

Endogenous Second Moments: A Unified Approach to Fluctuations in Risk, Dispersion, and Uncertainty*

Ludwig Straub

MIT

Robert Ulbricht

Toulouse School of Economics

March 7, 2018

Abstract

Many important statistics in macroeconomics and finance—such as cross-sectional dispersions, risk, volatility, or uncertainty—are *second moments*. In this paper, we explore a mechanism by which second moments naturally and endogenously fluctuate over time as *nonlinear* transformations of fundamentals. Specifically, we provide general results that characterize second moments of transformed random variables when the underlying fundamentals are subject to distributional shifts that affect their means, but not their variances. We illustrate the usefulness of our results with a series of applications to (1) the cyclical nature of the cross-sectional dispersions of macroeconomic variables, (2) the dispersion of MRPKs, (3) security pricing, and (4) endogenous uncertainty in Bayesian inference problems.

Keywords: Cross-sectional dispersion, endogenous uncertainty, monotone likelihood ratio property, nonlinear transformations, risk, second moments, volatility.

JEL Classification: C19, D83, E32, G13.

*We would like to thank Harry Di Pei and Christian Hellwig for valuable comments. This paper supersedes an earlier working paper that circulated in 2013 under the title “Endogenous Second Moments: Variance Transformation Theorems and Applications”. Straub acknowledges financial support from the Macro-Financial Modeling Group. Ulbricht acknowledges financial support from the Horizon 2020 Program under grant agreement No. 649396. *Email Addresses:* straub@mit.edu, robert.ulbricht@tse-fr.eu.

1 Introduction

Many important statistics in macroeconomics and finance—such as cross-sectional dispersions, risk, volatility, or uncertainty—are *second moments*. For example, dispersions can be measured by the cross-sectional variance, risk by the variance across future states with an objective probability measure, volatility by the variance of the realized path over time, and uncertainty by the variance across unknown states with respect to a possibly subjective probability measure. The recent financial crisis was a stark reminder that these second moments are nowhere near constant over the business cycle and allowing them to vary can help explain the economic fluctuations during the crisis.¹

In this paper, we explore a mechanism by which second moments can naturally and endogenously fluctuate across states or time, when economically interesting variables are *nonlinear* transformations of some fundamental. Specifically, we consider a general setup where there is a fundamental shock θ whose realization is either different across economic regions or agents (dispersion), not realized yet (risk), varies over time (volatility), or unknown to agents (uncertainty). We distinguish between the variance of θ , which defines the second moment of the fundamental, and the variances of certain nonlinear transformations of θ , which characterize the second moments of interest to us.

In standard stochastic business cycle models, for instance, the variables of economic interest are commonly approximated as linear (or linearized) functions of the fundamentals. In order to use these models to explain movements in second moments of endogenous variables one must therefore rely on exogenous shocks to the second moments of fundamentals. Moreover, to capture the apparent cyclicity of the second moments of these variables, the exogenous shocks to second moments need to be correlated with the corresponding first-order shocks to fundamentals. Here, we go another route and consider a setting where some variable of interest, y , is a convex or concave function of the fundamental θ .² We provide general theorems that characterize the behavior of the variance of y as the distribution of θ shifts up or down, *without* changing the fundamental variance of θ . Our framework hence provides a theoretical underpinning for a class of models where a single “first-moment” change in fundamentals causes fluctuations in first *and* second moments in endogenous variables.

Applied contribution. The usefulness of our results is illustrated in a series of four applications. The main application is a stylized business cycle model, in which we take θ as the productivity of distinct economic units (e.g., firms, plants, or regions). We study how aggregate, *variance-preserving* shifts of the distribution of θ across these units (i.e., fluctuations in aggregate productivity) translate into endogenous fluctuations in the cross-sectional dispersion of key macroeconomic quantities, such as output, employment, investment, and Solow residuals. The novel feature in this application is the focus on non-unit elasticities between factor inputs at the firm level, which is key for generating a

¹See, e.g., [Christiano, Motto and Rostagno \(2014\)](#), [Bloom \(2009\)](#) and [Bloom et al. \(2016\)](#) for business cycle theories based on risk or uncertainty shocks. For empirical evidence regarding the cyclicity of second moments, see, e.g., [Bachmann, Elstner and Sims \(2013\)](#), [Berger and Vavra \(2010\)](#), [Higson, Holly and Kattuman \(2002\)](#) and [Kehrig \(2015\)](#).

²Throughout this paper, we follow the convention that “convex” (“concave”) means weakly convex (weakly concave).

non-linear mapping from productivities to economic activity. In line with empirical cross-sectional patterns, the dispersions of output, employment and Solow residuals are shown to be countercyclical when employment and capital are gross complements. A simple calibration of the input elasticity to recent micro-data estimates suggests that this mechanism can account for a significant share of the empirical observed cyclical variation in various dispersion measures—without the need of introducing exogenous “volatility”-shocks.

The second application looks at a prominent proxy used by a recent literature to measure misallocation, the marginal revenue product of capital (MRPK), and explores how it changes with different shocks. We consider a simple model where firms face idiosyncratic borrowing constraints and are hit by idiosyncratic productivity and financial shocks. As expected, financial shocks lead to *counter*-cyclical fluctuations in the cross-sectional dispersion of MRPKs. In contrast, we show that productivity-driven fluctuations lead to *pro*-cyclical dispersions whenever the elasticity of borrowing limits to firm revenues is smaller than one.

The third application is a simple security market model where we explore how the comovement of a security’s risk with the underlying fundamental depends on the security’s payoff profile. In this context, we generalize the following two well-known results for a general class of underlying risk distributions: (i) concavely increasing securities (e.g., corporate debt) have return risk that is countercyclical to the underlying state of the corporation; and (ii) convexly decreasing securities (e.g., European Put options) have procyclical return risk.

Our final application illustrates how uncertainty in Bayesian inference problems can vary endogenously when signal structures have some degree of “non-linearity”. In particular, we study a set-up where agents receive a signal about some non-linear transformation $g(\theta)$ of the fundamental θ . Such non-linear transformation arises, for instance, when agents learn from financially constrained firms about their business potential. In this setup, when θ realizes in a range where g tends to be rather flat (e.g., firms are constrained), the signal endogenously loses some of its information content. This way, posterior uncertainty fluctuates with the realization of the signal and is thus determined by both, the fundamental θ and the exogenous noise in the signal itself.

Together, these applications illustrate how our results can be used to provide a unified perspective to explain endogenous fluctuations in risk, dispersion, and uncertainty. While Application 3 is fairly standard, Applications 1, 2, and 4 describe new perspectives on recent results in their respective literatures. For example, Application 1 contributes to the recent literature on the cyclicity of the dispersion of firm- or plant-level statistics.³ In particular, it complements [Ilut, Kehrig and Schneider \(2016\)](#) who develop a model where asymmetries in hiring and firing based on ambiguity aversion microfound nonlinear responses of firms to changes in productivities. These nonlinearities play a similar role to the nonlinearities emerging in our application from a non-unit elasticity between

³The cyclicity of the dispersions of firm- or plant-level variables such as output, productivity, employment, or investment is documented by [Bachmann, Elstner and Sims \(2013\)](#), [Bachmann and Bayer \(2013, 2014\)](#), [Berger and Vavra \(2010\)](#), [Bloom et al. \(2016\)](#), [Cui \(2014\)](#), [Döpke et al. \(2005\)](#), [Döpke and Weber \(2010\)](#), [Gourio \(2008\)](#), [Higson, Holly and Kattuman \(2002\)](#); [Higson et al. \(2004\)](#) and [Kehrig \(2015\)](#) among others.

factor inputs in that they introduce cyclical shifts in the dispersions of firm aggregates. Application 2 contributes to recent studies on the cyclicity of misallocation of capital (see, e.g., [Kehrig 2015](#) and [Gopinath et al. 2015](#)), in that it discusses—based on a simple static framework with heterogeneous borrowing constraints—under what conditions the dispersion of the marginal revenue product of capital can be expected to be cyclical. Finally, Application 4 introduces a general result about the cyclicity of Bayesian uncertainty when signals are nonlinear functions of the economy’s fundamental. Theories about the cyclicity of uncertainty have recently gained attention by a number of authors in the face of growing evidence and potential relevance of changes in uncertainty (see, e.g., [van Nieuwerburgh and Veldkamp 2006](#); [Orlik and Veldkamp 2014](#); [Straub and Ulbricht 2012, 2014](#); [Fajgelbaum, Schaal and Taschereau-Dumouchel 2017](#)).

Theoretical contribution. To be as broadly applicable as possible, the theory part of our paper is kept in a general and abstract form: Letting X and Y be real-valued random variables with equal variance, we compare the variances of $g(X)$ and $g(Y)$, where g is a monotone and convex or concave function.

Our results are easiest seen in the special cases where Y is either strictly larger than X —by which we mean the support of Y *strictly* exceeds the support of X without overlap—or when the distribution of Y is simply a positive translation of the distribution of X . Then as one would expect, the variance of $g(Y)$ (weakly) exceeds the variance of $g(X)$ if g is convexly increasing or concavely decreasing. However, these cases are quite special and may not translate well to real world examples.⁴

A natural question hence is whether and when these results carry over to the general case where the supports of X and Y are allowed to overlap and where the distributions do not have the same parametric shape. As it turns out, the answer is less straightforward than one may think. For instance, it is possible to construct examples where Y first-order stochastically dominates X —that is, the cumulative distribution function of Y is strictly below the one of X —yet the variance of $g(Y)$ is smaller than the one of $g(X)$, despite g being an increasing convex function (and X and Y having equal variance).

The main theoretical result in the paper states that when Y dominates X according to the monotone likelihood ratio property (MLRP)—that is, the ratio of the densities of Y and X is increasing—but shares the same variance as X , then the variance of $g(Y)$ indeed exceeds the variance of $g(X)$ if g is convexly increasing or concavely decreasing.⁵ This gives us a precise notion of when positive shifts in an underlying fundamental, say θ , translate into positive shifts to the second moment of a transformed variable $g(\theta)$, namely when g is convexly increasing or concavely decreasing.

⁴For instance, we expect that a positive shock to the aggregate productivity of an economy would typically not leave the exact shape of the productivity distribution unchanged, but may disproportionately affect firms at certain points of the productivity distribution. Similarly, the cross-sectional productivity distribution is very likely to always have some overlap.

⁵Analogously, it holds that the variance of $g(X)$ exceeds the one of $g(Y)$ when g is concavely increasing or convexly decreasing.

In addition to the non-parametric characterization in our main result, we provide a simple elementary proof for the above mentioned case where Y is a translation of X . Building on this proof, we further extend our results to the case where X and Y are not necessarily similar in shape yet their transformations $g(X)$ and $g(Y)$ are linked via an affine-linear relation⁶ with $\mathbb{E}g(Y) > \mathbb{E}g(X)$.

It is worth noting here that our results have nothing in common with Jensen’s inequality.⁷ Instead, the correct analogy is to the known transformation properties of first moments. Specifically, it is well-known that two random variables ordered by MLRP have means ordered the same way, and this property is preserved under increasing transformations. In analogy to these known results on first moments, our results imply that for two MLRP-ordered random variables, the order of variances is preserved under increasing convex and decreasing concave transformations

In relation to the literature, these results substantially extend existing theoretical results on the transformation behavior of variances under monotone concave or convex transformations. To the best of our knowledge, the most general predecessor of our results is found in [Bartoszewicz \(1985\)](#). In the above notation, [Bartoszewicz \(1985, Theorem 1\)](#) proves that if Y dominates X according to first order stochastic dominance (FOSD) and a convex stochastic order ([Van Zwet, 1964](#))⁸, then a similar result to ours holds, namely that $\text{Var } g(Y) > \text{Var } g(X)$ for any convexly increasing function g . Compared to [Bartoszewicz \(1985\)](#), our main result has the advantage that our stochastic orders (MLRP and $\text{Var}X \leq \text{Var}Y$) are in most applications easy to check while it can be hard to work with the convex stochastic order in [Bartoszewicz \(1985\)](#), especially in cases where Y is of a different parametric shape than X (i.e., Y is not a simple translation of X).

Layout. The layout of this paper is as follows: In [Section 2](#) we first introduce the necessary mathematical setup and then show how it can be used to prove our theoretical results. [Section 3](#) develops the main application on cross-sectional dispersions. [Section 4](#) contains the remainder applications. [Section 5](#) concludes. All formal proofs are contained in our mathematical Appendices [A](#), [B](#), and [C](#), where we prove the main results.

2 Variance transformation theorems

In this section, we present our formal results. We start out by explaining our main result in [Section 2.1](#), based on an MLRP stochastic order. Then, in the subsequent [Section 2.2](#), we explore two alternative stochastic orders and the variance transformation results they imply. In [Section 2.3](#) we

⁶That is, $g(Y) = \alpha_1 + \alpha_2 g(X)$ for real numbers α_1, α_2 .

⁷Jensen’s inequality states that under a convex transformation, the mean of $g(X)$ exceeds g applied to the mean of X , $\mathbb{E}g(X) > g(\mathbb{E}X)$. Apart from pertaining to variances, our results also differ in that they compare statistical measures of a transformed variable, $g(X)$, to the same statistical measures of *another* transformed variable, $g(Y)$. Jensen’s inequality compares a statistical measures of a transformed variable, $g(X)$ to the transformation of the statistical measures of the untransformed variable, X .

⁸The convex stochastic order that is meant here requires that $F_X^{-1} \circ F_Y$ be convex (well-defined if F_X is indeed invertible). Notice that this is *different* from the MLRP.

present two corollaries to our results. Section 2.4 presents simple parametric examples and discusses how they fit the assumptions of our theorems.

Consider the following set-up. Let X, Y be two real-valued and univariate random variables, defined over a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Denote by F_X and F_Y their cumulative distribution functions. Assume that there exists a measure μ over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with respect to which the distributions of X and Y admit density functions f_X and f_Y . Here, $\mathcal{B}(\mathbb{R})$ denotes the Borel sigma algebra.

The idea behind our results is to ask what order we can expect between $\text{Var}\{g(X)\}$ and $\text{Var}\{g(Y)\}$, depending on the function g . To answer this question in a meaningful way, the assumptions we make on the stochastic order between X and Y naturally have to involve the second or higher moments of X and Y themselves. For example, even if we were to assume that Y lies strictly above X without overlap in their respective supports, there is no way of ranking $\text{Var}\{g(X)\}$ and $\text{Var}\{g(Y)\}$ without any information on the variances of X and Y (or other measures of dispersion). For this reason, all of our stochastic orders will assume a weak ordering of the variances of X and Y .⁹

2.1 Main result: Variance transformation under MLRP

For our main result, we assume that X and Y are stochastically ordered in the following sense.

Assumption 1. (i) X is strictly dominated by Y in the sense of the MLRP; that is, $f_Y(z)/f_X(z)$ is weakly increasing for all z in the support of μ , and $F_X \neq F_Y$. (ii) The variance of X is less than or equal to the variance of Y , that is, $\text{Var}\{X\} \leq \text{Var}\{Y\}$, and both variances are finite and nonzero.

