	cv	CV min.	.280125	25

```
    Estimate:  dsadaptive
     Command:  dsregress
```

Variable	Model	Selection method	Selection criterion	lambda	No. of selected variables
htime	linear	adaptive	CV min.	11.90287	7
no2_class	linear	adaptive	CV min.	.0185652	20

Recommendations

- I provided lots of details, but here are some take always
  1. If you have time, use the cross-fit partialing-out estimator
     - `xporegress`, `xpologit`, `xpopoisson`, `xpoivregress`
  2. If the cross-fit estimator takes too long, use either the partialing-out estimator
     - `poregress`, `pologit`, `popoisson`, `poivregress`

     or the double-selection estimator
     - `dsregress`, `dslogit`, `dspoisson`
  3. Belloni, Chernozhukov, and Hansen (2014b) and Belloni, Chernozhukov, and Wei (2016) report simulations in which the DS estimator performed better than the PO estimator
  4. In simulations that I have run, the PO, DS, and XPO estimators perform better with plug-in than with CV or the adaptive lasso

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6): 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014a. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28(2): 29–50.

———. 2014b. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* 81(2): 608–650.

Belloni, A., V. Chernozhukov, and Y. Wei. 2016. Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics* 34(4): 606–619.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1): C1–C68.

Chernozhukov, V., C. Hansen, and M. Spindler. 2015a. Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments. *American Economic Review* 105(5): 486–90. URL `http://www.aeaweb.org/articles?id=10.1257/aer.p20151022`.

———. 2015b. Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach. *Annual Review of Economics* 7(1): 649–688.

Leeb, H., and B. M. Pötscher. 2005. Model Selection and Inference: Facts and Fiction. *Econometric Theory* 21: 21–59.

———. 2006. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics* 34(5): 2554–2591.

———. 2008. Sparse estimators and the oracle property, or the return of Hodges estimator. *Journal of Econometrics* 142(1): 201–211.

Pötscher, B. M., and H. Leeb. 2009. On the distribution of penalized

maximum likelihood estimators: The LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis* 100(9): 2065–2082.

Sunyer, J., E. Suades-Gonzlez, R. Garca-Esteban, I. Rivas, J. Pujol, M. Alvarez-Pedrerol, J. Forns, X. Querol, and X. Basagaa. 2017. Traffic-related Air Pollution and Attention in Primary School Children: Short-term Association. *Epidemiology* 28(2): 181–189.