FHHPS - Code Documentation

April 11, 2016

1 Overview

Main function definition

```
fhhps(Y1, Y2, X1, X2, Z1 = NULL, Z2 = NULL,
      shocks_bw = .1, mean_bw1 = .1, mean_bw2 = .1,
      cov_bw1 = .1, cov_bw2 = .1,
      mean_rcond_bnd = .1, cov_rcond_bnd = .1,
      q1_low = 0.01, q1_high = .99,
      q2_low = 0.00, q2_high = .98)
```

Required arguments

- `Y1, Y2, X1, X2`: $n \times 1$ `matrix` objects referring to the regressand and endogenous regressors at time 1 and 2.

Optional arguments

- `Z1, Z2`: $n \times k$ `matrix` objects referring to the exogenous regressors at time 1 and 2.
- `shocks_bw, mean_bw1, mean_bw2, cov_bw1, cov_bw2`: Set of scalar bandwidths, .
- `mean_rcond_bnd, cov_rcond_bnd`: Set of lower bounds to the reciprocal of the conditioning number of certain matrices.

- `q1_low, q1_high,q2_low, q2_high‘`: Lower and upper bounds on the quantiles of the estimated conditional moments. Values outside the thresholds are discarded.

Output A list consisting of:

- `shock_means`: $\begin{bmatrix} E[U_2] \\ E[V_2] \end{bmatrix}$
- `shock_variances`: $\begin{bmatrix} Var[U_2] & Var[V_2] \end{bmatrix}$
- `shock_covariance`: $\begin{bmatrix} Cov[U_2, V_2] \end{bmatrix}$
- `random_coeff_means`: $\begin{bmatrix} E[A_1] \\ E[B_1] \end{bmatrix}$
- `random_coeff_variances`: $\begin{bmatrix} Var[A_1] & Var[B_1] \end{bmatrix}$
- `random_coeff_covariance`: $\begin{bmatrix} Cov[A_1, B_1] \end{bmatrix}$
- `random_coeff_covariance`: $\begin{bmatrix} Cov[A_1, B_1] \end{bmatrix}$

2 Details

2.1 Algorithm

Shock moments The code following the Estimation section in the paper exactly.

$$\widehat{E}[U_2], \widehat{E}[V_2], \widehat{\beta}_1, \widehat{\beta}_2 = \arg \min_{\theta_U, \theta_V, \beta_1, \beta_2} \sum_i \left(\Delta Y_{i2} - \theta_U - \theta_V X_{i2} + Z'_{i1}\beta_1 - Z'_{i2}\beta_2\right)^2 K_{bw_0}(X_{i2} - X_{i2})$$

$$\begin{aligned} \widehat{Var}[U_2], \widehat{Var}[V_2], \widehat{Cov}[U_2, V_2] = \arg \min_{\theta_U, \theta_V, \theta_{UV}} \sum_i (\Big(\Delta Y_{i2} - \widehat{E}[U_2] - \widehat{E}[V_2] X_{2i} + Z'_{2i}\widehat{\beta}_2 - Z'_{1i}\widehat{\beta}_1\Big)^2 \\ -\theta_U - 2X_{i2}\theta_{UV} - \theta_V)^2 \cdot K_{bw_0}(X_{i2} - X_{i2}) \end{aligned}$$

The bandwidth of the kernel on the rightmost part of the equation is controlled by `shocks_bw`.

First moments of random coefficients Expanding on Bonhomme's suggestion. Let $\mathcal{I}_1 = \{X_1 = x_1, X_2 = x_2, Z = z\}$, and $\mathcal{I}_2 = \{X_1 = x_2, X_2 = x_2, Z = z\}$,

$$\begin{bmatrix} \widehat{E}[Y_1 \mid \mathcal{I}_1] \\ \widehat{E}[Y_2 \mid \mathcal{I}_1] - \widehat{E}[\Delta Y_2 \mid \mathcal{I}_2] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}}_{\Gamma_1(x_1,x_2)} \begin{bmatrix} \widehat{E}[A_1 \mid \mathcal{I}_1] \\ \widehat{E}[B_1 \mid \mathcal{I}_1] \end{bmatrix}$$

The nonparametric regressions in the formula above are computed using a simple leave-one-out Nadaraya-Watson regression, using the bandwidth parameter `mean_bw1` (for the terms involving $\mathcal{I}_1$) , and `mean_bw2` (for those invcolving $\mathcal{I}_2$)

When $x_1 \approx x_2$, the inverse $\Gamma_1(x_1, x_2)$ is numerically unstable, so the code offers the option to discard observations for which the reciprocal conditioning number[1] is small, by imposing a lower bound `mean_rcond_bnd`.

