Discussion of "A constrained nonparametric regression analysis of factor-biased technical change and TFP growth at the firm level" by Verschelde, Dumont, Merlevede and Rayp

Daniel J. Henderson

University of Alabama

October, 2014

- Recognize that functional form misspecification leads to biases and hence inaccurate TFP measurements
- Want flexibility, but not at the cost of violating economic assumptions (for example, monotonicity)
- Use constrained kernel regression to examine factor-biased technical change and TFP growth

- 'Reject' Hicks-neutrality across different manufacturing sectors via data-driven bandwidth selection
- Find the marginal productivity of low-skilled labor diminishes over time
- Find that materials bias is substantial in several sectors
- Argue that technical change is materials-using and low-skill labor-saving due to off-shoring and inclusion in global value chain networks

Does our regression model follow directly from the theory or are we making big assumptions?

$$Y = F\left[K, L, A(t)\right] exp(u)$$

Writing this in per-worker terms leads to

$$\frac{Y}{L} = \frac{F\left[K, L, A(t)\right] exp(u)}{L}$$

When is this equal to

$$y = F\left[k, A(t)\right] exp(u)$$

Can we simply take logs of our production function?

$$Y = F\left[K, L, A(t)\right] exp(u)$$

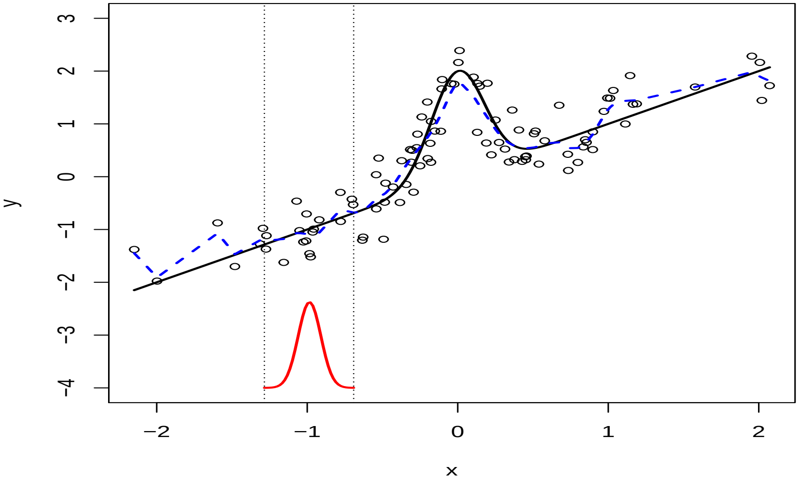
Taking logs of each side leads to

$$ln\left(Y\right) = ln\left\{F\left[K, L, A(t)\right] exp(u)\right\}$$

When is this equal to

$$ln\left(Y\right) = F\left\{ln\left(K\right), ln\left(L\right), ln\left[A(t)\right]\right\} + u$$

Using local information to construct the regression estimate



Cross-validation picks bandwidths such that the following is minimized

$$LSCV(\mathbf{h}) = \sum_{i=1}^{n} [y_i - \widehat{m}_{-i}(\mathbf{x}_i)]^2$$

Hall, Li and Racine (2007) show that if a (continuous) variable enters in linearly, then (asymptotically) cross-validation methods (for a local-linear regression) will give bandwidths equal to it's upper bound

Model	h_x	h_t
$y = x + t + u$	640902	7065746
$y = x + t + t^2 + u$	1650790	0.4504
$y = x + x^2 + t + u$	0.4377185	17949044
$y = x + x^2 + t + t^2 + u$	3125191	0.6440
$y = x + t + xt + u$	0.1763	5.7241
$y = x + t + \frac{500}{xt} + u$	4201154	95220477
Upper bound	1.9455	28.8906

- Hall, Li and Racine (2007) show that if a (continuous) variable is irrelevant, then (asymptotically) cross-validation methods (for a local-constant regression) will give bandwidths equal to it's upper bound
- What is rarely mentioned is that when an irrelevant variable is included in a local-linear regression, then cross-validation methods will also (asymptotically) give a bandwidth equal to it's upper bound

Model	h_x	h_t
LCLS		
$y = x + u$	0.2914	76063312
$y = x + x^2 + u$	0.1404	28835361
LLLS		
$y = x + u$	2576724	27650391
$y = x + x^2 + u$	0.4377	55067292
Upper bound	1.9455	28.8906

- It is impossible for the bandwidth on a continuous regressor to hit it's upper bound of infinity
- A large bandwidth is evidence (in a local-linear regression) that the regressor enters in linearly, but it is not a formal test
- A more formal method to test for linearity of a regressor (and hence, in this example, Hicks-neutrality) is to use statistical tests found in, for example, Du Parmeter and Racine (2013) or Li and Wang (1998)

- Du, Parmeter and Racine (2013) test for validity of restrictions placed on a nonparametric model. Here we could impose additive separability of the time variable: $ln(Y) = g\left[ln(X)\right] + m(t) + u$
- Li and Wang (1998) or the related test by Hsiao, Li and Racine (2007) test a restricted versus a completely nonparametric alternative: for example, we could estimate a partially linear model $ln(Y) = g\left[ln(X)\right] + m\left(t, \beta\right) + u$, where $m\left(t, \beta\right)$ is a (parametric) nonlinear function (with finite dimensional parameter β) and the model is estimated, say via Robinson (1988)

It is also arguable that one (or both) of these should be performed given that time is measured discretely in this application and thus should be considered as an ordered categorical variable. In a local-linear regression, cross-validation routines can only determine whether or not the discrete regressor is relevant, not whether or not it enters linearly (given the lack of Taylor expansions for discrete variables).

- I would like to see more (at least anecdotal) evidence on why some manufacturing sectors are impacted differently. For example, are food products more at risk from international competition than furniture?
- As a selling point, I would emphasize the materials bias result as this appears to be unique to the literature
- Off-shoring is a believable argument, but it is not backed by evidence in the paper (either citation or anecdotal) – this would strengthen the argument.
- How much off-shoring has happened in Belgium over the sample period? Is it really that large relative to the 10 years prior? I would really like to see a chart that shows this behavior over time?

- Are the translog estimates in the appendix constrained to have positive gradients?
- The claims (Table 3) that the distributions of types of workers have changed over time is testable: either via means or distributions.
- What is the distribution of returns-to-scale estimates? It is often the case that the average is near 1 (constant returns), but often the distribution varies from 1 significantly.
- Why are there no controls for sector or firm?