

**T. Kirk White**  
**Abstracts of Selected Working Papers**  
(See <http://www.econ.duke.edu/~tkw2> for abstracts of other papers)

**The Dynamics of Plant-level Productivity in U.S. Manufacturing**  
**(with Árpád Ábrahám, job market paper)**

In this paper, we provide the best and broadest set of estimates to date of the persistence and variability of plant-level productivity in U.S. manufacturing. In addition to being of independent interest, these estimates are useful for calibrating an important class of macroeconomic models (see, for example, Jovanovic [1982], Cooley and Quadrini [2001], and Covas and Den Haan [2006]). Previous estimates were not as useful for calibrating macro models because they focused on a single industry or a small number of industries. Further, previous estimates of the persistence of plant-level productivity overestimate the level of persistence because: (i) they use only quinquennial data, instead of annual data; (ii) they do not allow for within-industry heterogeneity in the plant-level productivity process; and/or (iii) they do not address the endogeneity of inputs to production and the endogeneity of plant exit. We address all of these problems, and we provide plant-specific estimates for roughly 100,000 plants in 415 industries in the manufacturing sector.

We use the Census Bureau's confidential Longitudinal Business Database linked to the Annual Survey of Manufactures and the Census of Manufactures to estimate plant-level productivity for the entire U.S. manufacturing sector from 1976 until 1999. We estimate industry-specific production functions using a modification of the Olley and Pakes (1996) estimator to allow for within-industry heterogeneity in the plant-level productivity process while addressing the endogeneity of inputs to production and the endogeneity of plant exit. We separate idiosyncratic shocks from aggregate (economy wide) and industry shocks. Then we characterize the time series properties (persistence and volatility) of these plant-level idiosyncratic shocks. We use simulated method of moments estimation to control for short sample bias and to distinguish between sampling variance and underlying plant heterogeneity.

We find that plant-level productivity shocks are less persistent than the previous literature suggests. We also show that the persistence of plant-level productivity is considerably lower than persistence measures obtained from data aggregated to the industry level. We also show that using quinquennial data instead of annual data and/or not allowing for within-industry heterogeneity in plant-level persistence significantly increases estimates of the average level of persistence. We find significant within-industry heterogeneity in the persistence and volatility of idiosyncratic plant-level productivity. These findings are robust to various specifications of the production function.

## **Multiple Imputation in the Annual Survey of Manufactures (with Jerry Reiter, working paper)**

Identifying and dealing with imputed data are important problems for researchers using confidential U.S. Census Bureau microdata. Large percentages of Economic Census microdata are imputed. Failure to identify and properly address imputed data typically leads to biased estimates and deflated standard errors. This paper uses new data and new methods to identify and properly address missing or imputed data in the confidential Annual Survey of Manufactures (ASM) and Census of Manufactures (CMF) data.

Using new item-level edit flags in the ASM and CMF, we find that about 30 percent of observations on key variables from surveyed plants are imputed by the Census Bureau, and there is considerable variation in the imputation rates across industries and variables. We apply the sequential regression multivariate imputation method (Raghunathan et al., 2001) to multiply impute missing values. We find that industry means based on our multiple imputations tend to be closer to the true mean (based on complete data in which we've created artificially "missing" observations) than estimates from the data imputed using the Census Bureau's cold-deck method. Further, we find that our estimated confidence intervals tend to be wider than the confidence intervals from the single cold-deck imputed data, and our confidence intervals are more likely to cover the confidence intervals from the complete data. In future work we plan to analyze the effects of imputations on plant-level and aggregate productivity estimates.

## **Who Gentrifies Poor Neighborhoods? (with Terra McKinnish and Randall Walsh, working paper)**

This paper uses confidential 1990 and 2000 Decennial Census Long-Form data to study the demographic processes underlying the gentrification of poor urban neighborhoods during the 1990's. In contrast to previous studies, the analysis is conducted at the more refined census-tract level with a narrower definition of gentrification and more narrowly defined comparison neighborhoods. The analysis is also richly disaggregated by demographic characteristic, uncovering differential patterns by race, education, age and family structure that would not have emerged in the more aggregate analysis in previous studies. The results provide little evidence of displacement of poor non-white households in gentrifying neighborhoods. The bulk of the income gains in gentrifying neighborhoods are attributed to white college graduates and black high school graduates. It is the disproportionate in-migration of the former and the disproportionate retention and income gains of the latter that appear to be the main engines of gentrification.