

CROSS-COUNTRY DIFFERENCES IN PRODUCTIVITY: THE ROLE OF ALLOCATION AND SELECTION

By Eric Bartelsman, John Haltiwanger and Stefano Scarpetta*

23 August 2011

Abstract

This paper investigates the effect of idiosyncratic (firm-level) policy distortions on aggregate outcomes. Exploiting harmonized firm-level data for a number of countries, we show that there is substantial and systematic cross-country variation in the within-industry covariance between size and productivity. We develop a model in which heterogeneous firms face adjustment frictions (overhead labor and quasi-fixed capital) and distortions. The model can be readily calibrated so that variations in the distribution of distortions allow matching the observed cross-country moments. We show that the differences in the distortions that account for the size-productivity covariance imply substantial differences in aggregate performance.

* Respectively: VU University Amsterdam, Tinbergen Institute and IZA, Bonn (e.j.bartelsman@vu.nl); University of Maryland, U.S. Census Bureau, and NBER (haltiwan@econ.umd.edu); the OECD and IZA, Bonn (stefano.scarpetta@oecd.org). We are indebted for many useful comments on earlier drafts and presentations to Susanto Basu, Mary Hallward-Driemeier, John Fernald, Penny Goldberg, Chang-Tai Hsieh, Pete Klenow, Ariel Pakes, Amil Petrin, John Sutton, and two anonymous referees, as well as to the participants of the World Bank Conferences on the "Microeconomics of Growth" and on "Firm Dynamics" and the NBER Summer Institute. We thank the International Growth Centre, the World Bank, EU COST, and the OECD for financial support. We thank Ana Abras and Brock Williams for excellent research assistance. The views expressed in this paper are those of the authors and should not be held to represent those of the OECD or its Member countries.

A vast theoretical and empirical literature has been devoted to identify the sources of the large and persistent differences in productivity across countries. At the same time, a parallel strand of research has emerged over the past decade suggesting large and persistent heterogeneity in firm-level productivity, even in narrowly-defined industries, in a variety of countries (e.g. Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta (2004)).

A few recent papers have tried to combine these two strands of research by shedding light on how cross-country differences in economic outcomes relate to differences in the within-industry productivity dispersion across firms. For example, recent papers explore the interplay between the heterogeneity in firm-level productivity, the business environment and aggregate economic performance (see e.g. Diego Restuccia and Richard Rogerson (2008), Chang-Tai Hsieh and Pete Klenow, (2009), Laura Alfaro, Andrew Charlton, and Fabio Kanczuk (2008), and Virgilio Midrigan and Daniel Yi Xu (2010)). A common element of this emerging literature is that heterogeneity in firm-level productivity performance may indicate misallocation of resources across firms with negative effects at the aggregate level. However, an open question is which measures of firm-level heterogeneity are most instructive for detecting possible misallocation. Hsieh and Klenow (2009) argue, for example, that higher dispersion of productivity across firms within a given industry in China and India relative to the U.S. reflects greater misallocation of resources in China and India.

Our paper provides a contribution to this line of research along two broad lines. First, we argue, both theoretically and empirically, that the within-industry covariance between size and productivity is a robust measure to assess the impact of misallocation. Second, we extend recent theoretical models to allow for distortions to play a role not only in the allocation of resources

amongst existing firms but also in the selection of firms in the market and in the degree of firm churning.

The motivation for our approach is that the empirical evidence from firm-level data has shown that the widespread heterogeneity in firm-level performance is accompanied by substantial heterogeneity in the size of firms, even within narrowly-defined industries. Moreover, consistent with core models of the size distribution of firms (e.g., Robert Lucas (1978) and Marc Melitz (2003)), there is evidence in the firm-level data that the distributions of productivity and size exhibit a positive correlation – that is to say, more productive firms tend to be larger than less productive ones. However, our cross-country data suggest that there is considerable variation in the strength of the link between productivity and size across countries, industries, and over time. The working hypothesis of this paper is that policy-induced distortions may be the source of the observed variation in this covariance across countries and that this contributes to explain observed differences in aggregate performance.

This misallocation hypothesis is not new (see, e.g., the handbook paper by Abhijit Banerjee and Esther Duflo (2005) for a review) but the development of firm-level databases in a variety of countries now permits exploring this hypothesis more directly. In this paper, we explore the misallocation hypothesis using a variety of moments drawn from a harmonized firm-level database for the U.S. and a number of European countries, including transition economies of Central and Eastern Europe. Our focus is to assess the extent to which distortions can account for the observed differences across countries and over time in the within-industry productivity dispersion, as discussed by Hsieh and Klenow (2009), and in our preferred measure of the covariance between productivity and size. Further, we emphasize that distortions not only affect the allocation of resources across firms, but also the selection of firms producing in each market.

To quantify the within-industry covariance between size and productivity, we use an established empirical decomposition of the *level* of industry productivity as proposed by Steven Olley and Ariel Pakes (hereafter OP, 1996). The OP decomposition splits an index of industry-level productivity, defined as the weighted average of firm-level (log-) productivity, into an *unweighted firm-level average* and a *covariance term*. The covariance term is a summary measure of the *within-industry* cross sectional covariance between size and productivity. In our analysis, we find that the OP covariance term for labor productivity averages about 50 log points within U.S. manufacturing industries. In an accounting sense, this implies that the industry index of labor productivity in the average U.S. manufacturing industry is 50 percent higher than it would be if employment shares were randomly allocated within industries. However, the OP covariance term only reaches 20-30 log points in Western Europe and it was close to zero, if not negative, in Central and Eastern European countries at the beginning of their transition to a market economy. Quite remarkably, however, in these latter countries the covariance term increased substantially in the 1990s as their transition to a market economy progressed.

The use of the OP covariance term to explore the role of market distortions is, of course, not new. In their seminal contribution, Olley and Pakes found that the covariance term (using a decomposition of industry Total Factor Productivity, TFP) increased substantially in the U.S. telecommunications equipment industry following the deregulation of the sector in the early 1980s. They argued that this was because the deregulation permitted outputs and inputs to be reallocated more readily from less productive to more productive U.S. firms.

We focus our analysis on three moments of the firm-level distributions, namely the within-industry standard deviations of labor productivity and of total factor productivity, and the within-industry covariance between labor productivity and employment shares. The individual

moments, as well as the relationship between them, prove instructive in assessing the role of misallocation and this class of models. We find that the within-industry dispersion of labor productivity is larger than the within-industry dispersion of total factor productivity.¹ Our database of harmonized moments shows that this finding is robust across the countries, industries and time periods studied.

This finding is difficult to reconcile with many of the standard models in the literature, insofar as they have specific features that do not allow for any dispersion in labor productivity. In particular, the production function is often assumed to be Cobb-Douglas or, more generally, has the property that the *average* product of labor is proportional to the *marginal* product of labor. Moreover, many models make assumptions so that profit-maximizing firms equate the marginal revenue product of labor to the market wage. These two assumptions together imply that there should be no dispersion in labor productivity within industries (in the absence of distortions), even if there is significant dispersion in physical TFP. Our model includes frictions such as overhead labor and quasi-fixed capital that allow for the dispersion in labor productivity to be larger than dispersion in TFP, even in the absence of distortions.

In accounting for these features, we are careful to distinguish between physical and revenue-based measures of productivity in the model and the data, as emphasized in the recent literature (Lucia Foster, Haltiwanger and Chad Syverson (2008) and Hsieh and Klenow (2009)). In particular, the frictions in our model imply that even in the absence of distortions there will be substantial dispersion in revenue-based measures of labor productivity and total factor productivity and that each of these alternative measures of productivity will be positively correlated with each other. A further implication is that the core prediction from models of firm

¹ Syverson (2004b) reports that, within narrowly-defined industries, the difference in the U.S. between the 90th and the 10th percentiles of the firm-level productivity distributions is about 99 log points for total factor productivity (TFP) and about 140 log points for labor productivity.

heterogeneity of a positive covariance between physical output and physical productivity extends to predictions of a positive covariance between measures of size and productivity using the revenue-based measures of productivity.

Furthermore, our simple model allows comparing and contrasting the effect of distortions on different moments in the data. For example, Hsieh and Klenow (2009) present evidence that dispersion in revenue-based measures of TFP is higher in China and India than in the United States. They use the quantitative variation in such measures of dispersion to back out the implied distortions that can account for these patterns. In our setting, we have similar measures of dispersion in revenue-based TFP for eight countries, but we also have measures of the dispersion of revenue-based labor productivity as well as the covariance between firm size and revenue labor productivity for these countries.

To preview our main results, we find that our model can be readily calibrated to match the cross-country, within-industry patterns of the covariance of productivity and size, while it is more difficult to match the cross-country, within-industry productivity dispersion patterns observed in the data. As will become clear, the reason is that the cross-country productivity dispersion patterns in the data and the model are less systematic than the cross-country covariance patterns.

The paper proceeds as follows. Section I describes the harmonized firm-level database used in our empirical analysis and presents some basic facts about the within-industry productivity dispersion (TFP and labor productivity) as well as the empirical OP decompositions of productivity. In Section II, we develop our model of allocative efficiency with idiosyncratic distortions in the selection and allocation process. Section III calibrates the model numerically to explore its implications in light of the empirical patterns discussed in Section I. Section IV

reports the results from the numerical simulations of the impact of distortions on the key moments and explore the extent to which we can match the cross-country patterns of key moments observed in the data. Section V briefly summarizes the extensive sensitivity analysis we have conducted to assess the robustness of our results, while Section VI compares our results with those presented in recent literature. Section VII presents our concluding remarks.

I. The harmonized indicators of dispersion and covariance

For the empirical analysis in this paper, we use a database of harmonized industry-level moments drawn from firm-level data for five industrial economies and three transition economies of Central and Eastern Europe.² The database was constructed using firm-level data from business registers, social security and corporate tax rolls, and enterprise surveys.³ In the construction of our database, particular attention was devoted to harmonizing key concepts (e.g. the definition of the unit of measurement) as well as to using common methods to compute the moments. A detailed technical description of the dataset and the measures of outputs, inputs, deflators and productivity measures can be found in Bartelsman, Haltiwanger, and Scarpetta (2009) but we provide a brief overview here.

² The database covers as many as 24 countries (a downloadable version of the data can be found at http://econweb.umd.edu/~haltiwanger/BHS_jobflows_productivity/). However, given the focus of this paper on within-industry dispersion of labor and multi-factor productivity and on the OP covariance term, we restrict our attention to the U.S. and European (including transition) economies for which we have the required industry-level data. One of the countries that we include in the analysis (Germany) does not have the required firm-level multi-factor productivity measures but we include it nonetheless given its relevance in the cross-country comparison. France only supplied moments from firm-level data for the early 1990s. We also note that the underlying data are firm-level as opposed to establishment-level data. In principle, both are interesting since misallocation might occur across establishments within the same firms. However, in practice between-firm reallocation accounts for most of the overall between-plant reallocation (see, Haltiwanger, Ron Jarmin and Javier Miranda (2009)).

³ The database of indicators was assembled as part of long-term research projects sponsored by the OECD, the World Bank and Eurostat. The methodology for collecting the country/industry/time indicators built up from underlying micro-level datasets has been referred to as “distributed micro-data analysis” (Bartelsman et al. 2004). Harmonized code for creating the indicators were distributed to experts with access to the relevant data in each country.

