THE NATURE OF FIRM GROWTH*

Benjamin W. Pugsley
Federal Reserve Bank of New York

Petr Sedláček
University of Bonn

Vincent Sterk
University College London

March 2017

Abstract

There are vast differences in the growth patterns of firms: high-growth, young businesses, or "gazelles", account for the vast majority of employment growth at incumbent firms. Based on a large administrative panel data set for the United States, this paper shows that a large fraction of size heterogeneity among firms, at a given age, is driven by ex-ante differences rather than ex-post shocks. We reach this conclusion after documenting the autocovariance structure of firm-level employment and estimating a reduced-form process that captures this structure. Next, we explore macroeconomic implications by matching a firm dynamics model to the empirical evidence. We show that, due to strong prevalence of ex-ante heterogeneity, firm selection creates sizeable gains in aggregate productivity. Nearly all of these gains derive from selection that takes place at the very beginning of firms' life cycles.

*We thank Mark Bils, Richard Blundell, Christian Bayer, Vasco Carvalho, Steve Davis, Urban Jermann, Greg Kaplan, Fabien Postel-Vinay and Kjetil Storesletten for helpful comments. We also are grateful for excellent research assistance by Harry Wheeler. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau, Federal Reserve Bank of New York or the Federal Reserve System. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

There are vast differences in the growth patterns of firms. While most businesses start small, relatively few high-growth, young businesses, sometimes called "gazelles", account for the vast majority of a cohort's employment growth (see e.g., Haltiwanger *et al.* (2014)), with the overwhelming majority of new and young firms experiencing little growth (see e.g., Hurst and Pugsley (2011)). One frequent explanation for these growth differences is that, following entry, firms are exposed to idiosyncratic shocks to marginal costs or to their product demand. According to this view, a firm outgrows its peers when it experiences a history of good shocks during its lifetime. An alternative view is that there are ex-ante differences between firm startups, with some types poised for growth and others destined to stay small. Under this view, heterogeneity in firms' growth paths are predictable, given their initial characteristics.

While it seems plausible that both views on firm growth are –to some extent– grounded in reality, little is known about their relative empirical relevance. However, the nature of growth differences may have important consequences for aggregate outcomes. For example, if there are large ex-ante differences in the growth potential of firms, then the process which selects those aspiring startups with the most potential to become actual producers, may have a large positive effect on aggregate productivity. By contrast, if a firms' growth paths are mostly determined by post-entry shocks, then the gains from selection may be much smaller.

In this paper, we present direct empirical evidence on the importance of a deterministic component in firms' growth patterns, vis-à-vis post-entry shocks. We then use this evidence to discipline a firm dynamics model, designed to quantify the impact of firm selection on aggregate productivity. We find that the a large fraction of the differences in firm size, conditional on age, can be attributed to ex-ante heterogeneity, ranging from 85 percent in the year of startup and 47 percent at old age. At the macro level, we find that firm selection increases aggregate productivity by nearly one quarter and that the bulk of this increase is driven by selection that takes at the very beginning of firms' life cycles, before they may have even started to produce.

The key piece of empirical evidence we present is the autocovariance matrix of employment at the firm (and establishment) level, up to 19 years after startup. We estimate this matrix from the Longitudinal Business Dynamics (LBD) database, which contains administrative information on the population of employers in the United States. In order to summarize the information contained in the autocovariance matrix, we propose a reduced-form employment process. Based on the estimated process, we then quantify the importance of ex-ante heterogeneity versus ex-post shocks. Moreover, we use the reduced-form process as a guideline for the type idiosyncratic shock process to be integrated into the structural model.

The proposed reduced-form process allows for heterogeneity in both initial and long-run "steady-state" employment levels, as well as heterogeneity in the speed at which this transition takes place. In addition, it allows for post-entry shocks. Moreover, the process nests various specifications that are commonly used in the firm dynamics literature to model the idiosyncratic shock process that firms are exposed to. However, we find that standard processes do not capture very well the autocovariance structure that we observe. For example, our process nests a simplified case, found in for example Melitz (2003) among many studies, in which there are no ex-post shocks and all heterogeneity is modeled as a firm-level type that is drawn ex-ante and remains constant over the life cycle, so that firms immediately reach their steady-state employment levels. This implies a flat autocovariance function, regardless of the age and horizon. In the data, however, autocovariances decline with the horizon, implying some role for ex-post shocks. Moreover, they increase with age, indicating a role for transitional dynamics. Another popular specification for the shock process is an AR(1) with a homogeneous constant, as found in for example Hopenhayn and Rogerson (1993). Such a process implies that firms gravitate towards the same steady-state levels and hence that ex-ante heterogeneity ultimately dies out. Thus, the implied autocovariances decline towards zero as the horizon is increased. In the data, however, autocovariances appear to stabilize at positive levels at longer horizons, suggesting an important role for heterogeneity

in steady-state levels.[1]

Not surprisingly, the generalization of the process turns out to be critical when quantifying the importance of ex-ante versus ex-post heterogeneity. A practical disadvantage of our process, however, is that it contains various state variables which may create computational difficulties when integrating the process into a structural model. The most general process introduces five exogenous state variables, versus only one in either Hopenhayn and Rogerson (1993) or Melitz (2003). To address this issue, we present restrictions that reduce the number of state variables from five to two, while preserving most of the dramatic improvement in fit, relative to the more standard processes.

Our next step is to explore implications of the empirical results for the macroeconomy using a structural firm dynamics model. While the nature of the idiosyncratic process is likely to matter in many applications of firm dynamics models, we focus on the effects of firm selection on aggregate productivity. Many studies, including Jovanovic (1982), Hopenhayn (1992), Hopenhayn and Rogerson (1993), Melitz (2003), Foster *et al.* (2008) have studied the selection of firms and have emphasized the importance of selection in the determination of aggregate outcomes. Our contribution is to demonstrate, qualitatively and quantitatively, the importance of the nature of the firm-level growth process in this dimension. The model we use for this purpose is an extension of the popular framework of Melitz (2003), but with an enriched idiosyncratic process which is flexible enough for the model to fit the empirical autocovariance structure of employment with reasonable accuracy.

Before conducting any quantitative exercises, we present two simplified cases which illustrates why the nature of the firm growth process is critical determinant of the strength of the selection channel. First, we consider a case in which all heterogeneity is permanent and determined ex ante, as in Melitz (2003). We show that the effect of selection on aggregate productivity depends positively on the amount of (ex-ante)

[1] Our process also nests specifications with heterogeneity in the constant, as commonly allowed for in the econometrics literature on panel data models. However, our process is still more general, since we allow different components of the process to have different persistence parameters. Allowing for the the latter turns out to be important to fit the autocovariance matrix well.

heterogeneity. Second, we illustrate a polar case in there is no ex-ante heterogeneity and all shocks are drawn ex-post and are purely transitory. In this example, selection effects are small (or even completely absent). Combining the two cases, only the ex-ante heterogeneity matters for the selection effects. Intuitively, when differences between startups are large and permanent, there are large productivity gains to be made from selecting the best startups. By contrast, the effect of transitory shocks of firm productivity is only short-lived and therefore have a very limited effect effects on a firm's expected value. Hence, firm's exit decision, which are based on expected firm values, are not much affected and selection effects are therefore small.

Finally, we take the model to the data. Having integrated an idiosyncratic shock process of the type proposed earlier, the model can provide a good fit of the observed autocovariance matrix of employment, as well as the profiles of average size and exit by age. We then use the model to quantify the effect of selection on aggregate productivity, by comparing the model to a counterfactual version in which selection effects are shut off. We find sizeable productivity gains from selection, in the order of 20 percent in the aggregate. This gain is almost entirely driven by selection in the very first period, at a point at which firms have observed their ex-ante parameters but have not actually started to produce. Interestingly, this is true even though a substantial amount of endogenous exit takes place in subsequent years. The reason why this subsequent exit has relatively limited effects on aggregate productivity, is that these firms tend to be close to indifferent between exit and continuation, whereas many of the startups who exit immediately are on average further away from the indifference point.

Relation to the literature. The importance of ex-ante conditions has been highlighted by Hurst and Pugsley (2011) who present survey evidence that many nascent entrepreneurs do not expect their business to grow large. Guzman and Stern (2011) present evidence that firm growth is partly predictable based on observable characteristics at the time of startup, e.g. whether or not the company is named after its owner or whether it is incorporated in the state of Delaware. Abbring and Campbell (2005) estimate an industry model with both transitory and persistent shocks, using

data on 305 bars in Texas. They find that ex-ante decision account for about 40 percent of the variation in ex-post outcomes. Sedláček and Sterk (2016) document the presence of strong cohort effects in employment data and estimate a firm dynamics model with ex-ante demand heterogeneity and aggregate shocks. They find that much of the differences across cohorts born can be attributed to the state of the economy in the year of startup, suggesting that cohorts differ in their composition with respect to ex-ante characteristics.[2] In the present paper, by contrast, we quantify the importance of ex-ante heterogeneity directly by exploiting within-cohort variation.

Our reduced-form analysis is inspired by a large empirical literature on earnings dynamics of workers, which traditionally derives identification from the autocovariance structure of earnings, see e.g. MaCurdy (1982), Abowd and Card (1989). A common assumption in this literature that earnings are the sum of an individual fixed effect, an age fixed effect, an AR(1) process with zero mean, and an i.i.d. shock. Some authors, however, have argued that allowing in addition for individual-specific trends helps to capture the autocovariance structure of earnings, see Guvenen (2009) for a discussion of this branch of the literature. The possible presence of such "Heterogeneous Income Profiles" (Guvenen (2007)), has received much attention since they may have large implications for the extent to which income changes should be expected to transmit to consumption, from the perspective of standard life-cycle models (see e.g. Guvenen and Smith (2014)). Somewhat surprisingly, the literature on firm-level employment dynamics does not have a similar tradition of estimating reduced-form processes. To the best of our knowledge, even the basic autocovariance structure of employment dynamics has not been systematically documented.

Our structural model builds on a large literature which uses firm dynamics models to understand the determinants of aggregate productivity. Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) quantify the effects of frictions that reduce aggregate productivity by creating misallocation of resources, but abstract from selection

[2] The importance of the composition of the firm population is also emphasized by Pugsley and Şahin (2016), who document a strong trend in the U.S. towards older firms, which is the result of accumulating startup deficits.

effects. Bartelsman *et al.* (2009) consider a framework but allow for both misallocation and selection effects. Importantly, they discipline their model using the observed covariance between firm size and productivity. Barseghyan and Dicecio (2011) use present a model to quantify the aggregate effects of variations in entry costs observed across countries,but abstract from post-entry shocks and restrict ex-ante heterogeneity to be constant, as in Melitz (2003). Our analysis complements these studies by highlighting the importance of matching the observed autocovariance matrix of employment when quantifying selection effects.

