Using proxy variables to control for unobservables when estimating productivity: A sensitivity analysis
Van Beveren, Ilke
Ornaghi, Carmine

Total factor productivity (TFP) is an important tool for researchers in evaluating the implications of various policy measures on firm performance. However, obtaining reliable measures of firm-level TFP is not easy, since firm-level productivity is typically unobserved by the econometrician. One approach that has been used to tackle this problem consists in estimating productivity as a residual of a production function. While it is possible to estimate a production function using Ordinary Least Squares (OLS), this results in an endogeneity bias due to the fact that productivity is known to the firms when they choose their inputs. As a consequence, TFP estimation has itself been the topic of a large and still growing literature¹. Semiparametric estimation of production function, where unobserved productivity is approximated using observable firm-level variables, has been particularly popular in recent years. It is on this class of estimators that the current paper focuses. The first semiparametric estimator was introduced by Olley and Pakes (1996). They address the simultaneity issue by developing a two-step estimator of a production function, whereby firm's (observed) investment was used to proxy for its unobserved productivity. Since its publication, their estimator has been applied in many papers (see for instance De Loecker (2007), Konings and Vandenbussche (2009) and Pavcnik (2002)). A citation analysis in Google Scholar in November 2010 revealed 1,667 references to the original paper of Olley and Pakes (1996). Several adaptations and extensions to the original estimator of Olley and Pakes (1996) have been developed meanwhile. Levinsohn and Petrin (2003) use intermediate inputs (e.g. materials and energy) rather than investment to proxy for unobserved productivity. The semiparametric estimator of Levinsohn and Petrin (2003) has also extensively been applied by applied econometricians, for instance Javorcik and Spatareanu (2008) and Fernandes (2007), among others. More recently, the timing assumptions underlying the semiparametric estimators of Olley and Pakes (1996) and Levinsohn and Petrin (2003) have been questioned by Akerberg et al. (2006) who suggest an alternative two-step estimator where all relevant parameters are recovered in the second stage (see Konings and Vanormelingen (2009) and Dumont et al. (2010) for some recent applications). Wooldridge (2009) on the other hand focuses on the inefficiencies associated with the two-step estimation procedure of existing methodologies and proposes a framework in which production function estimates can be obtained in one-step. His framework allows for the timing assumptions of both the original semiparametric estimators and of the adapted framework of Akerberg et al. (2006) and it has been used by Acharya and Keller (2009). The present paper focuses on the class of semiparametric estimators that have emerged since the seminal paper of Olley and Pakes (1996). We aim to contribute to the literature in two important ways. First we provide a structured overview of these estimators, their assumptions and their practical implementation. Akerberg, Benkard, Berry and Pakes (2007) (henceforth ABBP) provide an excellent summary of semiparametric estimators. But, in our view, there are some aspects concerning the implementation of these estimators that need to be clarified, particularly for the one-step estimator proposed by Wooldridge (2009). Second, and more importantly, we want to evaluate how these estimators perform in practice, whether there are relevant differences between their estimates and, if so, to explore what cause those differences. Moreover, this is the first paper, to the best of our knowledge, that tries to test the validity of the theoretical assumptions on which these estimators are built. As we show in Section 4, the assumption that the unobserved productivity is monotonically increasing in investment or intermediary inputs is often rejected, thus questioning the foundation of these estimators.