In part (i) of Assumption 1 we demand that X and Y are ordered by MLRP.¹⁰ To stress the importance of MLRP, in Example 1 below we show that our results do not carry over to cases where the random variables only obey a weaker form of stochastic order, e.g. first-order or second-order stochastic dominance. In part (ii) of Assumption 1, we require the variances to be ordered. As our results below will apply for increasing convex (and similarly decreasing concave) transformations, this is the direction of the inequality we need. If $\text{Var}\{X\} > \text{Var}\{Y\}$, it is straightforward to construct counterexamples where the results do not hold. However, in that case, an analogous theorem applies for increasing concave (and decreasing convex) transformations.

Our notation is general enough to nest both *continuous* and *discrete* random variables. If μ is equal to the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, f_Y and f_X are common density functions of continuous distributions over \mathbb{R} . For example, when X and Y are normally distributed with equal variance, $X \sim \mathcal{N}(m_X, \sigma^2)$ and $Y \sim \mathcal{N}(m_Y, \sigma^2)$ with $m_Y > m_X$, they satisfy Assumption 1. Choosing μ to have discrete support allows X and Y to be discrete. For example, a measure¹¹ $\mu = \sum_{i=1}^N \frac{1}{N} \delta_{z_i}$ puts equal weight on numbers $\{z_1, \dots, z_N\}$ (assumed to be in strictly increasing order), in which case

⁹It is worth pointing out here that the lack of basic first or second order dominance among our stochastic orders is coming from the fact that neither provide the right restrictions on X and Y to be able to rank $\text{Var}\{g(X)\}$ and $\text{Var}\{g(Y)\}$ for simple functions g (see Example 1 below).

¹⁰An alternative common characterization of the MLRP is that $F_Y \circ F_X^{-1}$ is (weakly) convex.

¹¹Here, δ_z denotes a Dirac measure which puts point mass 1 on $z \in \mathbb{R}$.

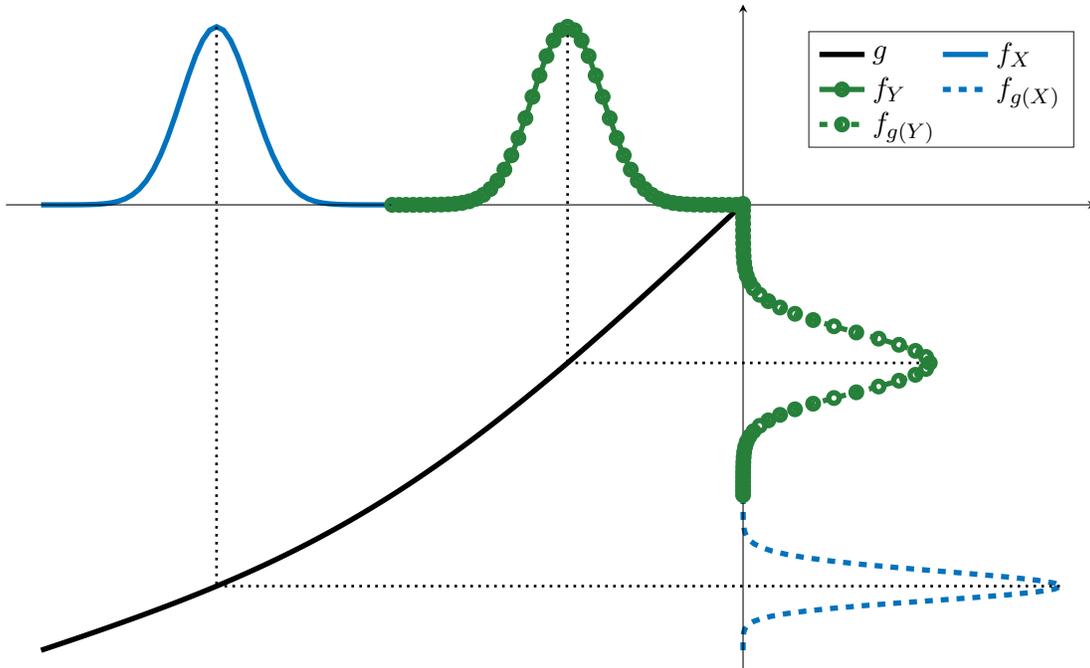


Figure 1: Convex transformation of a translation

Assumption 1(i) demands that $f_Y(z_i)/f_X(z_i) = \mathbb{P}\{Y = z_i\}/\mathbb{P}\{X = z_i\}$ be weakly increasing in i . In this discrete setup, standard formulae for the variances of discrete distributions can be used to impose Assumption 1(ii).¹²

Main result We now present our main transformation result for variances. The idea behind it is as follows. For two random variables X and Y which are stochastically ordered, we show that under strictly convex increasing transformations $g : \mathbb{R} \rightarrow \mathbb{R}$,¹³

$$\text{Var}\{X\} \leq \text{Var}\{Y\} \Rightarrow \text{Var}\{g(X)\} < \text{Var}\{g(Y)\}.$$

Figure 1 gives the straightforward intuition behind this result. Y exceeds X stochastically and therefore Y most likely realizes in regions where the slope of g is larger than for most realizations of X . This follows from the assumption that g be convex and increasing. With “on average” higher slope for Y , even if Y and X share the same variance, $g(Y)$ will have a larger variance than $g(X)$.

When X and Y are ordered according to the MLRP, we find the following result.

Theorem 1. *Suppose Assumption 1 holds. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and strictly increasing*

¹²Note that our definition of MLRP is *well-defined*, in the sense that it is independent of which measure μ one takes to construct the densities, as long as the distributions of X and Y are both absolutely continuous with respect to μ . For instance, in the above discrete example, introducing different weights on the numbers in the support $\{0, 1, 2\}$, merely causes proportional shifts in the densities, such that their ratios, $f_X(z)/f_Y(z)$, do not change.

¹³Throughout this paper, we use the term “strictly convex” to describe an a.e. differentiable function whose derivative is a.e. strictly increasing. (Here, “a.e.” is with respect to the Lebesgue measure.)

function. Then,

$$\text{Var}\{g(X)\} \leq \text{Var}\{g(Y)\}, \quad (1)$$

when these variances are finite. This holds with strict inequality if g is strictly convex.

To illustrate Theorem 1 consider the following example. If the support of X is strictly dominated by the support of Y (no overlap) and there is some point $z \in \mathbb{R}$ in the gap between the two supports, then it is straightforward to show that¹⁴

$$\text{Var}\{g(Y)\} \geq (g'(z))^2 \text{Var}\{Y\} \geq (g'(z))^2 \text{Var}\{X\} \geq \text{Var}\{g(X)\}. \quad (2)$$

The intuition behind the example is that the slope of g (which multiplies the variance) is greater for all possible realizations of Y than it is for the realizations of X . Clearly, this example is very sensitive to the assumption of non-overlapping supports. Still, one might wonder, why our proof of Theorem 1 is substantially more subtle than (2). The key to see why lies in understanding that the case where the supports of X and Y do overlap requires some notion of the slope of g being “on average” greater for realizations of Y than for realizations of X . Because Y is not necessarily larger than X for *any* possible realization—only stochastically so—we need to make sure Y dominates X in the “right” way. The following example illustrates this point, showing that Theorem 1 in fact fails if we relax the assumption that Y dominates X in the sense of MLRP, to the weaker notion of first and second order stochastic dominance.

Example 1. Let $\mu = \delta_0 + \delta_1 + \delta_2$, where δ_x is the Dirac measure with a point mass at x , and let X, Y be given by $f_X(0) = 1 - \epsilon$, $f_X(1) = 0$, $f_X(2) = \epsilon$ and $f_Y(0) = (1 - \epsilon)/2$, $f_Y(1) = (1 - \epsilon)/2$, $f_Y(2) = \epsilon$. Clearly, Y dominates X in the sense of both First and Second Order Stochastic Dominance (FOSD and SOSD). Assume $\epsilon = 0.1$. Then, the variances of X and Y are given by $\text{Var}\{X\} = 0.36$ and $\text{Var}\{Y\} = 0.4275$. In particular, $\text{Var}\{Y\} \geq \text{Var}\{X\}$.

Now define the convex function $g(x) = \max\{x, \lambda(x - 1) + 1\}$ with $\lambda > 1$. We can compute that

$$\text{Var}\{g(Y)\} - \text{Var}\{g(X)\} = 0.1575 - 0.09\lambda.$$

This shows that for sufficiently large values for λ , the variance of $g(X)$ will actually exceed the one of $g(Y)$.

Example 1 considers the case where Y does not dominate X in the sense of MLRP, but rather in the weaker senses of both FOSD and SOSD. In that case, Theorem 1 is shown to not necessarily hold.

This example is more general: Suppose X and Y are discrete and Y has a greater mean, $\mathbb{E}X < \mathbb{E}Y$. Suppose further that both have the same right tail, that is, $f_X(z) = f_Y(z)$ for all $z \geq \bar{z}$ where \bar{z} is in the support of either X or Y , and $\bar{z} > \mathbb{E}Y$. In that case, applying a convex kink function

¹⁴For this example, we tacitly assume that g is differentiable at z .

$g(x) = \max\{x, \lambda(x - \bar{z}) + 1\}$ like before with sufficiently large λ will ultimately lead $\text{Var}\{g(X)\}$ to be greater than $\text{Var}\{g(Y)\}$. This shows that it is not enough that the distribution of Y lies “to the right” of that of X over a subset of the support (e.g. below \bar{z}); in fact, what is needed is that Y is larger than X over the entire support. This is what is satisfied if Y MLRP-dominates X but not if Y merely dominates X in the sense of FOSD.

We now outline the main steps in the proof of Theorem 1. The actual proof can be found in Appendix A. For the proof, we derive the variance inequality (1) from two covariance inequalities:

$$\begin{aligned} \text{Cov}(X, X) \leq \text{Cov}(Y, Y) &\Rightarrow \text{Cov}(g(X), X) \leq \text{Cov}(g(Y), Y) \\ &\Rightarrow \text{Cov}(g(X), g(X)) \leq \text{Cov}(g(Y), g(Y)). \end{aligned} \quad (3)$$

Both steps are similar: Only one argument in the covariances changes at a time, while the fixed ones, e.g. X and Y in the first step, share an MLRP relationship.¹⁵ To see why these steps hold specifically when Y MLRP-dominates X , consider the first one as an example, with X, Y continuous random variables. Denoting by m_X and m_Y the means of X and Y , we can write

$$\text{Cov}(g(Y), Y) - \text{Cov}(g(X), X) = \int_{\mathbb{R}} g(x)(x - m_Y)f_Y(x) dx - \int_{\mathbb{R}} g(x)(x - m_X)f_X(x) dx.$$

The right hand side of this can be captured with a single integral and integrated by parts,

$$\begin{aligned} \int_{\mathbb{R}} g(x) \left((x - m_Y) \frac{f_Y(x)}{f_X(x)} - (x - m_X) \right) f_X(x) dx \\ = - \int_{\mathbb{R}} g'(x) \int_{-\infty}^x \left((z - m_Y) \frac{f_Y(z)}{f_X(z)} - (z - m_X) \right) f_X(z) dz dx. \end{aligned}$$

Interestingly, the first factor of the integrand of the interior integral $(z - m_Y) \frac{f_Y(z)}{f_X(z)} - (z - m_X)$ is first positive, then negative, then positive, precisely because $m_Y > m_X$ and $f_Y(z)/f_X(z)$ increases in z due to the MLRP. This means the entire interior integral can only cross 0 at a given $x_0 \in \mathbb{R}$. Convexity of g means g' is monotone and therefore this expression is bounded below by

$$-g'(x_0) \int_{\mathbb{R}} \int_{-\infty}^x \left((z - m_Y) \frac{f_Y(z)}{f_X(z)} - (z - m_X) \right) f_X(z) dz dx,$$

which, following the exact same steps backwards, is equal to $g'(x_0) (\text{Cov}(Y, Y) - \text{Cov}(X, X))$ and thus greater or equal to zero.

2.2 Two alternative stochastic orders

We now allow for two alternative stochastic orders and prove corresponding variance transformation theorems. Here, our first result, Theorem 2, formally establishes a commonly used alternative to our

¹⁵In the actual proof, we utilize this commonality and prove an auxiliary result that applies to both steps.

main result; our second result, Theorem 3, establishes a natural extension of it.

As our first alternative stochastic order, we consider the commonly used case where Y is an affine linear function of X . In this case, the assumptions on X and Y are as follows.

Assumption 2. (i) Y is an affine-linear function of X , $Y = \alpha_1 X + \alpha_2$, where $\mathbb{E}Y > \mathbb{E}X$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. (ii) The variance of X is less than or equal to the variance of Y , that is, $\text{Var}\{X\} \leq \text{Var}\{Y\}$, and both variances are finite and nonzero.

The first part of Assumption 2 demands Y to be an affine linear function of X , but with a higher mean so as to ensure that Y is larger than X on average. The case where Y is a simple positive translation of X , i.e. $\alpha_1 = 1$ and $\alpha_2 > 0$, is naturally covered. In fact, notice that Assumption 2(ii) is also satisfied for this case, since under a translation, $\text{Var}\{Y\} = \text{Var}\{X\}$. Assumption 2(ii) is slightly more general than that by allowing for Y to have a larger variance than X .

As our second alternative order, we consider a variation of Assumption 2, where the affine-linear transformation assumption is not imposed on X and Y itself; but rather after applying the transformation $g : \mathbb{R} \rightarrow \mathbb{R}$ to X and Y , under which one is interested in studying the variance transformation behavior.

Assumption 3. (i) $g(Y)$ is an affine-linear function of $g(X)$, that is, $g(Y) = \alpha_1 + \alpha_2 g(X)$, with $\mathbb{E}\{g(Y)\} > \mathbb{E}\{g(X)\}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. (ii) The variance of X is less or equal to the variance of Y ; that is, $\text{Var}\{X\} \leq \text{Var}\{Y\}$, and both variances are finite and nonzero.

The essential part of Assumption 3 is again part (i). It requires that $g(Y)$ and $g(X)$ be similar in shape, in the sense that $g(Y)$ is merely an affine-linear transformation of $g(X)$.

Results under the two alternative stochastic orders We now prove that a similar result to Theorem 1 holds under our two alternative stochastic orders. We start with the case where Y is a simple positive affine-linear translation of X .

Theorem 2. Suppose Assumption 2 holds. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and strictly increasing function. Then,

$$\text{Var}\{g(X)\} \leq \text{Var}\{g(Y)\}, \tag{4}$$

when these variances are finite. This holds with strict inequality if g is strictly convex.

