Bank of England

How do firms' financial conditions influence the transmission of monetary policy? A non-parametric local projection approach

Staff Working Paper No. 1,100

December 2024

Livia Silva Paranhos

Bank of England

Staff Working Paper No. 1,100

How do firms' financial conditions influence the transmission of monetary policy? A non-parametric local projection approach

Livia Silva Paranhos[(1)]

Abstract

How do monetary policy shocks affect firm investment? This paper provides new evidence on US non-financial firms and a novel non-parametric framework based on random forests. The key advantage of the methodology is that it does not impose any assumptions on how the effect of shocks varies across firms thereby allowing for general forms of heterogeneity in the transmission of shocks. My estimates suggest that there exists a threshold in the level of firm risk above which monetary policy is much less effective. Additionally, there is no evidence that the effect of policy varies with firm risk for the 75% of firms in the sample with higher risk. The proposed methodology is a generalisation of local projections and nests several common local projection specifications, including linear and nonlinear.

Key words: Local projection, impulse response estimation, nonlinearity, heterogeneity, firm investment.

JEL classification: C14, C23, E22, E52.

---

(1) Bank of England. Email: livia.silva-paranhos@bankofengland.co.uk

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email: enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

How do firms' financial conditions influence the transmission of monetary policy to investment? Since the seminal work of Bernanke et al. (1999), this question has gained importance in macroeconomic analysis, where now it is well understood that developments in credit market conditions can affect the real economy. Given the large heterogeneity in financial conditions across firms, a natural question is how this heterogeneity affects the transmission of monetary shocks. To answer this question, it is sometimes assumed in empirical and theoretical work that a change in the financial condition of firms would necessarily trigger a response in firm investment, and that this response is proportional to how financially constrained the firm is.[1] Although appealing for its simplicity, in reality the transmission mechanism across firms is likely to be more complex than this due to a variety of frictions that affect firms' decisions. One of the possible reasons is the presence of capital reallocation constraints that may impede firms to freely rebalance their portfolios (Leary and Roberts, 2005). It is then possible that some firms do not react at all to changes in market interest rates whenever this constraint is binding.

This paper provides new insight on this question by relaxing assumptions on how the effect of monetary shocks varies across firms. To achieve this, I consider a non-parametric estimation of the transmission mechanism based on a variation of the random forest model (Breiman, 2001). This approach gives more information on *what types* of firms react to monetary shocks, where here firms are defined by their financial position, and *how* they react—by how much their investment responds to monetary shocks. Both of these insights are valuable for policy analysis as well as economic modelling more generally.

I revisit the empirical application in Ottonello and Winberry (2020) on the role of financial heterogeneity in the investment channel of monetary policy, and show that the more flexible estimation proposed in this paper uncovers important nonlinearities in the transmission mechanism. In Ottonello and Winberry (2020), the authors proxy firms' financial position with either leverage or distance-to-default (a measure proposed in Gilchrist and Zakrajsek,

[1]From the theoretical side, see for example Zetlin-Jones and Shourideh (2017) and Ottonello and Winberry (2020). From the empirical side, specifications using interaction terms to capture heterogeneous effects implicitly assume a linear relation between the effect of shocks and firms' financial conditions.

2012), and consider in their main empirical exercise the universe of US non-financial firms between 1990Q1 and 2007Q4. They find that less risky firms, either firms that are one standard deviation less indebted than the average firm in the sample or firms that are one standard deviation more distant to default, tend to increase their investment by more given an unexpected decrease in interest rates.

In this context, I estimate non-parametrically the dynamic effects of monetary policy on investment for firms with distinct levels of leverage or distance-to-default. First, the results confirm the findings that firms with lower leverage and higher distance-to-default react more to the monetary shock on impact, and that the evidence of heterogeneity in the response disappears as we approach horizons larger than four quarters. For instance, firms at the 5^{th} percentile of leverage or at the 95^{th} percentile of distance-to-default present a positive significant semi-elasticity of investment of more than 4 on impact, while higher risk firms present an insignificant (or less pronounced) response.

Second, and more importantly, my estimations suggest that there exists a threshold in the level of firm risk above which monetary policy is much less effective. Specifically, the effect of the shock on firm investment at short horizons is *only* found to be positive and significant for firms below the first quartile of leverage, or firms above the third quartile of distance-to-default. Beyond such thresholds, that is for the 75% of firms in the sample with higher risk, there is no evidence that the effect of monetary policy varies with firm risk at all, and that the transmission to these firms is generally not significant. Crucially, this implies that the effect of monetary policy is different from what is predicted from more common specifications, in which the effect of the shock is assumed to be a linear function of firm risk.[2] My estimations predict that monetary policy is particularly less effective on middle-risk firms (firms between the first and third quartile of leverage or distance-to-default) than previously thought.

Although different frictions could help rationalize these findings, plausible reasons include the presence of binding capital reallocation constraints and/or fixed costs for issuing new debt. These features create a disincentive to engage in borrowing for firms with low enough

[2]Common specifications refer to the general class of regressions that include an interaction between the monetary policy shock and the firm-level variable used as proxy for firm risk.

net worth, which in turn makes the investment decisions of these firms indifferent to changes in borrowing rates triggered by monetary policy shocks. Capital reallocation frictions in particular are incorporated in e.g. Khan and Thomas (2008, 2013), and Koby and Wolf (2020) that study the dynamics of firm level investment and its aggregate implications.

A second contribution of this paper is to provide a novel methodology to estimate heterogeneous responses in a local projection framework. The methodology can be thought as a nonlinear extension of local projections (LPs; Jorda, 2005), which have become increasingly popular to estimate impulse response functions in macroeconomics. Local projections are very appealing for their simplicity, as they consist of a sequence of linear regressions of a future target variable on a current structural shock, each at a different prediction horizon. In its traditional version, LPs assume a homogeneous response to the shock across observations—in this context across firms. This paper generalizes local projections to accommodate *heterogeneous* impulse responses, i.e. responses that vary across firms, and refers to this new method as heterogeneous local projections (HLPs). In practice, this is achieved by conditioning the impulse response of firm investment on firms' default risk (e.g. leverage or distance-to-default). I show in simulations that HLPs present in general better coverage than common regression-based specifications in capturing heterogeneous effects across horizons. Additionally, HLPs can accommodate several identification schemes commonly applied in empirical work, including identification through controls and instrumental variables. Although this application only considers a single conditioning variable at a time, HLPs can also be estimated in higher dimensions, which is an important advantage of the method compared to other non-parametric techniques.

As mentioned above, the estimation of HLPs builds on a variation of the random forest (Breiman, 2001). Random forests are based on recursive partitioning, that is a model that sequentially partitions the data until "small enough" subsamples are reached. In the context of HLPs, we seek to exploit this recursive partitioning scheme until we find small enough subsamples in which the effect of the shock becomes homogeneous across firms. This means estimating a standard local projection of firm investment on a monetary shock in each subsample determined by the recursive partitioning scheme. Importantly, the data is partitioned according to the set of conditioning variables stipulated by the researcher, in this case firm risk. This process then yields local projection coefficients, or impulse responses, that depend on firm risk, which is the object of interest of the paper. In the context of random forests, we use decision trees for recursive partitioning (Breiman et al., 1984). Because individual trees tend to have high variance, we estimate instead a large number of trees and average their predictions, which loosely defines the random forest model.[3]

Related work. This paper relates to several topics of research in macroeconomics and econometrics. First, it relates to the broad literature interested in the heterogeneous effects of shocks in the economy. It connects more closely to papers studying how the effect of monetary policy varies across firms, for example according to size (Gertler and Gilchrist, 1994), age (Cloyne et al., 2018), liquidity (Jeenas, 2019), or default risk (Ottonello and Winberry, 2020). The paper extends this literature by proposing a novel estimation strategy that relaxes assumptions on how the effect of shocks varies across firms, providing new insights on the transmission mechanism. In particular, the paper extends the empirical analysis in Ottonello and Winberry (2020) and shows that there exists a threshold in the level of firm risk above which monetary policy is much less effective.