The firm-level data of the eight countries underlying our analysis cover the manufacturing sector and include variables permitting consistent measurement of output, employment, materials and capital inputs. The measure of gross output is based on sales data at the firm-level deflated with an industry-level deflator (2- or 3-digit industries). The measure of labor input is based on the number of employees, while the measure of capital is based on book values, with further adjustments described in detail in Bartelsman, Haltiwanger and Scarpetta (2009). The measure of materials is based on nominal materials expenditures deflated with an industry-level materials deflator. Total factor productivity calculations use expenditure shares for labor, capital and materials. Since all gross output and materials are deflated with industry-level deflators, the measures of real labor productivity (LPR) and real multi-factor productivity (TFPR) are revenue-based measures.⁴

Three basic moments from these harmonized data are used in this paper: the within-industry standard deviation of log revenue labor productivity STD(LPR); the within-industry standard deviation of log revenue total factor productivity STD(TFPR); and a measure of within-industry covariance between size and productivity (OP). To compute the covariance measure we exploit a cross-sectional decomposition of industry-level productivity developed by Olley and Pakes (1996). They note that an index of productivity for an industry, defined as the weighted average of firm-level productivity, can be decomposed as follows:

$$\Omega_t = \sum_i \theta_{it} \omega_{it} = \bar{\omega}_t + \sum_i (\theta_{it} - \bar{\theta}_t) (\omega_{it} - \bar{\omega}_t)$$

⁴ The measure of TFPR we use in this paper is denoted as MFP in the Bartelsman, Haltiwanger and Scarpetta (2009) paper and dataset. We follow the terminology of Foster et al. (2008) and Hsieh and Klenow (2009) and use TFPR to denote TFP based on nominal revenue deflated with industry-level deflators. Another issue that has arisen in the literature is the most appropriate way to estimate factor elasticities. We recognize this is an issue but we think it is not critical in our context since we are primarily focused on dispersion in TFPR. Johannes Van Biesebroek (2004) has shown that basic properties of TFPR (like standard deviations) are robust to alternative estimation approaches.

where Ω_i is the industry index, ω_{it} is firm-level productivity, θ_{it} is the share of activity for the firm, and a ‘bar’ over a variable represents the un-weighted industry average of the firm-level measure. The industry index includes two terms: the un-weighted average of firm-level productivity and a covariance term that reflects the extent to which firms with higher than average productivity have a higher than average share of activity. This second term is the OP covariance measure used in our analysis, with log labor productivity at the firm level for ω_{it} and the firm’s labor share in the industry for θ_{it} .

We acknowledge that the three moments featured in this paper are quite simple. First, revenue labor productivity, used for computing STD(LPR) and OP, has the advantage of being based on measures that are widely available in firm-level datasets across countries. While it is computed as revenue per worker, and thus does not control for hours or quality of labor, we find that LPR captures useful and systematic variation.⁵ Next, the revenue-based productivity measures, LPR and TFPR, do not control for within-industry variation in firm-level prices. In our theoretical model, firm-level heterogeneity occurs in physical productivity (labelled TFPQ, as in Foster et al. (2008) and Hsieh and Klenow (2009)). In the model, frictions will yield a high, positive correlation between LPR, TFPR, and TFPQ, across firms in the same industry. The empirical evidence for the U.S. supports such high correlations. For example, Foster,

⁵ As a cross check on this latter point, we have used U.S. firm-level data to conduct some robustness analysis. In particular, we considered an alternative measure of revenue labor productivity defined as revenue per unit of payroll. The latter has the advantage that payroll incorporates hours worked and, to the extent that wages reflect worker quality, it also controls for labor quality. We found that the correlation between the measure of LPR we use in this paper and this alternative to be 0.82, across plants within the industries in U.S. manufacturing. Moreover, when we computed the within industry OP covariance measure for each of the more than 450 industries in U.S. manufacturing (on a 4-digit SIC basis), we found the correlation between the OP measure we use in this paper and the alternative based on revenue per unit payroll (using payroll weights) to be 0.82.

Haltiwanger and Syverson (2008) find a correlation of about 0.75 for TFPR and TFPQ and 0.6 for TFPQ and LPR in the U.S.⁶

Further, measurement (e.g., index number) problems generally affect cross-country comparisons of productivity. We avoid some of these problems by focusing on within-industry measures of dispersion and covariance. Moreover, in order to facilitate cross-country comparisons and, in particular, to remove the possible influence of differences in the industry composition of the manufacturing sector in the cross country comparisons, we construct manufacturing average indexes for each country aggregating industry-level indicators using a common set of industry weights. In particular, for all of the data analysis in this paper, we use time-invariant U.S. industry labor shares as weights to aggregate over industries.⁷

Table 1 shows the three moments of interest for the sample of eight countries used in this analysis. Several observations emerge from the table. First, the within-industry dispersion of both revenue labor productivity and revenue total factor productivity is large in all countries. Second, the within-industry dispersion of revenue labor productivity always exceeds the within-industry dispersion of revenue total factor productivity.⁸ Thus, the finding by Syverson (2004*b*) is pervasive across countries and industries. Third, the covariance term is positive for almost all countries but exhibits systematic and notable cross-country patterns: in particular, the covariance term is the highest in the U.S. manufacturing, it is much lower on average in the

⁶ Marcela Eslava, Haltiwanger, Adriana Kugler, and Maurice Kugler (2011) report that the correlation between TFPR and TFPQ within industries is 0.70 in Colombia.

⁷ The patterns are robust to using country-average industry or country-specific industry weights. See the web appendix for details.

⁸ Of the nearly 1000 industry and year observations for the set of countries under study, the dispersion in revenue labor productivity exceeds the dispersion in revenue total factor productivity in all but 15 cases.

Western European countries and even lower in the transition economies of Central and Eastern Europe.⁹

It is also instructive to explore the within-country variation over time of these moments. These moments are available for a number of years spanning the period 1992-2001 (except France where data are available through 1995). Table 2 presents the within-country changes in the moments comparing the average for the period 1996-2001 and the average for the period 1991-1995.

A number of observations emerge. First, the covariance term increased substantially in the transition economies over time, while the increases were generally less marked in Western Europe and, in particular, in the U.S. This is consistent with the view that the transition to a market-based system has allowed Central and Eastern Europe countries to improve their allocation of resources. Over this period, the U.S. had more stable market structures and economic institutions. Second, the dispersion measures are relatively stable over time in all countries.

Overall, the covariance term exhibits systematic cross-country variation in terms of both levels and changes, while the within-industry dispersion measures are all large but relatively stable over time. In addition, there is a systematic pattern that the dispersion of labor productivity exceeds the dispersion of total factor productivity in all countries. These are the key features of the firm-level data that we confront with the model analysis below.

II. A Model of Idiosyncratic Distortions, Selection and Allocation

⁹ A slight exception to this pattern is that Hungary and the U.K. have about the same covariance term. For the U.K., this may be associated with the rather weak performance of aggregate productivity performance in the early 1990s (see e.g. Scarpetta (2003)). However, a new round of data collection for EU countries for the period 1998-2005 shows that the OP covariance term in the U.K. reached about 0.35 in the most recent period.

To guide our analysis of distortions and allocative efficiency we develop a model that shares some common features with Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). In common with both models, we have production units with heterogeneous productivity that face idiosyncratic distortions. Further, the curvature in the profit function, necessary for an equilibrium of firms with heterogeneous productivity, combines decreasing returns to scale and a downward sloping demand curve associated with a differentiated product environment. However, the model differs from the recent literature in a number of key dimensions. In particular, the model is developed to capture an important fact observed in the data in all countries: the dispersion in revenue labor productivity within industries exceeds the dispersion in revenue total factor productivity, even in economies with little or no distortions. To match this feature of the data, we include quasi-fixed capital in the presence of transitory productivity shocks and overhead labor.

A. *The model*

Starting with the behavior of firms, we assume that firms produce according to a production function given by:

$$(1) \quad Y_{it} = A_i \varepsilon_{it} (n_{it} - f)^{\gamma - \alpha} k_{it}^\alpha, \quad \gamma < 1$$

where Y_{it} is output for firm i in period t , A_i is the firm-specific, time-invariant productivity component for firm i , k_{it} is the amount of capital input of firm i at time t , n_{it} is employment at time t , f is overhead labor,¹⁰ and ε_{it} is an *iid* shock drawn from a time invariant distribution and

¹⁰ Incorporating fixed overhead labor in this fashion is consistent with the theoretical literature on incorporating overhead in factors (see, e.g., Hugo Hopenhayn (1992), Phillippe Aghion and Peter Howitt, (1994), Julio Rotemberg and Michael Woodford, (1999). We discuss below the calibration of this parameter and its relationship to the literature.

observed each period after k is chosen and the decision to produce has been made.¹¹ We also allow for decreasing returns to scale, possibly related to some unobserved fixed factor – such as managerial ability, as in Lucas (1978). The decreasing returns hypothesis is one of the factors that insure that the most productive firm/manager does not take over the market. The overhead labor implies that the distribution of labor productivity is not degenerate even in an economy without distortions (i.e., while the marginal revenue product of labor will be set equal to the wage rate, the average product of labor will vary with scale, given overhead labor). Moreover, since capital is quasi-fixed, only labor will absorb the transitory shocks which, in turns, yields heterogeneity in the marginal revenue product of capital.

Firms face a downward demand schedule that arises from a differentiated products environment. The final good is assumed to be a CES aggregator of intermediate goods produced by the individual firms. The final goods sector is assumed to be perfectly competitive with the only inputs coming from intermediate goods. In particular:

$$Y_{it} = N_t^{(\rho-1)/\rho} \left(\sum_i Y_{it}^\rho \right)^{1/\rho}$$

where $\rho < 1$. This implementation of the CES aggregator includes an adjustment factor to make the degree of substitution scale-free, as in George Alessandria and Horag Choi (2007) where N is

¹¹ It is for model tractability that we consider permanent, firm-specific, along with completely transitory productivity shocks. Identification of fixed effects vs. transitory effects requires the use of panel econometric techniques and our reading of the literature is that no consensus has yet emerged on reasonable empirical estimates of this decomposition. However, we note that in our later calibration of the model we have checked the implied persistence of TFPR from the combined permanent and transitory shocks. We find that our calibration implies a one year persistence of TFPR of about 0.78 which is not far from the estimates in the literature (see, e.g., Syverson (2004b) and Foster, Haltiwanger and Syverson (2008)).

the number of intermediate firms in operation.¹² This implies that the inverse demand for good i is given by:

$$P_{it} = P_t(\bar{Y}_t/Y_{it})^{1-\rho}$$

where P_t is the aggregate price for the final good and \bar{Y}_t is average output measured as final output divided by N .

Firms producing the intermediate goods maximize profits, within an environment with distortions to nominal output, in each period given by:

$$(2) \quad \pi_{it} = (1 - \tau_i - \kappa_{it})P_t\bar{Y}_t^{1-\rho}[A_i\varepsilon_{it}(n_{it} - f)^\gamma k_{it}^\alpha]^\rho - w_t n_{it} - R_t k_{it}$$

where w_t is the wage paid to homogenous workers, and $R_t = r_t + \delta_t$, is the user cost of capital which equals the real interest rate plus the rate of depreciation. The firm-specific and time invariant disturbance to revenue, τ_i , and the firm-specific time varying disturbance, κ_{it} , can be interpreted broadly to include any distortion that impacts the scale of a business. The inclusion of idiosyncratic components of distortion is consistent with evidence that certain regulations apply *de jure* differently to firms of different size,¹³ whereas others are enforced unevenly across firms. Evidence from enterprise surveys (e.g. the World Bank Investment Climate Survey, ICA, see World Bank, 2004) indicate that there is a substantial idiosyncratic component in the market and

¹² As first noted by Jean-Pascal Benassy (1996) and discussed by Alessandria and Choi (2007), including this adjustment factor permits distinguishing between the love of variety effect and the impact of market power. In our modelling, we chose not to have the ‘selection effect’ change the degree of market power for the remaining firms. The welfare losses of distortions would be larger if utility were reduced through this channel. This is not critical for the main conclusions of the paper.