The idea behind Theorem 2 is simple: Since g is convex, translating a distribution X to the right increases the ‘‘average slope’’ g has over the support of the distribution. This increases the variance. Clearly, when Y is not only a translation of X , but also wider, that is, $\alpha_1 > 1$ in Assumption 2, this can only increase the variance of $g(Y)$ compared to the one of $g(X)$. The formal proof of Theorem 2, which is relegated to Appendix B, proceeds along these lines and studies the function $G_{g,X}(\beta_1, \beta_2) \equiv \text{Var}\{g(\beta_1 + \beta_2 X)\} - \text{Var}\{g(X)\}$, formally defined in Appendix B. The subscripts g and X of $G_{g,X}$ emphasize the dependance on g and X . Notice that by construction, $G_{g,X}(0, 1) = 0$

and $G_{g,X}(\alpha_1, \alpha_2) = \text{Var}\{g(Y)\} - \text{Var}\{g(X)\}$. Since $G_{g,X}$ is shown to increase as we vary (β_1, β_2) from $(0, 1)$ to (α_1, α_2) , (4) follows.

Finally, we consider the case where $g(X)$ and $g(Y)$ are ordered in an affine-linear way. The same conclusion holds there.

Theorem 3. *Suppose Assumption 3 holds, with $g : \mathbb{R} \rightarrow \mathbb{R}$ a convex, strictly increasing function. Then,*

$$\text{Var}\{g(X)\} \leq \text{Var}\{g(Y)\}, \quad (5)$$

with strict inequality if g is strictly convex.

Theorem 3 is a simple variation of Theorem 2. Rather than parametrically linking the distributions of X and Y , it does so for $g(X)$ and $g(Y)$. Hence, it is suitable for applications where it is not X and Y that stem from a “similar” parametric class of distributions (e.g. normal distributions) but rather $g(X)$ and $g(Y)$. Example 2 illustrates this.

Example 2. Let X, Y be such that $g(X) \sim \mathcal{N}(\mu_{g(X)}, \sigma_{g(X)}^2)$ and $g(Y) \sim \mathcal{N}(\mu_{g(Y)}, \sigma_{g(Y)}^2)$, with $\mu_{g(X)} < \mu_{g(Y)}$. The question Theorem 3 asks is what we can say about the relationship between $\sigma_{g(Y)}^2$ and $\sigma_{g(X)}^2$ if we know that $\text{Var}\{X\} \leq \text{Var}\{Y\}$. The answer is that $\sigma_{g(X)}^2 < \sigma_{g(Y)}^2$ for strictly convex g .

Similar to the proof of Theorem 2, the idea for the proof of Theorem 3 is to study the properties of the function $G_{g^{-1},g(X)}(\beta_1, \beta_2) = \text{Var}\{g^{-1}(\beta_1 + \beta_2 g(X))\} - \text{Var}\{X\}$. Again, $G(0, 1) = 0$, but now it also holds by assumption that $G(\alpha_1, \alpha_2) \geq 0$. To prove (5) it needs to be shown that $\alpha_2 > 1$, which follows by proving that $G(\alpha_1, \alpha_2)$ can only be larger than $G(0, 1)$ if the second argument increased. The formal proof is in Appendix C.

2.3 Corollaries

All three results, Theorems 1, 2, and 3, have a common straightforward corollary.

Corollary 1. *Suppose either Assumption 1(i), Assumption 2(i), or Assumption 3(i) holds.*

1. *If $\text{Var}\{X\} \leq \text{Var}\{Y\}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a concave, decreasing function, then $\text{Var}\{g(X)\} \leq \text{Var}\{g(Y)\}$.*
2. *If $\text{Var}\{X\} \geq \text{Var}\{Y\}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a concave, increasing function, then $\text{Var}\{g(X)\} \geq \text{Var}\{g(Y)\}$.*
3. *If $\text{Var}\{X\} \geq \text{Var}\{Y\}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a convex, decreasing function, then $\text{Var}\{g(X)\} \geq \text{Var}\{g(Y)\}$.*

Here, all variances are assumed to be finite. The respective second inequalities hold strictly if g is strictly convex or concave.

These generalizations are simple consequences of Theorems 1, 2, and 3 that can be obtained by changing the signs of X and Y , or changing the sign of g , or both.

It is also worth noting that our results can also be combined in a straightforward manner. For example, Theorems 1 and 2 jointly imply the following corollary.

Corollary 2. *Suppose that (i) Y is an affine-linear function of a random variable \tilde{Y} , which MLRP-dominates X ; (ii) $\mathbb{E}Y > \mathbb{E}\tilde{Y}$; and (iii) $\text{Var}\{\tilde{Y}\} \geq \text{Var}\{Y\} \geq \text{Var}\{X\}$. Then,*

$$\text{Var}\{g(X)\} \leq \text{Var}\{g(\tilde{Y})\}.$$

This corollary is one of the most general results on the behavior of variances under nonlinear transformations. It includes the standard case where Y is a positive affine-linear shift of X , but frees Y from the necessity of having exactly the same (shifted) shape as X .

2.4 Simple examples

Each of the three theorems has its own set of examples. Here, we go over three examples that illustrate when which of the theorems can be applied. The examples show that Theorem 1 is applicable in a wide range of settings since it does not require a highly special *linear* relationship between X, Y or $g(X), g(Y)$.

Example 3. Suppose X, Y are normal distributions with equal variance and ordered means, $\mathbb{E}Y > \mathbb{E}X$. In that case, Theorems 1 and 2 apply for any convex or concave, monotone function g .

Example 4. Suppose X, Y are log-normal distributions with equal variance and ordered means, $\mathbb{E}Y > \mathbb{E}X$. In that case, Theorem 1 applies for any convex or concave, monotone function g . Theorem 3 applies only if g is equal to \log (up to scale and a constant intercept).

Example 5. Suppose X, Y are Pareto distributions with equal scale parameter or exponential distributions. Denote by ξ_X, ξ_Y their respective tail parameters. If $\xi_X > \xi_Y > 0$, that is Y has a fatter tail than X , then Y MLRP-dominates X and $\text{Var}(Y) > \text{Var}(X)$. Again, we can use Theorem 1 to rank the variances of any convex or concave, monotone function g .

For what kind of distributions does Theorem 1 not apply? Whenever X and Y have bimodal distributions, but are linearly related as in Assumption 2 or Assumption 3, Theorem 1 does generally not apply even though the variances of $g(X)$ and $g(Y)$ are still ordered (Theorem 2 or Theorem 3).

3 Main application: Cross-sectional dispersion over the business cycle

Recently, the cyclicity of the dispersions of macroeconomic variables has gained a renewed interest. For example, the dispersions of plant-level output, employment growth and Solow residuals have been

shown to be countercyclical, while investment dispersion is procyclical.¹⁶ Here we link these facts to simple shifts in the distribution of idiosyncratic productivities after applying one of Theorems 1–3. The key element is a CES firm-level production function with an elasticity of capital-labor substitution different from one, which generates a non-linear mapping from firm-level productivities to production. We consider first a stylized, partial-equilibrium model that allows for analytical results. In the subsequent section, we enrich the model and use it to explore the quantitative implications.

3.1 Partial Equilibrium Framework

Assume there is a continuum of firms, labeled by $i \in [0, 1]$, with CES production technology

$$y_{it} = e^{a_{it}} f(k_{it}, n_{it}), \quad (6)$$

where $f(k, n) = (\alpha k^{(\sigma-1)/\sigma} + (1-\alpha)n^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)}$. Here, a_{it} is the realization of a firm-specific (log) productivity shock; k_{it} and n_{it} are firm i 's choices for capital and labor, taken as given the paths for interest rates $\{r_t\}$ and wages $\{W_t\}$; $\sigma > 0$ is the (firm-level) elasticity of substitution between capital and labor; and $\alpha \in (0, 1)$. The investment to capital ratio, or “investment rate”, ι_{it} is chosen at t and the law of motion of capital is $k_{it} = (1 + \iota_{it-1} - \delta)k_{it-1}$, with depreciation rate $\delta > 0$. To allow for an analytical characterization, we assume (for now) that interest rates and wages are exogenous at constant levels, $r_t = r > 0$ and $W_t = W > 0$, and that the distribution of productivities a_{it} is independent of a_{it-1} and drawn from a time-varying distribution \mathcal{F}_t .

The optimization problem of firm i in period τ is thus

$$v_{i\tau} \equiv \max_{\{n_{it}, k_{it+1}\}} \mathbb{E}_\tau \sum_{t=\tau}^{\infty} \frac{Q_t}{Q_\tau} \pi_{it}, \quad (7)$$

where profits are given by

$$\pi_{it} = y_{it} - Wn_{it} + (1 - \delta)k_{it} - k_{it+1} \quad (8)$$

with discount factor $Q_t \equiv (1 + r)^{-t}$ and y_{it} given by (6). Since labor is perfectly flexible, and the distribution of a_{it} is the same across i , the capital stock is the same across i ; that is, $k_{it} = k_t > 0$ for some $\{k_t\}$. The first order condition for labor is then,

$$f_n(1, n_{it}/k_t) = e^{-a_{it}} W. \quad (9)$$

Notice that for $\sigma < 1$, the left-hand side is bounded above by $(1 - \alpha)^{\sigma/(\sigma-1)}$ so that firms with a productivity smaller than $a_{it} \leq -\frac{\sigma}{\sigma-1} \log(1 - \alpha) + \log W$ set $n_{it} = 0$ resulting in zero production. In the following, we assume that these firms exit the market, focusing our analysis on active firms.¹⁷

¹⁶See, e.g., Bachmann, Elstner and Sims (2013), Bachmann and Bayer (2013, 2014), Berger and Vavra (2010), Bloom et al. (2016), Cui (2014), Döpke et al. (2005), Döpke and Weber (2010), Gourio (2008), Higson, Holly and Kattuman (2002); Higson et al. (2004) and Kehrig (2015) among others.

¹⁷This could be for instance the result of a marginal operating cost $\epsilon \rightarrow 0$ that accrues when remaining active.

Solving for n_{it} and taking logs, (9) becomes

$$\log n_{it} = \log k_t + g(a_{it} - w) \quad (10)$$

with $w \equiv \log W$ and $g(x) = \log (f_n(1, \cdot)^{-1}(e^{-x}))$.¹⁸ It is easy to verify that g is an increasing function and is strictly concave (strictly convex) whenever $\sigma < 1$ ($\sigma > 1$).¹⁹

Consider the following experiment akin to an aggregate TFP shock: Suppose the distribution of log productivities (among active firms) shifts from \mathcal{F}_t to a distribution $\mathcal{F}_{t'}$ at time t' , which dominates \mathcal{F}_t according to either of Assumptions 1–3(i), but has equal dispersion: $\text{Var}_{\mathcal{F}_t}\{a_{it}\} = \text{Var}_{\mathcal{F}_{t'}}\{a_{it'}\}$. Applying our corollary of Theorems 1–3 to (10) we have the following proposition.

Proposition 1. *In the partial-equilibrium firm-dynamics model, when the cross-sectional distribution of log productivities \mathcal{F}_t increases to a distribution $\mathcal{F}_{t'}$ (with equal dispersion) such that one of Assumptions 1–3(i) holds for \mathcal{F}_t and $\mathcal{F}_{t'}$, then the cross-sectional dispersions of log employment and log output decrease if $\sigma < 1$ and increase if $\sigma > 1$.*

This proposition shows that in our simple model, the cyclicity of the dispersions of log output and log employment is entirely driven by the elasticity of substitution between labor and capital. The intuition behind this result is as follows. Suppose $\sigma < 1$. In that case, the capital and labor are complementary. Thus, since capital is fixed in the short run, labor adjusts relatively more to changes in productivity when it is scarce compared to when it is abundant. The limit case of a Leontieff production technology is especially stark: There, labor is infinitely elastic to productivity when it is scarce, and entirely inelastic when it is abundant. This means the elasticity of labor to productivity declines with rising productivity (or declining wages), or in other words, $\log n_{it}$ is increasing and concave in $\log a_{it}$. This then implies that when log productivities experience an upward shift as specified in Proposition 1, the cross-sectional variance of log employment must fall. The opposite logic applies when $\sigma > 1$.

Proposition 1 makes two statements, one about the cyclicity of the dispersion in log employment and one about the cyclicity of the dispersion in log output. To prove the latter, note that (9) can be rearranged to

$$y_{it} = \left(\frac{W}{1 - \alpha} \right)^\sigma e^{(1-\sigma)a_{it}} n_{it}. \quad (11)$$

Observe that, given the definition of $g(x)$, $x \mapsto (1 - \sigma)x + g(x)$ is also an increasing function that is strictly concave (strictly convex) whenever $\sigma < 1$ ($\sigma > 1$). Therefore the cyclicity of the dispersion of log output is equal to the one of log employment.

What would an observer find for the behavior of the dispersion of log productivities themselves under a shift from \mathcal{F}_t to $\mathcal{F}_{t'}$? Clearly, if the observer knew the true production function and firm-level

¹⁸Here, $f_n(1, \cdot)^{-1}$ is shorthand for the inverse of the function $x \mapsto f_n(1, x)$.

¹⁹Here we assume for simplicity, and broadly in line with aggregate data, that real wages are fully sticky. This is w.l.o.g. in the sense that we can always reinterpret the shift in average productivities to include the cyclical part of wages.

employment and capital stocks, he could recover the true shift in the productivity distribution and find that this dispersion did not change. If, however, the observer were to compute a standard Solow residual—as commonly done in the literature—based on a Cobb-Douglas production function with an empirical capital share of $\beta \in (0, 1)$ he would find

$$a_i^{Solow} = \log y_{it} - \beta \log k_t - (1 - \beta) \log n_{it}.$$

This can be simplified to

$$a_i^{Solow} = \sigma \log w_t + (1 - \sigma) a_{it} - \sigma \log(1 - \alpha) - \beta \log k_t + \beta \log n_{it}.$$

which proves the following result.

Proposition 2. *In the partial-equilibrium firm-dynamics model, when the cross-sectional distribution of log productivities \mathcal{F}_t increases to a distribution $\mathcal{F}_{t'}$ (with equal dispersion) such that one of Assumptions 1–3(i) holds for \mathcal{F}_t and $\mathcal{F}_{t'}$, the cross-sectional dispersion of Solow residuals decreases if $\sigma < 1$.*

Propositions 1 and 2 show that when $\sigma < 1$, our model exhibits the empirically correct cyclicalities of the dispersions of log employment, log output and the Solow residuals. Several recent papers suggest that labor and capital inputs are indeed complements, with estimated *firm-level* elasticities ranging from 0.25 to 0.53 (Chirinko, Fazzari and Meyer, 1999, 2011; Raval, 2015).²⁰ This motivates a more serious quantitative exploration, which is what we do in the next section.

3.2 Quantitative Exploration

We now explore the quantitative potential of the mechanism, embedding it into a general equilibrium model and allowing for persistence of a_{it} at the firm-level.