The paper also contributes to the literature of local projections for impulse response estimation in macroeconomics (Jorda, 2005; Ramey, 2016; Stock and Watson, 2018; Plagborg-Moller and Wolf, 2021). In this context, the paper relates more closely to studies that propose nonlinear specifications of local projections, of which a prominent example is state-dependent LPs. These include parametric specifications of local projections, using e.g. smooth transition functions (Auerbach and Gorodnichencko, 2013; Tenreyro and Thwaites, 2016) or threshold functions (Ramey and Zubairy, 2018), as well as semi-parametric specifications (Angrist et al., 2018 use propensity score methods). More recently, Cloyne et al. (2023) proposed a formal framework to study heterogeneous effects of an intervention across time and states of the economy in a local projection framework with interaction terms. Although the method accommodates a large number of applications in economics, the implied responses are linear on the conditioning information: an LP with an interaction between a monetary shock and the output gap assumes that the effect of monetary policy on the target is linear with respect

[3]This averaging process yields estimates with lower variance without increasing the bias. I refer to Hastie et al. (2001) for a more detailed introduction of trees and random forests.

to the state of the economy. One of the main innovations of the current paper is precisely to relax this linearity assumption of the response with respect to conditioning variables.

This paper differs from the above literature in two important ways. First, HLPs depart from the time series setting and explore both cross-sectional and time variation to estimate the impulse responses in a panel data framework, allowing for potentially different responses across individuals or firms, as opposed to macroeconomic conditions. Second, HLPs are non-parametric, so they accommodate general forms of nonlinearities in the response of shocks with respect to the conditional information. In this context, the paper more closely connects to Mumtaz and Piffer (2022) that consider a non-parametric estimation of local projections using Bayesian additive regression trees, although in a time series setting.

Finally, the paper connects to the growing literature on heterogeneous treatment effect estimation using random forests (Green and Kern, 2012; Hill and Su, 2013; Athey and Imbens, 2016; Athey and Wager, 2018; Athey et al., 2019; Friedberg et al., 2021). The paper contributes to this literature by adapting the framework to the context of impulse response estimation and by accommodating the use of panel data. The specific random forest used in HLPs' estimation is an extension of causal forests (Athey and Wager, 2018) shown to be consistent and asymptotically Gaussian. I explore the similarities between treatment effect estimation and impulse response estimation to borrow the structure of causal forests for heterogeneous impulse response estimation, provided standard assumptions on the exogeneity of the monetary shock. This strategy has the advantage of disciplining the asymptotic properties of the HLPs' estimator. From a practical perspective, this type of random forest is as simple to calibrate as traditional forests in that almost no hyperparameter tuning is required. As in causal forests, I apply the jackknife variance estimator for the construction of confidence intervals following Efron (2014) and Efron et al. (2014). An important difference between HLPs and causal forests is the data type, where the latter assumes i.i.d. samples while HLPs are estimated using dependent data (panel). I rely on the theoretical foundations from Davis and Nielsen (2020), who prove consistency of forests built on nonlinear autoregressive processes, to estimate HLPs in a panel setting.

Outline. The paper is structured as follows. Section 2 defines heterogeneous local projections and compares it to alternative methods. Section 3 describes the estimation procedure of HLPs. Section 4 provides a simulation study of HLPs, comparing the method with regression-based specifications. Section 5 extends the empirical application in Ottonello and Winberry (2020) and applies HLPs to understand the role of firms' financial conditions in the transmission of monetary policy, and Section 6 concludes.

2 Heterogeneous Local Projections

This section defines heterogeneous local projections as a non-parametric generalization of local projections (Jorda, 2005). The goal is in estimating the impulse response to a shock of interest as a flexible function of observables. I start the section by defining the object of interest and later discuss comparisons with other methods commonly used, highlighting some advantages of Heterogeneous Local Projections.

2.1 Definition

I assume the researcher has access to the data $\mathcal{Y} = (Y_{it}, W_t, \boldsymbol{X_{it}}, \boldsymbol{C_{it}})$ for $i = 1, ..., N$ denoting individuals or firms, and $t = 1, ..., T$. Consider first the panel local projection where the interest is in estimating the response of the individual-level variable $Y_{i,t+h}$ after an impulse of W_t,

$$Y_{i,t+h} = b^h \; W_t + \sum_{j=1}^{P} \gamma_j^h \; C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H. \tag{1}$$

$\boldsymbol{C_{it}} = (C_{1,it}, ..., C_{P,it})'$ is a $P \times 1$ vector that serves as a generic set of controls and may include lags of the dependent variable Y_{it}, lags of W_t, other individual-level variables denoted by the $K \times 1$ vector $\boldsymbol{X_{it}}$, and macroeconomic controls. δ_i^h is the individual fixed effect and $u_{i,t+h}$ the prediction error. Note that in this model the impulse response b^h captures the average causal effect of W_t on the target across all observations it. Notice also that if one wants to have a structural interpretation of b^h, some additional assumptions would be required, for instance having W_t to represent a "shock" variable, or through a careful selection of control variables (more on identification below).

This paper assumes that the response b^h at horizon h may vary depending on individual characteristics. More precisely, I assume that the dynamic causal effect of shocks is a function of the set $\boldsymbol{X_{i,t-1}}$ of size K. The specification of interest therefore extends (1) to

$$Y_{i,t+h} = b^h(\boldsymbol{X_{i,t-1}})\ W_t + \sum_{j=1}^{P} \gamma_j^h\ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H, \tag{2}$$

where $b^h(\boldsymbol{X_{i,t-1}})$ is a flexible function of $\boldsymbol{X_{i,t-1}}$. I denote this specification Heterogeneous Local Projections (HLPs) in reference to the concept of heterogeneous treatment effects from the microeconometrics literature (Willke et al., 2012 and Powers et al., 2018 provide surveys). Note that the conditioning implied by $\boldsymbol{X_{i,t-1}}$ allows for potentially different effects of W_t on (i) different individuals/firms, and on (ii) individuals/firms that changed over time. Additionally, note that only the coefficient of W_t is assumed to change with $\boldsymbol{X_{i,t-1}}$. This is a simplification assumption. It emphasizes the interest in capturing heterogeneous effects with respect to the variable W_t only, while being consistent with commonly used specifications that assume a linear relation between the target and control variables.[4]

More formally, the object of interest is the horizon h-response evaluated at specific values $\boldsymbol{X_{i,t-1}} = \boldsymbol{x}$ of individual characteristics, denoted as

$$b^h(\boldsymbol{x}) = \mathbb{E}\left[b^h(\boldsymbol{X_{i,t-1}}) \mid \boldsymbol{X_{i,t-1}} = \boldsymbol{x}\right]. \tag{3}$$

This can be viewed as the average causal effect of W_t on the target for individuals with characteristics similar to $\boldsymbol{x}$, as the conditional expectation is taken with respect to observations it that are "close" to $\boldsymbol{x}$. The object $b^h(\boldsymbol{x})$ is referred as the heterogeneous impulse response at $\boldsymbol{x}$.

In this paper, I propose to estimate the responses $\{b^h(\boldsymbol{x})\}_{h=0}^{H}$ non-parametrically using a modified random forest (Breiman, 2001). To bring some intuition to the functional form of $b^h(\boldsymbol{x})$, consider the simple case where (i) the conditioning set is composed only by a single

[4]The decision of conditioning the response b^h on the lagged values $t-1$ of the set $\boldsymbol{X_{it}}$, instead of contemporaneous values, is to avoid endogeneity issues with respect to W_t, as is common in empirical work. The number of lags in which to condition, however, is set to one only for simplification of exposition, and can be extended as necessary.

variable, e.g. firm leverage $X_{i,t-1}$, and (ii) the model consists of a single tree $\mathcal{T}$ with only two splits. In this case, the heterogeneous impulse response at a given value x of firm leverage, for horizon h and $\{c_1, c_2\} \in \mathbb{R}$, can be written as

$$\hat{b}^h_{\mathcal{T}}(x) = \hat{b}^h_1 \, \mathbf{1}\left[x \leq c_1\right] \mathbf{1}\left[x \leq c_2\right] + \hat{b}^h_2 \, \mathbf{1}\left[x > c_1\right] \mathbf{1}\left[x \leq c_2\right] + \hat{b}^h_3 \, \mathbf{1}\left[x > c_2\right], \tag{4}$$

where $\hat{b}^h_1, \hat{b}^h_2, \hat{b}^h_3$ are responses from the local projection (1) over subsamples defined by x and the threshold values c_1 and c_2. For example, $\hat{b}^h_1$ is the estimated response for the subsample $\{it \mid X_{i,t-1} \leq c_1 \cup X_{i,t-1} \leq c_2\}$. In this example, it is straightforward to see that the tree effectively conditions the response to be a function of firm leverage. If the conditioning set is empty, the tree prediction recovers the unconditional response in (1) for all $X_{i,t-1} = x$.

