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

Ludwig Straub
MIT

Robert Ulbricht
Toulouse School of Economics

June 15, 2016

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 cyclicality 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". Ludwig Straub acknowledges financial support from the Macro-Financial Modeling Group. *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 they are *nonlinear* transformations of some fundamental. Specifically, we consider a mathematical 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, 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 cyclicality 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. We provide mathematical 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 mathematical underpinning for a class of models where a single "first-moment" change in fundamentals causes fluctuations in first *and* second moments in endogenous variables.

The usefulness of our results is illustrated in a series of four applications. The first of these 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. 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. If the economy further exhibits sufficiently non-

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

convex marginal adjustment costs, the dispersion of investment is procyclical. 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 "uncertainty"-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, an aggregate shift in the distribution of financial shocks leads to counter-cyclical fluctuations in the cross-sectional dispersion of MRPKs. In contrast, we show that productivity-driven fluctuations may lead to pro-cyclical dispersions whenever the elasticity of borrowing limits to firm revenues is smaller than one in the cross-section.

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. In this setup, when θ realizes in a range where g tends to be rather flat, 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.

Mathematical results To be as broadly applicable as possible, our main results are 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 $g(Y)$ 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.[2]

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

[2]For 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.

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.[3] 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.

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 relation[4] 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.[5] 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

Contribution to the Literature. In sum, the contribution of our paper is twofold: First, we 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)[6], then a similar result to ours holds, namely that $\mathrm{Var}\, g(Y) > \mathrm{Var}\, g(X)$ for any convexly increasing function g. Compared to Bartoszewicz (1985), our main result has the advantage that our stochastic orders (MLRP and $\mathrm{Var}X \leq \mathrm{Var}Y$) are in most applications easy to check while it can be hard to work with the convex stochastic order in

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

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

[5] Jensen'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.

[6] The 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.

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).

Second, we argue using a sequence of applications that these theoretical 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 cyclicality of the dispersion of firm- or plant-level statistics.[7] In particular, it naturally complements Ilut, Kehrig and Schneider (2014) 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 factor inputs in that they introduce cyclical shifts in the dispersions of firm aggregates. Application 2 contributes to recent studies on the cyclicality 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 cyclicality of Bayesian uncertainty when signals are nonlinear functions of the economy's fundamental. Theories about the cyclicality of uncertainty have recently gained attention by a number of authors (see, e.g., van Nieuwerburgh and Veldkamp 2006; Orlik and Veldkamp 2014; Straub and Ulbricht 2012, 2014; Fajgelbaum, Schaal and Taschereau-Dumouchel 2015) in the face of growing evidence and potential relevance of changes in uncertainty.

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. In Section 3 we apply these results to four simple models from macroeconomics and finance. Section 4 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 present two corollaries to our results.

For all our three general results, we focus on 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, $\mathcal{B}(\mathbb{R})$

[7] The cyclicality 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. (2014), 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.

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.[8]

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. 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.

Two remarks are in place. First, our notation is general enough to nest both *continuous* and *discrete* random variables (as well as "mixed" continuous and discrete variables, of course). In the one extreme, choosing μ to be equal to the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ allows for continuous distributions. In that case, f_Y and f_X are just the 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. In the other extreme, choosing μ to have discrete support, the setup can account for cases where X and Y are discrete. For example, a measure[9] $\mu = \sum_{i=1}^{N} \frac{1}{N} \delta_{z_i}$ puts equal weight on numbers $\{z_1, \ldots, z_N\}$ (assumed to be in strictly increasing order), in which case 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).

Second, 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

[8] It 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).