2 The nature of firm growth: empirical evidence

2.1 Data description

We use data on establishment-level employment in the United States, taken from the from Census Longitudinal Business Database. The data cover the population of employers over the period between 1979 and 2012. We construct a panel of employment in the year of startup (age zero) up to age 19. Prior to the analysis, we take out a fixed effect for the birth year of the establishment and for its industry classification at the 4-digit level.

2.2 The autocovariance structure of firm-level employment

Figure 1 presents our central piece of empirical evidence: the cross-sectional autocovariance structure of logged employment. In order to understand this structure more easily, we break down the autocovariances into standard deviations and autocorrelations. Figure 1 presents this information both for a balanced panel, including establishments surviving up to at least age 20, and an unbalanced panel, including all establishments in our data set.

The left panel of Figure 1 shows that the cross-sectional standard deviations of log employment, conditional on age, range between 1 and 1.1. This reflects large size differences across establishments, even at young ages. The differences between the balanced

and the unbalanced panel are moderate, at least in comparison to the levels. The fact that differences are small even at young ages may strike one as somewhat surprising, given that small establishments are known to be relatively unlikely to survive.[3] Another pattern visible in the left panel is that standard deviations increase between age 0 and age 19. While the shape of the age pattern differs across the two panels, the magnitude overall increase is very similar. This suggests that the fanning out of firm size in the unbalanced panel is not purely driven by small establishments terminating operations. Indeed, even among the survivors, size differences tend to grow with age.

The right panel of Figure 1 presents the cross-sectional autocorrelations. That is, the figure displays the correlation between log employment at age h and age $a \geq h$. The results for the balanced and the unbalanced panel turn out to be extremely similar, again suggesting a moderate role for selection after entry. As expected, the autocorrelations decline convexly with the horizon $(a - h)$. This may happen because establishments are hit by unanticipated shocks as they age.

Interestingly, the autocorrelations remain high even at long horizons. For example, the correlation between at age 0 and 19 is about 0.44, while the correlation between age 9 and 19 is about 0.74. The figure suggests that as the horizon increases towards infinity, the autocorrelation stabilizes at positive levels. This provides some indication for the presence of ex-ante heterogeneity that does not die out with age. As we will show formally below, autocorrelations would converge to zero without such heterogeneity. Finally, we observe that autocorrelations are increasing concavely in age, given a certain horizon. Thus, size differences become less mutable as establishments mature.

2.3 Employment process

We now take a more formal approach in analyzing the autocovariance structure of establishment-level employment. We do so by proposing and estimating a reduced-

[3] Exit of small firms trims the left tail of the size distribution in the balanced panel, which in turn lowers the amount of size heterogeneity, in comparison to the unbalanced panel. While this effect is visible in the figure, its magnitude appears moderate. Possibly, the overal relation between size and exit probability is not very strong. Alternatively, it might be the case that the size heterogeneity may be dominated by the right tail of the distribution, and that the relation between size and exit among large firms is relatively weak.

form stochastic process. We then use the estimated process to quantify the importance of ex-ante versus ex-post heterogeneity.

Let $n_{i,a}$ be the employment level of an individual firm of age a. We propose the following process:

$$\begin{aligned}
\ln n_{i,a} &= u_{i,a} + v_{i,a} + w_{i,a} + z_{i,a} \\
u_{i,a} &= \rho_u u_{i,a-1} + \theta_i,\ u_{i,-1} \sim iid(0, \sigma_u^2),\ \theta_i \sim iid(\mu_\theta, \sigma_\theta^2) \\
v_{i,a} &= \rho_v v_{i,a-1},\ v_{i,-1} \sim iid(\mu_v, \sigma_v^2) \\
w_{i,a} &= \rho_w w_{i,a} + \varepsilon_{i,a},\ w_{i,-1} = 0,\ \varepsilon_{i,a} \sim iid(0, \sigma_\varepsilon^2) \\
z_{i,a} &\sim iid(0, \sigma_z^2)
\end{aligned}$$

Here, $u_{i,a}$ and $v_{i,a}$ jointly capture the *ex-ante* component of the process, both of which are governed by stochastic, firm-specific parameters that are drawn just prior to startup, at age $a = 0$. Specifically, $u_{i,-1}$ and $v_{i,-1}$ are the initial levels of, $u_{i,a}$ and $v_{i,a}$, whereas θ_i pins down the long-run steady-state level of $u_{i,a}$ which is given by $\overline{u}_{i,\infty} = \frac{\theta_i}{1-\rho_u}$. The steady-state level of $v_{i,a}$, by contrast, is zero. We will therefore refer to $u_{i,a}$ as the *permanent* part of the ex-ante component and $v_{i,a}$ as the *transitory* part. The parameters ρ_u and ρ_v are common across firms and govern the speed at which the steady-state levels of the permanent and transitory part are reached. Further σ_u, σ_θ, σ_v, σ_ε and σ_z denote the standard deviations of the draws, which all come from iid distributions, which all have mean zero except for the distribution of, θ_i which has mean μ_θ.

The variables $w_{i,a}$ and $z_{i,a}$ capture the *ex-post* component and are governed by shocks that take place *after* the firm starts. The first of these, $w_{i,a}$, has an autoregressive structure with an autocorrelation coefficient given by ρ_w and an initial level normalized to zero. The second, $z_{i,a}$, is an pure iid component, which may possibly capture measurement error.

Note that the steady-state level of the overall process, i.e. the level that would be reached in the absence of ex-post shocks, is given by $\overline{\ln n}_{i,\infty} = \overline{u}_{i,\infty} = \frac{\theta_i}{1-\rho_u}$. Thus, the

process allows for heterogeneity in the steady-state levels. Since the process also allows for heterogeneity in the two initial levels, $u_{i,-1}$ and $v_{i,-1}$, it admits rich heterogeneity in firm-level ex-ante growth profiles. At the same time, the process allows for ex-post shocks with mixed degrees of persistence, via $w_{i,a}$ and $z_{i,a}$.

Our proposed process nests various specifications commonly used in the firm dynamics literature to model firm-level shocks. For example, Hopenhayn and Rogerson (1993) assume an AR(1) for firm-level productivity, with a common constant across firms and heterogeneous initial draws. In their baseline model, without firing taxes, the firm-level shocks map one-for-one into employment. We obtain their specification by setting $\sigma_u = \sigma_\theta = \sigma_z$ and $\rho_v = \rho_w$. By contrast, Melitz (2003) and Hsieh and Klenow (2009) also allow for heterogeneity in steady-state levels, but abstract from ex-post shocks and assume that steady-states are immediately reached. We obtain their process by setting $\sigma_u = \sigma_v = \sigma_\varepsilon = \sigma_z = 0$ and $\rho_u = 0$, which implies that $\ln n_{i,a} = \theta_i$ at any age.

Finally, our process nests specifications commonly assumed in the econometrics literature on dynamic panel data models, see for example Arellano and Bond (1991). This literature typically assumes an autoregressive process and, like Hopenhayn and Rogerson (1993), but allow for heterogeneity in the constant θ_i and thus in steady-state levels. Commonly, however, θ_i is differenced out and hence no estimate is provided for σ_θ, a key parameter in our application. Moreover, the panel data econometrics literature commonly assumes that $\rho_u = \rho_v = \rho_w$. In our application, it turns out that this assumption is too restrictive to provide a good fit of the observed autocovariance matrix. Our results thus caution against the use of standard panel data estimators when applied to employment dynamics of young establishments.

2.4 Parameter identification and variance decomposition

All parameters, except for μ_θ and μ_v, can be identified from the autocovaraince matrix of logged employment. In the appendix, we show that the covariance of employment of

a firm at age a and at age $a-j$ can be expressed as:

$$\begin{aligned} Cov\left(\ln n_{i,a}, \ln n_{i,a-j}\right) &= \rho_u^j \rho_u^{2(a-j+1)} \sigma_u^2 + \left[\rho_u^j \left(1-\rho_u^{a-j+1}\right)^2 + \left(1-\rho_u^j\right)\left(1-\rho_u^{a-j+1}\right)\right] \frac{\sigma_\theta^2}{\left(1-\rho_u\right)^2} \\ &\quad + \rho_v^j \rho_v^{2(a-j+1)} \sigma_v^2 + \rho_w^j \frac{1-\rho_w^{2(a-j+1)}}{\left(1-\rho_w\right)^2} \sigma_\varepsilon^2 + 0^j \sigma_z^2. \end{aligned}$$

From this equation it can be seen that the autocovariance function is a nonlinear function of the persistence and variance parameters of the components of the underlying process. Given that in total there are eight such parameters, we need an autocovariance matrix with at least eight elements for identification.

A key object of our interest is the amount of heterogeneity in long-run steady-state levels, which has a cross-sectional standard deviation given by $Std(\overline{\ln n}_{i,\infty}) = \frac{\sigma_\theta}{1-\rho_u}$. To better understand how the steady-state heterogeneity identified, it is useful to derive the autocovariance between employment at age a and at infinity. Provided that ρ_u, ρ_v, and ρ_w are all smaller than one in absolute value is given by:

$$Cov\left(\ln n_{i,\infty}, \ln n_{i,a}\right) = \frac{1-\rho_u^{a+1}}{\left(1-\rho_u\right)^2} \sigma_\theta^2.$$

Note in the absence of ex-ante heterogeneity in steady-state levels, the long-horizon autocovariances is zero. With ex-ante heterogeneity in steady-state levels, the autocorrelation stabilizes at a positive level, as the lag-length is increased towards infinity. Figure 1 shows suggest that this is indeed the case.