2.2 Advantages of HLPs and comparison with other methods

The idea of conditional impulse response functions is not new and relates to the broad literature on nonlinear models in macroeconomics. A common example are state-dependent models, where the effect of shocks is assumed to be different across the business cycle (see e.g. Auerbach and Gorodnichenko, 2013 and Ramey and Zubairy, 2018 for applications on the effects of fiscal policy, and Tenreyro and Thwaites, 2016 and Angrist et al., 2018 for monetary policy applications). I depart from this literature by assuming that shocks can have different effects across the cross-section of individuals/firms, and explore both cross-sectional and time variation in individual characteristics to compute impulse responses in a panel data setting. The model allows however for more restricted settings that would condition the impulse responses to depend on cross-sectional variation only or on time variation only.

HLPs are also related to the concept of generalized impulse response functions (GIRFs), commonly applied in the context of nonlinear models (Gallant et al., 1993; Koop et al., 1996; Gourieroux and Jasiak, 2005). Assuming a dataset of the type $\mathcal{Y} = (Y_{it}, W_t, \boldsymbol{X_{it}})$, the causal effect on $Y_{i,t+h}$ of a unit intervention in W_t conditional on $\boldsymbol{X_{i,t-1}}$ assuming specific values $\boldsymbol{x}$ is defined as

$$E\big[Y_{i,t+h} \mid W_t = 1, \boldsymbol{X_{i,t-1}} = \boldsymbol{x}\big] - E\big[Y_{i,t+h} \mid W_t = 0, \boldsymbol{X_{i,t-1}} = \boldsymbol{x}\big]. \tag{5}$$

In this paper, we assume that $Y_{i,t+h}$ follows the model in (2), reproduced below abstracting from controls and fixed effects without loss of generality,

$$Y_{i,t+h} = b^h(\boldsymbol{X_{i,t-1}})\, W_t + u_{i,t+h}.$$

The above model implies that $E\big[Y_{i,t+h} \mid W_t = 1, \boldsymbol{X_{i,t-1} = x}\big] - E\big[Y_{i,t+h} \mid W_t = 0, \boldsymbol{X_{i,t-1} = x}\big] = b^h(\boldsymbol{x})$, provided that $E\big[u_{i,t+h} \mid W_t, \boldsymbol{X_{i,t-1}}\big] = 0$, or equivalently W_t is exogenous conditional on $\boldsymbol{X_{i,t-1}}$ (Section 3.2 discusses exogeneity assumptions in more details). We can then interpret the object $b^h(\boldsymbol{x})$ as the generalized impulse response at $\boldsymbol{x}$ given model (2).

GIRFs and HLPs differ however in some aspects. First, GIRFs are commonly estimated in the time series domain, while HLPs are designed to accommodate panel data. Second, in HLPs only the coefficient of the shock varies (nonlinearly) over the conditioning set. GIRFs can be less restrictive in this respect and may assume a general nonlinear model. The simpler, more restrictive, structure of HLPs makes the comparison with standard approaches for the estimation of impulse responses with panel data more direct, such as regression-based specifications. This restriction however can be easily relaxed such that all coefficients in (2) would vary with respect to the conditioning set (see Section 3.1 for more details). Finally, GIRFs have been widely used to detect asymmetric effects of policy by conditioning the responses on the size or sign of shocks, a conditioning set of the form X_{t-1}. This paper emphasizes responses that vary in the cross-section as well as across time by conditioning on a specific firm characteristic $X_{i,t-1}$ in a framework that accommodates panel data. However it is also possible to study asymmetric effects by conditioning responses on X_{t-1}, similarly to GIRFs, as long as the sign or size of past shocks are not affected by the current policy shock W_t (see Assumption 2 in Section 3.2).

Compared to common methods that test for heterogeneity of responses across individuals, such as regressions with interactions or in groups, HLPs relax assumptions on how shocks vary in the cross-section. As an example, consider expanding the baseline local projection in (1) to include an interaction between the shock W_t and an individual characteristic of interest

$X_{i,t-1} \in \mathbb{R}$,

$$Y_{i,t+h} = a^h \ (X_{i,t-1} \times W_t) + b^h \ W_t + \sum_{j=1}^{P} \gamma_j^h \ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H.$$

Alternatively, one could define *ex-ante* G groups of individuals and run

$$Y_{i,t+h} = \sum_{g=1}^{G} b_g^h \ \mathbf{1}[X_{i,t-1} \in g] \ W_t + \sum_{j=1}^{P} \gamma_j^h \ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H.$$

In the first case, the researcher assumes that the effect of W_t is linear on X, and in the second that it is constant within each group g (see Section 4 for a more detailed discussion). HLPs generalize the above methods by relaxing these restrictions, and estimate instead the flexible function $b^h(X_{i,t-1})$ from (2).

Heterogeneous local projections can also be related to quantile local projections and interpreted as a special case of the latter. To be more concrete, let $Q_{it}(\tau) = Q_{it}$ be the quantile τ of a target Y_{it}. Borrowing from the general impulse response interpretation of HLPs from (5), we can similarly write the response of quantile Q_{it} of Y_{it} at horizon h to a unit impulse of a specific shock of interest W_t from the system, conditional on covariates $\boldsymbol{X_{it}}$, as

$$E\big[Q_{i,t+h} \mid W_t = 1, \boldsymbol{X_{it}}\big] - E\big[Q_{i,t+h} \mid W_t = 0, \boldsymbol{X_{it}}\big].$$

Hence, while quantile analysis centers around distributional aspects of the target Y_{it}, HLPs focus on the mean of Y_{it}.

HLPs can also accommodate several identification schemes usually encountered in macroeconomic applications. Specifically, it is compatible with (i) identification through exogenous shocks $W_t = \varepsilon_t$, where the vector of controls $\boldsymbol{C_{it}}$ may be empty, (ii) identification through controls, where the structural shock can be recovered through the inclusion of an appropriate set of controls, in which case one would regress $Y_{i,t+h}$ on the endogenous variable W_t and controls $\boldsymbol{C_{it}}$, and (ii) identification through instrumental variables (IV) where a suitable instrument Z_t for the shock is available, in which case a first-stage estimation would regress W_t on Z_t (and controls), and a second-stage would regress $Y_{i,t+h}$ on the fitted values of the first-stage (and controls). The forest model then recovers the heterogeneous responses by computing the above steps on different subsamples as defined by a tree structure, as in e.g. (4).

HLPs can also easily handle the common issues of serial correlation in the residuals that arise in local projection estimation. Specifically, one can follow Olea and Plagborg-Moller (2021) and disregard the need to correct for serial correlation in the residuals as long as the local projection includes a sufficient number of lags of the variables of interest. Regarding the panel structure of HLPs, the proposed variance estimator can be interpreted as a type of sandwich estimator that accounts for heteroskedasticity and serial correlation within individuals/firms in the same spirit as the robust variance matrix estimator proposed by Arellano (1987). This is equivalent to cluster standard errors at the individual level in a panel setting for each estimated horizon.

A paper closely related to my work is Mumtaz and Piffer (2022) that considers Bayesian additive regression trees (BART) of Chipman et al. (2010) to estimate nonlinear local projections in a time series specification.[5] As random forests, BART is a non-parametric method and in this context can accommodate general types of nonlinearities between the conditioning information and the transmission of shocks. The main difference with the current paper is that HLPs are designed to handle panel data applications and consequently can detect heterogeneous responses of shocks across individuals or firms, while Mumtaz and Piffer (2022) focus on time series applications.