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

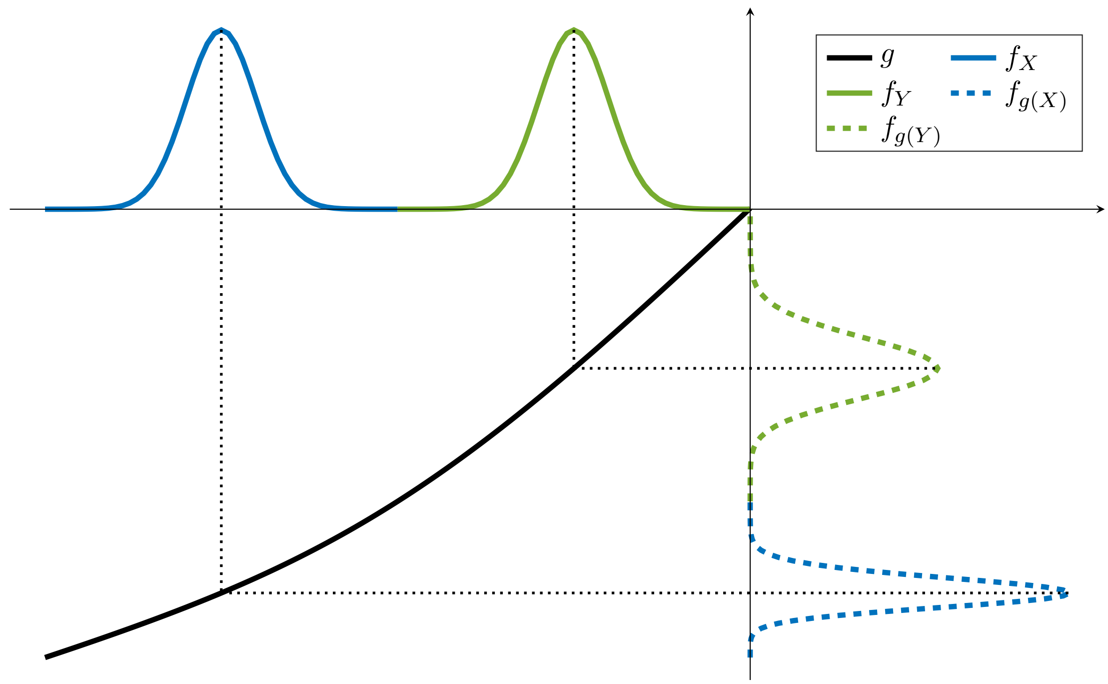

Figure 1: Convex transformation of a translation

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.

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 $g : \mathbb{R} \to \mathbb{R}$,[10]

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

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 $g : \mathbb{R} \to \mathbb{R}$ be a convex and strictly increasing function. Then,*

$$\mathrm{Var}\{g(X)\} \leq \mathrm{Var}\{g(Y)\}, \tag{1}$$

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

[10] Throughout this paper, we use the term "strictly convex" to describe an a.e. differentiable function whose derivative is a.e. strictly increasing.

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 that[11]

$$\mathrm{Var}\{g(Y)\} \geq \left(g'(z)\right)^2 \mathrm{Var}\{Y\} \geq \left(g'(z)\right)^2 \mathrm{Var}\{X\} \geq \mathrm{Var}\{g(X)\}. \tag{2}$$

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. Cleary, 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 $\mathrm{Var}\{X\} = 0.36$ and $\mathrm{Var}\{Y\} = 0.4275$. In particular, $\mathrm{Var}\{Y\} \geq \mathrm{Var}\{X\}$.

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

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

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.

We now outline the main steps in the proof of Theorem 1. The actual proof can be found in Appendix A. The proof works in two steps: First, we prove a version of inequality (1) for covariances, rather than variances. In particular, we show that when a random variable Z_X is perfectly rank-correlated with X, and similarly another random variable Z_Y is perfectly rank-correlated with Y, and Z_Y MLRP-dominates Z_X, then:

$$Cov(X, Z_X) \leq Cov(Y, Z_Y) \Rightarrow Cov(g(X), Z_X) \leq Cov(g(Y), Z_Y). \tag{3}$$

[11]For this example, we tacitly assume that g is differentiable at z.

Establishing (3) before proceeding to (1) simplifies the proof since both inequalities in (3) are *linear* in X, Y and $g(X), g(Y)$—compared to (1), which is *quadratic* in $g(X), g(Y)$. Equipped with (3), we can then essentially apply the inequality twice to complete the proof: Since Y MLRP-dominates X, it follows that:

$$Cov(X, X) \leq Cov(Y, Y) \Rightarrow Cov(g(X), X) \leq Cov(g(Y), Y) \tag{4}$$

where (3) was applied to the case of $Z_X = X$ and $Z_Y = Y$. Similarly, since the MLRP order is inherited by $g(Y)$ and $g(X)$, (3) can be applied once more, this time using $Z_X = g(X)$ and $Z_Y = g(Y)$, to get:

$$Cov(X, g(X)) \leq Cov(Y, g(Y)) \Rightarrow Cov(g(X), g(X)) \leq Cov(g(Y), g(Y)). \tag{5}$$

Together, (4) and (5) prove Theorem 1:

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

2.2 Two alternative stochastic orderings

We now allow for two alternative stochastic orderings and prove corresponding variance transformation theorems. Here, our first result, Theorem 2, formally establishes a commonly used alternative to our 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 $g : \mathbb{R} \to \mathbb{R}$ 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 orderings We now prove that a similar result to Theorem 1 holds under our two alternative stochastic orderings. We start with the case where Y is a simple positive affine-linear translation of X.

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

$$\mathrm{Var}\{g(X)\} \leq \mathrm{Var}\{g(Y)\}, \tag{6}$$

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 $G_{g,X}(\beta_1, \beta_2) \equiv \mathrm{Var}\{g(\beta_1 + \beta_2 X)\} - \mathrm{Var}\{g(X)\}$, 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 $G_{g,X}(\alpha_1, \alpha_2) = \mathrm{Var}\{g(Y)\} - \mathrm{Var}\{g(X)\}$. Since $G_{g,X}$ is shown to increase as we vary (β_1, β_2) from $(0, 1)$ to (α_1, α_2), (6) 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 $g : \mathbb{R} \to \mathbb{R}$ a convex, strictly increasing function. Then,*