Additionally, the code also allows for the option of trimming out the upper and lower quantiles of the estimated conditional moments before taking their unconditional mean. This is controlled by `q1_low` and `q1_high`.

Second moments of random coefficients

$$\begin{bmatrix} \widehat{E}[Y_1^2 \mid \mathcal{I}_1] \\ \widehat{E}[Y_2^2 \mid \mathcal{I}_1] - \widehat{E}[\Delta Y_2^2 \mid \mathcal{I}_2] - 2\widehat{E}[A_1 + B_1 x_2 \mid \mathcal{I}_1]\widehat{E}[\Delta Y_2 \mid \mathcal{I}_2]) \\ \widehat{E}[Y_1 Y_2 \mid \mathcal{I}_1] - \widehat{E}[Y_1 \mid \mathcal{I}_1]\widehat{E}[\Delta Y_2 \mid \mathcal{I}_2]) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1^2 & 2x_1 \\ 1 & x_2^2 & 2x_2 \\ 1 & x_1 x_2 & x_1 + x_2 \end{bmatrix}}_{\Gamma_4(x_1,x_2)} \begin{bmatrix} \widehat{E}[A_1^2 \mid \mathcal{I}_1] \\ \widehat{E}[B_1^2 \mid \mathcal{I}_1] \\ \widehat{E}[A_1 B_1 \mid \mathcal{I}_1] \end{bmatrix} \tag{1}$$

Similarly to the case of first moments, the bandwidth used in the non-parametric regressions are `cov_bw1`, `cov_bw2`. Moreover, the numerical conditioning can be controlled by parameters `cov_rcond_bnd`, and the user can trim the lower and upper quantiles of the expected conditional expectation by `q2_low`, `q2_high`.

[1] The reciprocal conditioning number of a matrix is defined as the ratio between its smallest and singular. Numerically stable matrices have reciprocal conditioning numbers closer to one.

3 Example

Inside the file `functions.r`, you will also find `create_data`, a function that takes a number of observations `n_obs` as argument, and outputs a list of variables that conform to the assumptions of the model, namely:

- Random coefficients `A`, `B` corresponding to:

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right)$$

- Shocks `U`, `V`, corresponding to:

$$\begin{bmatrix} U_2 \\ V_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} .5 \\ .5 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

- Exogenous regressors `Z1`, `Z2`, each ($n\_obs \times 2$)-matrices corresponding to:

$$Z_t \sim \mathcal{N}\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \right)$$

- Endogenous regressors `X1`, `X2` (section 4 explains the reason for demeaning):

$$\begin{aligned} \tilde{X}_1 &\sim 0.2A_1^2 + 0.5A_1^2 + 0.2B_1 - 0.5B_1^2 + \mathcal{N}(0, \sqrt{5}) \\ X_1 &= \tilde{X}_1 - \widehat{E}[\tilde{X}_1] \\ X_2 &\sim \mathcal{N}(0, \sqrt{5}) \end{aligned}$$

- Regressand `Y1`, `Y2`:

$$\begin{aligned} Y_1 &= A_1 + B_1 X_1 \\ Y_2 &= A_1 + U_2 + (B_1 + V_2) X_2 \end{aligned}$$

The file `example.r` successively simulates the data using the create_data function and then applies `fhhps` to it, and outputs the result of each iteration to the file `sim_results.txt`.

4 Known issues

Algorithm works better when $E[X] \approx 0$ [2]

To see why this is the case, recall the step that estimates second moment of shocks involves regressing the dependent variable on X_2, X_2^2 and other variables.

$$\min \sum_i \left((\Delta Y_{2i} - c_1)^2 - \theta_U - 2\theta_{UV} X_{i2} - \theta_V X_{i2}^2\right) - c_2)^2$$

where c_1, c_2 are irrelevant for this explanation.