From the practical side, HLPs are also relatively easy to estimate, as they are simply an adaptation of the plain random forest available from standard implementations, such as the `scikit-learn` library in `Python`, or the `randomForest` package in `R`. Importantly, this adaptation preserves the attractiveness of random forests in that almost no tuning is required.[6] At a high level, the main differences with respect to standard random forests are

[5]In a forecasting experiment in Chipman et al. (2010), BART shows superior forecast performance compared to random forests. I make two comments: (i) the focus of this paper is on statistical inference rather than forecasting, and the performance of the specific random forest used in HLPs for impulse response estimation is evaluated in simulation experiments in Section 4, (ii) the authors suggest using BART with a default set of parameters as a ready-to-use method, but the performance of the random forest in their experiment is very similar to the former (I note that the random forest is only slightly tuned).

[6]In particular, here we focus on fully grown trees, while other parameters are *ex-ante* appropriately scaled or fixed.

(i) the actual estimate at the subsample level (here we estimate a local projection instead of estimating the sample mean), (ii) a concept called honesty (Athey and Wager, 2018), in which different samples are used for either estimating the sample splits or computing the impulse responses in each subsample, but not both, and (iii) a few restrictions on how the trees are estimated, for example it is imposed that the bootstrap subsampling rate be scaled appropriately with respect to the total number of cross-sectional units (as discussed below).

3 Estimation of HLPs

In this section, I discuss why tree methods are appealing for the estimation of heterogeneous local projections, and describe the estimation strategy of HLPs based on causal forests proposed in Athey and Wager (2018).

From a pure methodological perspective, it is natural to think of HLPs as a local linear regression, that is a model that fits a linear regression over "small" subsamples of the data. This is because, for given values of individuals' characteristics $\boldsymbol{X_{i,t-1}} = \boldsymbol{x}$, the heterogeneous local projection in (2) becomes a linear model. Common non-parametric approaches to estimate these type of models are k-nearest neighbors, kernel smoothing, or series methods. This paper takes a different perspective and considers estimating (2) using a modified random forest, and interprets it as a local linear regression with adaptive weights as in Athey and Wager (2018), Athey et al. (2019) and Friedberg et al. (2021).

The forest estimation adopted in this paper departs from the original random forest version (Breiman, 2001) in many ways. In particular, I follow the adaptations proposed in Athey and Wager (2018) (AW henceforth) such that the resulting model inherits desirable statistical properties. In particular, it disciplines the variance of the estimator while controlling the bias. The current model however differs from AW regarding the type of data employed and the specification at the subsample level. In what follows, I describe the estimation in more detail (Section 3.1), and highlight the way this model can be viewed as an extension of AW (Section 3.2). For a more general introduction of random forests and regression trees, I refer to Hastie et al. (2001).

3.1 Trees and forest estimation

We seek a model that creates subsamples of the data and estimates a local projection in each of these subsamples. A tree model is particularly suited for this purpose as per its data partitioning design. Given an initial data set, the tree partitions the data sequentially into binary splits until specific stopping conditions are met. Trees are however intrinsically noisy, as slightly different data points can lead to substantially different partitions due to their hierarchical nature. Trees are then usually combined in an ensemble fashion, i.e. into forests, so that their intrinsic variability is averaged out, yielding a more stable model. In this context I estimate random forests, following the adaptations in AW, that yield estimates of the impulse responses $b^h(\boldsymbol{x})$, for $h = 0, ..., H$, with desirable statistical properties.

In this section I consider a data set of the form $(Y_{it}, W_t, \boldsymbol{X_{it}})$. In case of a non-empty vector of controls $\boldsymbol{C_{it}}$, we can retrieve this data format by orthogonalizing the variables with respect to $\boldsymbol{C_{it}}$, as discussed below. Assuming we have access to observations $i = 1, ..., N$ and $t = 1, ..., T$, consider grouping them into output-input pairs for each horizon h,

$$\mathcal{Y}_{NT} = \{(Y_{i,t+h}, W_t, \boldsymbol{X_{i,t-1}})\}_{i=1,...,N;t=1,...,T}.$$

The interest is in growing a tree that estimates heterogeneous local projections using the data $\mathcal{Y}_{NT}$, where heterogeneity is measured in terms of the set $\boldsymbol{X_{i,t-1}}$. In practice, this is achieved by sequentially finding splits of $\mathcal{Y}_{NT}$ with respect to some variable in $\boldsymbol{X_{i,t-1}}$ that minimize the squared error of the unconditional panel local projection in (1), reproduced here for convenience:

$$Y_{i,t+h} = b^h \; W_t + \sum_{j=1}^{P} \gamma_j^h \; C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H.$$

Note that in case of a non-empty vector of controls, there are in principle two options: (i) one could run (1) over different splits of the data $\mathcal{Y}_{NT}$, yielding coefficients that are functions of observables $\boldsymbol{X_{i,t-1}} = \boldsymbol{x}$, that is $b^h(\boldsymbol{x})$ and $\gamma_j^h(\boldsymbol{x})$, or (ii) one could orthogonalize both the dependent $Y_{i,t+h}$ and the shock W_t variables with respect to $\boldsymbol{C_{it}}$ in a first step, and then run (1) over splits of (an orthogonalized version of) $\mathcal{Y}_{NT}$ such that only $b^h(\boldsymbol{x})$ is estimated to depend on $\boldsymbol{X_{i,t-1}} = \boldsymbol{x}$. In both cases, the within estimator is applied to eliminate fixed effects. I proceed with the latter option to allow only the coefficient of the shock to vary over $\boldsymbol{X_{i,t-1}}$, as it facilitates the comparison with linear specifications intended to capture heterogeneous effects of shocks, as discussed previously. I therefore estimate HLPs using the transformed data $\mathcal{Y}_{NT}^{\perp} = \{(Y_{i,t+h}^{\perp}, W_t^{\perp}, \boldsymbol{X_{i,t-1}})\}_{i=1,\dots,N;t=1,\dots,T}$, with $Y_{i,t+h}^{\perp} = Y_{i,t+h} - \mathbb{E}\left[Y_{i,t+h} \mid \boldsymbol{C_{it}}\right]$ and $W_t^{\perp} = W_t - \mathbb{E}\left[W_t \mid \boldsymbol{C_{it}}\right]$, where the expectation terms are estimated assuming a linear model.[7]

Consider a variable $j \in \boldsymbol{X_{i,t-1}}$ and a splitting value c, and define the pair of subsamples,

$$L_1(j,s) = \{it \mid X_{i,t-1}^j \leq c\} \text{ and } L_2(j,c) = \{it \mid X_{i,t-1}^j > c\}.$$

Departing from the full data set, and given a random draw j over $\boldsymbol{X_{i,t-1}}$ (such that each variable is selected with probability $\pi = 1/K$), we seek the splitting value c that solves

$$\min_c \left\{ \min_{b_1^h} \sum_{\{it:\, X_{i,t-1}^j \in L_1\}} \left[\left(Y_{i,t+h} - \bar{Y}_i\right) - b_1^h \left(W_{it} - \bar{W}_i\right)\right]^2 + \min_{b_2^h} \sum_{\{it:\, X_{i,t-1}^j \in L_2\}} \left[\left(Y_{i,t+h} - \bar{Y}_i\right) - b_2^h \left(W_{it} - \bar{W}_i\right)\right]^2 \right\}. \tag{6}$$

Here $\bar{Y}_i$ and $\bar{W}_i$ denote averages over time and the demeaning serves to eliminate the individual fixed effects. I note that the indexing of W_t over i's is an abuse of notation to emphasize that the averaging is performed for each cross-sectional unit.

This process is recursive, and the tree continues growing from a given subsample until either of the following conditions holds: (i) the total number of observations it reaches a minimum value of k, or (ii) the number of cross-sectional units i reaches a minimum fraction ω of the number of cross-sectional units in the previous subsample. The first condition is standard in random forest implementations, where here we allow the trees to be fully grown while guaranteeing a minimum number of observations k in each subsample.[8] The second condition is more specific and intuitively guarantees that the splits are not too imbalanced.

[7]Note that the linear model used to center the outcome and shock variables is consistent with the linear assumption from (2).

[8]In Section C in the appendix, I show that the pointwise estimates of forests built on trees of different depths k are very similar and do not change the qualitative interpretation of results.

Following AW, ω is fixed at 0.2.