$$\mathrm{Var}\{g(X)\} \leq \mathrm{Var}\{g(Y)\}, \tag{7}$$

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 $g(X) \sim \mathcal{N}(\mu_{g(X)}, \sigma^2_{g(X)})$ and $g(Y) \sim \mathcal{N}(\mu_{g(Y)}, \sigma^2_{g(Y)})$, with $\mu_{g(X)} < \mu_{g(Y)}$. The question Theorem 3 asks is what we can say about the relationship between $\sigma^2_{g(Y)}$ and $\sigma^2_{g(X)}$ if we know that $\mathrm{Var}\{X\} \leq \mathrm{Var}\{Y\}$. The answer is that $\sigma^2_{g(X)} < \sigma^2_{g(Y)}$ 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 $G_{g^{-1},g(X)}(\beta_1, \beta_2) = \mathrm{Var}\{g^{-1}(\beta_1 + \beta_2 g(X))\} - \mathrm{Var}\{X\}$. Again, $G(0, 1) = 0$, but now it also holds by assumption that $G(\alpha_1, \alpha_2) \geq 0$. To prove (7) 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* $\mathrm{Var}\{X\} \leq \mathrm{Var}\{Y\}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *is a concave, decreasing function, then* $\mathrm{Var}\{g(X)\} \leq \mathrm{Var}\{g(Y)\}$.
2. *If* $\mathrm{Var}\{X\} \geq \mathrm{Var}\{Y\}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *is a concave, increasing function, then* $\mathrm{Var}\{g(X)\} \geq \mathrm{Var}\{g(Y)\}$.
3. *If* $\mathrm{Var}\{X\} \geq \mathrm{Var}\{Y\}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *is a convex, decreasing function, then* $\mathrm{Var}\{g(X)\} \geq \mathrm{Var}\{g(Y)\}$.

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)* $\mathrm{Var}\{\tilde{Y}\} \geq \mathrm{Var}\{Y\} \geq \mathrm{Var}\{X\}$*. Then,*

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

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.

3 Four Applications

In the following, we present four applications that illustrate the economic importance of the transformation behavior of second moments. In particular, the first two examples are about the *dispersion* of macroeconomic aggregates. The first example asks the question whether observed cyclicalities of the dispersions of output, employment, and investment across firms or establishments can be consistent with first-order fundamental shifts that do not exhibit changes in the dispersion.

The second 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 third 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.

3.1 Cross-sectional dispersion over the business cycle

Recently, the cyclicality of the dispersions of macroeconomic variables has gained a renewed interest. For example, Bachmann and Bayer (2014) find in a panel of German non-financial firms that investment dispersion is positively correlated with detrended GDP. This stands in contrast to the countercyclicality of the dispersions of plant-level output, employment growth and other macro variables, such as productivity or prices.[12]

Here we explore how many of these facts can be explained through simple shifts in the distribution of idiosyncratic productivities and applying one of Theorems 1–3. For this purpose, consider the following stylized, single-period, partial equilibrium model:[13] Assume there is a continuum of firms, labeled by $i \in [0,1]$, with production technologies $Y_i = A_i f(K_i, N_i)$. Denoting by lower case letters the respective logarithms, firm i has log productivity $a_i \in \mathbb{R}$. Let $\mathcal{F}$ be the distribution of log productivities across firms. We assume that $f(K,N)$ is a CES production function, equal across firms, with an elasticity of substitution of $\sigma > 0$, i.e.

$$f(K,N) = \left(\alpha K^{(\sigma-1)/\sigma} + (1-\alpha) N^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)},$$

where $\alpha \in (0,1)$. Labor N_i is hired at market wage $W > 0$. We assume every firm starts with an ex-ante capital stock $K_{0,i}$ and invests an idiosyncratic amount I_i so that the capital stock used in production is $K_i = (1-\delta)K_{0,i} + I_i$, where $\delta \in (0,1)$ is the rate of depreciation. Funds for investment need to be rented at rate $r > 0$, and there are convex capital adjustment costs $K_{0,i}\phi(I/K_{0,i})$. The optimization problem of firm i is given by:

$$\max_{\{I_i, N_i\}} \left\{ Y_i + (1-\delta)K_i - W N_i - (1+r)I_i - K_{0,i}\phi(I_i/K_{0,i}) \right\}.$$

We approach this problem first in the case where the capital stock is predetermined, that is, $\phi = \infty$ for $I_i > 0$, and $K_i = (1-\delta)K_{0,i}$. In this case, the only firm choice variable is employment

[12] The cyclicality 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. (2014), 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.

[13] The model is easily transformable into a general equilibrium model. The partial equilibrium nature of this application merely permits us to demonstrate our theorems in the most accessible way.