¹³ For example, regulations affecting the hiring and firing of workers only apply to firms above a certain threshold in a number of countries (see e.g. Danielle Venn, 2009).

policy-driven obstacles, such as start-up costs, or enforcement of regulations and taxes, faced by different firms.¹⁴

To make the model and analysis tractable, we assume a simple *ex ante* and *ex post* timing of information and decisions. Ex ante, before new firms enter, we assume that they do not know their productivity and distortion draws but they know the distribution of these idiosyncratic variables. New firms face a fixed cost of entry, c_e , that they have to pay to enter the market and to learn their draws from the joint *ex ante* distribution of productivity and distortions, $G(A, \tau)$. Once a firm learns its draws of A and τ , their values remain constant over time. In each period, the firm is subject to a further transitory idiosyncratic productivity shock from an ex-ante, known distribution, and a transitory idiosyncratic distortion shock that it learns after deciding whether to produce and choosing k .¹⁵

Firms discount the future at rate $1/(1+r)$ and face an exogenous probability of exiting in each period given by λ . Given free entry and the assumptions about the arrival of information, new firms enter up to the point where the expected discounted value of profits is equal to the entry fee. Moreover, given that the draws are time invariant in the steady state, the present discounted value for an incumbent firm i ex post is given simply by:

$$(3) \quad W(A_i, \tau_i) = E[\pi(A_i, \tau_i, \kappa_{it}, \varepsilon_{it})]/(1 - \chi), \quad \text{where: } \chi = (1 - \lambda)/(1 + r)$$

¹⁴ In many developing and emerging economies, labor and other regulations are *de facto* enforced to a different degree across firms of different size, sectoral affiliation etc. (see e.g. Gaëlle Pierre and Scarpetta, 2006; World Bank, 2004; Reyes Aterido, Mary Hallward-Driemeier and Carmen Pagés, 2007).

¹⁵ The fixed entry costs are denominated in output, which is reasonable since firms face these costs before starting production. They likely reflect a variety of costs including the time and effort of the entrepreneur but also the bureaucratic and other transactions costs (attorney fees, licensing and filing fees, and the like) for setting up a business. In terms of measurement of national accounts, costs that are paid in the market (attorney, licensing, filing fees) are captured as expenditures but likely not properly allocated to the correct sector for industry measures of productivity. Moreover, the time and effort of entrepreneurs in the nascent entrepreneur phase are unlikely to be well captured in national accounts and associated productivity statistics. The approach we take in the paper is consistent with much of the literature including Hopenhayn (1992), Hopenhayn and Rogerson (1993), and Restuccia and Rogerson (2008). In Hopenhayn (1992) there is a further discussion of this issue and possible alternatives.

In turn, the free entry condition is given by:

$$(4) \quad W^e = \int_{A,\tau} \max(0, W(A, \tau)) dG(A, \tau) - c_e$$

where c_e is the sunk entry cost. In evaluating (3), firms know their permanent draws and take expectations based on the distributions of the transitory shocks. In contrast, for (4) the firms don't know their permanent or temporary draws and take expectations over the distribution of all the disturbances.

New firms with a low productivity and/or a high-scale distortion draw will exit immediately upon learning their draws, if they cannot cover their operating costs. In what follows, we find that the share of firms that survive upon learning their productivity and distortion draws is an important factor for assessing the overall effects of distortions. The surviving subset, $M(A, \tau)$, is comprised of the firms that entered based on their knowledge of the distribution $G(A, \tau)$, and continued to operate owing to a positive present discounted value of operating profits, $W(A, \tau)$. Distortions thus influence the pace of churning of firms, which is captured in our model by the pace of entry (the number of firms deciding to pay the entry fee) and exit (the number of firms that exit upon learning their draws plus those hit by the exogenous exit shock). More importantly, the distortions and their potential correlation with idiosyncratic productivity affect which firms survive.

Conditional on survival, the distribution $M(A, \tau)$ and the equilibrium input prices will determine optimal firm-level capital input. In addition, given the optimal amount of capital, the labor choice will depend upon the realization of the transitory shocks. It is useful to start backwards within a period considering optimal employment for a given capital stock, which must satisfy:

$$(5) \quad \gamma p(1 - \tau_i - \kappa_{it})(P_i \bar{Y}_t^{1-\rho} [A_i \varepsilon_{it} k_{it}^\alpha]^\rho (n_{it} - f)^{\gamma p - 1}) = w_t$$

In turn the optimal capital stock must satisfy:

$$(6) k_{it} = \left[\frac{\alpha A_i^\rho P_t E \left(\bar{Y}_t^{1-\rho} (1-\tau_i - \kappa_{it}) (\varepsilon_{it} (n_{it} - f)^\gamma)^\rho \right)}{R_t} \right]^{1/(1-\alpha\rho)}$$

Output and profits for the operating firm are given by (1) and (2). Even though the firm is subject to transitory productivity and distortion shocks each period, the expected profits of the firm are the same every period and the optimal capital stock is the same every period. The firm adjusts to the *ex-post* transitory productivity and distortion draws, $\varepsilon_{it}, \kappa_{it}$, by adjusting employment. The firm is of course deciding whether to produce and with how much capital, conditional on the distribution of these transitory disturbances.

To close the model we must describe labor supply and the behavior of households and workers. A fixed number of households are assumed to supply labor inelastically so that aggregate labor supply is equal to N^s . Aggregate labor demand is given by the sum of labor demands for operating firms from (5). In equilibrium the number of firms and wages must satisfy both the free entry condition and the equality of labor demand and aggregate labor supply.

$$W^e = 0, \quad N_t^d = N_t^s$$

An aggregate resource constraint ensures that aggregate consumption plus resources spent on entry and depreciation will equal aggregate output in the stochastic steady state:¹⁶

$$C_t + E_t c_e + \delta K_t = Y_t$$

where K_t is the aggregate of capital of *ex-post* operating firms, and E_t is the mass of entry.

¹⁶ In steady state, gross investment is equal to replacement investment and net investment is zero.

The interest rate is pinned down by the production technology and utility maximization by the households. We assume a representative household that supplies labor inelastically and chooses consumption to maximize:

$$\sum_{t=0}^{\infty} \beta^t U(C_t)$$

Subject to the budget constraint:

$$\sum_{t=0}^{\infty} p_t (C_t + K_{t+1} - (1 - \delta)K_t) = \sum_{t=0}^{\infty} p_t (w_t N_t + R_t K_t + \pi_t)$$

where p_t is the time zero price of period t consumption, w_t and R_t are the period t rental prices of labor and capital measured relative to period t output, and π_t is the total profit from the operations of all plants. A standard result emerges from the first order conditions of this problem given by:

$$r_t = R_t - \delta = (1/\beta) - 1$$

So the real interest rate and rental cost of capital are pinned down by the discount factor for utility and the capital depreciation rate.

The model is deliberately simple so that steady-state inferences can be easily made. Model features that are at odds with the data include the capital stock being fixed over the firm's life, exit being exogenous after the decision to produce has been made, and the absence of business cycles or growth.

B. Allocation and Selection in our heterogeneous firm model

The model provides two different channels through which distortions affect output and welfare, namely resource allocation and firm selection. The first channel, resource allocation, is present in the models of Restuccia and Rogerson (2008), Hsieh and Klenow (2009) or Midrigan

and Xu (2010). In these models, allocation is distorted if the marginal revenue product of a production factor does not equal its marginal cost. Assuming wage rates to be constant across firms, Cobb-Douglas technology, and isoelastic demand, this would for example imply that in the absence of distortions the marginal revenue product of labor is equalized. In this environment, the average revenue product is proportional to the marginal revenue product so this also implies equalization of the average revenue product of labor. With an idiosyncratic revenue distortion added to this environment, the dispersion in the measured marginal and average revenue products will be directly proportional to the dispersion in the idiosyncratic tax/subsidy. In Hsieh and Klenow (2010), $STD(LPR)$ and $STD(TFPR)$ are both proportional to the dispersion in the idiosyncratic distortion.

In our model, LPR and TFPR both exhibit dispersion and are correlated with TFPQ even without distortions. Overhead labor, quasi-fixed capital and selection effects all play a role in influencing these patterns. With overhead labor, there is a wedge between marginal and average revenue products even in the absence of distortions. Further, overhead labor interacts with productivity heterogeneity and distortions to generate an endogenous exit threshold for firms.¹⁷

To help highlight the intuition for the role of overhead labor, it is useful to consider briefly a simplified version of our model without transitory productivity shocks (and thus no quasi-fixed capital). Specifically, consider the production function in equation (1), without the transitory productivity term, ε_{it} . In turn, consider the profit function of equation (2), again without the transitory productivity shock, and isolate the ‘profit shock’ that is a product of the idiosyncratic tax and the productivity shock: $S_i = (1 - \tau_i)A_i^\rho$. The capital and labor allocations

¹⁷ Overhead capital, or other per-period fixed operating expenses generate similar endogenous exit thresholds.

that satisfy the FOCs for profit maximization can be expressed as (where time subscripts have been suppressed for ease of exposition):

$$k_i^* = S_i^{1/(1-\rho\gamma)} K_x(w, \bar{Y}) \quad \text{and} \quad (n_i^* - f) = S_i^{1/(1-\rho\gamma)} \frac{R}{w} \frac{\gamma - \alpha}{\alpha} K_x(w, \bar{Y})$$

where K_x is a function of the equilibrium wage rate, average output, and parameters of the model that do not vary across firms. The equilibrium wage rate, average output, and the cutoff of S below which firms do not produce, S_i , can be solved for simultaneously. The equilibrium conditions imply that the flow profit at the cutoff S_i equals zero and the expected discounted profits of operating firms equal the entry fee.

The simple expressions for capital and labor yield a number of basic properties. First, factor inputs will be increasing functions of the profit shock. The implication is that in the absence of distortions, the firms with the highest TFPQ will have the highest output and inputs. It is this property that yields a positive OP covariance term if measured in terms of TFPQ and physical output or TFPQ and a composite input. The latter prediction is robust to a wide range of models. In our setting with overhead labor, such key predictions for TFPQ carry over to alternative measures of productivity such as LPR as we discuss now.

For the firms with idiosyncratic distortions and productivity above S_i , the following relation must hold:

$$\frac{P_i Y_i}{n_i^* - f} = \frac{1}{1 - \tau_i} \bar{Y}^{1-\rho} \left(\frac{R}{w} \frac{\gamma - \alpha}{\alpha} \right)^{\rho(\gamma - \alpha) - 1} K_x^{\rho\gamma - 1} = \frac{1}{(1 - \tau_i)} C \propto \frac{1}{(1 - \tau_i)}$$

Without idiosyncratic distortions, revenue per effective unit of labor is a constant, C , across firms. However, the moment of interest, LPR, will vary across firms, even without idiosyncratic distortions:

$$LPR_i = \frac{P_i Y_i}{n_i^*} = \frac{1}{(1 - \tau_i)} C - \frac{Cf}{(1 - \tau_i)n_i^*}$$

Once the model incorporates overhead labor, average revenue products of labor will vary across firms even without distortions. It is also evident that LPR will be increasing in firm size as measured by employment in the absence of distortions. This relationship has a number of related implications. First, in the absence of distortions, the positive relationship between LPR and size will yield a positive OP covariance using LPR and employment to measure shares. Second, there will be a positive correlation between TFPQ and LPR. It is also evident that distortions will have an impact on these relationships.