Given an estimated tree $\mathcal{T}$, we identify the subsample $L_{\mathcal{T}}(\boldsymbol{x})$ containing the individual characteristics of interest $\boldsymbol{x}$, and compute the impulse response estimate as

$$\hat{b}^h_{\mathcal{T}}(\boldsymbol{x}) = \frac{\sum_{\{it:\ \boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\}} \left(Y_{i,t+h} - \bar{Y}_i\right)\left(W_{it} - \bar{W}_i\right)}{\sum_{\{it:\ \boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\}} \left(W_{it} - \bar{W}_i\right)^2}. \tag{7}$$

In turn, the forest estimate is defined as the average impulse response estimate at x over many trees,

$$\hat{b}^h(\boldsymbol{x}) = \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \hat{b}^h_{\mathcal{T}}(\boldsymbol{x}), \tag{8}$$

where $|\mathcal{T}|$ is the number of trees in the forest. I set $|\mathcal{T}| = N$ following Efron et al. (2014) who show that this is sufficient to guarantee a negligible Monte Carlo approximation error. (Note that I use N instead of NT as the relevant number of observations for the bootstrap, as discussed below).

As usual in bagging techniques, each tree is estimated on a bootstrap sample of the original data. The typical bagging approach applied to random forests would resample from $\mathcal{Y}_{NT} = \{(Y_{i,t+h}, W_t, \boldsymbol{X_{i,t-1}})\}_{i=1,\dots,N;t=1,\dots,T}$, with each sampled row $(Y^*_{i,t+h}, W^*_t, \boldsymbol{X^*_{i,t-1}})$ being randomly chosen with replacement from $\mathcal{Y}_{NT}$. Note that this bootstrap method samples from the original data and always selects the independent and dependent variables in pairs (hence referred to as pairwise bootstrap). By resampling from the data directly, as opposed to ressampling from the residuals, this method accommodates general forms of heteroskedasticity (MacKinnon, 2006).

In this paper I rely on a few adaptations of the bagging process to account for the use of panel data as well as to inherit statistical properties for the heterogeneous impulse response following AW. First, to deal with the time dependence, I draw randomly across cross-sectional units only, and collect all the time periods corresponding to the sampled units. This technique is simple to implement and it tends to improve the approximation properties of bagging compared to e.g. block bootstraping (Kapetanios, 2008). It also permits the use of the jackknife for variance estimation, as discussed below. Note that by resampling across the cross-section, this bootstrap method preserves any form of within-units error correlations.

Second, I draw subsamples without replacement of size $s < N$, where s scales appropriately with respect to N.[9] Finally, I rely on the concept of honesty during tree estimation, found to be crucial to establish centered asymptotic normality for the heterogeneous treatment effect estimator proposed in AW. The idea is to separate tree construction—how to split the data—from tree prediction—the estimation of impulse responses at the subsample level. Each bootstrap sample of size s is first split into two equal parts, $\mathcal{I}$ and $\mathcal{J}$ (along with their respective time periods). The tree then employs sample $\mathcal{J}$ to construct the splits using (6), and sample $\mathcal{I}$ to estimate the impulse responses in (7), where the stopping conditions governed by k and ω are applied on the $\mathcal{I}$ sample. AW show that honest trees can avoid bias at the edges of the X-space, as opposed to traditional regression trees that are pointwise biased in these cases.

It is easy to see the similarity between the forest estimator for the heterogeneous impulse response $b^h(\boldsymbol{x})$ and a natural alternative that uses a local linear model with a pre-specified kernel $\mathcal{K}_{it}(\boldsymbol{x})$. Given (7) and (8), the forest estimate can be written as

$$\hat{b}^h(\boldsymbol{x}) = \frac{\sum_i \sum_t \ \alpha_{it}(\boldsymbol{x}) \left(Y_{i,t+h} - \bar{Y}_i\right) \left(W_{it} - \bar{W}_i\right)}{\sum_i \sum_t \ \alpha_{it}(\boldsymbol{x}) \left(W_{it} - \bar{W}_i\right)^2}. \tag{9}$$

The alternative kernel-based estimate takes the same form as above, except with weights $\mathcal{K}_{it}(\boldsymbol{x})$ instead of $\alpha_{it}(\boldsymbol{x})$. Here the weights are defined as $\alpha_{it}(\boldsymbol{x}) = \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \mathbf{1} \left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right] / |L_{\mathcal{T}}(\boldsymbol{x})|$ (see Appendix A for the derivation), and can be interpreted as the relative frequency in which observation it falls into the same subsample as observation $\boldsymbol{x}$ across trees. In this perspective, the methods are similar, as both weighting functions relate to a concept of distance between observations. While both methods have its merits, forest weights are adaptive in that they depend on the strength of the signal across the conditioning set $\boldsymbol{X}$. This means that if a given conditioning variable $X_{i,t-1} \in \boldsymbol{X_{i,t-1}}$ largely explains the difference in responses across individuals, the forest is able to detect it quickly and create subsamples accordingly. As a consequence, forests can in general handle larger conditioning sets compared to kernel methods that are generally estimated in two or three dimensions.[10]

[9] Following AW, I set $s = N^\gamma$, with $\gamma = 1 - \left(1 + \frac{\pi^{-1}\log(\omega^{-1})}{\log(1-\omega)^{-1}}\right)^{-1} < 1$.

[10] In kernel regressions, the convergence rate of the estimator slows as the dimension of the conditioning set increases (Hansen, 2022).

Variance estimation. As in AW, I consider the infinitesimal jackknife (IJ) variance estimator developed by Efron (2014) and Efron et al. (2014) to compute the variance of $\hat{b}^h(\boldsymbol{x})$. The IJ variance estimator, also known as the "nonparametric delta method", is an alternative to the ordinary jackknife (see Efron, 1982) in that we study the statistic of interest by changing each observation by an infinitesimal amount. It turns out to be appropriate in this setting since $b^h(\boldsymbol{x})$ is a smooth function of $\boldsymbol{x}$. One of the advantages of jackknife methods in the context of forests is that it relies on the same bootstrap samples used to compute the forest itself, therefore economizing in computational time.

Let $J_{i\mathcal{T}}$ be an indicator of whether observation i is in the bootstrap sample of tree $\mathcal{T}$, and let $\hat{b}^h_{\mathcal{T}}(\boldsymbol{x})$ be the tree $\mathcal{T}$ estimate at $\boldsymbol{x}$. Then the variance of $\hat{b}^h(\boldsymbol{x})$ can be computed as

$$\widehat{V}_{IJ}\left(\hat{b}^h(\boldsymbol{x})\right) = \frac{N(N-1)}{(N-s)^2} \sum_{i=1}^{N} \widehat{Cov}_i \left[J_{i\mathcal{T}}, \hat{b}^h_{\mathcal{T}}(\boldsymbol{x})\right]^2, \tag{10}$$

where the covariance is applied over the set of trees in the ensemble. The term in front of the summation is a correction for subsampling without replacement, where I recall s is the subsample size and N the total number of cross-sectional units.

It is worth mentioning the relation of jackknife variance estimators with other, more common variance estimators in econometrics. Efron (1982) shows that the IJ can be expressed as the popular heteroskedasticity-robust Eicker–Huber–White variance estimator. In the same spirit, MacKinnon and White (1985) show that the ordinary jackknife is asymptotically equivalent to the heteroskedasticity-robust estimator, and superior in small samples. Here, the estimator in (10) is a variation from the IJ proposed in Efron et al. (2014) in that the bootstrap samples used to construct the trees, and consequently to construct the IJ estimator, are randomly drawn across cross-sectional units only, where I collect all time steps from each selected unit. This guarantees independence across draws, as long as we assume no cross-correlation among units, and preserves the asymptotic properties of the jackknife estimator. In other words, each unit i can essentially be interpreted as a cluster consisting of several observations of that same unit over time. This extension to panel data is similar in spirit to the robust variance matrix estimator proposed by Arellano (1987), robust to any form of heteroskedasticity or serial correlation within units. The later is often interpreted as a cluster-robust variance estimator when each unit represents a cluster.