Once the process is estimated, it can be used to decompose employment heterogeneity, at a given age, into the contributions of ex-ante and ex-post factors. Towards this end, let us express the variance at age a as:

$$Var\left(\ln n_{i,a}\right) = \rho_u^{2(a+1)} \sigma_u^2 + \frac{\left(1-\rho_u^{a+1}\right)^2}{\left(1-\rho_u\right)^2} \sigma_\theta^2 + \rho_v^{2(a+1)} \sigma_v^2 + \frac{1-\rho_w^{2(a+1)}}{\left(1-\rho_w\right)^2} \sigma_\varepsilon^2 + \sigma_z^2.$$

The first three expressions on the right-hand side capture the contribution of the ex-ante component, whereas the last two terms capture the contribution of the ex-post components. Thus, the equation can be used to disentangle the contributions of ex-ante

and ex-post heterogeneity to the overall variance of employment, conditional on age. Note further that, as we let age approach infinity, the variance simplifies to:

$$Var\left(\ln n_{i,\infty}\right) = \frac{\sigma_\theta^2}{\left(1-\rho_u\right)^2} + \frac{\sigma_\varepsilon^2}{\left(1-\rho_w\right)^2} + \sigma_z^2.$$

The first of the three terms on the right-hand side captures the contribution of heterogeneity in the steady-state state levels, which are determined ex ante, and last two terms capture the contribution of ex-post shocks.

2.5 Estimation procedure

We estimate the parameters of the process using a minimum distance procedure, as proposed by Chamberlain (1984). Specifically, we minimize the sum of squared deviations of the upper triangular parts of the autocovariance matrix implied by the process, from its counterpart in the data. Because there is a very large number of observations underlying each element in the empirical autocovariance matrix, we assign equal weights to all elements in the estimation procedure.

2.6 Results

We estimate the baseline model as well as several restricted versions. The left two panels of Figure 2 illustrate the fit of the baseline process. Both for the balanced and unbalanced panel, the autocovariance structure is matched very well. In particular, the process is able to match the fact that autocovariances are convexly declining in the horizon. The left columns of Table plot associated parameter values and statistics. For the balanced panel, the estimate for the key parameter, σ_θ^2 equals 0.3637, which is substantially above zero. The estimation also reveals a substantial difference between ρ_u on the one hand, with an estimated value around 0.25, and ρ_v and ρ_w on the other hand, with estimated values around 0.9. The parameter estimates imply a substantial amount of heterogeneity in steady state levels. Table 1 shows that $Std(\ln_\infty)$, i.e. the standard deviation in long-run steady-state levels, is given by 0.76 in the balanced

panel and 0.77 in the unbalanced panel. As made clear by Figure 1, the cross-sectional standard deviation of employment of firms up to age twenty ranges between 1 and 1.2. In this light, the amount of steady-state heterogeneity is substantial.

Direct insight into the contribution of ex-ante heterogeneity to overall size heterogeneity can be obtained from Figure 3, which decomposes the overall variance of employment into the contributions of ex-ante heterogeneity and ex-post shocks. The dashed lines indicate age groups that were not used in the estimation. The figure shows that this contribution ranges between about 85 percent in the year of startup to 45 percent at old age, for both the balanced and the unbalanced panel. The estimated process thus reveals an important role for ex-ante heterogeneity as well as ex-post shock.

For illustrative purposes, Figure 3 also plots the decomposition for the restricted specifications following Melitz (2003) and Hopenhayn and Rogerson (1993). The figure suggests that neither of these two processes captures very well the amount of ex-ante heterogeneity present in the data, in particular for older firms. In the Melitz case, 100 percent of the variance is, by construction, accounted for by the ex-ante component. In the Hopenhayn-Rogerson specification, the contribution is about 90 percent in the year of startup, but completely dies out as with age. The latter is a direct consequence is that the process does not allow for permanent ex-ante heterogeneity. Thus, for older firms the contributions under the Melitz- and the Hopenhayn and Rogerson specifications are at two extremes. The baseline process is somewhere in the middle.

Figure 4 conducts the variance decomposition for employment growth between age h and age $a > h$, rather than for the level of employment. The figure reveals that ex-ante heterogeneity contributes importantly to employment growth. Out of the growth between age $h = 0$ and any age $a > 0$, about 40-45 percent is driven by the ex-ante component. At older age groups ($h > 0$) this contribution is lower between zero and 25 percent. At high ages (i.e. high levels of h) the contribution is close to zero. This is consistent with the idea that at some age, firms have reached their steady-state levels and any subsequent growth is due to ex-post shocks that make firms fluctuate around those steady states. Note further that at many age groups, the contribution is either

stable or increasing in the length of the horizon $(a-h)$.

The right panels of Figure 4 repeat the decomposition for the Hopenhayn-Rogerson specification.[4] The figure reveals a very different decomposition, with much lower contributions of ex-ante heterogeneity for growth of young firms. As for the baseline process, the contribution of ex-ante heterogeneity is declining in age (h). However, under the Hopenayn-Rogerson specification the contribution of ex-ante heterogeneity is are by and large increasing in the horizon.

The Baseline process has 5 state variables (θ,u, v, w, z). This may create a substantial computational burden when integrated into a quantitative structural model. We now explore alternative ways of reducing the number of state variables (see Table 1). Restricted model 1 sets $\sigma_\theta = 0$. This version fits data much less well than baseline, with a Root Mean Squared Error (RMSE) that is about 3 times as high. Restricted model 2 sets $\rho_u = \rho_v$, so u and v are no longer separate state variables. Fit worsens somewhat relative to baseline, but still much better than Restricted model 1. This versions appears to undershoot a bit on the amount of steady-state heterogeneity (especially in balanced panel). Restricted model 3 to sets $\rho_u = \rho_v$ and $\sigma_z = 0$, i.e. it drops another state variable by additionally dropping down the iid shock. Some additional worsening of fit but still better than baseline with $\sigma_\theta = 0$, even though it has one state variable less. Restricted 4 model is the AR(1) specification Hopenhayn and Rogerson, but this time allowing for heterogeneity in the fixed effect $(\sigma_\theta > 0)$. The fit as measured by the RMSE worsens further, to a level comparable with the standard Hopenhayn-Rogerson specification with $\sigma_\theta = 0$. However, the amount of steady-state heterogeneity is more in line our baseline process.

[4] Under the Melitz specification there is no ex-post growth. Hence, the variance is zero, so a decomposition is not possible.

3 Firm growth and aggregate productivity: a structural model

We now explore the macroeconomic implications of the nature of the firm growth process by evaluating how this process alters the effects of firm selection on aggregate productivity. We do so by enriching a standard model of firm dynamics with a firm-level shock process that entails both ex-ante and ex-post heterogeneity. We use stylized examples to illustrate how the importance of ex-ante versus ex-post heterogeneity has a critical impact on aggregate productivity, vis-a-vis its effect on firm selection. Specifically, we show that aggregate productivity gains from firm selection are particularly large in an economy in which firm size heterogeneity is mostly driven by ex-ante factors. By contrast, in an economy with only ex-post shocks such gains may be completely absent.

Next, we quantify the effects by matching the model to the empirical evidence presented in the previous subsection. We find that firm selection elevates aggregate productivity by about twenty percent. Moreover, we find that nearly all of this productivity gain is due to selection that happens at the very beginning of firms' life cycles, before they may have even started to produce.

3.1 The model

The model is an extension of the closed-economy model presented in Melitz (2003), and features heterogeneous and monopolistically competitive firms and endogenous entry and exit. Unlike Melitz, however, we allow not only for heterogeneity in a fixed, ex-ante productivity parameter, but also for heterogeneity in ex-ante growth profiles (depending on age) and on ex-post shocks. This extension will allow the model to match the autocovariance structure of firm-level employment, as well as the age profiles of average size and exit. Additionally, we allow for stochastic fixed costs of production.

Households. The economy is populated by is an infinitely-lived representative household who owns the firms and supplies a fixed amount of labor in each period, denoted

by $\overline{N}$. Household preferences are given by $\sum_{t=0}^{\infty} \beta^t C_t$, where $\beta \in (0,1)$ is the discount factor. C_t is a Dixit-Stiglitz basket of differentiated goods given by:

$$C_t = \left(\int_{i \in \Omega_t} a_{i,t}^{\frac{1}{\eta}} c_{i,t}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}},$$

where Ω_t is the measure of goods available in period t, $c_{i,t}$ denotes consumption of good i, η is the elasticity of substitution between varieties, and $a_{i,t} \in [a_{\min}, \infty)$ is a stochastic and time-varying demand shifter specific to good i. We consider a stationary economy from now on and simplify notation by dropping time subscripts.

The household's budget constraint is given by $\int_{i \in \Omega} p_i c_i = W\overline{N} + \Pi$, where p_i denotes the price of good i, W denotes the wage and Π denotes firm profits. Utility maximization implies a demand schedule given by $c_i = (p_i/P)^{-\eta} a_i C$, where P is a price index given $P \equiv \left(\int_{i \in \Omega} a_i p_i^{1-\eta} \right)^{\frac{1}{1-\eta}}$, so that total expenditure satisfies $PC = \int_{i \in \Omega} p_i c_i$.

Incumbent firms. There is an endogenous measure of incumbent firms, each of which produces a unique good. Firms are labeled by the goods they produce $i \in \Omega$. The production technology of firm i is given by $y_i + f_i = n_i$, where y_i is the output of the firm, n_i is the amount of labor input (employment) and f_i is a firm-specific fixed cost of operation, which is stochastic and drawn from an i.i.d. distribution in each period. It follows that firms face the following profit function:

$$\pi_i = p_i y_i - W n_i.$$

Additionally, given the market structure, each firm faces a demand constraint given by

$$y_i = (p_i/P)^{-\eta} a_i C, \tag{1}$$

which is the demand schedule of the household combined with anticipated clearing of goods markets, which implies $c_i = y_i$.

At the beginning of each period, a firm may be forced to exit exogenously with

probability $\delta \in (0,1)$. If this does not occur, the firm learns its fixed cost f_i and has the opportunity to exit endogenously. If the firm exits, it avoids paying the fixed cost, but it is permanently shut down. If the firm chooses to remain in businesses, it then pays its fixed cost and learns its demand shifter a_i. Given its production technology and demand function, the firm sets its price p_i (and implicity y_i, n_i and π_i) to maximize the net present value of profits. The price-setting problem is static and the firms sets prices as a constant markup over marginal costs W:

$$p_i = \frac{\eta}{\eta - 1} W.$$

We let labor be the numeraire so that $W = 1$, and define the real wage $w \equiv W/P$ as the price of labor in terms of the Dixit-Stiglitz consumption basket C. Using this result, we can express profits as $\pi_i = w^{-\eta} C \chi a_i - f_i$, where $\chi \equiv \frac{(\eta-1)^{\eta-1}}{\eta^{\eta}}$, and labor demand as $n_i = \left(\frac{\eta}{\eta-1}\right)^{-\eta} w^{-\eta} C a_i + f_i$.