N_i, characterized by[14]

$$f_N(1, N_i/K_i) = W/A_i. \tag{8}$$

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 $A_i \leq (1-\alpha)^{-\sigma/(\sigma-1)} W$ set $N_i = 0$ resulting in zero production. In the following, we assume that these firms exit the market, focusing our analysis on active firms.[15] Solving for N_i and taking logs this becomes

$$n_i = k_i + g(a_i - w) \tag{9}$$

with $g(x) = \log\left(f_N(1, \cdot)^{-1}(e^{-x})\right)$.[16] It is easy to verify that g is an increasing function and is concave (convex) whenever $\sigma < 1$ ($\sigma > 1$).[17]

Consider the following experiment akin to an aggregate TFP shock: Suppose the distribution of log productivities shifts from $\mathcal{F}$ to a distribution $\mathcal{F}'$, which dominates $\mathcal{F}$ according to either of Assumptions 1–3(i), but has equal dispersion: $\text{Var}_i\{a_i\} = \text{Var}_i\{a_i'\}$ (with slight abuse of notation). Applying our corollary of Theorems 1–3 to (9) we have the following proposition.

Proposition 1. *In the model with predetermined capital, when the cross-sectional distribution of log productivities $\mathcal{F}$ increases to a distribution $\mathcal{F}'$ such that one of Assumptions 1–3 holds for $\mathcal{F}$ and $\mathcal{F}'$, 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 cyclicality 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. It is well known that a firm's log wage bill $n_i + w$ is increasing in log wages w (and hence also in log "effective" wages $w - a_i$) precisely if $\sigma < 1$. For the result in Proposition 1, however, it is important that $n_i + w$ is in fact *convexly* increasing in $w - a_i$ if $\sigma < 1$. This then implies that when log productivities experience an upward shift as specified in Proposition 1, the variances of log employment and similarly log output must fall.

Proposition 1 makes two statements, one about the cyclicality of the dispersion in log employment and one about the cyclicality of the dispersion in log output. The former is an immediate consequence of the properties of g in equation (9). To prove the latter, note that (8) can be rearranged to

$$Y_i = (1-\alpha)^{-\sigma} A_i^{\sigma} N_i, \tag{10}$$

and therefore the cyclicality of the dispersion of log output is equal to the one of log employment.

[14]Here f_N and f_K denote partial derivatives of f.

[15]This could be for instance the result of a marginal operating cost $\epsilon \to 0$ that accrues when remaining active.

[16]Here, $f_N(1, \cdot)^{-1}$ is shorthand for the inverse of the function $x \mapsto f_N(1, x)$.

[17]Here 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.

What would an observer find for the behavior of the dispersion of log productivities themselves under a shift from $\mathcal{F}$ to $\mathcal{F}'$? Clearly, if the observer knew the true production function and firm-level 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, based on a Cobb-Douglas production function with an empirical capital share of $\beta \in (0, 1)$, he would find

$$a_i^{Solow} = y_i - \beta k_i - (1 - \beta) n_i.$$

This can be simplified to

$$a_i^{Solow} = -\sigma \log(1 - \alpha) + \sigma a_i - \beta k_i + \beta n_i,$$

which proves the following result.

Proposition 2. *In the model with predetermined capital, the cross-sectional dispersion of Solow residuals is countercyclical if $\sigma < 1$ and procyclical 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, including Antràs (2004), Klump, McAdam and Willman (2007), and Oberfield and Raval (2014), suggest that labor and capital inputs are indeed complements at the firm level (with σ ranging from 0.5 to 0.9).

So far, we assumed the capital stock to be predetermined. Relaxing this, suppose now that ϕ is a finite and differentiable function. The first order condition for labor is still given by $n_i - k_i = g(a_i - w)$ and the one for capital reads

$$A_i f_K(K_i/N_i, 1) = r + \delta + \phi'(I_i/K_{0,i}),$$

or in logs

$$\iota_i - k_{0,i} = h_1 \left(a_i + h_2 \left(n_i - k_i\right)\right)$$

where $h_2(x) = \log f_K(e^{-x}, 1)$ and $h_1(x) = \log(\phi')^{-1}(e^x - r - \delta)$. Using this representation, we have the following result.

Proposition 3. *For adjustment cost functions $\phi : \mathbb{R} \to \mathbb{R}$ for which $x \mapsto h_1(x + h_2(g(x - w)))$ is a convex function, it follows that the cross-sectional dispersion of investment is procyclical.*

Again, this proposition is a direct consequence of Corollary 1, once $x \mapsto h_1(x + h_2(g(x - w)))$ is a convex function. Because the standard assumption of convexity of ϕ places no restrictions on h_1 other than to be monotone increasing, this requirement can always be satisfied for sufficiently concave marginal adjustment cost functions ϕ'. As long the procyclicality of capital dispersion is not too strong, labor dispersion will continue to be countercyclical.

	$S.d.[n_i]$	$S.d.[y_i]$	$S.d.[a_i^{Solow}]$
$1\sigma^{\text{avg}}$ boom	1.30	1.46	0.55
	(-6.2%)	(-5.4%)	(-4.1%)
$1\sigma^{\text{avg}}$ recession	1.48	1.63	0.60
	$(+6.6\%)$	$(+5.7\%)$	$(+4.3\%)$

Table 1: Cross-sectional standard-deviation of firm-level statistics in numerical example. Differences (in percent) to the cross-sectional standard deviation in normal times are reported in parenthesis.