3.2 Tree construction and relation to Athey and Wager (2018)

In AW, the authors are interested in estimating a heterogeneous treatment effect, where they assume a randomly assigned binary treatment W conditional on covariates $\boldsymbol{X}$. They define the heterogeneous treatment effect at $\boldsymbol{X} = \boldsymbol{x}$ as $\tau(\boldsymbol{x}) = \mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid \boldsymbol{X} = \boldsymbol{x}\right]$, where $Y^{(1)}$ and $Y^{(0)}$ are the responses with and without treatement respectively. In this context, they construct trees that estimate treatment effects in each subsample, and the paper further establishes asymptotic guarantees for forests based on this type of trees. I hereby discuss how to adapt this framework in the context of impulse response estimation and highlight the important assumptions for the present application.

First, I rely on the analogy between treatment effect estimation as commonly defined in microeconomics and impulse response estimation as studied in macroeconomics. As argued in Stock and Watson (2018), the above concepts can be regarded as equivalent as long as we assume a certain exogeneity condition that would identify the macroeconomic shock of interest, here expressed by the variable W_t. In the present context, such equivalence implies that we can relate the impulse response recovered from (1) in each subsample of the tree to the treatment effect estimated in the same fashion in AW. For our purposes, two assumptions of model (2) are needed to grant causal meaning to $b^h(\boldsymbol{X_{i,t-1}})$, as defined below.

Assumption 1 (Conditional exogeneity.) *Consider the HLPs' specification in (2), and assume that the conditioning set $\boldsymbol{X_{i,t-1}}$ is contained in $\boldsymbol{C_{it}}$, that is $\boldsymbol{X_{i,t-1}} \subset \boldsymbol{C_{it}}$.*[11] *We say that W_t is exogeneous conditional on $\boldsymbol{C_{it}}$ if*

$$E\left[u_{i,t+h} \mid W_t, \boldsymbol{C_{it}}\right] = 0, \text{ for } h = 0, ..., H.$$

This assumption is equivalent to unconfoundedness (also known as selection on observables) in the context of treatment effect estimation, since they both imply random assignment of the treatment or shock: the treatment is randomly assigned across units i, or equivalently the macroeconomic shock W_t is randomly assigned over time, after conditioning on observables

[11] This is the case in the empirical application.

$\boldsymbol{C_{it}}$. Assumption 1 is also similar to stating that W_t is predetermined given $\boldsymbol{C_{it}}$, in the sense that W_t is allowed to be correlated with past errors, but not present or future errors, while holding $\boldsymbol{C_{it}}$ fixed.

Attaching causal meaning to $b^h(\boldsymbol{X_{i,t-1}})$ also depends on the relationship between the policy intervention W_t and the conditioning set $\boldsymbol{X_{i,t-1}}$ since the response to W_t varies with $\boldsymbol{X_{i,t-1}}$ by definition in HLPs. In recent work, Gonçalves et al. (2024) study the validity of the local projection estimator when the response is allowed to depend on the state of the economy in a time series setting. They show that, assuming a population impulse response as in (5), the LP estimator only recovers the true response function when the state of the economy is exogenous with respect to macroeconomic shocks. We will assume a slightly weaker assumption than exogeneity here which is nonetheless sufficient to interpret $b^h(\boldsymbol{X_{i,t-1}})$ causally:

Assumption 2 (Hierarchical causality.) *The intervention W_t is allowed to depend on the conditioning set $\boldsymbol{X_{i,t-1}}$, but the conditioning set $\boldsymbol{X_{i,t-1}}$ is not allowed to depend on the intervention W_t.*

This is a common assumption in applied microeconomics (Fortin et al., 2011) as it ensures that the stratification variable is not affected by the intervention. Note that this assumption does not contradict Assumption 1 as the latter implies that, unconditionally, W_t may still depend on $\boldsymbol{X_{i,t-1}}$. In the current application, $X_{i,t-1} \in \mathbb{R}$ is the financial position of firm i at time $t-1$, and given Assumption 1, W_t captures unanticipated monetary policy given $\boldsymbol{C_{it}}$. Assumption 2 is then plausible because current policy may be influenced by past firm conditions but past firm conditions are likely not influenced by current or future monetary policy disturbances (or their expectations).

The second important divergence between this application and the method developed in AW is the data structure, where the later assumes i.i.d. samples. Several works have studied the consistency properties of random forests assuming independent data (see e.g. Biau, 2012; Scornet et al., 2015), as well as dependent data (Davis and Nielsen, 2020). In particular, Davis and Nielsen (2020) prove consistency of forests built on nonlinear autoregressive processes, hence providing theoretical justification for growing trees using e.g. $\mathcal{Y}_{NT} = \{(Y_{i,t+h}, \boldsymbol{X_{i,t-1}}, W_t)\}_{i=1,...,N;t=1,...,T}$. Regarding the construction of confidence intervals, the cross-sectional variation in panel data is convenient as it allows the use of the jackknife estimator, as discussed above.

Finally, to guarantee consistency of the treatment effect $\tau(\boldsymbol{x})$, AW also need to assume Lipschitz continuity of the conditional mean functions $\mathbb{E}\left[Y^{(1)} \mid \boldsymbol{X} = \boldsymbol{x}\right]$ and $\mathbb{E}\left[Y^{(0)} \mid \boldsymbol{X} = \boldsymbol{x}\right]$. In this setting, where an OLS is performed at the subsample level, this translates into assuming continuity in $\boldsymbol{x}$ of the functions $\mathbb{E}[Y_{i,t+h} \mid \boldsymbol{X_{i,t-1}} = \boldsymbol{x}]$, $\mathbb{E}[W_t \mid \boldsymbol{X_{i,t-1}} = \boldsymbol{x}]$, $\text{Cov}[W_t, Y_{i,t+h} \mid \boldsymbol{X_{i,t-1}} = \boldsymbol{x}]$, and $\text{Var}[W_t \mid \boldsymbol{X_{i,t-1}} = \boldsymbol{x}]$. This continuity assumption is intuitive, as it means assuming smoothness of impulse responses along individuals' characteristics, and is also standard for consistency results in the literature (Meinshausen, 2006; Biau, 2012; Scornet et al., 2015; Wager and Walther, 2015).

4 Simulations

The object of interest in this paper is the heterogeneous impulse response at x, $b^h(x)$, which intends to capture different responses to shocks across individuals or firms at horizon h, if they exist. In this section, I carry out a simulation exercise to verify if (i) $b^h(x)$ is able to capture responses that vary both linearly and nonlinearly with respect to individual characteristics, and (ii) the confidence intervals of $b^h(x)$ based on (10) are asymptotically valid, which makes inference possible.

The data generating process is

$$
\begin{aligned}
Y_{it} &= \rho_y \; Y_{i,t-1} + \; b^0(X_{it}) \times W_t + \sigma_i (1+\theta^2)^{-1/2} u_{it}, \quad \text{with} \\
&\quad u_{it} = q_{it} + \theta q_{i,t-1}, \\
&\quad W_t \sim iid\mathcal{N}(0,1), \quad \sigma_i^2 \sim iid(1+\mathcal{X}_1^2)/2, \quad q_{it} \sim iid\mathcal{N}(0,24),
\end{aligned} \tag{11}
$$

for cross-sectional units $i = 1, ..., N$ and $t = 1, ..., T + 500$, where the first 500 time periods are discarded. ρ_y is fixed to 0.8 and θ is fixed to 0.4.

I consider three data generating processes for how the response to W_t varies across the cross-section of individuals (and across time) according to some characteristic $X_{it} \in \mathbb{R}$. I denote this response by $b^0(X_{it})$ in (11) to emphasize that it is a function of the variable X_{it} and that the implied estimation horizon is 0. Note that here there is no need to condition $b^0(\cdot)$ on lagged values of X, as usually implemented in empirical analysis to avoid endogeneity concerns, since W_t is an independent process by construction. I consider the following three cases:

i. Linear,

$$b^0(X_{it}) = X_{it} + \epsilon_{it},$$

ii. Piecewise linear,

$$b^0(X_{it}) = \begin{cases} 0 + \epsilon_{it} & \text{if } X_{it} \leq 0 \\ X_{it} + \epsilon_{it} & \text{if } X_{it} > 0, \end{cases}$$

iii. Quadratic,

$$b^0(X_{it}) = X_{it}^2 + \epsilon_{it},$$

with $\epsilon_{it} \sim iid\mathcal{N}(0, 8)$.