The demand shifter a_i, for reasons that will be clear in a moment, may not be Markov. However, we can write a_i as a function of an underlying Markov state vector $\mathbf{s}_i$. Let $V(\mathbf{s}_i, f_i)$ be the value of a firm at the moment it chooses whether or not to exit. At this point it has survived the exogenous exit probability δ and observed its fixed cost f_i , but it has not yet observed its demand shifter $a_i'(\mathbf{s}_i')$ for the current period. The value of a firm at the moment the exit decision is taken, denoted V, can now be expressed as:

$$V(\mathbf{s}_i, f_i) = \max\left\{\mathbb{E}\left[\pi\left(\mathbf{s}_i', f_i\right) + \beta(1-\delta) V\left(\mathbf{s}_{i,}', f_i'\right) \middle| \mathbf{s}_i, f_i\right], 0\right\}$$

In the above equation $\mathbf{s}_i'$ denotes the value of the state after the continuation decision is taken and new shocks are realized, and f_i' denotes the fixed cost at the beginning of the next period. Accordingly, we can express the profit, output, employment and exit policies as $\pi_i = \pi(\mathbf{s}_i', f_i)$, $y_i = y(\mathbf{s}_i', f_i)$, $n_i = n(\mathbf{s}_i', f_i)$, and $x_i = x(\mathbf{s}_i, f_i)$, respectively.

Firm entry. Firm entry is endogenous and requires paying an entry cost f^e, denominated in units of labor. After paying the entry cost at the beginning of a period, the firm observes its initial level of $\mathbf{s}_i$ and f_i, at which point the firm becomes like an incumbent. That is, the firm may decide to exit immediately or pay f_i, observe $\mathbf{s}_i$, and commence production. Free entry implies the following condition:

$$wPf^e \geq \int V(\mathbf{s}, f)\, G(d\mathbf{s})\, H(df)$$

where G is the distribution from which the initial levels of $\mathbf{s}_i$ are drawn, and H is the distribution from which the fixed cost f_i is drawn.

Aggregation and market clearing Let $\mu(\mathbf{S})$ be the measure of firms in $\mathbf{S} \in \mathcal{S}$, where is $\mathcal{S}$ is the Borel $\sigma-$algebra generated by $\mathbf{s}$. Given the exit policy, $\mu(\mathbf{S})$ satisfies:

$$\mu\left(\mathbf{S}'\right) = \iint \left((1 - x(\mathbf{s}, f))\, F\left(\mathbf{S}'|\mathbf{s}\right) H(df) \left(\mu(d\mathbf{s}) + M^e G(d\mathbf{s})\right)\right)$$

where M^e denotes the measure of entrants and $F(\mathbf{S}'|\, s)$ is consistent with the transition law for $\mathbf{s}_i$. The total measure of active firms is given by:

$$\Omega = \int \mu(d\mathbf{s}).$$

Labor market clearing implies:

$$\bar{N} = \int y\left(\mathbf{s}'\right) \mu\left(d\mathbf{s}'\right) + \iint f(1 - x(\mathbf{s}, f)) \left(\mu(d\mathbf{s}) + M^e G(d\mathbf{s})\right) H(df) + M^e f^e.$$

3.2 Selection and aggregate productivity in two simplified cases

Before we evaluate the model quantitatively, we study two extreme cases which illuminate the importance of nature of the exogenous process in the determination of aggregate productivity.

Simple case 1: only ex-ante heterogeneity. In the first case, we assume that the firm-level fundamental is time-invariant and drawn ex-ante from a distribution. That is, $s_i = a_i$ is a scalar which is drawn from the ex-ante distribution with CDF G . This is precisely the assumption made by Melitz (2003). For simplicity, we set $\beta = 1$ in this example, as in Melitz (2003). The equilibrium can now be characterized in a simple way, by defining a cutoff level a^* such that any firm exits if and only if $a_i < a^*$. As a result, the productivity distribution of active firms is given by $\mu(a_i) = \frac{G(a_i)}{1-G(a^*)}$ for $a_i \geq a^*$, and $\mu(a_i) = 0$ for $a_i < a^*$. The free-entry condition can now be expressed as a relation between average profits, $\overline{\pi} \equiv \int \mu(a)\pi(a)dG(a)$, and the cutoff a^*:

$$\overline{\pi} = \frac{f_e \delta}{1 - G(a^*)}.$$

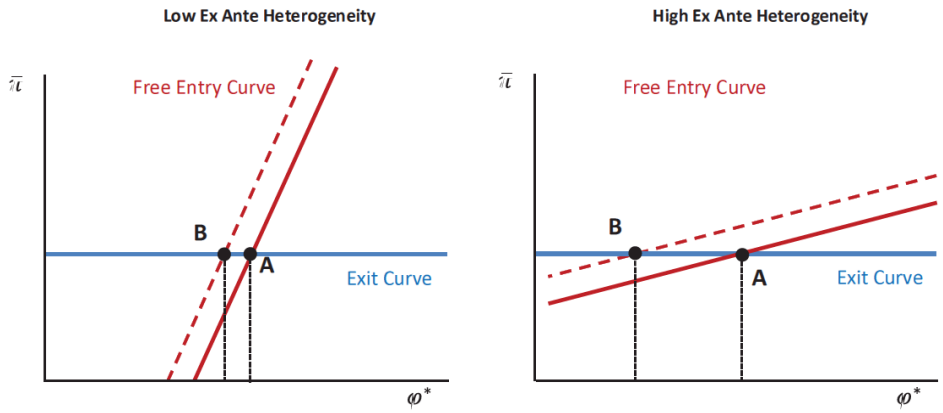
Now define $\widetilde{a}(a^*) \equiv \left[\int a^{\sigma-1}\mu(a)ds\right]^{\frac{1}{1-\eta}}$, i.e. a weighted average of firm-level productivity. Given that $\mu(a)$ is determined by the cutoff a^*, $\widetilde{a}$ is implicitly a function of a^*. As shown by Melitz (2003), $\widetilde{a}(a^*)$ coincides with aggregate productivity (CHECK), so the cutoff directly pins down aggregate productivity. We can now express the exit condition as another relation between $\overline{\pi}$ and a^*:

$$\overline{\pi} = k(a^*) f,$$

where $k(a^*) \equiv \left(\frac{\widetilde{a}(a^*)}{a^*}\right)^{\eta-1} - 1$. The equilibrium is at the intersection of the curves defined by the exit condition and the free-entry condition.

Combining, the two equations, the equilibrium must satisfy $\frac{f_e}{f/\delta} = (1 - G(a^*))\, k(a^*)$, which makes clear that the equilibrium cutoff is determined as a function of entry cost f_e, *relative* to the present value fixed costs to be paid, i.e.f/δ. Note also that if $\frac{f_e}{f/\delta}$ is sufficiently high, the cutoff may be driven down to the point that it hit its minimum i.e. $a_{\min} = a^*$.

For simplicity, let us further assume that the ex-ante draw comes from a Pareto distribution with scaling parameter α. In this case, it can be shown that $k(a^*) = \left(\frac{\alpha}{\alpha-1}\right)^{\eta-1} - 1$. Thus, the exit condition then pins down $\overline{\pi}$ independently of the level of



a^*. The slope of the free-entry condition is given by:

$$\frac{d\overline{\pi}}{da^*} = \frac{\delta f_e}{1 - G(a^*)} \frac{G'(a^*)}{1 - G(a^*)} = \overline{\pi}\alpha.$$

Thus, the slope of the free-entry curve is proportional to the Pareto parameter α. Letting α increase towards infinity, the variance of the Pareto distribution reduces to zero. Thus, the lower the variance of the productivity distribution (and hence the firm size distribution), the steeper the slope of the free-entry condition. This is illustrated by the left and right panel of the Figure above.

The figure also illustrates what happens after an increase in the entry cost, which shifts up the free-entry curve.[5] In equilibrium, the cutoff declines, which reduces aggregate productivity. Thus, selection effects increase aggregate productivity in this model. Intuitively, an increase in the entry cost increases the cost of sampling from the distribution of ex-ante draws of the fundamental. This induces firms to sample less often, i.e. entry decreases, which in turn has a positive effect on firm profits. This, however, makes firms more willing to continue operation under relatively low draws of the fundamental. That is, exit declines and the cutoff a^* shifts down, which pushes down firm profits. Under a Pareto distribution, the latter effects completely offsets the

[5] As mentioned above, a decrease in the fixed cost f or an increase in the exogenous exit rate δ as the same effect on the cutoff as an increase in the entry cost f_e.

increase increase in firm profits induced by the decline in entry. The decline in the cutoff, in turn, reduces aggregate productivity.

As illustrated by the illustration above, however, the magnitude of the decline depends critically on the slope of the free-entry condition. In the case with a large amount of ex-ante heterogeneity (right panel), the reduction in the cutoff is particularly large. Given that average profits are pinned down by the exit condition, the free-entry condition implies that an increase in the entry cost f_e must be offset by a proportional decline in the probability of successful entry $1-G(a^*)$ must adjust downwards. Under a large amount of heterogeneity, the distribution is spread out which means that a given change in the cutoff a^* has a relatively small effect on the probability of successful entry. Thus, a large decline in the cutoff is required to push down the entry success probability sufficiently in order to restore equilibrium.

Simple case 2: only ex-post heterogeneity. We now consider an opposite case in which there is no ex-ante heterogeneity. Specifically, we now assume that a_i is now determined as an ex-post i.i.d. shock. Recall that, in each period, the exit decision is made before observing the new shock. Since shocks are i.i.d, this implies that no firm has any specific information on its productivity when they make their exit decision. It follows that in equilibrium no firm voluntarily exits, given that new entrants pay the entry cost plus the fixed cost and that exits do still occur for exogenous reasons. As a result, there is no selection in the model with only i.i.d. ex-post shocks. Hence, aggregate productivity is not affected at all by firms' exit decisions.

It is possible combine Case 1 and 2, that is, to consider a model with both permanent ex-ante heterogeneity and i.i.d. ex-post shocks. It is straightforward to verify (Appendix?) that in this case only the ex-ante component of the process affects firm selection. Hence, the strength of selection effects is critically determined by the degree to which overall heterogeneity is determined by mix of ex-ante vis-à-vis the ex-post component. The literature typically makes ad hoc assumptions and extreme assumptions on this mix, like the ones we made above. Our empirical results suggest, however, that reality is more subtle. In the next section, we quantify the strength of selection

effects taking into account a realistic mix between ex-ante and ex-post heterogeneity, as revealed by the autocovariance structure of firm-level employment.