Model. The model is a version of the one presented in the previous subsection, with the following modifications. First, firms' log productivities are given as the sum of an aggregate and a firm-specific TFP component, $a_{it} = z_t + \epsilon_{it}$. Both components are modeled as AR(1) processes, $z_t = \rho_z z_{t-1} + e_t$ and $\epsilon_{it} = \rho_\epsilon \epsilon_{it-1} + u_{it}$ with persistence parameters $\rho_z, \rho_\epsilon \in (0, 1)$. The innovations $\{e_t, u_{it}\}$ are mutually independent iid shocks. Denote by G_ϵ the stationary distribution of $\{\epsilon_{it}\}$ and by G_u the distribution of u_{it} . Second, we now assume that firms' production functions have decreasing returns to scale,

$$y_{it} = e^{a_{it}} f(k_{it}, n_{it})^\nu, \tag{12}$$

where $\nu \in (0, 1)$. This ensures that the distribution of capital does not become degenerate with a positive persistence ρ_ϵ . Finally, we assume that labor N_t is supplied and state-dependent assets are

²⁰Firm-level and plant-level elasticities of substitution are generally smaller than their analog at the macro level, because the macro elasticity also includes shifts of production across plants and firms (e.g., Oberfield and Raval, 2014).

traded by a representative household with preferences

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - \theta \frac{1}{1+\zeta} N_t^{1+\zeta} \right) \quad (13)$$

and budget constraint

$$C_t + \mathbb{E}_t \left(\frac{Q_{t+1}}{Q_t} V_{t+1} \right) \leq W_t N_t + V_t, \quad (14)$$

where V_t is a random variable chosen by the agent, measurable with respect to aggregate information at time t . V_t captures the value of wealth, i.e. the number of Arrow-Debreu securities, the agent saves for a given state at time t . Capital does not enter the household's budget constraint directly as it is accumulated within firms.

Definition of Equilibrium. Given exogenous productivity shocks $\{z_t, \epsilon_{it}\}$, an equilibrium in this economy consists of stochastic processes of aggregate quantities $\{C_t, K_{t+1}, N_t, V_t\}$, firm-specific quantities $\{y_{it}, k_{it}, n_{it}, \pi_{it}, v_{it}\}$, prices $\{Q_t, W_t\}$, such that (a) each firm i solves (7) with production function (12), (b) the representative household maximizes (13) subject to (14), (c) firm i 's profits are given by (8), aggregate capital is defined as $K_t \equiv \int_0^1 k_{it} di$, and (d) markets clear, that is, $Y_t = C_t + \int_0^1 v_{it} k_{it} di$, $N_t = \int_0^1 n_{it} di$, and $V_t = \int_0^1 v_{it} di$.

Analytical simplification. Despite the large number of firms in the model and the presence of idiosyncratic productivity shocks, the equilibrium is still efficient and equivalent to the outcome of a planning problem. We formalize this in the following lemma.

Lemma 1. *There is an equilibrium with aggregate quantities $\{C_t, K_{t+1}, N_t\}$ if and only if they solve*

$$\max_{\{C_t, K_{t+1}, N_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - \theta \frac{1}{1+\zeta} N_t^{1+\zeta} \right) \quad (15)$$

$$C_t + K_{t+1} \leq e^{z_t} F(K_t, N_t) + (1 - \delta) K_t$$

where the aggregate production function F is defined as

$$\begin{aligned} F(K, N) &\equiv \max_{k(\tilde{\epsilon}), n(\tilde{\epsilon}, \tilde{u})} \iint e^{\rho \tilde{\epsilon} + \tilde{u}} f(k(\tilde{\epsilon}), n(\tilde{\epsilon}, \tilde{u}))^\nu dG_\epsilon(\tilde{\epsilon}) dG_u(\tilde{u}) \\ \text{s.t.} \quad &\int k(\tilde{\epsilon}) dG_\epsilon(\tilde{\epsilon}) = K, \quad \iint n(\tilde{\epsilon}, \tilde{u}) dG_\epsilon(\tilde{\epsilon}) dG_u(\tilde{u}) = N. \end{aligned} \quad (16)$$

Lemma 1 simplifies the solution of the aggregate economy to that of a standard real business cycle (RBC) model. The key intuition behind the lemma is that in the absence of capital adjustment costs at the firm-level, the distribution of next period capital stocks k_{it+1} is independent of the distribution of existing capital stocks k_{it} . Thus, the allocation of next period capital stocks is entirely

Table 1: Calibrated parameters.

Parameter	Value	Rationale
β	0.99	Discount factor; periods are interpreted as quarters.
δ	0.025	Yearly depreciation is 10 percent.
θ	4.29	Scaling parameter. Set so that $N = 1/3$ at steady state.
ζ	0.5	Frisch elasticity of labor supply is 2.
σ	0.4	Chirinko, Fazzari and Meyer (1999, 2011); Raval (2015).
α	0.71	Labor share at steady state is 0.64.
ν	0.73	Capital-output ratio at steady state is 2.5.
ρ_z	0.72	Persistence of HP-filtered Solow residuals.
ρ_ϵ	0.8	Gilchrist, Sim and Zakrajšek (2016).
σ_ϵ	0.15	Gilchrist, Sim and Zakrajšek (2016).

determined by current idiosyncratic firm productivities ϵ_{it} , and total production thus only depends on the aggregate amount of available capital, not its distribution last period.

Calibration and simulation. We use a standard calibration of both the aggregate and the micro-level economy. Periods are interpreted as quarters. We set the quarterly discount factor to $\beta = 0.99$, corresponding to an equilibrium real rate of around 4%. We set depreciation to $\delta = 0.025$ corresponding to a 10% annual depreciation rate. We normalize $\theta = 4.29$, setting steady state labor supply N to one third. We set $\zeta = 0.5$, equivalent to a Frisch elasticity of labor supply of 2. The persistence of aggregate TFP shocks, ρ_z , is calibrated to match the one of HP-filtered Solow residuals. At the micro level, we calibrate α to match an aggregate steady-state labor share of 0.64, ν to match a steady-state capital-output ratio of 2.5,²¹ and $\rho_\epsilon, \sigma_\epsilon$ according to the estimates of Gilchrist, Sim and Zakrajšek (2016). Finally, for the micro-level elasticity of substitution σ , we set $\sigma = 0.4$, falling in the middle of several recent estimates. Chirinko, Fazzari and Meyer (1999, 2011) find point estimates of 0.25 and 0.40 using firm-level data, and Raval (2015) finds plant-level estimates in the range 0.34 to 0.53. The calibration is summarized in Table 1.

We simulate the economy in three steps: First, we solve for the steady state of the model.²² Second, we simulate a log-linearized version of the RBC model to determine the first-order movements of aggregate variables. Finally, we use the underlying definition of the aggregate production function F to trace out the (highly nonlinear) behavior of firms at the micro level as response to the movements of aggregate variables.

²¹The implied value for ν is 0.73, which is broadly in line with Barkai (2017) who documents that returns on labor and capital amount to a combined output share of 75.5 percent.

²²This requires computing the aggregate production function F at the steady state of the planning problem in Lemma 1, which we find by approximating the idiosyncratic productivity shocks using Rouwenhorst’s method (c.f., Kopecky and Suen, 2010).

Dispersion of levels vs. dispersion of growth rates. In the partial equilibrium example with iid idiosyncratic productivity shocks, we derived results for the dispersion of *levels*, e.g. the cross-sectional variance of log employment or log output. In the data, however, many economists have focused on the dispersions of *growth rates* of these variables, e.g. the cross-sectional variance of the first difference of log employment or log output. In the context with iid shocks, any growth rate dispersion is exactly twice the corresponding level dispersion, so all the above analytical results are independent of which of the two approaches we use. In the following model, with non-iid shocks, we go with the literature and focus on growth rate dispersions.²³

Impulse responses. The blue lines in Figure 2 show the impulse responses of first and second moments to a positive innovation in aggregate productivity. The innovation is normalized to lead to a 1% increase in aggregate output. The response of first moments in Panel (a) is standard: hours, investment, and wages increase. According to our results in the previous subsection (see (10)), the key determinant for the cyclicity of dispersions are effective wages, $e^{-z_t}W_t$, shown in the last plot. As can be seen, the strong procyclical response in wages dominates the productivity shock after around 2.5 quarters, implying that $e^{-z_t}W_t$ is countercyclical for only around 2.5 quarters.

Panel (b) shows the response of growth rate dispersions to the technology shock. On impact, the standard deviations of first differences in log output, hours, and the firm-level Solow residual are all countercyclical, while the standard deviation of the investment rate $\iota_{it} \equiv (k_{it+1} - (1 + \delta)k_{it})/k_{it}$ is procyclical on impact. Due to the short-lived response of $e^{-z_t}W_t$, the four dispersion responses are, however, relatively shortly lived and overshoot after around 2.5 quarters.

As is well known, the strong procyclical response of wages implied by our basic RBC model is largely at odds with the data. In addition, there is ample evidence against technology shocks as the driving force behind business cycles. For instance, [Chari, Kehoe and McGrattan \(2007\)](#) stress the importance of labor wedges for business cycle fluctuations, in line with the idea that recessions are demand-driven. As a short-cut to demand shocks, we simulate the economy's response to a shock to the labor wedge τ_t^N , which we model in the standard way, as a wedge in the first order condition for labor,

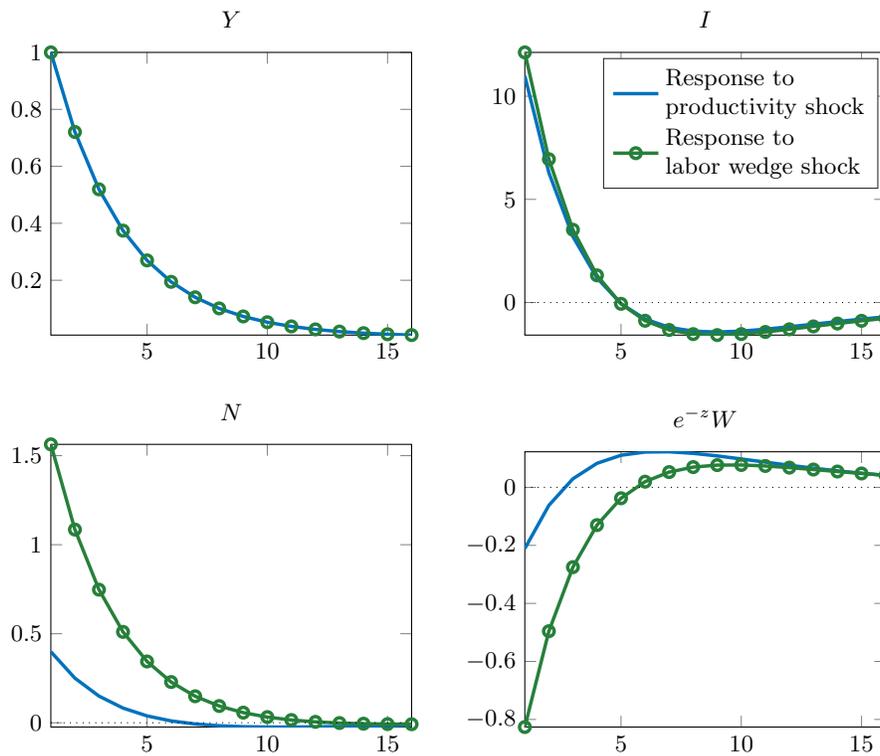
$$\theta C_t N_t^\zeta = (1 - \tau_t^N) e^{z_t} F_N(K_t, N_t).$$

We choose the path for τ_t^N to match the output response of the economy after a technology shock to allow a direct comparison. The resulting impulse responses are depicted as green circles in Figure 2. Naturally, the labor wedge shock yields a large response of hours and a more countercyclical response of the effective wage $e^{-z_t}W_t$.²⁴ This explains a more pronounced response of second moments in response to the labor wedge shock.

²³One exception is investment, which is frequently zero in the data (and can become negative in the model) so that both logs and growth rates are ill-defined. To avoid this problem, we compute the dispersion of investment-to-capital ratios, $\iota_{i,t}$, similar to the approach in [Bachmann and Bayer \(2014\)](#).

²⁴In the presence of shocks to the labor wedge, our notation assumes that W_t is the marginal product of labor, which no longer has to equal the observed wage (e.g. the observed wage could be sticky).

(a) Response of aggregates.



(b) Response of growth rate dispersions.

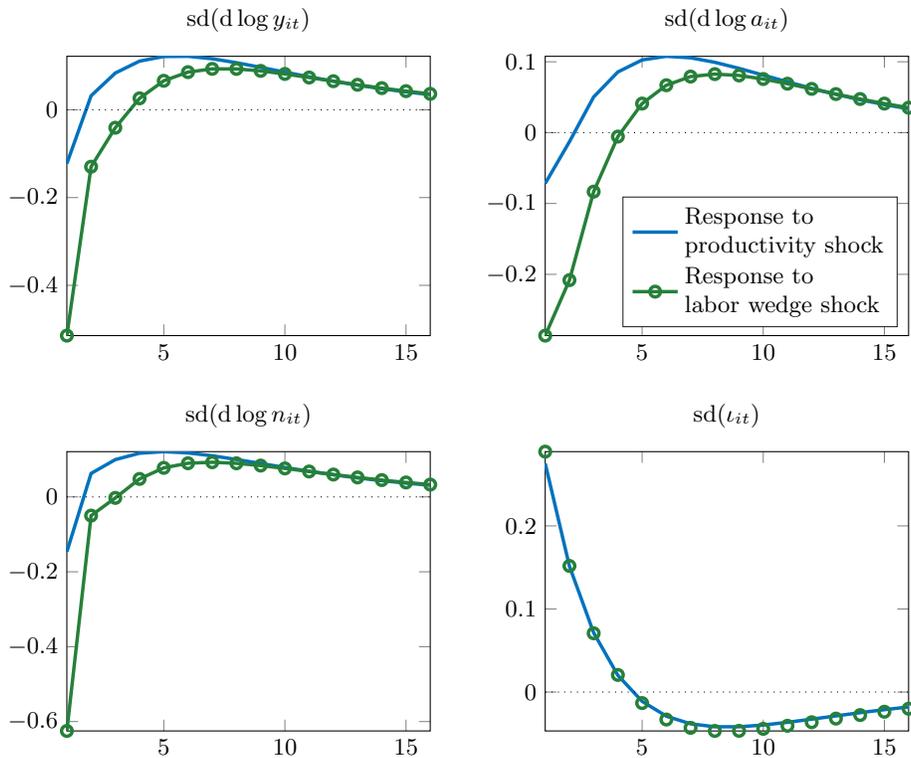


Figure 2: The response of the quantitative model to productivity and labor wedge shocks (% relative to steady state).

Table 2: Correlations of dispersions with output.

	correlation with cycle ($\log y_t$)			
	$\text{sd}(\text{d} \log y_{it})$	$\text{sd}(\text{d} \log n_{it})$	$\text{sd}(\text{d} \log a_{it})$	$\text{sd}(l_{it})$
model (linearly detrended)	-0.58	-0.57	-0.59	+0.72
model (high-pass filtered)	-0.43	-0.43	-0.44	+0.61
data	-0.45	-0.50	-0.47	+0.45

How important could this be quantitatively? The gap in GDP between average NBER booms and recessions is 5.7%. If the entire gap is driven by fluctuations in the labor wedge, this would correspond to a 3.6% peak-to-trough gap in employment growth dispersion. This is roughly one third of the peak-to-trough variation in employment growth dispersion document by [Ilut, Kehrig and Schneider \(2016\)](#).