Finally, I model X_{it} as the sum of two components, the first accounting for variation across individuals i and the second describing the dynamics:

$$\begin{aligned} X_{it} = \mu_i + \xi_{it}, \quad & \text{with} \\ & \mu_i \sim iid\mathcal{N}(0,3) \\ & \xi_{it} = \rho_\xi \; \xi_{i,t-1} + \left(1 - \rho_\xi^2\right)^{1/2} \upsilon_{it}, \quad \upsilon_{it} \sim iid\mathcal{N}(0,1), \end{aligned} \tag{12}$$

where I fix $\rho_\xi = 0.4$. The above modelling choices are set to match moments of the data used in the empirical application. Specifically, the time variation in X_{it} is set to be equal to the variation in W_t, i.e. $Var(\xi_{it}) = Var(W_t)$, while the cross-sectional variation in X_{it} is set to be three times its time variation, i.e. $Var(\mu_i) = 3\ Var(\xi_{it})$.[12]

Consider iterating forward equation (11) until horizon h, assuming for illustration the

[12] Note that $Var(\xi_{it}) = 1$.

linear case, $b^0(X_{it}) = X_{it}$, where here we abstract from the error ϵ_{it} to simplify notation,

$$Y_{i,t+h} = \rho_y^{h+1} Y_{i,t-1} + \sum_{j=0}^{h} \rho_y^{h-j} \ X_{i,t+j} \times W_{t+j} + \sum_{j=0}^{h} \rho_y^{h-j} \ \sigma_i (1+\theta)^{-1/2} u_{i,t+j}. \tag{13}$$

Note that when $h = 0$ we recover (11). The purpose of the simulations is to assess the ability of HLPs to capture the true response of $Y_{i,t+h}$ to an impulse in W_t, denoted $b^h(X_{it})$, for horizons $h = 0, ..., H$. In the linear case above, the true response is given by $b^h(X_{it}) = \rho_y^h \ X_{it}$. Similarly, for a piecewise linear process of $b^0(X_{it})$, the true response is $b^h(X_{it}) = \rho_y^h \ X_{it}$ if $X_{it} > 0$, and $b^h(X_{it}) = 0$ if $X_{it} \leq 0$, and for a quadratic process, the true response is $b^h(X_{it}) = \rho_y^h \ X_{it}^2$.

A common way of testing for heterogeneous impulse responses with respect to individual characteristics $\boldsymbol{X_{it}}$ is running local projections with interaction terms between the shock of interest W_t and $\boldsymbol{X_{it}}$ (and controls). For the case where $X_{it} \in \mathbb{R}$, the local projections are

$$Y_{i,t+h} = a^h \ (X_{it} \times W_t) + b^h \ W_t + \sum_{j=1}^{P} \gamma_j^h \ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H. \tag{14}$$

Note that the total effect of W_t on the target $Y_{i,t+h}$ is linear on X since $\delta Y_{i,t+h}/\delta W_t = a^h \ X_{it} + b^h$. One can then infer the response at a specific $X_{it} = x$ simply by computing $\hat{b}^{h,\mathrm{LP}}(x) = \hat{a}^h \ x + \hat{b}^h$. The object $b^{h,\mathrm{LP}}(x)$, which corresponds to the heterogeneous impulse response implied by model (14), is a natural benchmark to the forest-based impulse response $b^h(x)$ of HLPs, which does not assume linearity with respect to X.

Alternatively, one can incorporate power terms to (14) to also accommodate non-linear responses to W_t, as in

$$\begin{aligned} Y_{i,t+h} = a_1^h \ (X_{it} \times W_t) + a_2^h \ (X_{it}^2 \times W_t) + a_3^h \ (X_{it}^3 \times W_t) + \\ b^h \ W_t + \sum_{j=1}^{P} \gamma_j^h \ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H. \end{aligned} \tag{15}$$

In this case, the response at a specific $X_{it} = x$ is $\hat{b}^{h,LP-AUG}(x) = \hat{a}_1^h \ x + \hat{a}_2^h \ x^2 + \hat{a}_3^h \ x^3 + \hat{b}^h$. Note that the regression with power terms encompasses all elements from the true DGP when

$h = 0$ in both the linear and quadratic cases.

In the simulations, I therefore consider the following estimating equations: (i) HLPs from (2), (ii) local projections from (14), and (iii) local projections augmented with quadratic and cubic interactions terms from (15). All specifications control for one lag of the dependent variable and one lag of the individual characteristic of interest, hence $\boldsymbol{C_{it}} \equiv (Y_{i,t-1}, X_{i,t-1})'$. The standard errors of the local projection specifications are estimated using the robust variance matrix estimator from Arellano (1987) which is equivalent to cluster at the unit level. Standard errors of HLPs can be interpreted analogously, as discussed in Section 3.1. Additionally, both the dependent and the shock variables are orthogonalized with respect to the controls before the forest estimation (see Section 3.1 for more details). I consider k, the minimum number of observations at the subsample level, to be 5% of the total number of observations used for tree estimation.

For each data generating process considered for $b^0(X_{it})$, I report the root mean squared error, empirical coverage rates and the median length of confidence intervals. I also vary the number of time periods, $T = \{15, 30, 45\}$, as well as the number of cross-sectional units, $N = \{20T, 50T\}$ (this grid of choices is applied to horizon 0, while I fix $T = 30$ for higher horizons). I consider $M = 300$ Monte Carlo repetitions, where for each repetition I evaluate the responses at 500 values of individual characteristics X_{it}. The reported simulation statistics are computed over a total of $300 \times 500 = 150,000$ instances, and the nominal level of confidence intervals is 0.90.

Table 1 reports the root mean squared error (RMSE), average coverage rates and median lengths of confidence intervals for the heterogeneous impulse response $b^0(x)$ according to the different cases considered, for horizon 0. Table 2 shows the same results for $b^h(x)$, with $h = 4, 8$ and 12. I discuss a few number of findings.

First, HLPs recover well the true shape of the response at $h = 0$ and present relatively better coverage independently of the DGP considered and sample size. At $h = 0$, although the regression is the preferred specification for a linear DGP and the regression with power terms for a quadratic DGP, HLPs still perform well in those cases, albeit with larger confidence intervals. Note that for the piecewise linear DGP, both regressions show similar errors (slightly smaller) than HLPs but significantly lower coverage for all horizons. This flexibility

Table 1: Root mean squared error for the heterogeneous impulse response at horizon 0, along with average coverage and median length of 90% confidence intervals, from a regression, a regression augmented with power terms and HLPs on out-of-sample predictions.

			Regression			Regression w/ powers			HLPs		
DGP of $b^0(X_{it})$	T	N	RMSE	Cov	Length	RMSE	Cov	Length	RMSE	Cov	Length
Horizon 0											
Linear	15	20T	0.16	0.89	0.43	0.23	0.86	0.51	0.47	0.98	3.02
	15	50T	0.10	0.88	0.28	0.15	0.87	0.33	0.38	0.98	1.96
Pcwise linear	15	20T	0.62	0.28	0.45	0.32	0.67	0.50	0.42	0.99	2.91
	15	50T	0.61	0.18	0.29	0.26	0.52	0.32	0.30	0.99	1.87
Quadratic	15	20T	5.67	0.13	1.53	0.23	0.88	0.51	2.76	0.97	3.75
	15	50T	5.70	0.08	0.98	0.16	0.87	0.32	2.70	0.95	2.43
Linear	30	20T	0.07	0.90	0.21	0.10	0.90	0.25	0.33	0.98	1.49
	30	50T	0.05	0.91	0.13	0.07	0.89	0.16	0.30	0.97	0.98
Pcwise linear	30	20T	0.61	0.14	0.23	0.24	0.45	0.25	0.27	0.99	1.43
	30	50T	0.60	0.09	0.15	0.22	0.26	0.16	0.22	0.98	0.90
Quadratic	30	20T	5.69	0.09	1.05	0.11	0.89	0.25	2.50	0.96	1.90
	30	50T	5.71	0.06	0.68	0.07	0.88	0.16	2.52	0.93	1.39
Linear	45	20T	0.05	0.91	0.14	0.07	0.90	0.16	0.29	0.97	1.02
	45	50T	0.03	0.89	0.09	0.05	0.88	0.11	0.28	0.95	0.69
Pcwise linear	45	20T	0.61	0.10	0.16	0.23	0.28	0.17	0.23	0.99	0.94
	45	50T	0.60	0.06	0.10	0.22	0.17	0.11	0.20	0.98	0.60
Quadratic	45	20T	5.68	0.07	1.01	0.08	0.89	0.17	2.49	0.96	1.43
	45	50T	5.64	0.04	0.55	0.05	0.87	0.11	2.43	0.91	1.14

Simulations of the regressions in (14) and (15), and HLPs in (2) at $h = 0$. I run 300 Monte Carlo repetitions and evaluate the responses at 500 different values of $X_{it} = x$ for each repetition. Reported statistics are computed over $300 \times 500 = 150,000$ instances. I consider k, the minimum number of observations at the subsample level, to be 5% of the total number of observations used for tree estimation.