Numerical example To illustrate the potential quantitative importance of the ideas described above, we consider the following numerical example. For simplicity, we focus on the case with capital being predetermined. We set both the model's capital share α and the one used to compute the Solow residual β to 0.3. The elasticity of substitution between input factors σ is set equal to 0.7 which is in the middle of the range of the above-mentioned micro-data estimations. Our choice of productivity parameters is based on Gilchrist, Sim and Zakrajšek (2014). In particular, the cross-sectional distribution of productivities is normal with a standard deviation of $\sigma^{\text{cross}} = 0.25$, whereas the mean of the distribution is taken to be the realization of an aggregate productivity shock that has a standard deviation of $\sigma^{\text{avg}} = 0.038$.[18] Based on this, we define a boom and a recession as a $1\sigma^{\text{avg}}$ increase and decrease in the average cross-sectional productivities, respectively. Finally, we set the mean of the aggregate productivity shock so that 1 percent of firms exit the market ($N_i = 0$) at the median realization.

Table 1 shows the cross-sectional variance of employment, output and the Solow residuals in a $1\sigma^{\text{avg}}$ boom compared to a $1\sigma^{\text{avg}}$ recession. Clearly, a non-unit elasticity between factor inputs can generate a sizable cyclicality in the cross-section of firms. To evaluate the quantitative importance, we report the percentage difference to the cross-sectional dispersion in the absence of an aggregate productivity shift in parenthesis. An one standard-deviation shift in aggregate productivity can account for a cyclical variation of about 4-6 percent in the reported statistics. For comparison, Bachmann and Bayer (2009) report an annual fluctuation in the cross-sectional standard deviation of output growth rates and Solow residuals of 3.7 and 2.7 percent, whereas Kehrig (2015) finds that Solow residuals are on average 12 percent more dispersed during recessions compared to booms. The latter statistic suggests that a non-unit elasticity of factor inputs can account for roughly 70% (8.3 out of 12 percent) of the observed cyclical dispersion in Solow residuals without the need of an exogenous shock to second moments. Figure 2 visualizes the findings, reporting the cross-sectional distributions in output and Solow residuals for a large $2\sigma^{\text{avg}}$-shift in the productivity distribution. In addition to the aforementioned increase in dispersion, it can also be seen how recessions increase the skewness of the cross-sectional distribution consistent with recent evidence reported by, e.g., Salgado, Guvenen and Bloom (2015).

Overall we showed in this application how even a very simple framework can match the stylized

[18] Since our setup is static, we use the unconditional standard deviations in Gilchrist, Sim and Zakrajšek (2014) given by $0.15/\sqrt{1-0.8^2} = 0.25$ and $0.0075/\sqrt{1-0.98^2} \simeq 0.038$ to parametrize the model.

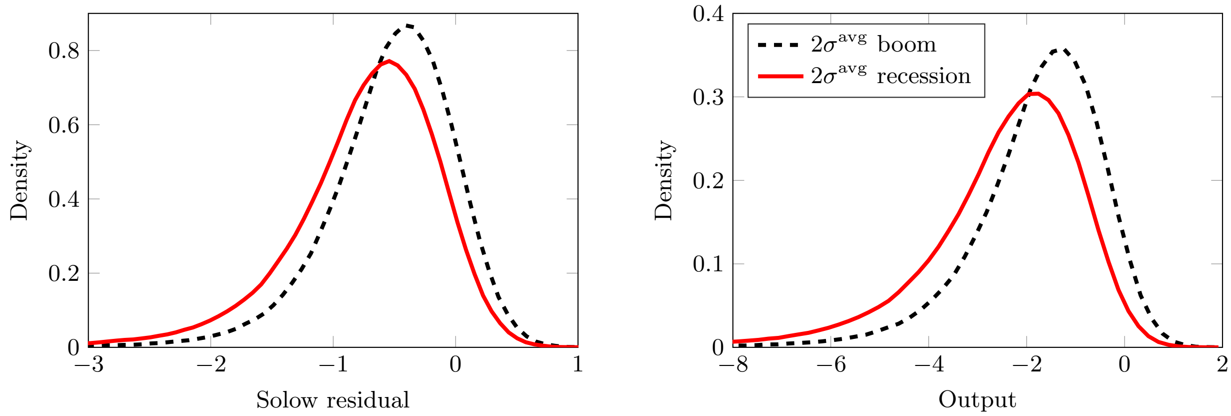

Figure 2: The cross-sectional variance of Solow residuals and output for large productivity shifts

facts about the dispersions of business cycle aggregates. The two key ingredients to match shifts in the cross-sectional distribution of productivities to shifts in the dispersions of macroeconomic variables were convex (or concave) relationships and the application of Theorems 1–3.

3.2 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.[19] With some slight abuse of terminology, we still refer to A_i as productivity

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

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 $\overline{K}_i$ which may depend on i. Thus firm i solves

$$\max_{K_i \le \overline{K}_i} A_i K_i^{\alpha} - rK_i.$$

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

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

which yields the following expression for the MRPK,

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

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

Proposition 4. *The dispersion of log MRPK increases if, other things equal,*

(i) all firms' productivities improve, and the cross-sectional distribution of log productivities shifts up in the sense of MLRP without changing its variance.