Dispersion cyclicity in the model and the data. To provide a more direct comparison between model and data, we now extend the model to exactly replicate US aggregate time series on output, employment, consumption and investment. Using the implied time series for second moments, we then compute the correlations of model implied second moments and compare them with evidence on their cyclicity from the literature.

To replicate the data, we follow [Chari, Kehoe and McGrattan \(2007\)](#) and add government and investment wedges to the model (see Appendix E.2 for details). Since our model does not feature growth, we eliminate the growth trend from the data by either detrending linearly or applying a high-pass filter with 40 years. Using the extended model, we infer a process for the aggregate shocks that match the data perfectly and then compute the firm-level dynamics given the inferred path of the aggregate economy. Figure 3 plots the time series for the resulting dispersion measures. Table 2 computes the implied cyclicalities and compares them to the evidence in [Bachmann and Bayer \(2014\)](#). For both detrending methods, the model-implied cyclicalities are quantitatively of similar magnitudes to the data.

Takeaway. In our main application, we have illustrated how a standard CES production function with capital-labor complementarity generates an increasing concave relationship between log productivity and log employment across firms. Thus, shifts in the distribution of log productivity that preserve its variance and satisfy one of the Assumptions 1-3(i) translate into countercyclical shifts in log employment dispersion, as well as several other variables such as log output and Solow residuals. Quantitatively, the model-implied cyclicalities are of similar magnitude as in the data.

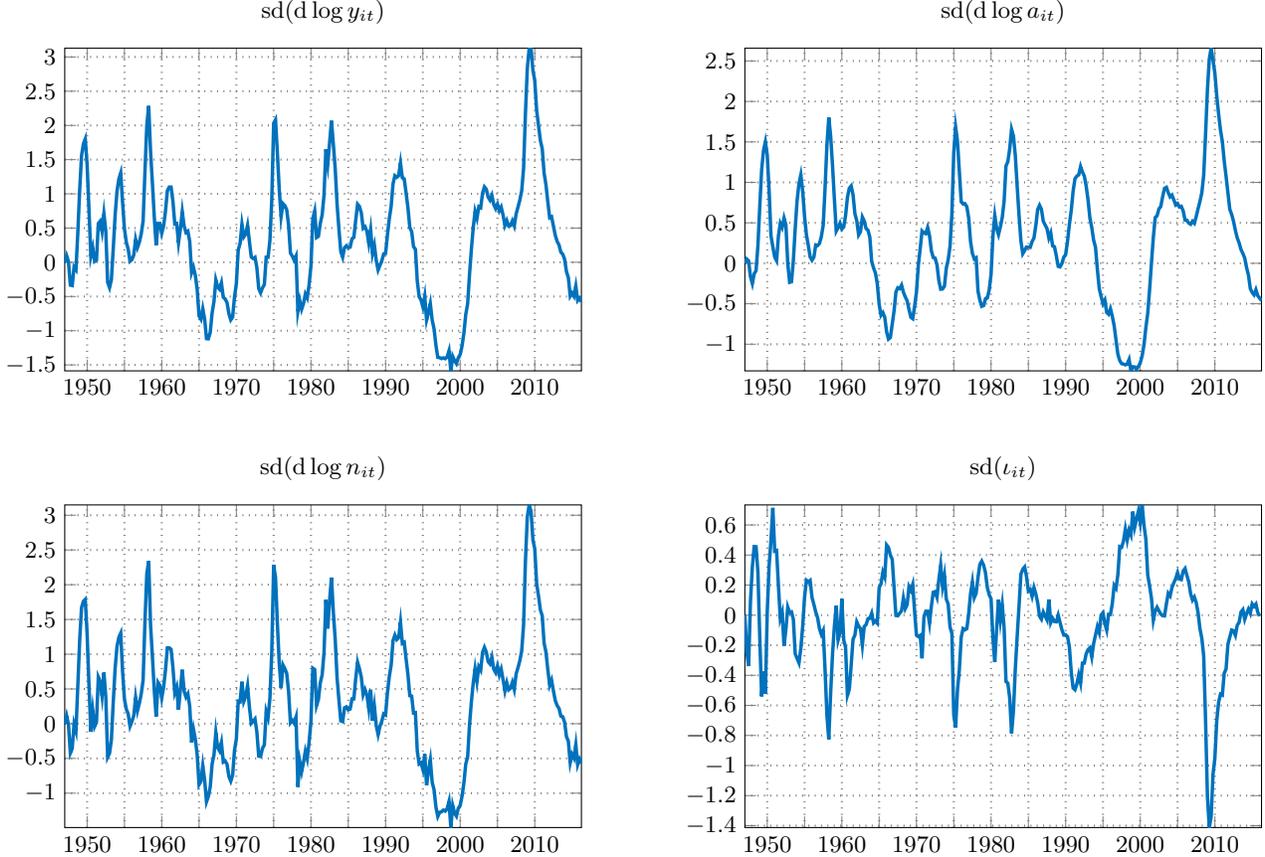


Figure 3: Time series of dispersions from business cycle accounting exercise.

4 Other Applications

We now present three additional, smaller applications that illustrate the broad applicability of our theoretical results. The first example relates the *dispersion* of marginal revenue products of capital—a prominent proxy for misallocation used in the recent literature—to the nature of the shocks that cause cyclical fluctuations. The second example revolves around *risk*. In particular, it revisits the classic finance question of how the payoff profile of a security affects its payoff and return risk. In our final example, we illustrate how posterior *uncertainty* can fluctuate endogenously as soon as learning takes place over a non-linear function of the fundamental.

4.1 MRPK and the business cycle

A recent literature on misallocation following [Hsieh and Klenow \(2009\)](#) uses the cross-sectional dispersion of marginal revenue products of capital (MRPK) to identify misallocation of capital across firms. One common rationale to explain this are financial frictions, which essentially prevent productive plants from accumulating the efficient amount of capital (e.g., [Moll, 2014](#); [Buera and Moll, 2015](#); [Kehrig, 2015](#); [Gopinath et al., 2015](#)). In this application, we explore a simple static

setup where firms face borrowing constraints, and show how the dispersion of MRPK depends on the distribution of productivity relative to the distribution of borrowing constraints.

Assume there is a unit interval of firms, indexed by $i \in [0, 1]$. Firms are perfectly competitive and sell in a market with price normalized to 1. Each firm operates a production technology exhibiting decreasing returns to scale,

$$Y_i = A_i K_i^\alpha.$$

Here, A_i is firm i 's productivity and K_i is the capital rented by firm i . One way to rationalize such a production function is that there is a fixed factor such as land or entrepreneurial labor which is owned by the entrepreneur that runs the firm. We excluded workers from the production function to streamline the exposition. All results go through when firms use labor from a common labor market and labor and capital enter the production function in a Cobb-Douglas way with decreasing returns to scale. In this sense, we interpret A_i more broadly than just firm i 's productivity: Even when productivity itself does not improve, an increase in A_i might still happen due to a boom in employment or hours.²⁵ With some slight abuse of terminology, we still refer to A_i as productivity in the following paragraphs.

The crucial ingredient to our analysis is the imposition of a borrowing constraint: Firm i can only rent capital up to an exogenous limit \bar{K}_i which may depend on i . Thus firm i solves

$$\max_{K_i \leq \bar{K}_i} A_i K_i^\alpha - r K_i.$$

The MRPK of firm i may then exceed r ,

$$MPRK_i = A_i K_i^{\alpha-1} = r + \mu_i,$$

if and only if the borrowing constraint binds, $\mu_i > 0$. We can express the multiplier as

$$\mu_i = \max \left\{ 0, A_i \bar{K}_i^{\alpha-1} - r \right\}$$

which yields the following expression for the MRPK,

$$\log MRPK_i = \max \left\{ \log r, \log A_i - (1 - \alpha) \log \bar{K}_i \right\}.$$

Noticing that $\max\{\log r, \cdot\}$ is a convex and weakly increasing function, the following result is an immediate consequence of Theorem 1.

Proposition 3. *The dispersion of log MRPK increases if*

- (i) *the cross-sectional distribution of log productivities shifts up such that one of Assumptions 1-3(i) holds without changing its variance and the shifts are orthogonal to $\log \bar{K}_i$.*

²⁵Such a boom could come about due to an increase in labor supply or in aggregate demand.

- (ii) the cross-sectional distribution of log constraints shifts down such that one of Assumptions 1–3(i) holds without changing its variance and the shifts are orthogonal to $\log A_i$.

Proposition 3 illustrates how the degree of misallocation is essentially pinned down by the relative position of productivity and borrowing constraints. In a world with only productivity improvements, misallocation would increase, while in a world with more relaxed borrowing constraints, misallocation falls.

More realistically, borrowing constraints are correlated with a firm’s business conditions. For instance, suppose that \bar{K}_i is a function of each firm’s revenue, $A_i K_i^\alpha$, and some credit supply-driven factor u_i :

$$\log \bar{K}_i = \epsilon \log(A_i K_i^\alpha) + u_i, \quad (17)$$

where $\epsilon < 1/\alpha$ is the elasticity of \bar{K}_i to firm-revenues. In this case, the dispersion in MRPKs increases during productivity-driven booms (or, similarly, in high-productivity sectors) if and only if $\epsilon < 1$, which could reflect, say, a lack of information by financial markets about A_i . In contrast, a “financially-driven” crisis akin to a downward shift in the distribution of u_i would unambiguously increase the dispersion of MRPKs. The following proposition summarizes the results.

Proposition 4. *Suppose \bar{K}_i is given by (17).*

- (i) *If the distribution of log productivities $\{\log A_i\}$ shifts up such that one of Assumptions 1–3(i) holds without changing its variance and shifts are orthogonal to u_i , then $\text{Var}\{\text{MRPK}_i\}$ increases (decreases) if $\epsilon < 1$ ($\epsilon > 1$).*
- (ii) *If the distribution of credit factors $\{u_i\}$ shifts up such that one of Assumptions 1–3(i) holds without changing its variance and shifts are orthogonal to $\log A_i$, then $\text{Var}\{\text{MRPK}_i\}$ decreases.*

Propositions 3(i) and 4(i) provide a possible explanation for why booms can raise the degree of misallocation as documented, e.g., by Garcia-Santana, Moral-Benito and Pijoan-Mas (2015) and Gopinath et al. (2015) in the case of Spain between 1995 and 2007.

Numerical example. We quantify the model using a simple and very stylized numerical example with productivities given by $a_i = z + \epsilon_i$. The cross-sectional distribution of productivities is normal with the same (unconditional) standard deviation as in the previous subsection. The (unconditional) standard deviation of z is set to 0.038, matching the (unconditional) standard deviation of aggregate productivity shocks in Gilchrist, Sim and Zakrajšek (2016). The capital share α is set equal to 0.3, the cost of capital r is set equal to 0.05, and the elasticity of the constraint to revenues ϵ is set equal to 0.8.²⁶ Finally, we consider two scenarios for the tightness of financial constraints. In the “healthy” baseline scenario, we set $u_i = \bar{u}_{5\%}$ where $\bar{u}_{5\%}$ is so that a fraction of 5 percent of firms

²⁶Since we do not have a strong prior on ϵ it is worth mentioning that all results discussed below are robust to the exact choice of ϵ . While different values affect the level of the MRPK-dispersion, they have virtual no effect on the cyclical changes discussed below (in percentage-terms) as long as $\epsilon < 1$.

	Financial constraints	
	healthy	tight
$1\sigma^{\text{avg}}$ boom	0.98 (+21.2%)	1.50 (+84.9%)
$1\sigma^{\text{avg}}$ recession	0.66 (−18.3%)	1.06 (+30.7%)

Table 3: Cross-sectional standard-deviation of MRPK in numerical example. All standard deviations are rescaled by a factor of 100. The two columns display two regimes for the financial constraint: In the left (right) column, 5 (10) percent of firms are constrained absent a productivity shock. In parentheses, we report the differences (in percent) to the case in which aggregate productivity is at its median value and financial constraints are healthy.

are constrained in the absence of an aggregate productivity shock. In the “tight” scenario, that is supposed to capture a negative credit supply shock, we set $u_i = \bar{u}_{10\%}$, so that a fraction of 10 percent of firms are constrained (without aggregate productivity shock).

Table 3 displays the cross-sectional standard deviation of MRPK across firms. The left column shows the effects of a change in the average firms’ productivity when financial conditions are healthy (5 percent of firms being constrained in the absence of a productivity shock), whereas the right right column shows the same statistics for the case where financial conditions imply a twice-as-high fraction of firms being constrained (without aggregate productivity shock). In both cases a one standard deviation increase in the average productivity increases the dispersion of MRPK by roughly 20 percent. Comparing the two columns further illustrates a significant increase in the dispersion induced by an increase in financial pressure. Compared to the baseline case, even a $1\sigma^{\text{avg}}$ reduction in average productivity increases the dispersion by 31 percent when combined with financial pressure. The results are broadly in line with the evidence given in [Gopinath et al. \(2015\)](#) who find that the dispersion of MRPKs in Spain increased by almost 30 percent between 2000 and 2012.

4.2 Risk in financial derivatives

Our third example illustrates how the payoff profile of a security influences the cyclicity of the risk premium associated with it. It is shown for instance that derivatives with concave increasing payoff profiles as function of their underlying (e.g. debt contracts) have countercyclical return risk, while derivatives with convex decreasing payoff profiles (e.g. Put options) have procyclical return risk. The cyclicity here is determined with respect to constant-risk shifts in the underlying and without strong distributional assumptions. While these results may not be entirely novel, viewing them through our theoretical framework gives them a new level of generality.

The economy we study is a simple financial market setup, consisting of two periods, $t = 0, 1$, and a continuum of identical traders with standard mean-variance preferences over $t = 1$ wealth W ,

$$\mathbb{E}W - \frac{\alpha}{2}\text{Var}(W),$$

where $\alpha > 0$ measures the degree of risk aversion. We assume that the agents trade a real asset (i.e. the market portfolio) with stochastic payoff X_0 in unit supply, as well as derivative assets with stochastic payoffs X_i in zero net supply. The representative agent’s holdings of asset $k \in \{0, 1, \dots, K\}$ are denoted by a_k and the price of asset k is denoted by p_k . This implies that $t = 1$ wealth is given by

$$W = X_0 + \sum_{k=1}^K a_k (X_k - (1+r)p_k),$$

where we use $r > 0$ to denote the risk-free interest rate.²⁷

Our goal in the following is to characterize equilibrium prices for a given derivative asset X_k , whose payoff can be described as a monotone function of the market, $X_k = g(X_0)$. The starting point for the analysis is, of course, the first order optimality condition with respect to a_k , yielding

$$\mathbb{E}X_k - (1+r)p_k = \alpha Cov(X_0, X_k).$$

Notice that this is almost like a Capital Asset Pricing Model (CAPM), which holds in our mean-variance environment, only that the equation is in terms of actual payoffs rather than returns.²⁸ We now investigate how the “risk term” in the pricing equation, $\alpha Cov(X_0, X_k)$, behaves under changes in X_0 .