The problem with running that regression is best understood from the graphs on Figure 1. That is, as we move the support of X away from zero, the relationship between X and its square very quickly becomes nearly linear. This collinearity then drives down the performance of the estimator.

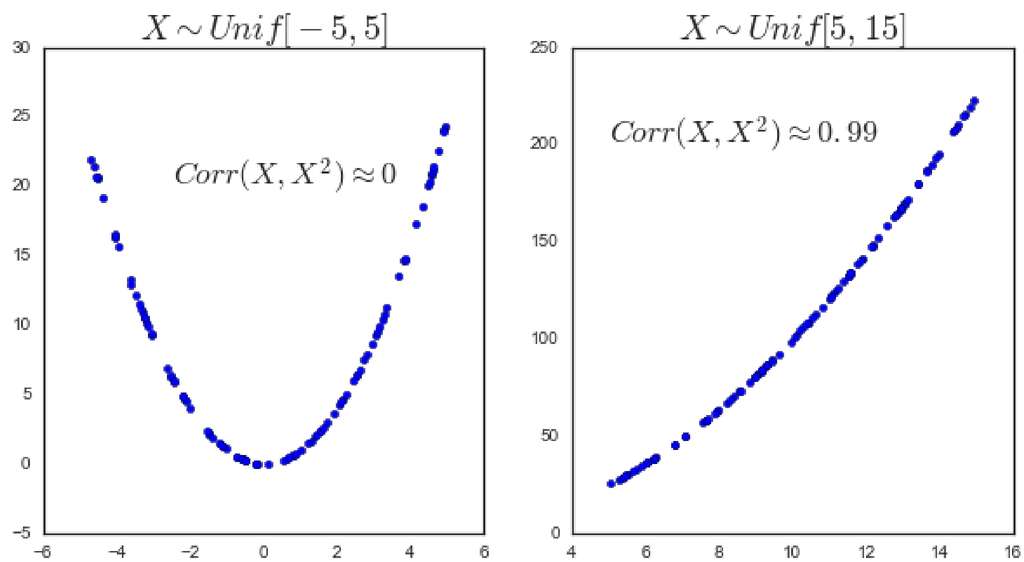

Figure 1: Variables X and X^2 become highly correlated when the support of X lies far from zero.

[2]Assuming a roughly symmetric distribution. Otherwise, the results in this section apply to the median.

When $\mu_{X_1} \neq \mu_{X_2}$, demeaning does not help Suppose that the model is given by

$$\begin{aligned} Y_1 &= A_1 + B_1 \tilde{X}_1 \\ Y_2 &= A_2 + B_2 \tilde{X}_2 \end{aligned}$$

where $E[\tilde{X}_j] \neq 0$. We can rewrite $\tilde{X}_1 = \mu_{X_1} + X_1$, where $E[X_1] = 0$. Similarly for $\tilde{X}_2$.

$$\begin{aligned} Y_1 &= A_1 + B_1 \tilde{X}_1 \\ &= A_1 + B_1 X_1 + B_1 \mu_{X_1} \\ &= \tilde{A}_1 + B_1 X_1 \qquad \text{where } \tilde{A}_1 := A_1 + B_1 \mu_{X_1} \end{aligned}$$

Similarly,

$$\begin{aligned} Y_2 &= A_2 + B_2 \tilde{X}_2 \\ &= A_2 + B_2 X_2 + B_2 \mu_{X_2} \\ &= \tilde{A}_2 + B_2 X_2 \qquad \text{where } \tilde{A}_2 := A_2 + B_2 \mu_{X_2} \end{aligned}$$

$$\begin{aligned} \Delta \tilde{A}_2 &= A_2 + B_2 \mu_{X_2} - A_1 + B_1 \mu_{X_1} \\ &= A_1 + U_2 + B_2 \mu_{X_2} - A_1 + B_1 \mu_{X_1} \\ &= U_2 + \underbrace{B_2 \mu_{X_2} - B_1 \mu_{X_1}}_{\star} \end{aligned}$$

The problem is that since B_1, B_2 are *not* independent of X_1, X_2, the $\star$ term cannot be independent of X_1, X_2 either. To conclude: if $\mu_{X_1} \neq \mu_{X_2}$, intercepts will not follow a random walk anymore