(ii) all firms' borrowing constraints tighten, and the cross-sectional distribution of log constraints shifts down in the sense of MLRP without changing its variance.

Proposition 4 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. For instance, suppose that borrowing constraints are a function of each firm's revenue, $A_t K_i^{\alpha}$, and some credit supply-driven factor u_i:

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

where $\epsilon < 1/\alpha$ is the elasticity of $\overline{K}_i$ to firm-revenues. Then Proposition 4 says that 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 downshift in the distribution of u_i would unambiguously increase the dispersion of MRPKs.

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

Table 2: 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.

Proposition 4(i) is 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. In our framework, this is not necessarily driven by large, productive firms that are already constrained and become more constrained during the boom. In fact, when the schedule of constraints $\overline{K}_i$ is steep enough, it is large productive firms that are unconstrained, while small firms might be constrained. A boom in productivity, labor supply, or aggregated demand, as in Proposition 4(i), then increases the number of small firms that are hitting their constraint, while large firms might still be far away from doing so.

Numerical example We again quantify the model using a simple numerical example with the same productivity distribution as in the previous application. The remaining parameters are set as follows. The capital share α is set equal to 0.3, the cost of capital r ist set equal to 0.05, and the elasticity of the constraint to revenues ϵ is set equal to 0.8.[20] 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 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 $u_i = \overline{u}_{10\%}$, so that a fraction of 10 percent of firms are constrained (without aggregate productivity shock).

Table 2 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*

[20]Since 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 $\epsilon < 1$.

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. The narrative suggested by our calibration is that during the boom years 2000–2008, an increase in the productivity term A_i could have been the main driver behind that increase. Interpreting our model more broadly, this could reflect an increase in labor supply (e.g. due to strong immigration into Spain) or an increase in employment caused by high aggregate demand, rather than a literal increase in productivity. During the financial crisis, our simple numerical exercise suggests that the dispersion could have increased due to financial pressure (possibly despite a reduction in the productivity term A_i). Given the comparatively strong effects of financial pressure, this narrative would be consistent with the acceleration in the increase in dispersion after 2008 documented by Gopinath et al. (2015).

The calibrated numbers are also consistent with cross-country evidence that consistently finds that the dispersion in MRPK is higher in countries that are likely to have less developed financial systems (e.g., Romania, Mexico and Slovenia have a MRPK distribution that is 46, 48 and 64 percent more dispersed compared to the U.S.; see Asker, Collard-wexler and Loecker, 2014 for details). While this prediction is not unique to our model, it suggests that tighter financial constraints can account for these differences.

3.3 Risk in financial derivatives

Our third example illustrates how the payoff profile of a security influences the cyclicality 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 cyclicality 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 $k \in \{0, 1, \ldots, K\}$ 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.[21]

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.[22] 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 Assumption 1 holds for X_0 and X_0', then the risk term in the price of asset k*

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, $\mathrm{Var}\{X_k\} = \mathrm{Var}\{g(X_0)\}$ 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 (3) 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 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 Assumption 1 holds for X_0 and X_0', then the risk premium of asset k*

[21] To 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 $(1+r)p_k$.

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

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.

3.4 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), Kozeniauskas, Orlik and Veldkamp (2014), and Albagli, Hellwig and Tsyvinski (2015). There are many natural reasons why one might want to consider nonlinear signal structures. For example, Straub and Ulbricht (2012, 2014) argue that financial frictions naturally lead to nonlinearities in the mapping from economic fundamentals to economic observables.

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. 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 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.

To be concrete, we focus in the following on the case where it satisfies the two conditions of Assumption 1: The posterior $\phi|s$ increases in the signal realization of signal s in the sense of the MLRP; and, the variance of the posterior $\text{Var}\{\phi|s\}$ is nondecreasing in s. Both assumptions are natural and satisfied in key examples (e.g., Straub and Ulbricht, 2012, 2014; and Orlik and Veldkamp, 2014). The first is a simple monotonicity requirement, demanding that higher signal realizations correspond to higher values of ϕ. The second is a precision requirement, demanding that the transformation g, and the transformed variable ϕ itself, are chosen such that the posterior variance does not shrink as the agent observes higher signal realizations.

What can our Theorems tell us about the posterior belief $\theta|s$? In the case where we impose Assumption 1, we already know that the posterior belief $\theta|s$ is increasing in s in the sense of the MLRP—the MLRP is preserved under monotone transformations such as g. While this implies that the *first moment* $\mathbb{E}\{\theta|s\}$ increases with s, this does not provide us with any information about the behavior of the *second moment.* Theorem 1 however, 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.

We conclude with a simple example that illustrates the results parametrically.

Example 3. Let $\phi \sim \mathcal{N}(\mu, \sigma^2)$ be a normally distributed random variable and let $\theta = \exp \phi$ be the (log-normally distributed) variable of interest, that is, $g = \log$ in this example. Suppose for simplicity the observer starts with an uninformative prior before receiving a signal $s = \phi + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. It is straightforward to check that the posterior belief over ϕ is again normally distributed with mean s and standard deviation σ_ϵ, so that both Assumption 1 and 2 are satisfied in this case.