Proposition 5. *As we move from X_0 to a random variable X'_0 such that either Assumption 1 or Assumption 2 holds for X_0 and X'_0 , then the risk term in the price of asset k with monotone g*

1. *increases whenever g is convex,*
2. *decreases whenever g is concave.*

The basic intuition of Proposition 5 is very much along the lines of Theorem 1: When Assumption 1 holds, the distribution X'_0 is a positive equal-variance shift of X_0 . Hence, when g is convex, $\text{Var}\{X_k\} = \text{Var}\{g(X_0)\}$ must increase as we move from X_0 to X'_0 . This already indicates that the *covariance* between X_0 and $X_k = g(X_0)$ should also increase, as long as the correlation is increasing, or at least not decreasing too fast. The logic of the formal proof of Proposition 5 in Appendix D is similar to the proof strategy of Theorem 1, described in equation (??) above.

There are a number of well-known examples for the behavior described in Proposition 5. Perhaps most prominently, if X_k is a Call or Put option with X_0 as its underlying asset, then g would be convex, explaining why the risk term increases for both Calls and Puts, with the difference that the Put’s risk term is negative and rises towards zero while the Call’s is positive. In a corporate finance context, one could think of X_0 as the (random) asset side of a firm’s balance sheet and of $X_k = g(X_0)$ a corporate debt contract. In this situation, the risk in the debt contract is countercyclical with

²⁷To be precise, one can imagine this to be the subproblem of a full 2-period maximization problem over consumption at both dates, $t = 0, 1$. Given a risk-free rate of r , the $t = 1$ price paid for asset k is then $(1+r)p_k$.

²⁸One could easily derive the CAPM $\mathbb{E}R_k - (1+r) = \alpha Cov(X_0, R_k)$ in our model. See Corollary 3 below.

respect to the firm's assets: In bad times, when the asset side is potentially below the debt's principal value, credit risk rises.

Finally, notice that our Proposition 5 was stated for the risk term $\alpha Cov(X_0, X_k)$ as opposed to the risk premium $\mathbb{E}R_k - (1 + r) = \alpha Cov(X_0, R_k)$. We can derive a similar result, albeit less sharp, for the risk premium.

Corollary 3. *As we move from X_0 to a random variable X'_0 such that either Assumption 1 or Assumption 2 holds for X_0 and X'_0 , then the risk premium of asset k*

1. *increases whenever g is convexly decreasing,*
2. *decreases whenever g is concavely increasing.*

The key observation behind Corollary 3 is that the risk premium can be expressed as

$$\mathbb{E}R_k - (1 + r) = (1 + r) \frac{\alpha Cov(X_0, X_k)}{\mathbb{E}X_k - \alpha Cov(X_0, X_k)}.$$

Thus, when g increases concavely (e.g. like a debt contract), then $\mathbb{E}X_k$ rises as well as X_0 moves to X'_0 , while, according to Proposition 5, $\alpha Cov(X_0, X_k)$ falls, explaining a drop in the risk premium. A similar logic applies to the case where g is convexly increasing, for example to Put options.

Summing up, this application showed how our results can be applied to a basic asset pricing context. We illustrated that the concavity or convexity of a derivative's payoff structure crucially determines the comovement of the risk inherent in the derivative with the underlying asset.

4.3 Nonlinear learning and uncertainty

Our final example is a simple Bayesian learning problem where nonlinearities in the signal structure cause posterior uncertainty to be a function of the signal realization. In past years, such learning problems have spurred a considerable research interest among macroeconomists, e.g. see [Straub and Ulbricht \(2012, 2014\)](#), [Orlik and Veldkamp \(2014\)](#), [Kozeniaskas, Orlik and Veldkamp \(2016\)](#), and [Albagli, Hellwig and Tsyvinski \(2015\)](#).

To see how the aforementioned theorems are useful to study such a problem, suppose θ is the unknown variable an observer seeks to gather information about. She observes a signal s about a nonlinear transformation of θ , $\phi = g(\theta)$. Suppose g is increasing and concave. To fix ideas, suppose for instance that θ defines the optimal investment level of a firm, summarizing a firm's private information regarding future business conditions. Also suppose that the firm faces financial constraints, limiting its investments to satisfy $k \leq \bar{k}$. Relabeling $\phi = k$, we then have $g(\theta) = \min\{\theta, \bar{k}\}$, giving rise to an increasing concave mapping.²⁹

There are two steps to the agent's updating problem: First, she forms a posterior over the values of ϕ . Then, she maps her posterior from values of ϕ to values of the variable of her interest, θ . In

²⁹This example is inspired by a previous, unpublished working paper [Straub and Ulbricht \(2012\)](#).

the following, we assume the agent solved her first problem and is left with a situation where the posterior $\phi|s$ satisfies either Assumption 1, 2 or 3 above. Arguably, these are natural assumptions in the context of a signal extraction problem. Condition (i) of Assumptions 1–3 represent a simple monotonicity requirement, demanding that higher signal realizations correspond to higher values of ϕ in the sense of the respective stochastic order. Condition (ii) is a precision requirement, demanding that the transformation g , and the transformed variable ϕ itself, are such that the posterior variance does not shrink exogenously as the agent observes higher signal realizations.

What can our Theorems tell us about the posterior belief $\theta|s$? It is relatively straightforward to see that, based on condition (i) of Assumptions 1–3, the *first moment* $\mathbb{E}\{\theta|s\}$ increases with s . However, this does not yet provide us with any information about the behavior of the *second moment*. Theorem 1 implies that, in fact, the second moment, $\text{Var}\{\theta|s\}$ is also increasing in s .

The intuition for this application is as follows: As the agent observes a large realization of signal s , she rationally updates that large values of ϕ are relatively likely. For the correspondingly large values of θ the nonlinear function g is relatively flat, implying that the posterior variance over ϕ translates into relatively large posterior variances over θ . With the interpretation of ϕ as firm investments, this means that learning from financially constrained firms leads to higher uncertainty, because financially constrained firms respond less to changes in business fundamentals, so that for large realization of the signal (relative to the financial constraint), the signal is more likely to be driven by signal noise.

In sum, this application illustrates how a simple *nonlinear* signal structure can give rise to endogenous, signal-dependent movements in uncertainty.³⁰

5 Conclusion

In this paper we studied the behavior of second moments under nonlinear transformations. We showed that when a random variable shifts up, in one of three senses (MLRP, affine linear, or affine-linear in transformed variables) but keeps the same variance, the variance of any convex increasing or concave decreasing transformation of it necessarily increases. We see our theoretical contribution in proving the MLRP result and providing elementary proofs in the two other settings.

The economic implications of our theoretical results were illustrated in four different settings. The first and main application provided a (to the best of our knowledge) novel explanation for the cyclicalities of the dispersions of macroeconomic variables; the second application extended these insights to the cross-sectional dispersion of MRPKs which is often used to measure capital misallocation; the third application studied the cyclicalities of derivative risks; and the final application shows how our results can be used to generate endogenously fluctuating uncertainty.

An understanding of how and why second moments move over the business cycle is directly important for policymaking: With exogenous second moment shocks, understanding and possibly

³⁰For alternative approaches to modeling endogenous variations in uncertainty, see, e.g., Van Nieuwerburgh and Veldkamp (2006), Nimark (2014), Fajgelbaum, Schaal and Taschereau-Dumouchel (2017), and Senga (2016).

mitigating the sources of these shocks becomes important. Endogenous second moment shocks, however, may not carry any policy consequences at all. As an example, the quantitative model in our first application is efficient and nonetheless can produce some of the cyclical properties of cross-sectional dispersions that we see in the data. Understanding the policy implications in models with endogenous second moments more broadly is a promising avenue for future research.

A Proof of Theorem 1

We prove Theorem 1 using a sequence of lemmas and propositions. The main proposition that we establish below relates the difference between covariances of X and Y with some other random variables to the difference between the covariances of $g(X)$ and $g(Y)$ with the same other random variables (Proposition 6). The advantage of proving this result as a first step is that it is “linear”, in the sense that X and Y only enter linearly, whereas the result in Theorem 1 is quadratic. By essentially applying Proposition 6 twice, we can then establish Theorem 1.

A.1 Notation and Assumptions

We introduce the following notation for our proof. For any real random variable Z which admits a density with respect to measure μ we denote by f_Z the density function of Z , by F_Z the right-continuous cumulative density of Z , and by $Q_Z : (0, 1) \rightarrow \mathbb{R}$ the right-continuous quantile function, defined as

$$Q_Z(u) = \inf_{x \in \mathbb{R}} \{x : F_Z(x) \geq u\}.$$

Further, we denote by μ_Z the mean of Z (as long as it exists) and let $q_Z : (0, 1) \rightarrow \mathbb{R}$, $q_Z(u) = Q_Z(u) - \mu_Z$, be a “normalized” quantile function, as well as $\bar{q}_Z : (0, 1) \rightarrow \mathbb{R}$, $\bar{q}_Z(u) \equiv \int_0^u q_Z(\tilde{u}) d\tilde{u}$, be an “integral” quantile function. Notice that whenever the mean of Z exists, the integral of q_Z over $[0, u]$ exists for all $u \in [0, 1]$ and $\bar{q}_Z(0) = \bar{q}_Z(1) = 0$, by the property of the quantile function that $\int_0^1 Q_Z(u) du = \mu_Z$. Furthermore, because q_Z is weakly increasing from negative to positive values, \bar{q}_Z is quasiconvex (i.e. u-shaped).

In the following we let Z_X and Z_Y be two real random variables, satisfying the following assumptions.

Assumption 4. *The random variables Z_X and Z_Y satisfy the following assumptions:*

1. Z_X and X are perfectly rank-correlated, that is, there exists a random variable $U \sim \text{unif}[0, 1]$ that is uniformly distributed over the unit interval such that $Z_X = Q_{Z_X}(U)$ and $X = Q_X(U)$. Similarly, Z_Y and Y are perfectly rank-correlated.

2. Z_X and Z_Y have finite mean and nontrivial variance.

3. The ratio

$$\frac{q_{Z_X}(F_X(x))}{q_{Z_Y}(F_Y(x))},$$

defined for values of x for which it is positive and finite, is weakly decreasing in x .

4. Let I_X (I_Y) denote the interval of zeros of the function $x \mapsto q_{Z_X}(F_X(x))$ ($x \mapsto q_{Z_Y}(F_Y(x))$). Then, $\inf I_X < \inf I_Y$ and $\sup I_X < \sup I_Y$.

While these assumptions appear technical at first, they will help us prove the necessary results about covariances in the next section.

A.2 Covariances

We now state the main result of this appendix.

Proposition 6. *Let Z_X and Z_Y be two random variables satisfying Assumption 4, and X and Y two random variables satisfying Assumption 1 (i). Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex, strictly increasing, continuously differentiable function. Then it holds that*

$$\text{Cov}(X, Z_X) \leq \text{Cov}(Y, Z_Y) \Rightarrow \text{Cov}(g(X), Z_X) \leq \text{Cov}(g(Y), Z_Y),$$

where the latter inequality is strict if g is strictly convex.

We prove the result in a sequence of smaller steps. In our first step, we prove some properties of the function $\delta : \text{supp}(X) \cup \text{supp}(Y) \rightarrow \mathbb{R}$, $\delta(x) = q_{Z_Y}(F_Y(x))f_Y(x) - q_{Z_X}(F_X(x))f_X(x)$, which is defined on the supports of X and Y .

Lemma 2. *We have the following results about δ :*

1. If $x \in \text{conv}(I_X \cup I_Y)$, then $\delta(x) \leq 0$.³¹
2. If $x > I_Y$ and $\delta(x) > 0$, then $\delta(x') > 0$ for all $x' > x$.
3. If $x < I_X$ and $\delta(x) > 0$, then $\delta(x') > 0$ for all $x' < x$.

Proof. We treat each case in turn.

1. In this case, by the definitions of I_X and I_Y , either $\delta(x) = 0$ —if $x \in I_X \cap I_Y$ —or $\delta(x) < 0$ —if $x > I_X$ or $x < I_Y$, since the normalized quantile functions q_{Z_Y} and q_{Z_X} are weakly decreasing.³²
2. Fix such a $x > I_Y$ with $\delta(x) > 0$, and a $x' > x$ in the support of X or Y . Then, $\delta(x') > 0$ is a direct consequence from the following set of inequalities,

$$\frac{f_Y(x')}{f_X(x')} \geq \frac{f_Y(x)}{f_X(x)} > \frac{q_{Z_X}(F_X(x))}{q_{Z_Y}(F_Y(x))} \geq \frac{q_{Z_X}(F_X(x'))}{q_{Z_Y}(F_Y(x'))},$$

³¹Here, $\text{conv}(A)$ denotes the convex hull of set $A \subset \mathbb{R}$.

³²Here, $x > I_X$ means $x > a$ for all $a \in I_X$. Similar for $x < I_Y$.

where the first inequality follows from the MLRP of X and Y , the second inequality follows from $\delta(x) > 0$, and the third inequality follows from the fact that for $x > I_Y$ the ratio in Assumption 4.3 is positive and hence weakly decreasing. Comparing the first and last terms of this set of inequalities, this implies $\delta(x') > 0$.

3. Analogously to the previous case, if $x < I_X$, the ratio in Assumption 4.3 is positive and hence weakly decreasing. Thus, for any $x' < x$ in the support of X ,

$$\frac{f_Y(x')}{f_X(x')} \leq \frac{f_Y(x)}{f_X(x)} < \frac{q_{Z_X}(F_X(x))}{q_{Z_Y}(F_Y(x))} \leq \frac{q_{Z_X}(F_X(x'))}{q_{Z_Y}(F_Y(x'))},$$

where again we used the MLRP of X and Y and the fact that $\delta(x) > 0$. Taken together, this implies $\delta(x') > 0$. □

These results turn out to be very helpful in characterizing the behavior of the function $\Delta : \mathbb{R} \rightarrow \mathbb{R}$, $\Delta(x) = \bar{q}_{Z_Y}(F_Y(x)) - \bar{q}_{Z_X}(F_X(x))$. Notice that, by definition of Δ ,

$$\Delta(x) = \int_{(-\infty, x]} \delta(\tilde{x}) \mu(d\tilde{x}), \quad (18)$$

which is to be understood as a Lebesgue integral with measure μ , restricted to $(-\infty, x]$ (so that it nests all three cases, where μ is continuous, discrete, or anything in between). We can characterize Δ as follows.