Because of overhead labor, a firm facing a low value of S will exit rather than pay the wage associated with f and incur negative flow profits. Without idiosyncratic distortions, the cutoff will occur only for firms with productivity below a given threshold. With idiosyncratic distortion, some highly productive firms with a bad distortion draw will exit, while some low productivity firms will be able to operate (leading to a misallocation of inputs). In other words, τ distorts selection as well. The threshold S_l is given by:

$$S_l = \frac{\bar{Y}^{1-\rho} R^{\rho\alpha} W^{1-\rho\alpha} f^{1-\rho\gamma}}{\alpha^{\rho\alpha} (\gamma - \alpha)^{\rho(\gamma-\alpha)} \rho^{\rho\gamma} (1 - \rho\gamma)^{1-\rho\gamma}}$$

The threshold depends on the equilibrium wage rate and average output, which are simultaneously determined. A higher dispersion in distortions will worsen selection based on productivity and will lower the overall efficiency of allocation of the operating firms.

III. Calibration of Benchmark “U.S.” Model

We set a number of key parameters of our model drawing from empirical evidence in the literature and choose the remaining benchmark parameters to match the U.S. moments from Table 1. For the purpose of this “benchmark” calibration, we make the assumption that the U.S. is a non-distorted economy---i.e. no idiosyncratic distortions ---and choose the parameters to match U.S. moments. In the analysis in the next sections, we assume that all countries face the

same technology, the same distribution of technology shocks, the same curvature parameters of the profit function and then seek to account for the cross-country variations observed in Tables 1 and 2 using variation in the distortion parameters alone.

In exploring the model simulations it is useful to note that there are a number of possible measures of firm-level productivity that are interesting to examine. The measure of physical TFP, TFPQ, is given in the model by the product $A_i \varepsilon_{it}$ with the permanent component of physical productivity given by A_i . The measure of revenue TFP, TFPR, and the associated measure of revenue labor productivity, LPR are also of interest. It is important to note that in our numerical analysis, the moments and decompositions we report are based on $\log(\text{TFPQ})$, $\log(\text{TFPR})$ and $\log(\text{LPR})$.¹⁸

We also note that in our numerical analysis of the theoretical model we consider an OP decomposition of revenue labor productivity that corresponds to what we measure in the data. That is, one of the OP decompositions we consider is based on the employment-share weighted average of firm-level log revenue labor productivity. Using the simulated moment that matches what we measure in the data permits us to benchmark to U.S. patterns and in turn seek to account for differences across countries on this moment. However, theoretically, it is also interesting to consider alternative moments capturing the covariance between size and productivity. In the simulations, we consider the OP covariance using TFPQ and physical output as well as the OP covariance using TFPR and an input weight.¹⁹

¹⁸ In the discussion that follows in the text, when we refer to TFPQ, TFPR and LPR we typically omit the reference to logs for expositional convenience but in all cases log based measures are used. In the web appendix, we include additional details about relating the model to the data moments.

¹⁹ In the appendix, we show and discuss results for alternatives. As we noted in section II.B, the most robust prediction is that in the absence of distortions higher TFPQ firms will have more physical output. In our model, this prediction carries over to other correlations/covariances such as those between LPR and employment, TFPQ and inputs, TFPR and revenue.

For our calibration of the non-distorted economy, we select the key parameters drawing from empirical evidence. In particular,²⁰

- $\gamma = 0.95$, (returns to scale – much of the evidence points towards close to constant returns to scale – see the discussion in Martin Baily, Charles Hulten and David Campbell (1992), Syverson (2004a), Foster, Haltiwanger and Syverson (2008)). We include some modest decreasing returns to scale since this has been used as an important source of curvature in the profit function in the literature (e.g., Restuccia and Rogerson (2008) and Cooper and Haltiwanger (2006)). However, as is apparent below, we also include curvature in the profit function through the demand structure. We need curvature from one source or the other but not necessarily both.
- $\alpha = 0.3$, (capital output elasticity) so the implied labor elasticity γ is 0.65 (these are in the range of standard values for these parameters).
- $\lambda = 0.10$, this is consistent with evidence of exit rates in the United States (see the Business Dynamic Statistics from the U.S. Census Bureau)²¹ and other OECD countries for businesses more than five years old (Bartelsman, Haltiwanger, and Scarpetta (2004) and $r = .02$, and $\delta = .10$, consistent with long-run real interest rates in OECD countries and typical depreciation rates from national accounts.
- $\rho = .8$, this is in the Christian Broda and David Weinstein (2006) range and implies a markup of 25 percent (which is broadly consistent with evidence from the cross-country empirical literature, e.g. Joaquim Oliveira-Martins and Scarpetta (2002)).

²⁰ We have conducted robustness analysis on each of these parameters and the findings in the paper are robust to modest variations in them.

²¹ See http://www/ces.census.gov/index.php/bds/bds_home.

The remaining parameters include the overhead labor parameter f , the entry cost c_e and the variances of the permanent and transitory productivity shocks. To pin these parameters down for the benchmark model, we use the information in Table 1 for the U.S. along with auxiliary information from the harmonized database on survival rates. Our model is not well suited to explore the full dynamics of young businesses, with rich selection and learning dynamics over the first part of their life (see, e.g., Foster, Haltiwanger and C.J. Krizan (2001, 2006)). However, our model has a simple form of endogenous selection of young businesses as entrants pay the entry fee, learn their productivity and distortion draws and decide whether or not to produce. To approximate this process, we use information on the exit rates of young businesses in the U.S. In our harmonized data, about 55 percent of businesses survive after five years in the U.S. This is about the same as found from other U.S. sources, such as the Business Dynamic Statistics from the U.S. Census Bureau. We match this moment in our calibration so that in the benchmark non-distorted economy, 55 percent of the businesses that pay the sunk entry cost survive. We then seek to match the U.S. moments in Table 1. Not surprisingly, the survival, dispersion and covariance moments are all closely connected. For example, setting a high overhead labor parameter yields lower survival rates as fewer firms can cover their fixed operating costs and, as we shall see, this selection effect impacts the dispersion and covariance moments.

In practice, we have found that we can match exactly the survival rate and one of the moments in Table 1 for the U.S. while we can respect the qualitative properties of the other remaining moments in Table 1 for the U.S.. The reason, as discussed in more detail below, is that the moments are nonlinear functions of key parameters. For our analysis, we have chosen to match exactly the OP covariance term so that in our benchmark calibration we have a survival rate of 55 percent and an OP covariance term of 0.51 (the value in Table 1 for the U.S.). For the

standard deviation of revenue total factor productivity (TFPR) the benchmark calibration has 0.47 compared to 0.38 for the U.S. in Table 1. For the standard deviation of revenue labor productivity the benchmark calibration has 0.75 compared to 0.58 for the U.S. in Table 1. Thus, while we do not obtain exact matches for the two moments of productivity dispersion, they have the same qualitative properties as those observed in the data, i.e. the standard deviation of revenue labor productivity exceeds the standard deviation of revenue total factor productivity as observed in the data.²² We show below that measurement error may account for the difficulty of matching dispersion moments.

Some other properties of the benchmark model are worth noting since they are relevant for the analysis of distortions. In the calibrated model, we also compute the OP covariance term for revenue total factor productivity (using a composite input weight with factor elasticities as weights) and obtain an OP covariance term for TFPR of 0.40. Likewise, we compute the OP covariance term using physical productivity in the model and obtain an OP covariance term of TFPQ of 0.91 (using physical output weights). Thus, the calibrated model has the property that more productive businesses are larger on a number of different dimensions. This is not surprising, since the source of heterogeneity in the benchmark model is idiosyncratic productivity variation and more productive businesses employ more workers, use more capital and produce more output. However, it is important to emphasize that the frictions yield the dispersion in TFPR and LPR that would otherwise not be present even with dispersion in TFPQ. An accompanying implication is that the frictions imply that there is a high positive correlation between TFPQ, TFPR, and LPR in the benchmark calibration.

²² We have also considered benchmark calibrations where we benchmark to the U.S. standard deviation of TFPR instead of to the OP covariance. The impact of distortions is quite similar both qualitatively and quantitatively to what we present here. We focus on the benchmark calibrated to the OP covariance term in the U.S. since we can match exactly the OP covariance term patterns in Table 1 by permitting only the distribution of distortions to vary.

Two other properties of the benchmark model are worth noting. First, in matching the moments, we obtain that overhead labor accounts for 14% of total employment. To put this in perspective, we note that in the U.S. manufacturing, non-production workers account for roughly 30% of total employment, and managers account for about 10% of the total. Classifying all non-production workers as overhead labor is probably too strong an assumption, while focusing only on managers does not fully account for overhead labor. Valerie Ramey (1991) reviews the evidence on the range of estimates of overhead labor and concludes that 0.20 is a reasonable consensus estimate. Interestingly, our estimate is close to these estimates and based on matching evidence on dispersion in LPR as well as survival rates.

Another instructive statistic from the calibrated model is the share of overall output accounted for by entry costs. Recall that output is used for consumption, capital accumulation (replacing depreciated capital in equilibrium) and entry costs. In the non-distorted economy, we find that about 15 percent of output goes to entry costs. This reflects the costs incurred *via* the ongoing churning process with some businesses exiting on an ongoing pace and other businesses paying sunk entry costs to learn their draws and deciding to exit conditional on bad news. As will become clear, this statistic is sensitive to the distortions and will be one of the factors that influence the relationship between distortions and consumption.

Before proceeding to the effects of the idiosyncratic distortions, it is worth highlighting the role that overhead labor plays in our exercise of matching the key moments in the U.S. data. Figure 1 shows the patterns for a number of key moments and outcomes as we change the overhead labor parameter f , keeping all other parameters constant.²³ The vertical line in Figure 1 shows the benchmark calibration for the U.S.. Lowering overhead costs relative to the

²³ Underlying Figure 1 is variation in f . We present the results showing the implied variation in overhead costs (share of labor going to overhead) since this is a more easily interpretable metric.

benchmark leads to a lower OP covariance term, a lower standard deviation of LPR and a much higher firm survival rate.²⁴ It is clear from Figure 1 that overhead labor plays a critical role in matching a number of key features of the U.S. data. First, it is critical for matching the finding that there is greater dispersion of revenue labor productivity relative to revenue total factor productivity. Second, overhead labor is critical for matching the magnitude of the OP covariance term. Third, overhead labor is critical for capturing the high pace of exit of recent entrants.

Figure 1 also helps illustrate the difficulty of matching all of the U.S. moments exactly. The figure shows that, as overhead labor rises, the covariance between size and productivity first rises sharply but then becomes flat and eventually declines with further increases in overhead labor. The underlying reason for the non-monotonicity is suggested by another key pattern in Figure 1 – that is, the share of entering firms that survive declines monotonically with overhead labor costs. Eventually overhead labor costs become sufficiently high that much of the distribution of firms is cut off by selection and this yields a decline in the covariance. Similarly, the standard deviation of LPR, while also initially rising sharply with overhead labor, eventually becomes flat with respect to further increases. This pattern also reflects the selection effects.

IV. The Impact of distortions

We now turn to assess the potential role of distortions to allocation and selection. Our objective is twofold. First, we explore the implications of such distortions for key outcomes of the economy. Second, we explore whether the patterns of dispersion and covariance displayed in Tables 1 and 2 can be accounted for and understood in terms of differences in the distribution of distortions across countries.

A. Uncorrelated distortions

²⁴ Note that even with low overhead, there is still positive covariance for LPR given other frictions.