Accordingly, the posterior of interest, $\theta|s$ is log-normally distributed with mean $\exp s$ and variance

$$\text{Var}\{\theta|s\} = (\exp\{\sigma_\epsilon^2\} - 1) \exp\{2s + \sigma_\epsilon^2\}.$$

The example illustrates parametrically the more general implications of Theorems 1 and 2: The posterior variance over θ increases with s.[23]

In sum, this application illustrates how a simple *nonlinear* signal structure can give rise to endogenous, signal-dependent movements in uncertainty.[24]

4 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 mathematical 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 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

[23] Of course, in the simple parametric case of this example the result can be computed directly without the need to refer to Theorem 1.

[24] For alternative approaches to modeling endogenous variations in uncertainty, see, e.g., Van Nieuwerburgh and Veldkamp (2006), Nimark (2014), Fajgelbaum, Schaal and Taschereau-Dumouchel (2015), and Senga (2015).

cross-sectional dispersion of MRPKs which is often used to measure capital misallocation; the third application studied the cyclicality of derivative risks; and the final application shows how our results can be used to generate endogenously fluctuating uncertainty.

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 *co*variances of X and Y with some other random variables to the difference between the *co*variances 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 establishes 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 $Q_Z : (0,1) \to \mathbb{R}$ 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 $q_Z : (0,1) \to \mathbb{R}$, $q_Z(u) = Q_Z(u) - \mu_Z$, be a "normalized" quantile function, as well as $\overline{q}_Z : (0,1) \to \mathbb{R}$, $\overline{q}_Z(u) \equiv \int_0^u q_Z(\tilde{u}) d\tilde{u}$, 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 $\overline{q}_Z(0) = \overline{q}_Z(1) = 0$, 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, $\overline{q}_Z$ 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 $U \sim unif[0,1]$ 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 $g : \mathbb{R} \to \mathbb{R}$ be a convex, strictly increasing, continuously differentiable function. Then it holds that*

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

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 $\delta : \text{supp}(X) \cup \text{supp}(Y) \to \mathbb{R}$, $\delta(x) = q_{Z_Y}(F_Y(x))f_Y(x) - q_{Z_X}(F_X(x))f_X(x)$, which is defined on the supports of X and Y.

Lemma 1. *We have the following results about δ:*

1. *If $x \in conv(I_X \cup I_Y)$, then $\delta(x) \leq 0$.*[25]
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.[26]
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,

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

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 $\delta(x') > 0$.

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

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

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 $\Delta : \mathbb{R} \to \mathbb{R}$, $\Delta(x) = \overline{q}_{Z_Y}(F_Y(x)) - \overline{q}_{Z_X}(F_X(x))$. Notice that, by definition of Δ,

$$\Delta(x) = \int_{(-\infty,x]} \delta(\tilde{x})\mu(d\tilde{x}), \tag{11}$$

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 2. *Δ has limits $\lim_{x\to-\infty} \Delta(x) = \lim_{x\to+\infty} \Delta(x) = 0$. 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 $\overline{q}_{Z_X}$ and $\overline{q}_{Z_Y}$ are continuous and zero for arguments 0 and 1, and the cumulative distribution functions F_X and F_Y have the usual limits for $x \to \pm\infty$. 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 (11) we can write

$$\Delta(x_2) - \Delta(x_1) = \int_{(x_1,x_2]} \delta(x)\mu(dx) \tag{12}$$

and

$$\Delta(x_1) = \int_{(-\infty,x_1]} \delta(x)\mu(dx) \quad \text{and} \quad 0 - \Delta(x_2) = \int_{(x_2,\infty)} \delta(x)\mu(dx). \tag{13}$$

Using (12) 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 (13) 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 1: If $a_1 < I_X$, then $\delta(a_0)$ would have to be positive. If $a_1 \in \mathrm{conv}(I_X \cup I_Y)$, $\delta(a_1)$ 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 $\mu-$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 2. □

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,[27]

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

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

$$Cov(g(Y), Z_Y) - Cov(g(X), Z_X) = \int_{\mathbb{R}} g(x)\delta(x)\mu(dx). \tag{14}$$

Using integration by parts, we further rewrite this as

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

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

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

where the last equality follows by the existence and finiteness of the covariances in (14). Consider now $x \to -\infty$. Focus first on the case where g eventually turns negative for sufficiently small x. Then,

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

where, again, the last equality follows by the existence and finiteness of the covariances in (14). If g stays positive for all x, then the limit $\lim_{x\to-\infty} g(x)$ exists and is finite. Hence,

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

This proves that

$$Cov(g(Y), Z_Y) - Cov(g(X), Z_X) = -\int_{\mathbb{R}} g'(x)\Delta(x)dx. \tag{15}$$

[27] For 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 $Cov(Z_1, Z_2) = \mathbb{E}[Z_1(Z_2 - \mu_{Z_2})] = \mathbb{E}[Q_{Z_1}(U)q_{Z_2}(U)] = \int_0^1 q_{Z_2}(u)Q_{Z_1}(u)du$.