3.3 Quantifying the effects of selection on aggregate productivity

We now integrate a more realistic shock process into the model. In line with the reduced-form we postulate the following process:

$$\begin{aligned}
\ln a_{i,t} &= u_{i,t} + v_{i,t} + w_{i,t} + z_{i,t} \\
u_{i,t} &= \rho_u u_{i,t-1} + \theta_i,\ u_{i,-1} \sim iid(0, \sigma_u),\ \theta_i \sim iid(\mu_\theta, \sigma_\theta) \\
v_{i,t} &= \rho_v v_{i,t-1},\ v_{i,-1} \sim iid(0, \sigma_v) \\
w_{i,t} &= \rho_w w_{i,t} + \varepsilon_{i,t},\ w_{i,-1} = 0,\ \varepsilon_{i,t} \sim iid(0, \sigma_\varepsilon) \\
z_{i,t} &\sim iid(0, \sigma_z)
\end{aligned}$$

Note that the firm-level state is given by $s_{i,t} = [u_{i,t}, v_{i,t}, w_{i,t}, z_{i,t}]$. The components $u_{i,t}$ and $v_{i,t}$ jointly capture the ex-ante component of the process, whereas $w_{i,t}$ and $z_{i,t}$ capture the ex-post shocks. Given the parameter values, we solve for the equilibrium using the following algorithm, which follows Hopenhayn and Rogerson (1993)

We solve the model using the following numerical method. Let us first normalize $W = 1$. The profit function can then be written as $\pi(\mathbf{s}', f) = \chi a(\mathbf{s}') w^{-\eta} Y - f$. Further, we can define $\widehat{\mu}(\mathbf{S}) \equiv \frac{\mu(\mathbf{S})}{M^e}$, which evolves as:

$$\widehat{\mu}\left(\mathbf{S}'\right) = \iint \left((1 - x(\mathbf{s}, f)) F\left(\mathbf{S}'|\mathbf{s}\right) H(df) \left(\widehat{\mu}(d\mathbf{s}) + G(d\mathbf{s})\right)\right)$$

Also, the labor market clearing condition can now be written as:

$$\bar{N} = M^e \left(\frac{\eta}{\eta - 1}\right)^{-\eta} w^{-\eta} Y \widetilde{a} + M^e \widetilde{f} + M^e c_e,$$

where $\widetilde{a} \equiv \int a(\mathbf{s}') \widehat{\mu}(d\mathbf{s}')$ and $\widetilde{f} \equiv \iint f (1 - x(\mathbf{s}, f)) \left(\widehat{\mu}(d\mathbf{s}) + G(d\mathbf{s})\right) H(df)$. Note

further that $p_i = \frac{\eta}{\eta-1}$ and that the wage is given as

$$w = P^{-1} = \frac{\eta - 1}{\eta} \left(M^e \widetilde{a}\right)^{\frac{1}{\eta-1}}$$

We solve the model using the following algorithm (following Hopenhayn and Rogerson, 1993):

1. Solve for $w^{-\eta}Y$ from the free entry condition (i.e. guess $w^{-\eta}Y$, solve for the firm value functions, evaluate the free-entry condition, update the guess for $w^{-\eta}Y$ and iterate until the condition holds with equality).

2. Normalize $M^e = 1$, simulate the model and compute $\widehat{\mu}\left(S\right)$, $\widetilde{a}$ and $\widetilde{f}$.

3. Solve for M^e from the labor market clearing condition. Compute w, Y, and $\frac{Y}{N}$.

3.3.1 Calibration

The model period is one year. The parameter values are displayed in Table 2. The discount factor, β, is set to imply a real interest rate of four percent. The elasticity of substitution between goods, η, implies a markup of 11 percent over marginal costs. The ratio of the entry cost to the fixed cost, $\frac{f_e}{f}$ is set to 0.82, following an empirical estimate reported by Barseghyan and DiCecio (2011). Regarding the shock process, we ease the computational burden by setting $\sigma_z = 0$ and assuming $\rho_v = \rho_w$. The reduced-form evidence suggests that these restrictions are not very costly in terms of the ability to match the empirical autocovariance.

The remaining parameters are chosen to target jointly the autocovariance matrix of employment and the exit and average size profile by age. We assume that shock innovations are drawn from normal distributions. In order to fit the exit rate profile better, we introduce an iid shock to the entry cost, with mean zero and standard deviation σ_f. Figure 5 shows the model fit. The model fits very well the autocovariance function and the profile of the exit rate by age. Specifically, the model reproduces the fact that the exit rate initially declines convexly with age and then stabilizes. The

model also reproduces the increasing profile of average size, by age, although the profile predicted by the model is steeper than its empirical counterpart.[6] Comparing Tables 1 and 2 reveals that, compared to the reduced-form model, the variances pertaining to the ex-ante component (σ_θ^2, σ_u^2 and σ_v^2) are much larger in the structural model. The difference derives from selection at the very beginning of firms' life cycles, before have started to produce, which curtails the distribution of firms.

3.3.2 Results

Figure 6 plots exit rate, by age, in the data, the baseline model, and two counterfactuals: (i) no ex-ante heterogeneity, (ii) no ex-post shocks. The figure shows that without ex-post shocks, the model still predicts high and steeply declining exit rates between age 0 and 5. Without ex-ante heterogeneity, this is much less the case. This suggests that much of the "up-or-out" dynamics among young firms are due to ex-ante factors, i.e. by startups that are bound to have only a short life duration.

Figure 7 plots aggregate productivity as a function of the entry cost, in the baseline model and the two counterfactual versions. In each of the three economies, output is normalized to 1 under the baseline entry cost. Consider first the baseline. The figure shows that selection of firms has a substantial positive effect on aggregate productivity. Under high levels of entry costs, productivity can be more than 30 percent lower than in the baseline. Figure 8 shows that this is not driven by a change in the amount of labor used for productive purposes. The general-equilibrium elasticity of aggregate productivity with respect to the entry cost seems to be about 0.03 in the baseline (and half as large without ex-ante heterogeneity). The elasticity may not seem like a very high number, but the literature has documented be large differences in entry costs across countries. In the counterfacutal economy without ex-post shocks, the pattern in Figure 6 is almost exactly the same as in the baseline. This indicates selection based on ex-post shock is irrelevant in determining aggregate productivity. By contrast, ex-ante

[6] One possible way to improve the fit would be to allow for age-fixed effects, which would introduce additional state variables. Moreover, the age-fixed effects would be common across firms, and hence would have no direct impact on the amount of heterogeneity across firms at a given age.

heterogeneity does seem important: in the counterfactual economy without ex-ante heterogeneity, aggregate productivity is less sensitive to a change in the entry costs than in the baseline. These patterns are consistent with the simple examples given earlier.

4 Concluding remarks

We have documented the autocovariance structure of firm-level employment in the population of U.S. employers. Our results show that a large fraction of firm size heterogeneity, at any given age, is due to ex-ante differences in growth profiles. We further proposed a reduced-form employment process which generalizes popular specifications in the literature, and provides a much better fit.

Using a structural firm dynamics model following Hopenhayn (1992) and Melitz (2003), we have explored the implications of our empirical findings for firm selection and aggregate productivity. The model is able to capture the autocovariance structure of employment, as well as average employment and exit rates by age. The ex-ante heterogeneity emerges as a key margin of firm selection, and hence as a key determinant of aggregate productivity. Most of this selection takes place before have even started to produce. Moreover, we find that without such selection, aggregate productivity would be more than 30 percent lower. By contrast, the aggregate impact of selection based on ex-post shocks is negligible.

Our results thus imply that the entrepreneurial process of trying out business ideas is a key contributor to aggregate productivity and welfare. Factors that inhibit this process, related to for example financial frictions or government regulations, may have large negative effects on social welfare according to our model.

References

Jaap Abbring and Jeffrey Campbell. A firm's first year, 2005. Tinbergen Institute Discussion Paper 05-046/3.

John Abowd and David Card. On the covariance structure of earnings and hours changes. *Econometrica*, 57(2):4111–445, 1989.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and application to employment equations. *Review of Economic Studies*, 58(2):277–297, 1991.

Levon Barseghyan and Riccardo Dicecio. Entry costs, industry structure, and cross-country income and tfp differences. *Journal of Economic Theory*, 146:1828–1851, 2011.

Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Measuring and analyzing crosscountry differences in firm dynamics, 2009. In *Producer Dynamics: New Evidence from Micro Data*,NBER Chapters, pp. 15âĂŞ76. National Bureau of Economic Research, Inc.

Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425, 2008.

Fatih Guvenen and Anthony Smith. Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6):2085–2129, 2014.

Fatih Guvenen. Learning your earning: Are labor income shocks really very persistent? *American Economic Review*, 97(3):687–712, 2007.

Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79, 2009.

Javier Guzman and Scott Stern. Where is silicon valley? *Science*, 347(6222):606–609, 2011.

John Haltiwanger, Ron Jarmin, Robert Kulick, and Javier Miranda. High growth young firms: Contribution to job, output and productivity growth, 2014. working paper.

Hugo Hopenhayn and Richard Rogerson. Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, 101(5):915–938, 1993.

Hugo Hopenhayn. Entry, exit and firm dynamics long run equilibrium. *Econometrica*, 60(5):1127–1150, 1992.

Chiang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 4:1403–1448, 2009.

Erik Hurst and Benjamin Pugsley. What do small businesses do?, 2011. NBER working paper no. 17041.

Boyan Jovanovic. Selection and the evolution of industry. *Econometrica*, 50(3):649–670, 1982.

Thomas MaCurdy. The use of time-series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, 18:83–114, 1982.

Marc Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71:1695–1725, 2003.

Ben W. Pugsley and Aşsegül Şahin. Grown-up business cycles, 2016. working paper.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous plants. *Review of Economic Dynamics*, 11(4):707–720, 2008.

Petr Sedláček and Vincent Sterk. The growth potential of startups over the business cycle, 2016. working paper.