We begin with a relatively simple version of distortions, that is to say, permanent distortions that are uncorrelated with any fundamentals. Figure 2 shows the effect of increasing the dispersion of distortions on different moments. The metric of the horizontal axis is the standard deviation of the ex ante distortions (permanent plus transitory) – we discuss how to interpret the magnitudes of this metric below. Note that the non-distorted (U.S. benchmark) case is at the left of the figures (with zero distortions). We find that an increase in the dispersion of distortions yields modest and non-monotonic effects on the standard deviation of LPR and TFPR, a decrease in the OP covariance terms (for LPR, TFPR, and TFPQ) and in the fraction of surviving firms, with an accompanying increase in the cost of entry as a share of output, an increase in the capital-labor ratio and a decline in consumption. Both the capital-labor ratio and consumption are indexed to the non-distorted economy, so an index above (below) one indicates the variable has increased (decreased) relative to the non-distorted economy.²⁵

It is clear from Figure 2 that increasing the dispersion of the distortions has non-trivial effects on a number of indicators. One of them is the effect on the selection margin. As the “noise” from distortions increases, so does the churning of businesses that pay the sunk cost, and this has significant effects, for example, on consumption. In turn, the strong sensitivity of the selection margin implies that, while the dispersion in TFPR rises with the dispersion in distortions, as expected the quantitative increase is relatively small. Similarly, the quantitative decline in the OP covariance terms is relatively small. The reason for these outcomes is that the selection effect truncates the lower-end of the distribution of firms, which mitigates the impact on the other key moments.

²⁵ For consumption, the log of aggregate consumption is computed and then converted to index form. For the capital-labor ratio, the log of the aggregate capital-labor ratio is computed and then converted into index form. The use of indices here are more interpretable given that level (even log level) units in the calibration are not meaningful. The remaining moments are all unit free.

To interpret the magnitude of the dispersion in idiosyncratic distortions, it is useful to consider this in the context of the dispersion of profit shocks facing firms. Recall that profit shocks are given by the product of (one minus) distortions and productivity shocks (raised to the power of ρ). We calculate the ratio of the standard deviation of the distribution of profit shocks with idiosyncratic distortions to the standard deviation of the distribution of profit shocks without idiosyncratic distortions. This ratio provides perspective on the magnitude of the distortion dispersion reported in Figure 2. Over the range of the horizontal axis in Figure 2, the ratio of dispersion of distorted to non-distorted profit shocks ranges from 1 to 2.

B. Correlated distortions

A much stronger effect of distortions on the economy can be obtained by allowing their distribution to be positively correlated with the idiosyncratic productivity draw.²⁶ This hypothesis is consistent with empirical evidence (see e.g. the World Bank Investment Climate Survey, Aterido *et al.* 2007) suggesting that when regulations are very strict and enforcement more pronounced among large businesses and/or when corruption and rent seeking loom large, it may be better for productive firms to stay small and “fly beneath the radar screen”. Likewise, there is evidence, as best exemplified by the experience of the centrally-planned economies, suggesting that in countries characterized by widespread market- and policy-induced distortions, the largest businesses are not necessarily the most productive ones but rather those that receive preferential treatment.

Figure 3 shows the relationship between key moments and distortion dispersions when we allow the idiosyncratic productivity and the idiosyncratic distortion to be positively

²⁶ Restuccia and Rogerson (2008) also consider correlated distortions and find that they have a much more sizeable effect on misallocation. In their exercise, they calibrate the benchmark model (non-distorted) to the U.S. size distribution and then consider the impact of misallocation distortions. They do not seek to account for cross-country differences in observed moments.

correlated. We set the positive correlation very high (about 0.95) and vary the dispersion of the distortions for this high degree of correlation. We observe similar qualitative patterns to those in Figure 2 but the quantitative variation is much greater for specific moments. In particular, we find that increasing the dispersion of distortions yields: *i*) an increase in TFPR dispersion; *ii*) an initial decline and then an increase in LPR dispersion; *iii*) substantial reductions in the OP covariance terms for TFPQ, TFPR and LPR; *iv*) a decline in survival with an accompanying increase in the entry cost as a share of aggregate output; *v*) an increase in the capital-labor ratio index; and *vi*) a decline in the consumption index. Even though the correlated case yields substantially larger quantitative variation in the covariance moments, it is interesting that the impact on the consumption index in the correlated case is roughly similar to what we obtain in the uncorrelated case.

There are opposite effects of distortions on the dispersion of profit shocks in this context. On the one hand, as the dispersion of distortions increases, this tends to increase the dispersion of profit shocks. On the other hand, the positive correlation between distortions and productivity shocks lowers the dispersion in profit shocks. A high productivity firm is more likely to draw a high distortion which lowers profits while a low productivity firm is more likely to draw a low distortion (or positive subsidy) which raises profits. Over the range depicted in Figure 3, the ratio of the dispersion of total profit shocks including distortions to dispersion in profit shocks without distortions begins at 1 at the far left of Figure 3, it falls to about .5 when the dispersion in distortions itself is around 1.5, and then rises to about 0.7 at the far right of Figure 3. These patterns along with Figure 3 itself highlight that both the dispersion of distortions and the correlation of distortions with productivity are important in this context. These patterns also suggest some caution in comparing the magnitudes of the dispersion of distortions in Figures 2

and 3 since in terms of the metric of the ratio of distorted to undistorted profit shocks, Figure 2 exhibits a range of 1 to 2 while Figure 3 exhibits a range from 1 to .5.

What lessons can we draw from these calibrations? First, consistent with the intuition from Olley and Pakes (1996), we find that economies characterized by sizeable scale distortions have distorted size-productivity relationships, as measured by the OP covariance terms for TFPQ, TFPR, and LPR. We think it is striking that the impact of distortions on the covariances is similar across these alternative indicators of productivity. The ideal measure is the OP covariance using TFPQ and physical output since the canonical prediction of firm heterogeneity models is that, in the absence of distortions, this covariance should be positive. Moreover, it makes intuitive sense that scale distortions will reduce this covariance. Given the frictions we have included in our model, these patterns carry over to the covariance measures using TFPR (with input weights) and LPR (and employment weights).

Second, we find that, other than the size-productivity relationship, many other margins are affected by distortion, including the productivity dispersion, the selection margin and the capital-labor ratio. Third, increases in the dispersion of distortions across firms reduce consumption. Fourth, the impact of the increase in the dispersion of distortions on the OP covariance terms and on productivity dispersion is much greater if the distortions are positively correlated with productivity. This makes intuitive sense as distortions in this case have greater bite as they are not only add noise but actively induce the most productive firms to stay smaller and the least productive firms to survive and be larger than would have been the case in a distortion-free environment.

C. Matching the observed moments

We now turn to the second objective of our analysis where we try to match the observed cross-country differences in our key moments shown in Tables 1 and 2 by using differences in distortions alone. Figures 2 and 3 already suggest that it will not be feasible to match all of the patterns shown in Tables 1 and 2. Matching the OP covariance term patterns is feasible since there is strong monotonic relationship between increasing distortions and reducing the size-productivity relationship and Tables 1 and 2 show systematic patterns in the OP covariance terms across countries. However, Figures 2 and 3 also show that increasing distortion dispersion systematically yields some increase in the dispersion of TFPR but they do not show systematic patterns for the dispersion of LPR. Thus, matching these patterns with the distribution of distortions alone is not feasible.

Given this obvious challenge, we think it is instructive to show the patterns consistent with exactly matching the OP covariance terms in Table 1. Table 3 shows the model and data moments in this case along with the implied model implications for consumption. The implied cross-country variation in the standard deviation for LPR in the model only roughly matches the patterns in the data. In particular, the model manages to replicate the similar LPR dispersions observed in the data for the Western European economies but does not come close to matching the very high standard deviation of LPR in the transition economies. The model is also not effective in accounting for the cross-country patterns in the standard deviation of TFPR.

Taken at face value, the model yields interesting implications for the consumption patterns across countries. The general pattern is that the implied consumption index for most Western European countries is about 0.95 of that in the U.S., while in some of the transition economies the consumption index is only 0.85 of that in the US.

The calibrated model is also capable of capturing the changes in covariance terms in Table 2 for the Central and Eastern European countries during the transition. Using Figure 3 and Table 3 together, an increase in the covariance term from around 0 to 0.20 is associated with an increase in the consumption index of about 0.05 to 0.10. So again taken at face value, our results suggest that the improved size-productivity relationship in the transition economies can be associated with a substantial increase in consumption as a result of an improved allocative efficiency.

Our approach to matching moments is a form of indirect inference in the sense that we identify key parameters by matching data and model moments. That is, we choose key parameters to first match the U.S. data moments to model moments in Table 1 as our benchmark and then in turn choose the distribution of distortions to match the cross-country variation in data moments in Table 1. However, we readily acknowledge that our approach for parameter estimation is not a formal estimation approach to indirect inference such as in Russell Cooper and Haltiwanger (2006). In the latter approach, parameters are also chosen to minimize the difference between actual and simulated moments but the estimation of the parameters is through the method of simulated moments. We leave such formal estimation for future work but would also note that this would likely require dealing with some of the other specification issues discussed in more detail below.

V. Robustness

This section summarizes the main results of our extensive sensitivity analysis aimed at assessing the robustness of our results to changes in the key assumptions and parameterization of our model. In particular, we consider: i) lower levels of overhead labor; ii) measurement errors that may affect differently the key moments used in the analysis; iii) heterogeneity in factor

elasticity of labor; and iv) a negative, instead of positive, correlation between distortions and productivity draws. We provide a brief overview of our findings here and provide more details in the web appendix.

A. *Lower overhead labor*

A distinguishing feature of our model and analysis is to allow for the effect of overhead labor on endogenous selection. To assess the implications for our model simulations of a lower overhead labor, we set it sufficiently low so that all firms from the ex-ante productivity distribution produce in the absence of distortions. From Figure 1 we know that this implies that we cannot come close to match the U.S. moments for labor productivity dispersion or selection. Accordingly, we benchmark the non-distorted economy to match the standard deviation of TFPR for the U.S and consider the correlated case. A number of interesting results emerge as seen in Figure 4. First, with very low overhead labor there is a substantial range over which increasing distortions has no impact on selection. However, for sufficiently large distortion dispersion, selection begins to bite. Second, productivity dispersion both in terms of TFPR and LPR rises much more rapidly with distortion dispersion when selection is not playing a role. Third, we find a general pattern of the covariance between size and productivity declining with distortion dispersion even with low overhead labor, although the patterns are somewhat more complex for LPR. For TFPQ, we find that the covariance between size and productivity falls monotonically with distortion dispersion. For LPR, the covariance first increases over some range and then declines. The increasing portion reflects the fact that in the absence of distortions and overhead labor, there is relatively low LPR dispersion. The magnitude of the covariance depends on the magnitude of dispersion and thus over some range dispersion in LPR is so low that the covariance with LPR is also low. This is not surprising that the patterns for LPR are less

systematic in a low overhead labor environment with low distortion dispersion. With low overhead labor, we cannot match the observed pattern that LPR has greater dispersion than TFPR. Again, recall that this is a robust finding for all countries in Table 1 (and also a robust finding in Syverson (2004a)). With low overhead labor, this pattern can only be met with sufficiently large dispersion in idiosyncratic distortions.

B. Measurement Error

Measurement error in firm-level data is a ubiquitous challenge for empirical analyses and also affects significantly the recent misallocation literature that relies on within-industry second moments to identify distortions. Measurement error likely varies not only across countries but also across specific variables, within a country. Sales or revenue are likely to suffer from greater measurement errors than employment, while capital is likely to suffer from greater measurement error than both sales and employment. To illustrate the potential effect of measurement error, we have considered in the sensitivity analysis a multiplicative measurement error in revenue (where the multiplicative factor is centered at one) and assessed its impact on two key moments: the OP covariance term, using revenue labor productivity, and the standard deviation of TFPR. We find (see Figure 5) that the OP covariance term is much less sensitive to revenue measurement error than is the standard deviation of TFPR. The reason is straightforward. Multiplicative measurement error in revenue yields increased dispersion in measured revenue relative to actual revenue translating directly into the standard deviation of measured TFPR. However, multiplicative measurement error in revenue that is classical does not change the measured covariance between LPR and employment. This follows from the property that the expectation of the product of uncorrelated variables is the product of the expectations. However, we note that in unreported results we show that multiplicative measurement error in employment has a

substantial impact on both dispersion and the covariance. The effect on the covariance results from the fact that measurement error influences the denominator of LPR and the numerator of the employment weight, yielding a negative contribution to the covariance. However, as noted above, we think it is more plausible that revenue is measured with error than is employment.