A completely analogous computation establishes that

$$0 \leq Cov(Y, Z_Y) - Cov(X, Z_X) = \int_{\mathbb{R}} x\delta(x)\mu(dx) = -\int \Delta(x)dx. \tag{16}$$

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

$$-\int_{\mathbb{R}} g'(x)\Delta(x)dx \geq -g'(x_0)\int \Delta(x)dx \tag{17}$$

where we used the x_0 with $\{\Delta \geq 0\} \leq x_0 \leq \{\Delta \leq 0\}$ shown to exist in Lemma 2. Using (15) and (16), this establishes the weak inequality $Cov(g(Y), Z_Y) - Cov(g(X), Z_X) \geq 0$.

For the case where $g'(x)$ is strictly increasing, notice that according to Lemma 2 $\Delta \neq 0$ on a non-zero measure set and so the inequality in (17) is strict. In that case, $Cov(g(Y), Z_Y) - Cov(g(X), Z_X) > 0$. This proves Proposition 6. □

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), $Cov(X, X) = \text{Var}\{X\} \leq \text{Var}\{Y\} = Cov(Y, Y)$, Proposition 6 implies that

$$Cov(g(X), X) \leq Cov(g(Y), Y),$$

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

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

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

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 $G_{g,X} : [0, \alpha_1] \times [0, \alpha_2] \to \mathbb{R}$ by

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

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.[28]

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 $U \sim unif[0,1]$ that is uniformly distributed over the unit interval such that $Z_i = Q_{Z_i}(U)$, for some weakly increasing measurable functions $Q_{Z_i} : [0,1] \to \mathbb{R}$.*

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

Lemma 3. *Suppose random variables Z_1 and Z_2 satisfy Assumption 5. Suppose further that $h : supp(Z_2) \to \mathbb{R}$ is a continuous function with $h(z) - z$ increasing in z. Then,*

$$Cov(Z_1, Z_2) \leq Cov(Z_1, h(Z_2)) \tag{18}$$

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

Proof. Notice that

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

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.,

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

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, (18) can only be strict. □

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 $\mathrm{Var}\{Z\} < \mathrm{Var}\{Z'\}$. Define the function $h : (\underline{g}, \infty) \to \mathbb{R}$ by $h(z) = g(g^{-1}(z) + (\beta_1' - \beta_1))$ where $\underline{g} = \inf_{z \in \mathbb{R}} g(z) \in [-\infty, \infty)$. 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

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

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 3 twice, we find

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

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 $\mathrm{Var}\{Z\} < \mathrm{Var}\{Z'\}$.

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 3 twice again, 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 (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 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 $G_{g^{-1},g(X)} : [0, \alpha_1] \times [0, \alpha_2] \to \mathbb{R}$ 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 3.3, we need to prove that

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

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.

References

Albagli, Elias, Christian Hellwig, and Aleh Tsyvinski. 2015. "A Theory of Asset Pricing Based on Heterogeneous Information." *working paper.*

Antràs, Pol. 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution." *Contributions in Macroeconomics*, 4(1).

Asker, John, Allan Collard-wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis) Allocation." *Journal of Political Economy*, 122(5): 1013–1063.

Bachmann, Rüdiger, and Christian Bayer. 2009. "Firm-Specific Productivity Risk over the Business Cycle: Facts and Aggregate Implications." *mimeo.*

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.

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. 2014. "Really Uncertain Business Cycles." *mimeo.*

Buera, Francisco J., and Benjamin Moll. 2015. "Aggregate implications of a credit crunch." *American Economic Journal: Macroeconomics*, 7(3): 1–42.

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. 2015. "Uncertainty traps." *working paper.*

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. 2014. "Uncertainty, Financial Frictions, and Investment Dynamics." *mimeo.*

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.

Ilut, Cosmin, Matthias Kehrig, and Martin Schneider. 2014. "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.*

Klump, Rainer, Peter McAdam, and Alpo Willman. 2007. "Factor Substitution and Factor-Augmenting Technical Progress in the United States: A Normalized Supply-Side System Approach." *Review of Economics and Statistics*, 89(1): 183–192.

Kozeniauskas, Nicholas, Anna Orlik, and Laura L. Veldkamp. 2014. "Black Swans and the Many Shades of Uncertainty." *mimeo.*

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.*

Salgado, Sergio, Fatih Guvenen, and Nicholas Bloom. 2015. "Skewed Business Cycles." *working paper.*

Schmidt, Klaus D. 2003. "On the covariance of monotone functions of a random variable." *mimeo.*

Senga, Tatsuro. 2015. "A New Look at Uncertainty Shocks: Imperfect Information and Misallocation." *mimeo.*

Straub, Ludwig, and Robert Ulbricht. 2012. "Credit Crunches, Information Failures, and the Persistence of Pessimism." *Unpublished, MIT and Toulouse School of Economics.*

Straub, Ludwig, and Robert Ulbricht. 2014. "Endogenous Uncertainty and Credit Crunches." *MIT and Toulouse School of Economics 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.