5 Appendix

5.1 GMM estimation and overidentification from autocovariance

Repeating from above:

Generally, the autocovariance function for $a, j \geq 0$ is:

$$\begin{aligned}
\text{Cov}\left[\log n_{ia}, \log n_{ia+j}\right] =& \rho^{2(a+1)+j}\sigma_{\tilde{u}}^2 + \frac{1-\rho_u^{a+1}}{1-\rho_u}\frac{1-\rho_u^{a+1+j}}{1-\rho_u}\sigma_\theta^2 + \rho_v^{2(a+1)+j}\sigma_{\tilde{v}}^2 \\
&+ \sum_{k=0}^{a} \rho_w^{2k+j}\sigma_\varepsilon^2 + \sigma_\zeta^2 \mathbf{1}_{j=0} \\
=& \rho^{2(a+1)+j}\sigma_{\tilde{u}}^2 + \frac{1-\rho_u^{a+1}}{1-\rho_u}\frac{1-\rho_u^{a+1+j}}{1-\rho_u}\sigma_\theta^2 + \rho_v^{2(a+1)+j}\sigma_{\tilde{v}}^2 \\
&+ \rho_w^j \frac{1-\rho^{2(a+1)}}{1-\rho^2}\sigma_\varepsilon^2 + \sigma_z^2 \mathbf{1}_{j=0} \qquad (2)
\end{aligned}$$

5.1.1 Nonlinear GMM estimation

Let $\theta = \left(\rho_u, \rho_v, \rho_w, \sigma_\theta^2, \sigma_u^2, \sigma_v^2, \sigma_\varepsilon^2, \sigma_z^2\right)'$ be an arbitrary parameter vector in compact parameter space $\mathbb{P}$. Since we use ages 0 to A, we define the $\frac{A*(A+1)}{2}$ length vector valued function

$$f\left(n_i, \theta\right) = \left[\left(\log n_{ia} - E\left[\log n_{ia}\right]\right)\log n_{ij} - \text{Cov}\left[\log n_{ia}, \log n_{ia+j}; \theta\right]\right]$$

where $a = 0, \ldots, A$ and $j = a, a+a, \ldots, A$. Let θ_0 be the true parameter vector, so that identification follows from $E\left[f\left(n_i; \theta\right)\right] = 0$ iff $\theta = \theta_0$. The term $\text{Cov}\left[\log n_{ia}, \log n_{ia+j}; \theta\right]$ is a constant and equal to the formula from equation (2) computed for an arbitrary parameter vector θ.

Define the sample analog to $E\left[f\left(n_i; \theta\right)\right]$

$$g_N\left(\theta\right) \equiv \frac{1}{N}\sum_i f\left(n_i; \theta\right).$$

A law of large numbers implies $g_N\left(\theta\right) \to^p E\left[f\left(n_i; \theta\right)\right]$. Define the GMM estimator

$$\tilde{\theta}_N = \underset{\theta \in \mathbb{P}}{\text{argmin}}\, g_N\left(\theta\right)' W g_N\left(\theta\right)$$

for an arbitrary symmetric positive definite weighting matrix W. The asymptotic

distribution of the estimator $\tilde{\theta}_N$ is:[7]

$$\sqrt{N}\left(\tilde{\theta}_N - \theta_0\right) \to_d N\left(0, \Sigma\right)$$

where

$$\begin{aligned}
\Sigma &\equiv \left(d'Wd\right)^{-1}\left(d'WVWd\right)\left(d'Wd\right)^{-1} \\
d &\equiv \frac{\partial E f\left(n_i;\theta_0\right)}{\partial \theta'} \\
V &\equiv E\left[f\left(n_i;\theta_0\right) f\left(n_i;\theta_0\right)'\right].
\end{aligned}$$

Note V is not a covariance matrix for $\log n_{ia}$ since $E\left[f\right]$ is the unique elements of the covariance matrix.

Estimation To operationalize the estimator we have to estimate both V and the means $E\left[\log n_i\right]$. Define

$$\tilde{f}\left(n_i,\theta\right) \equiv \left[\left(\log n_{ia} - \frac{1}{N}\sum_{i'} \log n_{ia}\right)\log n_{ij} - \mathrm{Cov}\left[\log n_{ia}, \log n_{ia+j};\theta\right]\right]$$

[7] To see this, write $g_N\left(\tilde{\theta}_N\right)$ as

$$g_N\left(\tilde{\theta}_N\right) \approx g_N\left(\theta_0\right) + \frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}\left(\tilde{\theta}_N - \theta_0\right).$$

Multiplying through by $\frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W$ so that the LHS is equal to the first order condition $\frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W g_N\left(\tilde{\theta}_N\right) = 0$, then

$$\begin{aligned}
0 &\approx \frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W g_N\left(\theta_0\right) + \frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W\frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}\left(\tilde{\theta}_N - \theta_0\right) \\
\left(\tilde{\theta}_N - \theta_0\right) &\approx -\left(\frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W\frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}\right)^{-1}\frac{\partial g_N\left(\tilde{\theta}_N\right)}{\partial \theta'}W g_N\left(\theta_0\right).
\end{aligned}$$

Letting $N \to \infty$ then

$$\sqrt{N}\left(\tilde{\theta}_N - \theta_0\right) \to^d -\left(\frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}W\frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}\right)^{-1}\frac{\partial g_N\left(\theta_0\right)}{\partial \theta'}W\sqrt{N}g_N\left(\theta_0\right)$$

since $\tilde{\theta}_N \to^p \theta_0$. And from the CLT $\sqrt{N}g_N\left(\theta_0\right) \to^d N\left(0, V\right)$.

$$\tilde{g}_N(\theta) = \frac{1}{N}\sum_{i=1}\tilde{f}(n_i,\theta),$$

where $a = 0, \ldots, 10$ and $j = a, a+a, \ldots, 10$. Define the $A(A+1)/2 \times A(A+1)/2$ moment covariance matrix

$$\tilde{V}_N = \frac{1}{N}\sum_{i=1}\left[\tilde{f}(n_i,\theta)\,\tilde{f}(n_i,\theta)'\right] = \frac{1}{N}\sum_{i=1}\left(h(n_i)-\bar{h}\right)h(n_i)'.$$

Note that since $\tilde{f}(n_i,\theta) = h(n_i) - q(\theta) = 0$ then $E\left[\tilde{f}_i\tilde{f}_i\right] = \text{Cov}\left[\tilde{f}_i,\tilde{f}_i\right] = \text{Cov}\left[h(n_i),h(n_i)\right]$.

To deal with missing data. We can create an indicator variable λ_{iaj} for whether or not the observation is missing and then define

$$\tilde{f}(n_i,\theta,\lambda_i) \equiv \left[\lambda_{iaj}\left(\left(\log n_{ia} - \frac{1}{N}\sum_{i'}\log n_{ia}\right)\log n_{ij} - \text{Cov}\left[\log n_{ia}, \log n_{ia+j};\theta\right]\right)\right]$$

and use weighting matrix

$$A = \Pi^{-1}\Pi^{-1}$$

where

$$\Pi = \begin{bmatrix} \frac{N_{00}}{N} & & & \\ & \frac{N_{01}}{N} & & \\ & & \ddots & \\ & & & \frac{N_{AA}}{N} \end{bmatrix}$$

and N_{aj} is the number of non missing observations for that moment. This is equivalent to equally weighting but uses the correct number of observations when computing $\tilde{V}$

Tables and Figures

A. Balanced Panel

	Baseline	H&R	Melitz	Restricted 1	Restricted 2	Restricted 3	Restricted 4
ρ_u	0.2059	0	0	0.0000	0.4475	0.2441	0
ρ_v	0.8415	0.9752	0	0.9776	0.9726	0.9568	0.9599
ρ_w	0.9489	0.9752	0	0.9808	0.9726	0.9568	0.9599
σ_θ^2	0.3637	0	0.8519	0	0.0688	0.2123	0.3007
σ_u^2	4.1864	0	0	0.7568	1.5864	6.4723	0
σ_v^2	0.5444	0.8225	0	0.8090	0.6020	0.4507	0.5005
σ_ε^2	0.0652	0.0681	0	0.0590	0.0570	0.0723	0.0760
σ_z^2	0.0688	0	0	0.0897	0.0834	0	0
$RMSE$	0.0100	0.0387	0.1575	0.0311	0.0184	0.0270	0.0380
$Std(\ln \overline{n}_\infty)$	0.7594	0	0.9230	0	0.4747	0.6095	0.5483
# state vars.	5	1	1	4	4	3	2

B. Unbalanced Panel

	Baseline	H&R	Melitz	Restricted 1	Restricted 2	Restricted 3	Restricted 4
ρ_u	0.2604	0	0	0.7216	0.3204	0.2383	0
ρ_v	0.8942	0.9693	0	0.9814	0.9489	0.9176	0.9306
ρ_w	0.9341	0.9693	0	0.9555	0.9489	0.9176	0.9306
σ_θ^2	0.3242	0	0.9228	0	0.1984	0.3326	0.4846
σ_u^2	2.9212	0	0	0.2548	2.5225	5.9414	0
σ_v^2	0.5472	0.9786	0	0.8319	0.5721	0.4259	0.4479
σ_ε^2	0.0830	0.0820	0	0.0841	0.0781	0.1043	0.1054
σ_z^2	0.0800	0	0	0.0859	0.0890	0	0
$RMSE$	0.0131	0.0439	0.1710	0.0336	0.0158	0.0246	0.0403
$Std(\ln \overline{n}_\infty)$	0.7699	0	0.9606	0	0.6555	0.7572	0.6962
# state vars.	5	1	1	4	4	3	2

Table 1. Parameter estimates reduced-form model.

parameter	*value*	**parameter**	*value*	**parameter**	*value*
β	0.96	σ_θ^2	1.8578	μ_θ	-1.8714
η	10	σ_u^2	11.0790	ρ_u	0.4436
f	0.2949	σ_v^2	0.7656	ρ_v	0.9764
f^e	0.2418	σ_ε^2	0.0677	ρ_w	0.9764
δ	0.0594	σ_f^2	0.0095		

Table 2. Parameters structural model.