While this simulation of the role of measurement error within the context of our model is not definitive, it is suggestive as to one reason we might account for the finding that in the data we observe much less systematic variation in dispersion moments relative to the OP covariance moment across countries.

C. Dispersion in Factor Elasticity for Labor

In our model, we have relied on overhead labor to obtain a higher dispersion in revenue labor productivity relative to TFPR as observed in the data. There are, however, other factors that could yield this pattern between the dispersions in labor productivity and TFPR. An obvious one is to allow for heterogeneity in factor elasticities with respect to labor.

We considered a parameterization of the model with low overhead labor, no distortions, and no dispersion in factor elasticities (see Figure 6). With no dispersion in factor elasticity of labor, dispersion in TFPR matches that in the U.S. but is much larger than that of LPR. Moreover, the OP covariance term is relatively small and much smaller than that observed in Table 1 for the U.S. As we increase the dispersion in the factor elasticity of labor (along with that of capital since in this experiment we keep the return to scale constant), dispersion in LPR rises rapidly. But the OP covariance falls rather than rises. The intuition for the latter is straightforward. Firms with high factor elasticity of labor have a lower capital/labor ratio for a given level of TFPQ which in turn yields a lower LPR. Hence, matching the key U.S. moments from dispersion in factor elasticities is not feasible. We also note that other features of the data

work against this explanation. We find, for example, that persistence in labor shares in the U.S. over a five year horizon is only about 0.18. Presumably dispersion in factor elasticities would yield more persistence in labor shares at the plant level.

D. Negative correlation between distortions and productivity

In our framework, matching the cross-country patterns in the covariance between productivity and size requires a positive correlation between distortions and productivity. Uncorrelated distortions lower the size/productivity covariance but not by a sufficient amount. Given that the non-distorted economy is not a first-best in the presence of market power, an interesting question is whether it would be possible to increase allocative efficiency and consumption by introducing distortions that are negatively correlated with productivity. Model simulations suggest that allowing for a negative correlation between distortions and productivity can yield (at least modest) increases in the productivity/size covariance and consumption. However, the impact of changing either the degree of dispersion or the negative correlation of distortions and productivity yield non-monotonic results. It is beyond the scope of this paper to consider optimal policy in this context, but the lack of monotonicity is intuitively the interaction of opposing forces. On the one hand, distortions by their very nature affect incentives. On the other hand, the non-distorted economy has its own distortions from market power such that businesses are not producing as much as they should given their productivity and the given level of market prices.

VI. Discussion

What are our contributions relative to the results obtained using similar models of misallocation in the recent literature? First, we are the first to consider the size-productivity relationship in this context. We regard this relationship as critical to the canonical models of the

size distribution of activity that all of these recent papers build upon. That is, in the absence of distortions, these models have the property that the largest firms are the most productive ones. We think it is intuitive that distortions will have a non-trivial impact on the size-productivity relationship. We present novel empirical evidence on the size-productivity relationship using a sample of countries and explore the implications in the context of these models. We also show some suggestive evidence that the size-productivity relationship is more robust to measurement error than dispersion moments.

Second, we pay particular attention to one key empirical finding from our analysis (consistent with results obtained in the recent empirical literature), namely that the within-industry dispersion in labor productivity is larger than that of total factor productivity. We regard this basic feature of the firm-level data important since it highlights possible frictions that need to be taken into account even in the absence of distortions.

Third, and in a closely related manner, we consider the role of distortions where there is a non-trivial role for endogenous selection. None of the recent papers calibrate their models using information on survival and exit. In our setting, the role of overhead labor plays a critical role in matching key patterns in the non-distorted benchmark and also, in turn, serves as a key factor in permitting endogenous selection.

VII. Concluding Remarks

In this paper, we shed light on the impact of idiosyncratic distortions at the firm level on the allocation of resources and on aggregate outcomes. We show that the within-industry distributions of productivity and size are closely related to each other and that this relationship varies significantly across countries. The evidence presented suggests that the size/productivity relationship is stronger in the more advanced economies, although there are large differences

even within this group of countries, and became stronger in Central and Eastern European countries as they progressed in the transition to a market economy.

We confront these interesting patterns in the data with a model that seeks to account for variation in outcomes due to distortions. Our simple model with heterogeneous firms facing frictions (overhead labor and quasi-fixed capital) and idiosyncratic productivity shocks provides a reasonable match to key patterns in the U.S. indicators of productivity and size built up from firm-level data. In particular, the model captures a number of features observed in the data, namely: the high dispersion in revenue-based measures of total factor and labor productivity; the greater dispersion in labor productivity than in total factor productivity; the large positive covariance between size and productivity; and finally the high failure rate among new entrants in the early years of activity.

Using this model, we seek to match the cross-country variation in a number of key moments using differences in the distribution of distortions alone. We are successful in matching the significant within-industry differences in the size-productivity relationship across countries, with the U.S. experiencing a much higher covariance than countries in Western Europe and especially those in Central and Eastern Europe, but also the significant increase in the covariance observed in the latter group as they progressed in the transition to a market economy. In matching these patterns, our calibrated model implies a relatively large impact on consumption resulting from these distortions.

While we are quite successful in matching the size-productivity covariance patterns, we are less successful in matching other patterns in the data. In particular, the theoretical model implies that dispersion in revenue-based measures of total factor productivity should monotonically increase with an increase in the dispersion of idiosyncratic distortions to

businesses in the correlated dispersion case. While this pattern makes sense theoretically, there are not systematic patterns in the dispersion of revenue-based measures of total factor productivity in the data. One possible explanation is that this moment is less robust to measurement error. However, it is equally likely that the model is not capturing other factors that influence the dispersion of productivity.

In this latter respect, we think our analysis helps to shed further light on the strengths and weaknesses of this class of models of misallocation to track key firm-based productivity moments observed in the data. While we believe these models offer rich new insights, they are quite simple relative to the theoretical and empirical models of firm behavior in the literature. For one, they are steady-state models with no meaningful dynamics. A large literature models and empirically analyzes firm dynamics, including the role of selection and learning effects for young firms as well as the adjustment dynamics of capital and labor. The adjustment dynamics of capital and labor (as well as other adjustment frictions) are undoubtedly also related to distortions (e.g., firing taxes). A more refined or different approach will be needed to capture the potential distortions to these multiple margins. But there are undoubtedly other missing pieces as well. Imposing the same distribution of shocks and market structure across countries and in turn attributing all of the differences in cross country firm-level moments to differences in the distribution of distortions is obviously a heroic assumption. Of course that is understood in this literature as one can view the approach of asking how far can distortions go in accounting for key cross country difference. Our approach using multiple moments for many countries highlights some of the challenges of this approach.

References

- Aghion, Phillippe and Peter Howitt. 1994. "Growth and Unemployment" *The Review of Economic Studies*, 61(3), 477-49.
- Alessandria, George, and Horag Choi. 2007. "Do Sunk Costs of Exporting Matter for Net Export Dynamics?" *Quarterly Journal of Economics* 122(1):289-336.
- Alfaro, Laura, Andrew Charlton, and Fabio Kanczuk. 2008. "Plant-Size Distribution and Cross-Country Income Differences." National Bureau of Economic Research Working Paper Series No. 14060.
- Aterido Reyes, Mary Hallward-Driemeier and Carmen Pagés. 2007. "Investment Climate and Employment Growth: The Impact of Access to Finance, Corruption and Regulations Across Firms." IZA Discussion Papers 3138, Institute for the Study of Labor (IZA)
- Baily, Martin N., Charles Hulten and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity. Microeconomics* 1992:187-267.
- Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth*, Elsevier, p. 473-552.
- Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2004. "Microeconomic evidence of creative destruction in industrial and developing countries." The World Bank, Policy Research Working Paper No. 3464, December.
- Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2009. "Measuring and Analyzing Cross-Country Differences in Firm Dynamics." In *Producer Dynamics: New Evidence from Micro Data*, T. Dunne, J.B. Jensen and M. Roberts (eds.). National Bureau of Economic Research, Inc. University of Chicago Press.
- Benassy, Jean-Pascal. 1996. "Taste for variety and optimum production patterns in monopolistic competition." *Economics Letters* 52(1):41-47.
- Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121(2):541-585.
- Cooper, Russell and John Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs," *Review of Economic Studies*,. 73(3), 611-633.
- Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler, 2011. "Trade, Technical Change and Market Selection: Evidence from Manufacturing Plants in Colombia" NBER Working Paper No. 14935 (revised).
- Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis*, Chicago: University of Chicago Press.
- Foster, Lucia, John Haltiwanger, and C. J Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics* 88(4):748-758.

- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98, no. 1: 394-425.
- Haltiwanger, John, Ron Jarmin and Javier Miranda. 2009. "Who Creates Jobs? Small vs. Large vs. Young," NBER Working Paper No. 16300.
- Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101(5):915-938.
- Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60(5), 1127-1150 .
- Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4), 1403-1448.
- Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9(2):508-523.
- Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6):1695-1725.
- Midrigan, Virgiliu and Daniel Yi Xu. 2010. "Finance and Misallocation: Evidence from Plant-level Data," NBER Working Paper 15647.
- Oliveira-Martins, Joaquim, and Stefano Scarpetta. 2002. "The Cyclicity of Price-Cost Margins: A note", *OECD Economic Studies*, No. 34 2002/1.
- Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6):1263-1297.
- Pierre, Gaëlle, and Stefano Scarpetta. 2006. "Employment protection: Do firms' perceptions match with legislation?" *Economics Letters* 90(3):328-334.
- Ramey, Valerie. 1991. "Comment on Markups and the Business Cycle" *NBER Macro Annual 1991*, 134-39.
- Restuccia, Diego, and Richard Rogerson. 2008. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic Dynamics* 11(4):707-720.
- Rotemberg, Julio and Michael Woodford. 1999. "The Cyclical Behavior of Prices and Costs" *Handbook of Macroeconomics*, 1151-1135.
- Scarpetta, Stefano. 2003. *The sources of economic growth in OECD countries*. OECD Publishing.
- Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 112(6):1181-1222.
- Syverson, Chad. 2004b. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86(2):534-550.

Van Biesebroek, Johannes. 2004. "Robustness of Productivity Estimates." NBER Working Paper 10303.

Venn, Danielle. 2009. "Legislation, Collective Bargaining and Enforcement: Updating the OECD Employment Protection Indicators," OECD, Social, Employment and Migration Working Papers, No. 89, July.

World Bank. 2004. World Development Report 2005: A Better Investment Climate for Everyone. A World Bank Publication.

Table 1. Within-industry productivity dispersion and OP covariance term

(weighted averages of industry-level data, U.S. Industry Weights)

	STD in Revenue Labor Productivity	STD in Revenue Total Factor Productivity	OP covariance term
United States	0.58	0.38	0.51
United Kingdom	0.57	0.41	0.16
Germany	0.72	NA	0.28
France	0.53	0.22	0.24
Netherlands	0.56	0.15	0.30
Hungary	1.03	0.91	0.18
Romania	1.05	0.56	-0.03
Slovenia	0.80	0.22	0.05

Notes: Averages over 1992-2001 data. Industry-level firm based TFP measures not available for Germany.
Source: Firm-Level database; see Bartelsman, Haltiwanger and Scarpetta (2009).