Lemma 3. *Δ has limits $\lim_{x \rightarrow -\infty} \Delta(x) = \lim_{x \rightarrow +\infty} \Delta(x) = 0$. There exists a point $x_0 \in \mathbb{R}$ such that $\Delta(x) \leq 0$ for all $x \geq x_0$ and $\Delta(x) \geq 0$ for all $x \leq x_0$, with strict inequality for at least a non-zero measure of such x .*

Proof. First, note that the limits of Δ come from the fact that \bar{q}_{Z_X} and \bar{q}_{Z_Y} are continuous and zero for arguments 0 and 1, and the cumulative distribution functions F_X and F_Y have the usual limits for $x \rightarrow \pm\infty$. Second, we prove that an x_0 exists with $\sup\{x : \Delta(x) \geq 0\} \leq x_0 \leq \inf\{x : \Delta(x) \leq 0\}$. Suppose it did not. Then there must exist $x_1 < x_2$ with $\Delta(x_1) < 0 < \Delta(x_2)$. As in (18) we can write

$$\Delta(x_2) - \Delta(x_1) = \int_{(x_1, x_2]} \delta(x) \mu(dx) \quad (19)$$

and

$$\Delta(x_1) = \int_{(-\infty, x_1]} \delta(x) \mu(dx) \quad \text{and} \quad 0 - \Delta(x_2) = \int_{(x_2, \infty)} \delta(x) \mu(dx). \quad (20)$$

Using (19) and $\Delta(x_2) - \Delta(x_1) > 0$, there must exist a $a_1 \in (x_1, x_2]$ such that $\delta(a_1) > 0$. On the other hand, using (20) and $\Delta(x_1) < 0 < \Delta(x_2)$, there must exist $a_0 < x_1$ and $a_2 > x_2$ such that $\delta(a_0) < 0$ and $\delta(a_2) < 0$. This contradicts Lemma 2: If $a_1 < I_X$, then $\delta(a_0)$ would have to be

positive. If $a_1 \in \text{conv}(I_X \cup I_Y)$, $\delta(a_1)$ would have to be negative. And if $a_1 > I_Y$, $\delta(a_2)$ would have to be positive. This proves $\{\Delta \geq 0\} \leq x_0 \leq \{\Delta \leq 0\}$.

Finally, we show that there exists a non-zero measure of x such that $\Delta(x) \neq 0$. Suppose the contrary was true, $\Delta(x) = 0$ for μ -almost all x . Then $\delta(x)$ must be zero μ -almost every x . This cannot be the case given that we assumed in Assumption 4.4 that for example $\inf I_X < \inf I_Y$. This concludes the proof of Lemma 3. \square

Finally we are ready to prove Proposition 6.

Proof of Proposition 6. Notice that by Assumption 4.1 we can express covariances as integrals over the “normalized” quantile functions,³³

$$\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) = \int_0^1 q_{Z_Y}(u)g(Q_Y(u))du - \int_0^1 q_{Z_X}(u)g(Q_X(u))du.$$

Substituting out u for $F_Y(x)$ in the first integral and $F_X(x)$ in the second one, this simplifies to

$$\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) = \int_{\mathbb{R}} g(x)\delta(x)\mu(dx). \quad (21)$$

Using integration by parts, we further rewrite this as

$$\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) = \lim_{x \rightarrow \infty} g(x)\Delta(x) - \lim_{x \rightarrow -\infty} g(x)\Delta(x) - \int_{\mathbb{R}} g'(x)\Delta(x)dx.$$

We now consider each limit in turn, starting with the one for $x \rightarrow \infty$. Notice that for sufficiently large x , $g(x) > 0$ and $\delta(x) \geq 0$, hence

$$\lim_{x \rightarrow \infty} |g(x)\Delta(x)| = \lim_{x \rightarrow \infty} \int_{(x, \infty)} g(x)\delta(\tilde{x})\mu(d\tilde{x}) \leq \lim_{x \rightarrow \infty} \int_{(x, \infty)} g(\tilde{x})\delta(\tilde{x})\mu(d\tilde{x}) = 0$$

where the last equality follows by the existence and finiteness of the covariances in (21). Consider now $x \rightarrow -\infty$. Focus first on the case where g eventually turns negative for sufficiently small x . Then,

$$\lim_{x \rightarrow -\infty} |g(x)\Delta(x)| = - \lim_{x \rightarrow -\infty} \int_{(-\infty, x]} g(x)\delta(\tilde{x})\mu(d\tilde{x}) \leq - \lim_{x \rightarrow -\infty} \int_{(-\infty, x]} g(\tilde{x})\delta(\tilde{x})\mu(d\tilde{x}) = 0,$$

where, again, the last equality follows by the existence and finiteness of the covariances in (21). If g stays positive for all x , then the limit $\lim_{x \rightarrow -\infty} g(x)$ exists and is finite. Hence,

$$\lim_{x \rightarrow -\infty} |g(x)\Delta(x)| = \left(\lim_{x \rightarrow -\infty} g(x) \right) \left(\lim_{x \rightarrow -\infty} \Delta(x) \right) = 0.$$

³³For any two perfectly rank-correlated random variables $Z_1 = Q_{Z_1}(U)$ and $Z_2 = Q_{Z_2}(U)$ (for some common uniform-distributed U) it holds that $\text{Cov}(Z_1, Z_2) = \mathbb{E}[Z_1(Z_2 - \mu_{Z_2})] = \mathbb{E}[Q_{Z_1}(U)q_{Z_2}(U)] = \int_0^1 q_{Z_2}(u)Q_{Z_1}(u)du$.

This proves that

$$\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) = - \int_{\mathbb{R}} g'(x) \Delta(x) dx. \quad (22)$$

A completely analogous computation establishes that

$$0 \leq \text{Cov}(Y, Z_Y) - \text{Cov}(X, Z_X) = \int_{\mathbb{R}} x \delta(x) \mu(dx) = - \int_{\mathbb{R}} \Delta(x) dx. \quad (23)$$

By convexity of g , $g'(x)$ is weakly increasing and thus

$$- \int_{\mathbb{R}} g'(x) \Delta(x) dx \geq -g'(x_0) \int_{\mathbb{R}} \Delta(x) dx \quad (24)$$

where we used the x_0 with $\{\Delta \geq 0\} \leq x_0 \leq \{\Delta \leq 0\}$ shown to exist in Lemma 3. Using (22) and (23), this establishes the weak inequality $\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) \geq 0$.

For the case where $g'(x)$ is strictly increasing, notice that according to Lemma 3 $\Delta \neq 0$ on a non-zero measure set and so the inequality in (24) is strict. In that case, $\text{Cov}(g(Y), Z_Y) - \text{Cov}(g(X), Z_X) > 0$. This proves Proposition 6. \square

A.3 Variances

In order to prove Theorem 1 itself, we can now apply Proposition 6 twice. This is essentially the idea behind the following proof.

Proof of Theorem 1. Pick $Z_X = X$ and $Z_Y = Y$. Notice that Assumptions 4.1 and 4.2 are satisfied by construction. Assumption 4.3 holds because

$$\frac{q_{Z_X}(F_X(x))}{q_{Z_Y}(F_Y(x))} = \frac{x - \mu_X}{x - \mu_Y} = 1 + \frac{\mu_Y - \mu_X}{x - \mu_Y},$$

which is strictly decreasing. Also, the function $x \mapsto q_{Z_X}(F_X(x)) = x - \mu_X$ has a unique zero at $x = \mu_X$ which is clearly strictly less than $x = \mu_Y$, the unique zero of $x \mapsto q_{Z_Y}(F_Y(x)) = x - \mu_Y$. Therefore, Assumption 4.4 is satisfied as well and we can apply Proposition 6. Since by assumption Assumption 1(ii), $\text{Cov}(X, X) = \text{Var}\{X\} \leq \text{Var}\{Y\} = \text{Cov}(Y, Y)$, Proposition 6 implies that

$$\text{Cov}(g(X), X) \leq \text{Cov}(g(Y), Y),$$

with strict inequality if g is strictly convex.

In a second step, we apply Proposition 6 again, but this time using $Z_X = g(X)$ and $Z_Y = g(Y)$. Again, Assumptions 4.1 and 4.2 are satisfied by construction, and Assumption 4.3 holds because

$$\frac{q_{g(X)}(F_X(x))}{q_{g(Y)}(F_Y(x))} = \frac{g(x) - \mu_{g(X)}}{g(x) - \mu_{g(Y)}} = 1 + \frac{\mu_{g(Y)} - \mu_{g(X)}}{g(x) - \mu_{g(Y)}},$$

which is still strictly decreasing given that $g(Y)$ MLRP-dominates $g(X)$ and hence has a larger mean.

Assumption 4.4 is satisfied because the unique zero of $x \mapsto g(x) - \mu_{g(X)}$, $g^{-1}(\mu_{g(X)})$, is strictly smaller than $g^{-1}(\mu_{g(Y)})$, which is the unique zero of $x \mapsto g(x) - \mu_{g(Y)}$. In sum, the conditions of Proposition 6 are satisfied, and we obtain

$$\text{Var}\{g(X)\} = \text{Cov}(g(X), g(X)) \leq \text{Cov}(g(Y), g(Y)) = \text{Var}\{g(Y)\},$$

with strict inequality if g is strictly convex. This concludes the proof of Theorem 1. \square

B Proof of Theorem 2

In this proof, we assume without loss of generality that $\mathbb{E}X = 0$ (a simple horizontal translation of X and g establishes the general result). Also, we only proof the strict version of the theorem. The weak version follows by replacing strict inequalities with weak ones throughout the proof.

By Assumption 2(i) and $\mathbb{E}X = 0$ it follows that $\alpha_1 > 0$. By Assumption 2(ii) it follows that $\alpha_2 > 0$. Define the function $G_{g,X} : [0, \alpha_1] \times [0, \alpha_2] \rightarrow \mathbb{R}$ by

$$G_{g,X}(\beta_1, \beta_2) = \text{Var}\{g(\beta_1 + \beta_2 X)\} - \text{Var}\{g(X)\}.$$

For convenience we drop the subscripts of $G_{g,X}$ during the proof, yet since the proof of Theorem 3 is very similar, it is good to keep the explicit dependence of $G_{g,X}$ on g and X in mind.³⁴

We now prove that G is strictly increasing in (i) β_1 and (ii) β_2 . Notice that this also establishes the well-definedness (finiteness) of G . To prove parts (i) and (ii) we use a simple lemma which we study first. For this, let us introduce two real, non-degenerate random variables Z_1 and Z_2 which satisfy the following assumption.

Assumption 5. Z_1 and Z_2 are perfectly rank correlated, that is, there exists a random variable $U \sim \text{unif}[0, 1]$ that is uniformly distributed over the unit interval such that $Z_i = Q_{Z_i}(U)$, for some weakly increasing measurable functions $Q_{Z_i} : [0, 1] \rightarrow \mathbb{R}$.

Let $\text{supp}(Z_2)$ denote the support of Z_2 . The key lemma is as follows.

Lemma 4. *Suppose random variables Z_1 and Z_2 satisfy Assumption 5. Suppose further that $h : \text{supp}(Z_2) \rightarrow \mathbb{R}$ is a continuous function with $h(z) - z$ increasing in z . Then,*

$$\text{Cov}(Z_1, Z_2) \leq \text{Cov}(Z_1, h(Z_2)) \tag{25}$$

and strictly so if $h(z) - z$ is strictly increasing.

Proof. Notice that

$$\text{Cov}(Z_1, h(Z_2)) - \text{Cov}(Z_1, Z_2) = \text{Cov}(Z_1, h(Z_2) - Z_2) \geq 0$$

³⁴Essentially, the proof of Theorem 3 proceeds by changing g and X in the definition of $G_{g,X}$.

since both Z_1 and $h(Z_2) - Z_2$ can be expressed as weakly increasing functions of U , and it is known that the covariance of two perfectly rank-correlated random variables is non-negative (see, e.g., [Schmidt, 2003](#) for an elementary proof). The covariance $Cov(Z_1, h(Z_2) - Z_2)$ can only be equal to zero if Z_1 were degenerate, which is ruled out by assumption, or if $h(z) - z$ were (a.e.) constant over the support of Z_2 . Hence, if $h(z) - z$ is strictly increasing, (25) can only be strict. \square

Part (i): G strictly increasing in β_1 . Consider $\beta_1 \in [0, \alpha_1)$, $\beta_2 \in [1, \alpha_2]$ and $\beta'_1 \in (\beta_1, \alpha_1]$, i.e. $\beta'_1 > \beta_1$. For simplicity, we let $Z \equiv g(\beta_1 + \beta_2 X)$ and $Z' \equiv g(\beta'_1 + \beta_2 X)$ with the goal to show that $\text{Var}\{Z\} < \text{Var}\{Z'\}$. Define the function $h : (g, \infty) \rightarrow \mathbb{R}$ by $h(z) = g(g^{-1}(z) + (\beta'_1 - \beta_1))$ where $g = \inf_{z \in \mathbb{R}} g(z) \in [-\infty, \infty)$. By definition, h is strictly increasing and $h(Z) = Z'$.

We now prove that $h(z) - z$ is strictly increasing. First, rewrite $h(z) - z$ as

$$h(z) - z = g(g^{-1}(z) + (\beta'_1 - \beta_1)) - g(g^{-1}(z)).$$

Introduce the notation $\epsilon \equiv \beta'_1 - \beta_1 > 0$ and $a \equiv g^{-1}(z)$. By strict convexity of g , it holds for any $\Delta > 0$ that

$$\begin{aligned} g(a) + g(a + \Delta + \epsilon) &= \underbrace{\frac{\Delta}{\Delta + \epsilon} g(a) + \frac{\epsilon}{\Delta + \epsilon} g(a + \Delta + \epsilon)}_{>g(a+\epsilon)} \\ &\quad + \underbrace{\frac{\epsilon}{\Delta + \epsilon} g(a) + \frac{\Delta}{\Delta + \epsilon} g(a + \Delta + \epsilon)}_{>g(a+\Delta)} \\ &> g(a + \epsilon) + g(a + \Delta) \end{aligned}$$

and therefore, for any $z' > z$, defining $\Delta \equiv g^{-1}(z') - g^{-1}(z)$,

$$\begin{aligned} h(z') - z' &= g(a + \Delta + \epsilon) - g(a + \Delta) \\ &> g(a + \epsilon) - g(a) = h(z) - z, \end{aligned}$$

establishing that $h(z) - z$ is strictly increasing.

Notice that Z and Z' are by construction perfectly rank-correlated, using $U \equiv X$. Thus, applying Lemma 4 twice, we find

$$\text{Var}\{Z\} = Cov(Z, Z) < Cov(Z, \underbrace{h(Z)}_{Z'}) < Cov(Z', Z') = \text{Var}\{Z'\},$$

concluding the proof of part (i).

Part (ii): G strictly increasing in β_2 . Consider $\beta_1 \in [0, \alpha_1)$, $\beta_2 \in [1, \alpha_2)$ and $\beta'_2 \in (\beta_2, \alpha_2]$. Define $Z = g(\beta_1 + \beta_2 X)$ and $Z' = g(\beta_1 + \beta'_2 X)$. The goal is again to show that $\text{Var}\{Z\} < \text{Var}\{Z'\}$.

By construction, $h(Z) = Z'$, where we now define

$$h(z) = g\left(\frac{\beta'_2}{\beta_2}(g^{-1}(z) - \beta_1) + \beta_1\right).$$

In the following, denote $\chi \equiv \beta'_2/\beta_2 > 1$ and let z^* be the unique value such that $g^{-1}(z^*) = \beta_1$. Notice that $h(z^*) = z^*$. We now prove again that $h(z) - z$ is strictly increasing.