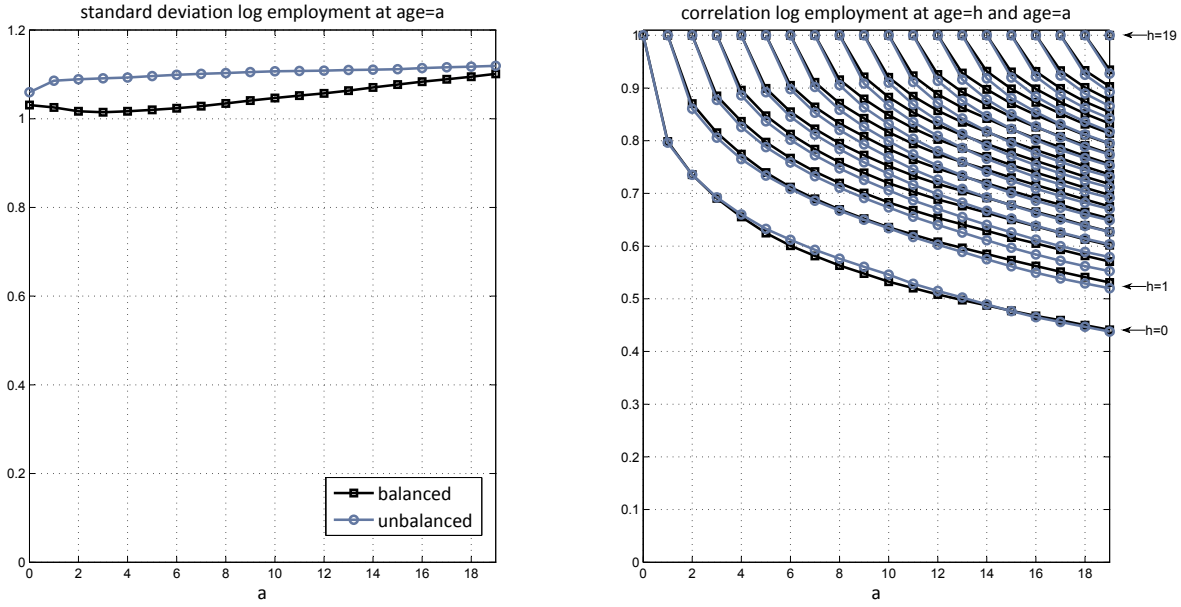


Figure 1: Standard deviations and autocorrelations of log employment. Balanced panel of survivors up to age 20 and unbalanced panel.

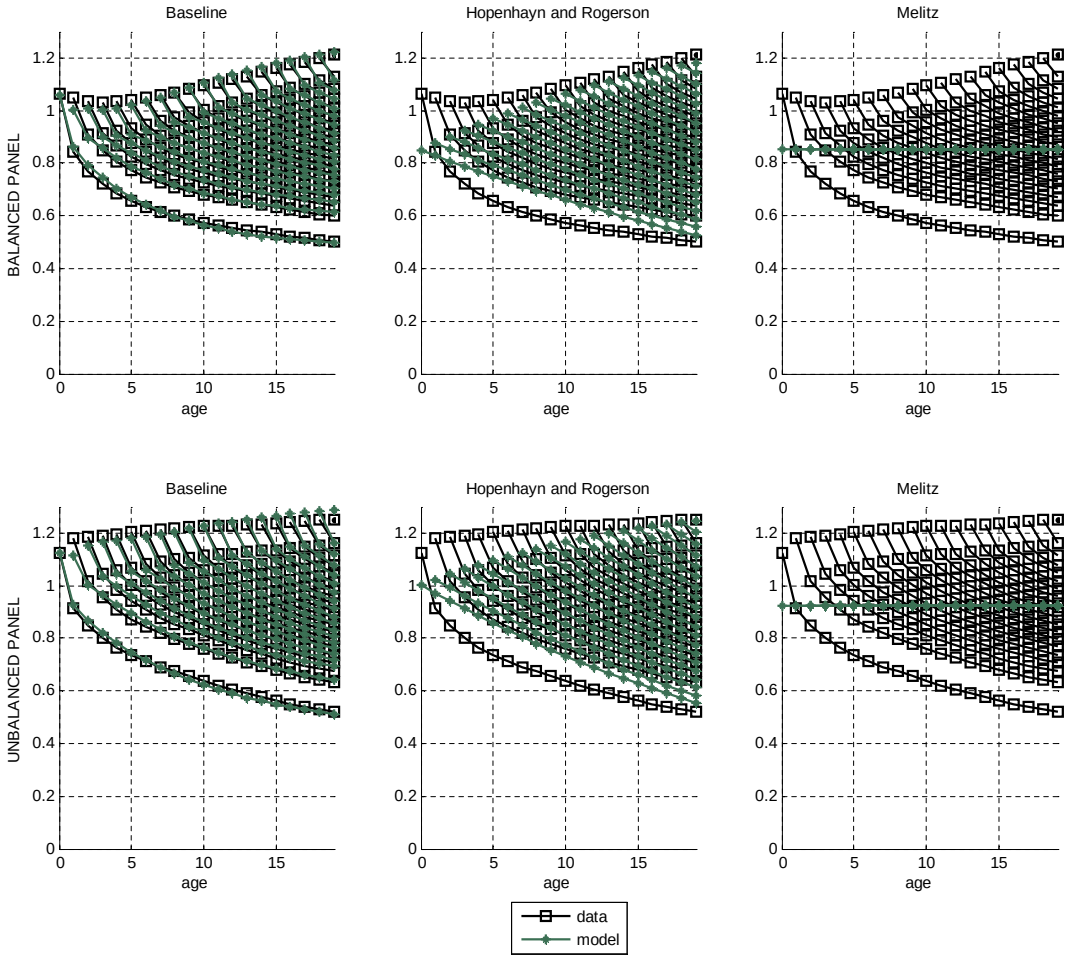


Figure 2: Autocovariance matrix: reduced-form model versus data.

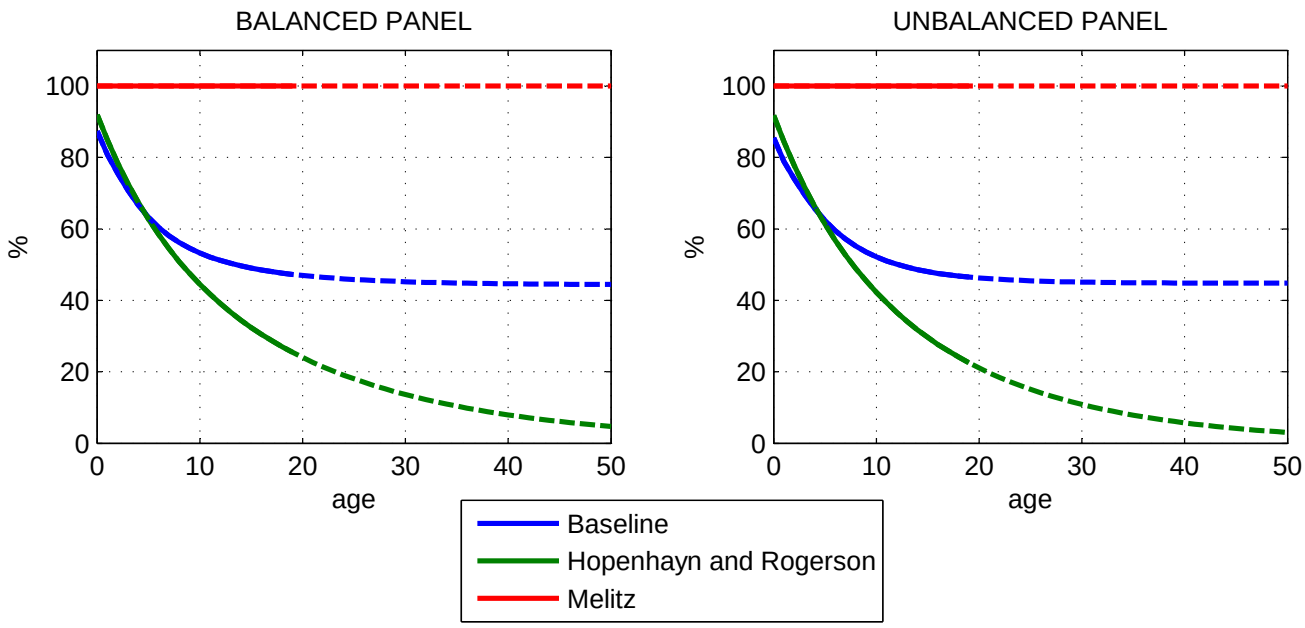


Figure 3: Variance decomposition: contribution of ex-ante heterogeneity to cross-sectional variance of employment by age.