Table 2. Changes in productivity dispersion and OP covariance term

(weighted averages of industry-level data, U.S. Industry Weights)

	STD in Revenue Labor Productivity	STD in Revenue Total Factor Productivity	OP covariance term
United States	0.02	0.00	0.09
United Kingdom	0.04	0.03	0.06
Germany	0.06	NA	0.15
France	NA	NA	NA
Netherlands	0.01	0.00	0.11
Hungary	-0.03	-0.03	0.21
Romania	0.03	-0.03	0.25
Slovenia	-0.06	-0.02	0.16

Note: Change is difference in moments between average value in 1996-2001 and the average value in 1992-1995. Data for France only available from 1992-1995.

Source: Firm-Level database; see Bartelsman, Haltiwanger and Scarpetta (2009).

Table 3. Key data moments and model moments when the model is calibrated to match the OP LPR covariance

Country	COV_LPR (Data)	COV_LPR (Model)	STD_LPR (Data)	STD_LPR (Model)	STD_TFPR (Data)	STD_TFPR (Model)	Consumption Index (Model)
United States	0.51	0.51	0.58	0.75	0.38	0.47	1.00
United Kingdom	0.16	0.16	0.57	0.65	0.41	0.69	0.93
Germany	0.28	0.28	0.72	0.59	na	0.64	0.97
France	0.24	0.24	0.53	0.60	0.22	0.66	0.96
Netherlands	0.30	0.30	0.56	0.59	0.15	0.63	0.97
Hungary	0.18	0.18	1.03	0.63	0.91	0.68	0.94
Romania	-0.03	-0.03	1.05	0.72	0.56	0.70	0.88
Slovenia	0.05	0.05	0.80	0.71	0.22	0.71	0.90

Figure 1. The impact of overhead costs on key moments

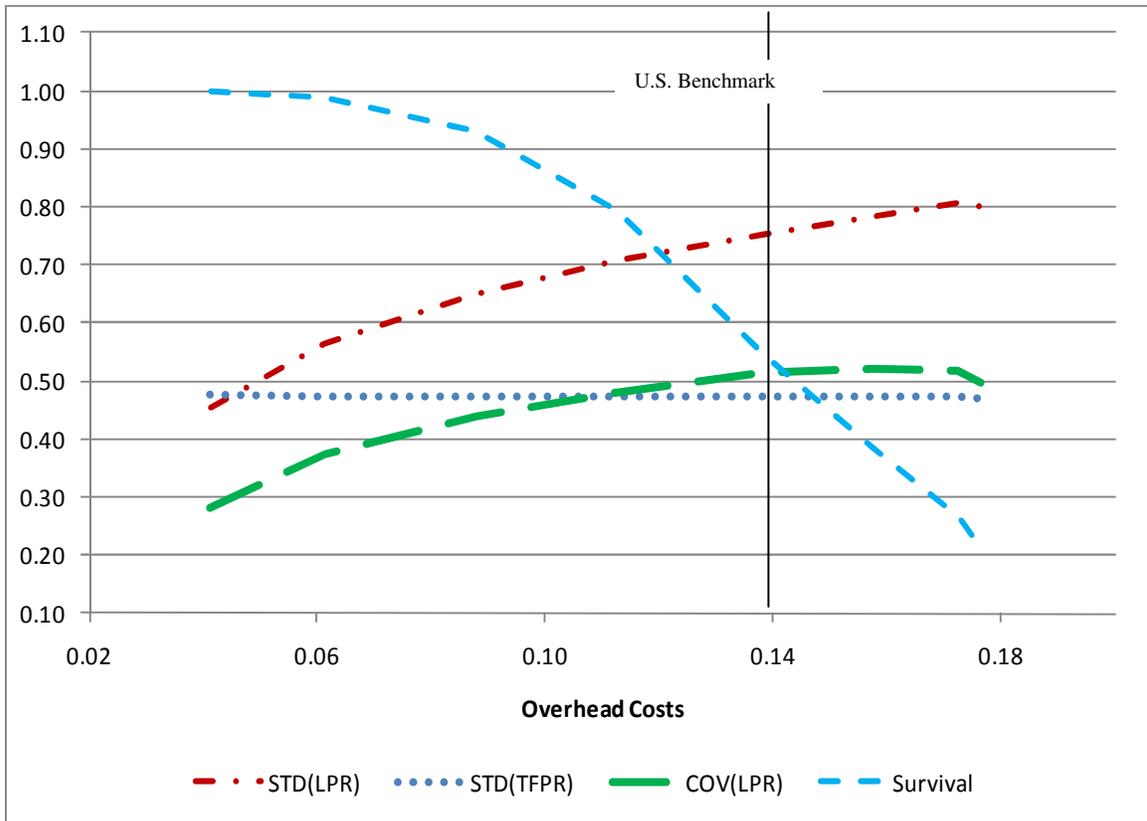


Figure 2: The Relationship Between Key Moments and Outcomes with Distortion Dispersion (Uncorrelated Case)

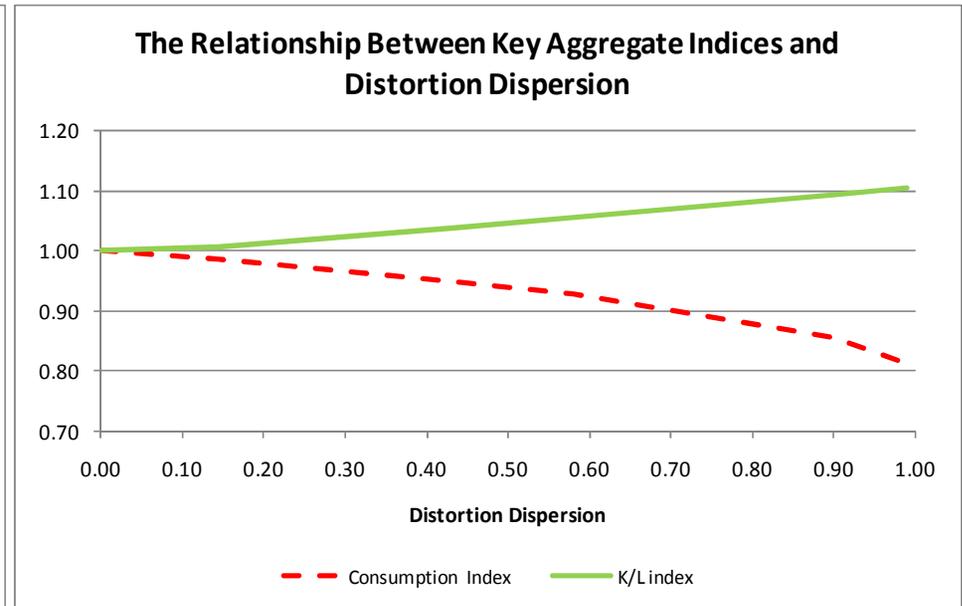
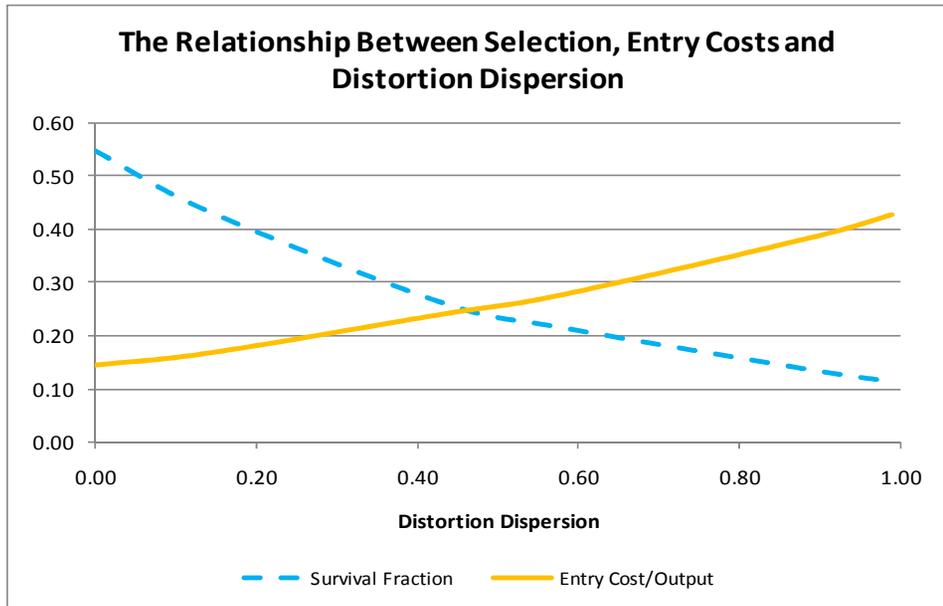
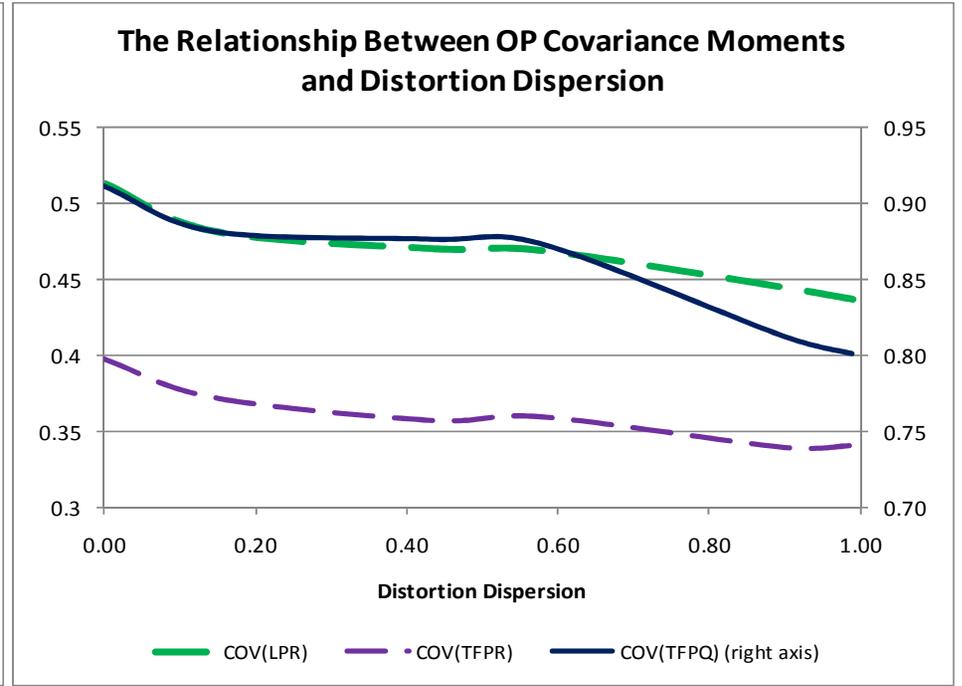
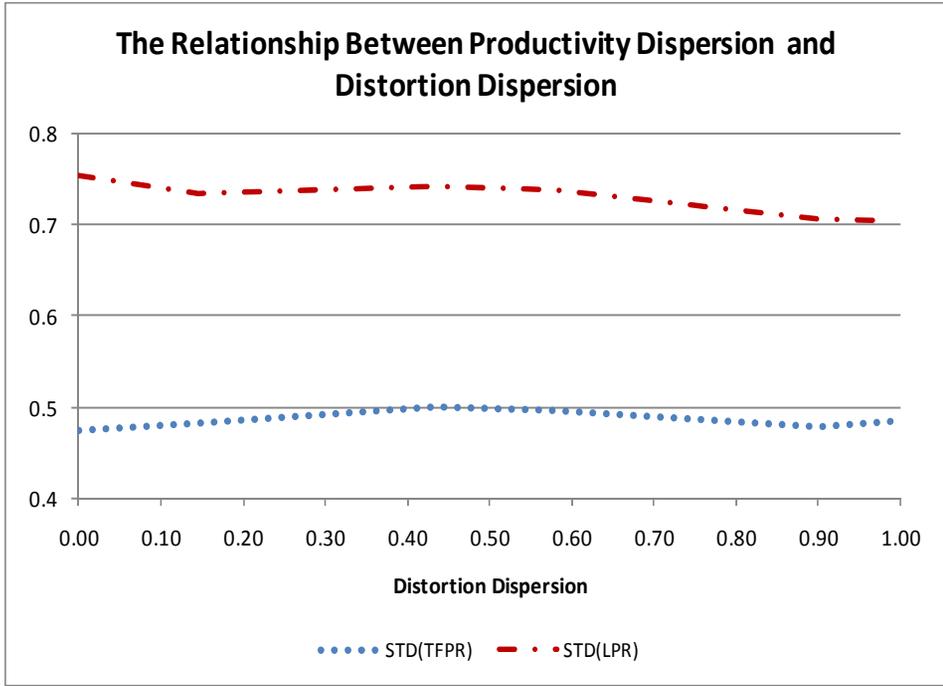