By definition of h it holds that

$$g^{-1}(h(z')) - g^{-1}(h(z)) = \chi(g^{-1}(z') - g^{-1}(z)),$$

and, defining $a \equiv g^{-1}(h(z'))$, $b \equiv g^{-1}(h(z))$, $c \equiv g^{-1}(z')$, $d \equiv g^{-1}(z)$, this implies that $\chi = (a - b)/(c - d)$ and

$$\begin{aligned} g(a) + \chi g(d) &= g(a) \left(\frac{b-d}{a-d} + \chi \frac{c-d}{a-d}\right) + \chi g(d) \left(\frac{a-c}{a-d} + \frac{a-b}{a-d} \chi^{-1}\right) \\ &\geq \chi g(c) + g(b), \end{aligned}$$

or rearranging this,

$$h(z') - z' \geq h(z) - z + (z' - z)(\chi - 1) > h(z) - z.$$

Thus, applying Lemma 4 twice again, we find

$$\text{Var}\{Z\} = \text{Cov}(Z, Z) < \underbrace{\text{Cov}(Z, h(Z))}_{Z'} < \text{Cov}(Z', Z') = \text{Var}\{Z'\},$$

concluding the proof of part (ii).

C Proof of Theorem 3

This proof uses a similar strategy as the proof of Theorem 2 in Appendix B above. We assume without loss that $\mathbb{E}\{g(X)\} = 0$ which can be achieved by a simple vertical translation of g . The assumption that $\mathbb{E}\{g(Y)\} > \mathbb{E}\{g(X)\}$ then implies that $\alpha_1 > 0$. Define the function $G_{g^{-1}, g(X)} : [0, \alpha_1] \times [0, \alpha_2] \rightarrow \mathbb{R}$ by

$$G_{g^{-1}, g(X)}(\beta_1, \beta_2) = \text{Var}\{g^{-1}(\beta_1 + \beta_2 g(X))\} - \text{Var}\{X\}.$$

Notice that $G_{g^{-1}, g(X)}(0, 1) = 0$, $G_{g^{-1}, g(X)}(\alpha_1, \alpha_2) = \text{Var}\{Y\} - \text{Var}\{X\} \geq 0$, and $\text{Var}\{g(Y)\} = \alpha_2^2 \text{Var}\{g(X)\}$, so it remains to be shown that $\alpha_2 > 1$. We now prove that $G_{g^{-1}, g(X)}(\beta_1, \beta_2)$ is strictly decreasing in β_1 and strictly increasing in β_2 , and thus, due to $\alpha_1 > 0$, $G_{g^{-1}, g(X)}(\alpha_1, \alpha_2) \geq 0$ necessitates that $\alpha_2 > 1$.

The function g^{-1} is increasing and concave, so the function $h(z) \equiv -g^{-1}(-z)$ is increasing and

strictly convex. Observe that

$$G_{g^{-1},g(X)}(\beta_1, \beta_2) = G_{h,h^{-1}(-X)}(-\beta_1, \beta_2).$$

As h is increasing convex and $\mathbb{E}\{h^{-1}(-X)\} = 0$, we can follow the exact steps of the proof of Theorem 2 to show that $G_{h,h^{-1}(-X)}$ is strictly increasing in both its first and second arguments. Therefore, $G_{g^{-1},g(X)}(\beta_1, \beta_2)$ strictly decreases in β_1 and strictly increases in β_2 , which concludes the proof of Theorem 3.

D Proof of Proposition 5

We only prove that the risk term increases when g is convexly increasing. All other cases can be reduced to this case by simple horizontal or vertical reflections of g . Hence, using the notation from Section 4.2, we need to prove that

$$\text{Var}\{X_0\} = \text{Var}\{X'_0\} \Rightarrow \text{Cov}(X_0, g(X_0)) \leq \text{Cov}(X'_0, g(X'_0)).$$

Upon replacing X_0 by X and X'_0 by Y , this is the exact first step in the proof of Theorem 1 in Appendix A.3.

E Appendix for the applications

E.1 Proof of Lemma 1

We prove both directions of this result.

Start with an allocation $\{C_t, K_{t+1}, N_t\}$. The claim is that if the allocation solves (15), there exists an equilibrium to support the allocation. Given the processes $\{z_t, \epsilon_{it}\}$, define the candidate equilibrium as follows: $W_t \equiv F_N(K_t, N_t)$, $r_t \equiv e^{z_t} F_K(K_t, N_t) - \delta$, $Q_t \equiv (1 + r_t)^{-1} Q_{t-1}$, $Q_0 \equiv 1$, $k_{it+1} \equiv k(\epsilon_{it})$, $n_{it} \equiv n(\epsilon_{it-1}, u_{it})$, y_{it} is set according to the production (12), profits π_{it} are set according to (8), v_{it} is set according to (7), and $V_t \equiv \int_0^1 v_{it} di$. Condition (a) is satisfied, since the first order condition for $k(\tilde{\epsilon})$ in (16) is

$$\mathbb{E}_{t-1} \frac{d}{dk_{it}} f(k_{it}, n_{it})^\nu = F_K(K_t, N_t),$$

which, due to the orthogonality of aggregate and idiosyncratic shocks immediately implies that

$$\mathbb{E}_{t-1} \left\{ \frac{Q_t}{Q_{t-1}} \left(\frac{d}{dk_{it}} e^{z_t} f(k_{it}, n_{it})^\nu + 1 - \delta \right) \right\} = \frac{e^{z_t} F_K(K_t, N_t) + 1 - \delta}{1 + r_t} = 1,$$

which is precisely the first order condition for firm i 's profit maximization problem (7). Condition (b) holds because the optimality conditions are the same as in the planning problem. The household's

intertemporal optimality condition is $C_t^{-1} = \beta \frac{Q_t}{Q_{t+1}} C_{t+1}^{-1}$ which is equivalent to the planner's by construction of Q_t . The intratemporal optimality condition is $\theta N_t^\zeta C_t = W_t$ which is again equivalent to the planner's by construction of W_t . Finally, the constraints are equivalent since from (7) and (8), it follows that

$$\begin{aligned} V_t &= \int_0^1 \pi_{it} di + \mathbb{E}_t \frac{Q_{t+1}}{Q_t} V_{t+1} \\ &= e^{z_t} F(K_t, N_t) - W_t N_t + (1 - \delta) K_t - K_{t+1} + \mathbb{E}_t \frac{Q_{t+1}}{Q_t} V_{t+1}. \end{aligned} \quad (26)$$

Substituting out V_t from (14) using (26) yields the resource constraint. Conditions (c) and (d) hold by construction.

The other direction of the lemma follows these exact same steps in reverse. Defining F as in (16) means that by virtue of the first order conditions, it has to be the case that $k(\epsilon_{it}) = k_{it+1}$ and $n_{it} = n(\epsilon_{it-1}, u_{it})$, as well as $W_t = F_N(K_t, N_t)$, $r_t = e^{z_t} F_K(K_t, N_t) - \delta$, and $e^{z_t} F(K_t, N_t) = \int y_{it} di$. The constraints and optimality conditions of the planning problem and the household optimization problem are then equivalent as before. This proves that $\{C_t, K_{t+1}, N_t\}$ satisfies 15.

E.2 Business cycle accounting

To do the business cycle accounting exercise, we introduce in addition to the labor wedge also an investment wedge $\tau_{x,t}$ and government spending G_t . The extended model is described by the following four equations:

$$\begin{aligned} C_t N_t^\zeta &= (1 - \tau_{n,t}) e^{z_t} F_N(K_t, N_t) \\ (1 + \tau_{x,t}) C_t^{-1} &= \beta \mathbb{E}_t \left\{ C_{t+1}^{-1} [e^{z_{t+1}} F_K(K_{t+1}, N_{t+1}) + (1 + \tau_{x,t+1})(1 - \delta)] \right\} \\ Y_t &= C_t + K_{t+1} - (1 - \delta) K_t + G_t \\ Y_t &= e^{z_t} F(K_t, N_t). \end{aligned}$$

We proceed similar to [Chari, Kehoe and McGrattan \(2007\)](#) and pick a path for $\{\tau_{n,t}, \tau_{x,t}, G_t, e^{z_t}\}$ to satisfy all four equations perfectly when evaluated at the realized paths of US aggregate data for output, hours, consumption, and investment from 1947 to 2016. Since our model does not feature growth, we first eliminate the growth trend from the data, detrending either linearly or applying a high-pass filter with a cutoff of 40 years. For the accounting exercise, we begin by conjecturing a VAR(1) process for the four exogenous shocks and solve for the equilibrium mapping from shocks to endogenous variables. We then filter through the data using the equilibrium mapping, extracting a path for the four wedges so that the model matches the (detrended) data exactly. In the final step, we verify that the resulting timeseries of wedges is well-described by the conjectured VAR(1)

process.³⁵

References

- Albagli, Elias, Christian Hellwig, and Aleh Tsyvinski.** 2015. “A Theory of Asset Pricing Based on Heterogeneous Information.” *working paper*.
- Bachmann, Rüdiger, and Christian Bayer.** 2013. “‘Wait-and-See’ business cycles?” *Journal of Monetary Economics*, 60(6): 704–719.
- Bachmann, Rüdiger, and Christian Bayer.** 2014. “Investment Dispersion and the Business Cycle.” *American Economic Review*, 104(4): 1392–1416.
- Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims.** 2013. “Uncertainty and Economic Activity: Evidence from Business Survey Data.” *American Economic Journal: Macroeconomics*, 5(2): 217–49.
- Barkai, Simcha.** 2017. “Declining Labor and Capital Shares.” *working paper*.
- Bartoszewicz, Jaroslaw.** 1985. “Moment Inequalities for Order Statistics from Ordered Families of Distributions.” *Metrika*, 32: 383–389.
- Berger, David, and Joseph Vavra.** 2010. “Dynamics of the US price distribution.” *mimeo*.
- Bloom, Nicholas.** 2009. “The Impact of Uncertainty Shocks.” *Econometrica*, 77(3): 623–685.
- Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen Terry.** 2016. “Really Uncertain Business Cycles.” *working paper*.
- Buera, Francisco J., and Benjamin Moll.** 2015. “Aggregate implications of a credit crunch.” *American Economic Journal: Macroeconomics*, 7(3): 1–42.
- Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan.** 2007. “Business cycle accounting.” *Econometrica*, 75(3): 781–836.
- Chirinko, Robert S., Steven M. Fazzari, and Andrew P. Meyer.** 1999. “How responsive is business capital formation to its user cost? An exploration with micro data.” *Journal of Public Economics*, 74: 53–80.
- Chirinko, Robert S., Steven M. Fazzari, and Andrew P. Meyer.** 2011. “A New Approach to Estimating Production Function Parameters: The Elusive Capital- Labor Substitution Elasticity.” *Journal of Business & Economic Statistics*, 29: 587–594.

³⁵At the fixed point to our algorithm, the coefficients of the conjectured process are numerically identical with VAR(1) estimates on the filtered wedge data and there are no signs of (auto) correlation among the filtered innovations.

- Christiano, Lawrence J., Roberto Motto, and Massimo Rostagno.** 2014. “Risk Shocks.” *American Economic Review*, 104(1): 27–65.
- Cui, Wei.** 2014. “Delayed Capital Reallocation.” *UCL working paper*.
- Döpke, Jörg, and Sebastian Weber.** 2010. “The within-distribution business cycle dynamics of German firms.” *Applied Economics*, 42(29): 3789–3802.
- Döpke, Jörg, Michael Funke, Sean Holly, and Sebastian Weber.** 2005. “The cross-sectional dynamics of German business cycles: a bird’s eye view.” *Deutsche Bundesbank Discussion Paper Series 1*.
- Fajgelbaum, Pablo, Edouard Schaal, and Mathieu Taschereau-Dumouchel.** 2017. “Uncertainty traps.” *Quarterly Journal of Economics*, 132(4): 1641–1692.
- Garcia-Santana, Manuel, Enrique Moral-Benito, and Josep Pijoan-Mas.** 2015. “Growin like Spain: 1995–2007.” *working paper*.
- Gilchrist, Simon, Jae W. Sim, and Egon Zakrajšek.** 2016. “Uncertainty, Financial Frictions, and Investment Dynamics.” *working paper*.
- Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez.** 2015. “Capital Allocation and Productivity in South Europe.” *working paper*.
- Gourio, Francois.** 2008. “Estimating firm-level risk.” *Boston University*.
- Higson, Chris, Sean Holly, and Paul Kattuman.** 2002. “The cross-sectional dynamics of the US business cycle: 1950–1999.” *Journal of Economic Dynamics and Control*, 26(9): 1539–1555.
- Higson, Chris, Sean Holly, Paul Kattuman, and Stylianos Platis.** 2004. “The Business Cycle, Macroeconomic Shocks and the Cross-Section: The Growth of UK Quoted Companies.” *Economica*, 71(282): 299–318.
- Hsieh, Chang-Tai, and Peter J. Klenow.** 2009. “Misallocation and Manufacturing Tfp in China and India.” *Quarterly Journal of Economics*, 124(4): 1–55.
- Iltut, Cosmin, Matthias Kehrig, and Martin Schneider.** 2016. “Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News.” *working paper*.
- Kehrig, Matthias.** 2015. “The Cyclical Nature of the Productivity Distribution.” *SSRN Working Paper*.
- Kopecky, Karen A., and Richard M H Suen.** 2010. “Finite state Markov-chain approximations to highly persistent processes.” *Review of Economic Dynamics*, 13(3): 701–714.

- Kozeniauskas, Nicholas, Anna Orlik, and Laura L. Veldkamp.** 2016. “The Common Origin of Uncertainty Shocks.” *NYU working paper*.
- Moll, Benjamin.** 2014. “Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?” *American Economic Review*, 104(10): 3186–3221.
- Nimark, Kristoffer P.** 2014. “Man-Bites-Dog Business Cycles.” *American Economic Review*, 104(8): 2320–2367.
- Oberfield, Ezra, and Devesh Raval.** 2014. “Micro Data and Macro Technology.” *NBER Working Paper No. 20452*.
- Orlik, Anna, and Laura L. Veldkamp.** 2014. “Understanding Uncertainty Shocks and the Role of Black Swans.” *mimeo*.
- Raval, Devesh.** 2015. “The Micro Elasticity of Substitution and Non-Neutral Technology.” *working paper*.
- Schmidt, Klaus D.** 2003. “On the covariance of monotone functions of a random variable.” *mimeo*.
- Senga, Tatsuro.** 2016. “A New Look at Uncertainty Shocks: Imperfect Information and Misallocation.” *working paper*.
- Straub, Ludwig, and Robert Ulbricht.** 2012. “Credit Crunches, Information Failures, and the Persistence of Pessimism.” *mimeo, Toulouse School of Economics*.
- Straub, Ludwig, and Robert Ulbricht.** 2014. “Endogenous Uncertainty and Credit Crunches.” *working paper*.
- van Nieuwerburgh, Stijn, and Laura L. Veldkamp.** 2006. “Learning asymmetries in real business cycles.” *Journal of Monetary Economics*, 53(4): 753–772.
- Van Zwet, William Rutger.** 1964. “Convex transformations of random variables.” *MC Tracts*, 7: 1–116.