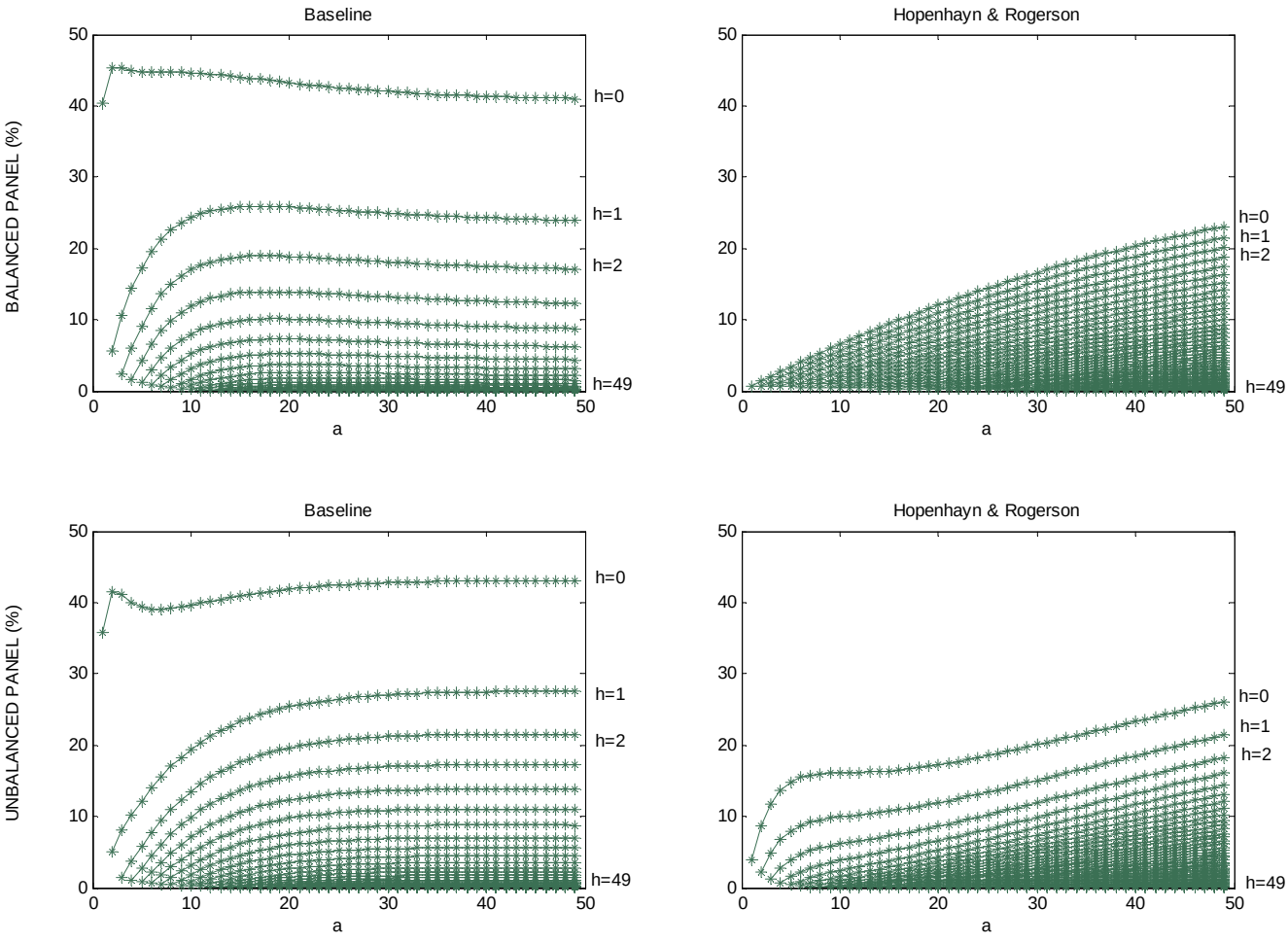


Figure 4: Variance decomposition of employment *growth* between age h and age $a > h$: contribution of ex-ante heterogeneity.

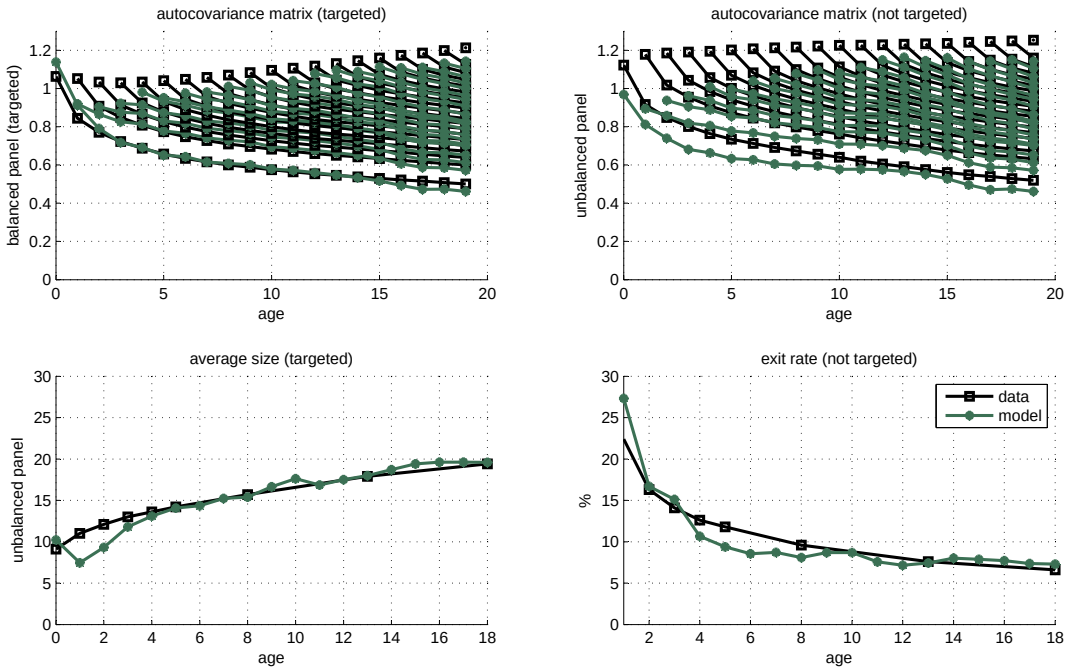


Figure 5: Structural firm dynamics model versus data.

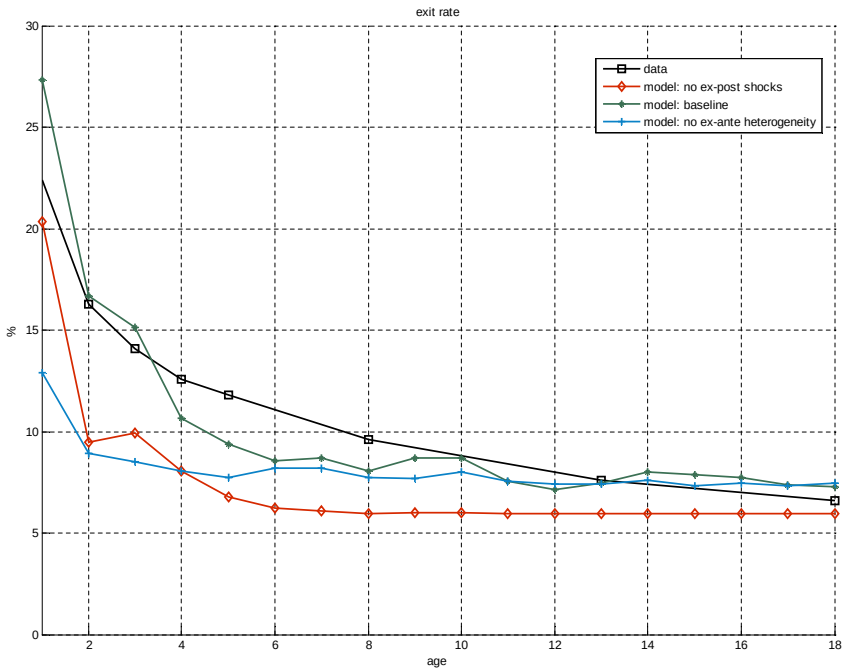


Figure 6: Exit rates

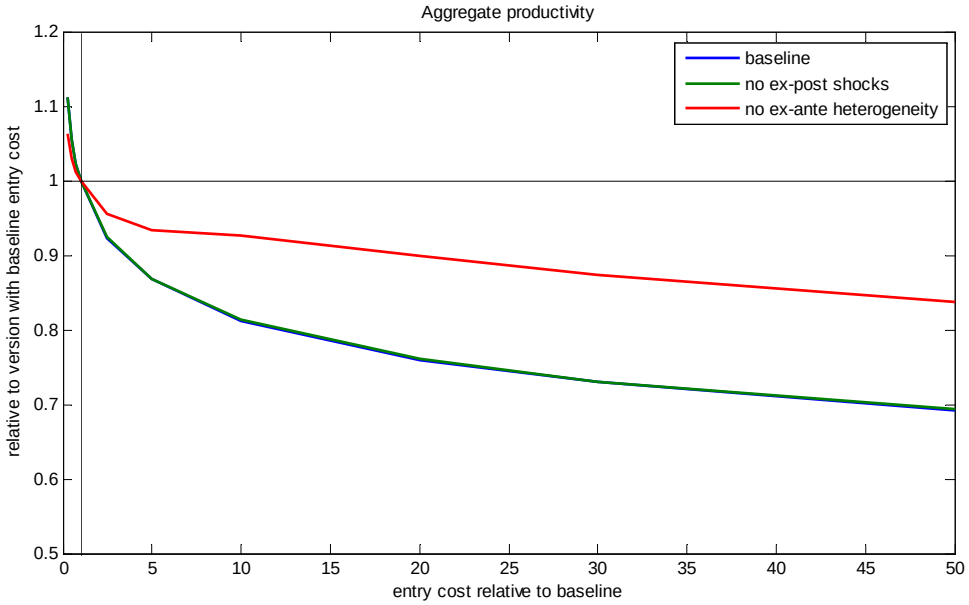


Figure 7: Aggregate productivity as a function of the entry cost. Results are plotted relative to baseline entry cost.

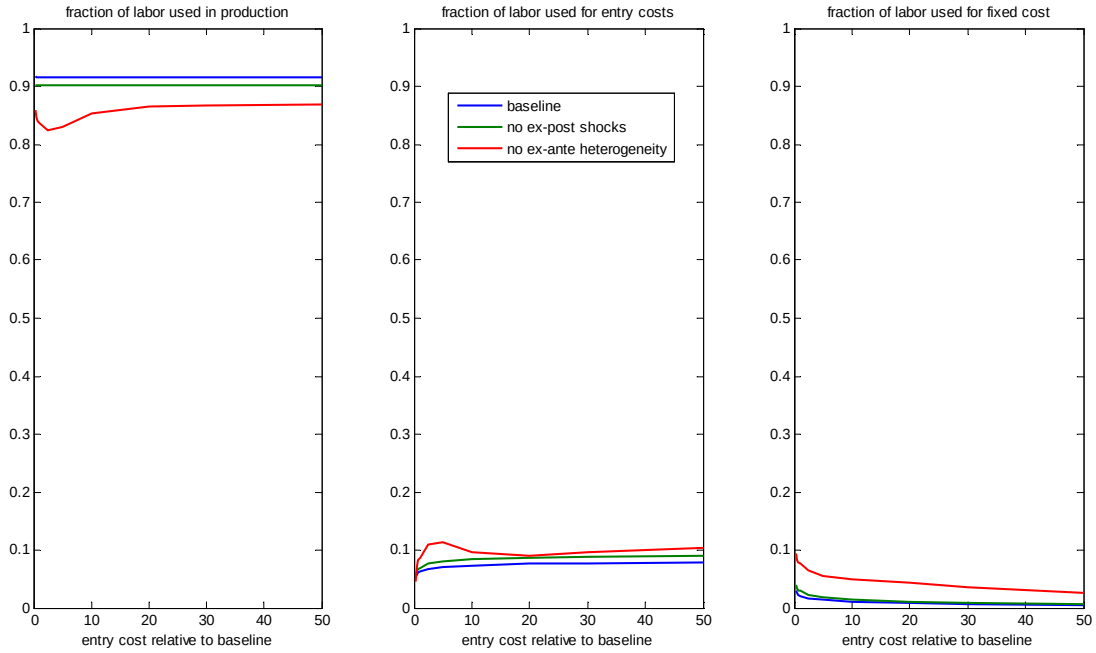


Figure 8: Allocation of labor as a function of the entry cost.