Figure 3: The Relationship Between Key Moments and Outcomes with Distortion Dispersion (Positive Correlation Case)

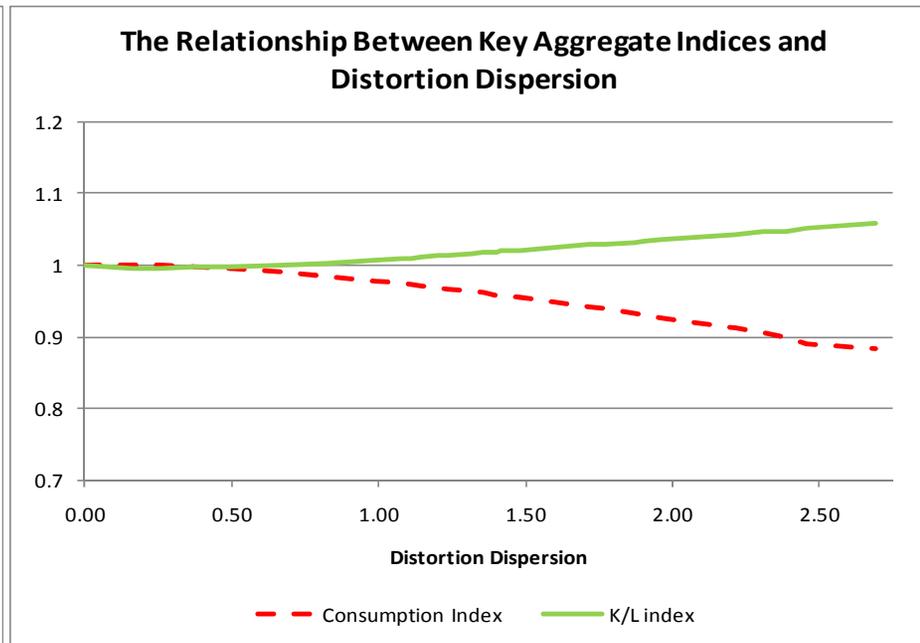
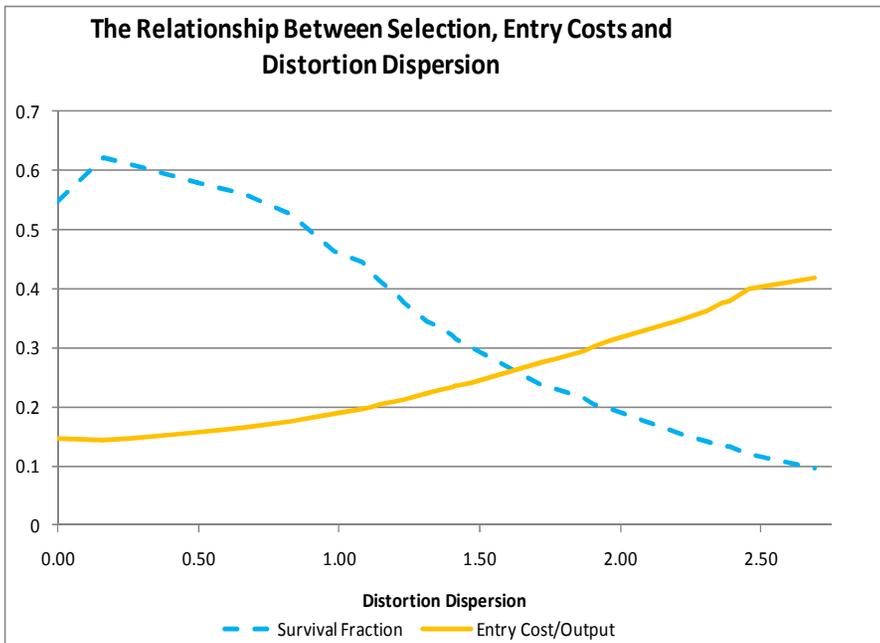
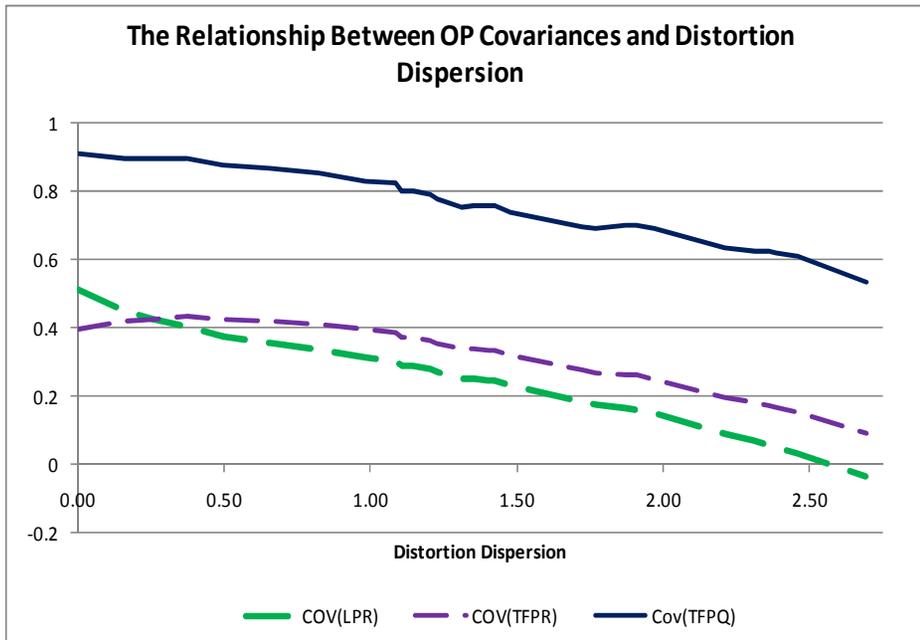
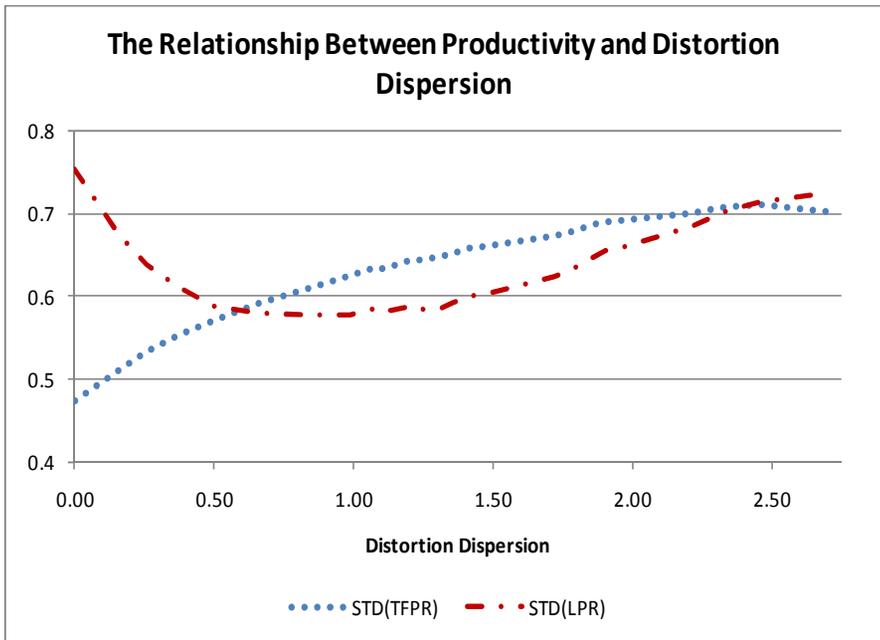


Figure 4: The Impact of Distortion Dispersion on Key Moments (Low Overhead Case)

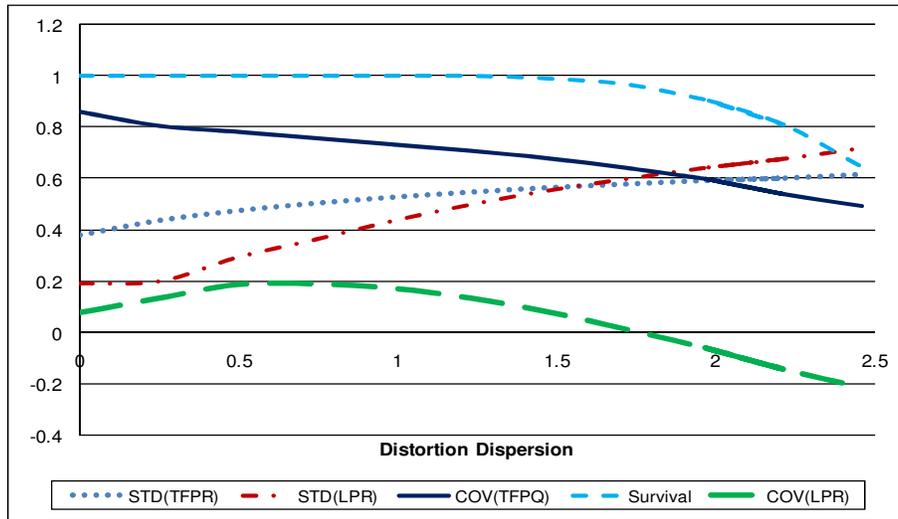


Figure 5: The Impact of Measurement Error on Dispersion of TFPR and Covariance(RLP)

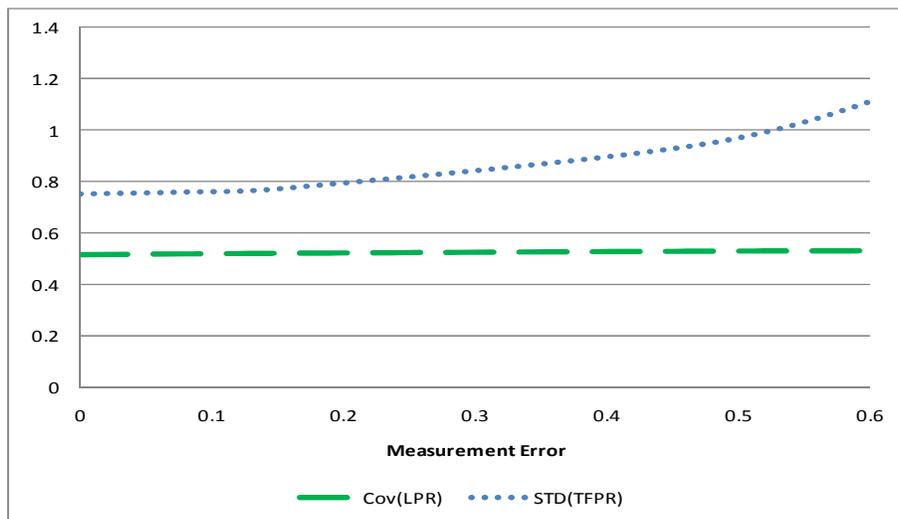


Figure 6: The Impact of Dispersion in Factor Elasticity of Labor on Key Moments