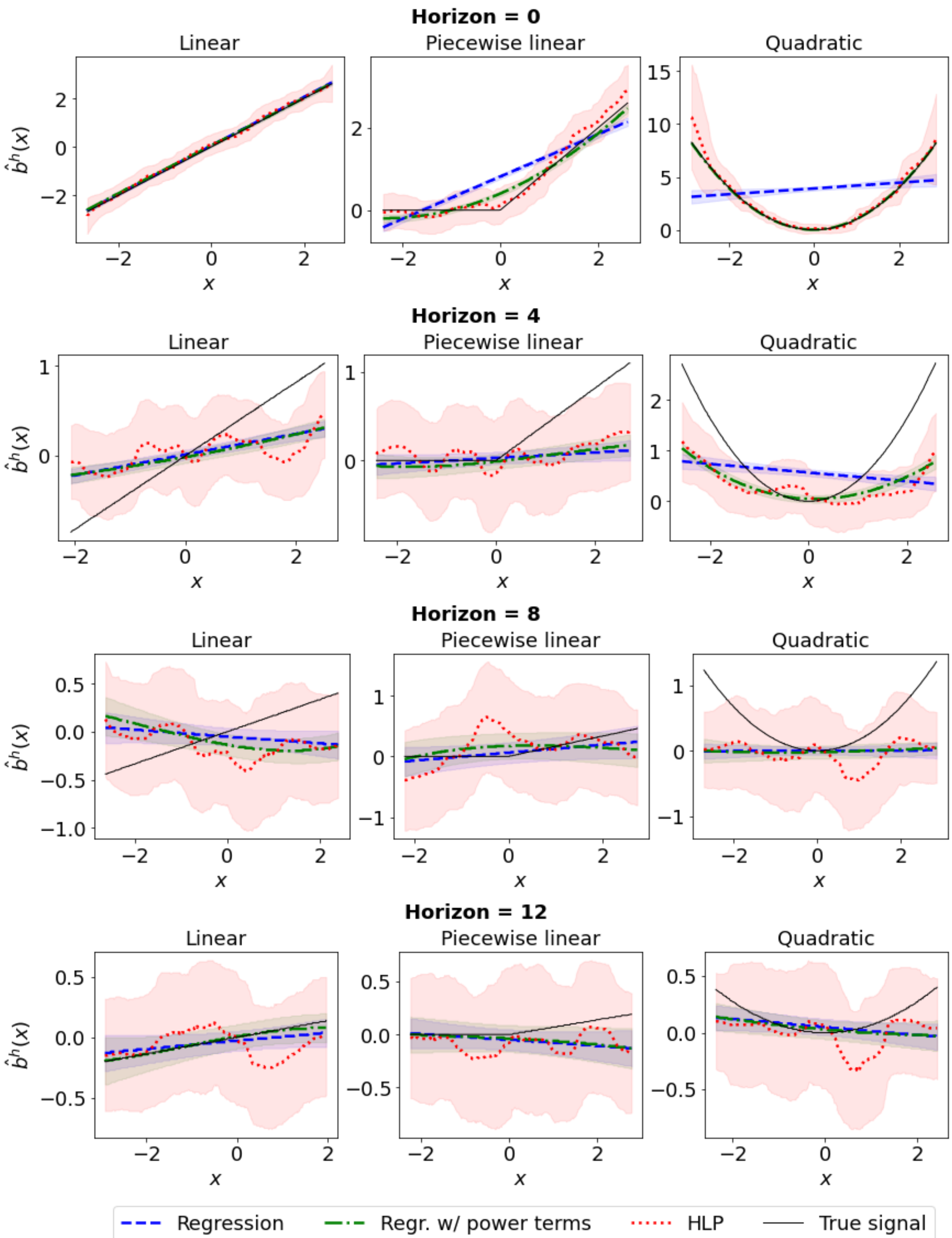
Table 2: Root mean squared error for the heterogeneous impulse response at horizons 4, 8 and 12, along with average coverage and median length of 90% confidence intervals, from a regression, a regression augmented with power terms and HLPs on out-of-sample predictions.

			Regression			Regression w/ powers			HLPs		
DGP of $b^0(X_{it})$	T	N	RMSE	Cov	Length	RMSE	Cov	Length	RMSE	Cov	Length
Horizon 4											
Linear	30	20T	0.65	0.21	0.32	0.66	0.25	0.38	0.70	0.83	1.97
	30	50T	0.65	0.14	0.20	0.65	0.17	0.24	0.68	0.67	1.22
Piecewise linear	30	20T	0.49	0.52	0.31	0.49	0.55	0.37	0.53	0.91	1.94
	30	50T	0.49	0.44	0.20	0.47	0.51	0.24	0.49	0.84	1.22
Quadratic	30	20T	2.66	0.13	0.35	2.17	0.23	0.40	2.34	0.67	2.08
	30	50T	2.66	0.08	0.23	2.18	0.15	0.26	2.34	0.55	1.31
Horizon 8											
Linear	30	20T	0.37	0.37	0.33	0.39	0.44	0.39	0.43	0.94	1.85
	30	50T	0.36	0.21	0.21	0.37	0.27	0.24	0.39	0.84	1.15
Piecewise linear	30	20T	0.27	0.64	0.33	0.30	0.67	0.39	0.34	0.93	1.86
	30	50T	0.27	0.54	0.20	0.28	0.58	0.24	0.30	0.92	1.14
Quadratic	30	20T	1.21	0.21	0.34	1.27	0.33	0.40	1.26	0.74	1.88
	30	50T	1.20	0.15	0.22	1.24	0.27	0.26	1.24	0.65	1.18
Horizon 12											
Linear	30	20T	0.21	0.69	0.35	0.25	0.76	0.42	0.28	0.99	1.76
	30	50T	0.18	0.55	0.23	0.20	0.64	0.27	0.22	0.98	1.13
Piecewise linear	30	20T	0.18	0.79	0.35	0.23	0.82	0.42	0.26	0.99	1.77
	30	50T	0.14	0.70	0.23	0.17	0.75	0.28	0.19	0.99	1.14
Quadratic	30	20T	0.52	0.44	0.36	0.58	0.56	0.43	0.59	0.88	1.80
	30	50T	0.52	0.29	0.23	0.57	0.43	0.27	0.56	0.79	1.12

Simulations of the regressions in (14) and (15), and HLPs in (2) at $h = 4, 8, 12$. I run 300 Monte Carlo repetitions and evaluate the responses at 500 different values of $X_{it} = x$ for each repetition. Reported statistics are computed over $300 \times 500 = 150,000$ instances. I consider k, the minimum number of observations at the subsample level, to be 5% of the total number of observations used for tree estimation.

Figure 1: Simulated heterogeneous impulse responses and relation to X

The figure plots heterogeneous impulse responses over one run of simulations for regressions (14) and (15) as well as HLPs in (2) for different values x of X_{it}. Shown are x values in between percentiles $10^{th} - 90^{th}$. Shaded areas are 90% confidence bands. Results for $T = 30$ and $N = 50T$.



in capturing general responses is the main advantage of HLPs over regression specifications. Figure 1 illustrates these findings. It plots heterogeneous impulse responses for a single repetition over 500 values of individual characteristics X_{it} for both regressions with and without power terms, as well as HLPs, assuming $T = 30$ and $N = 50T$. At $h = 0$, note that HLPs recover the shape of the true process with good coverage for all DGPs. The regression with power terms also performs well, especially in the linear and quadratic cases as expected, although it performs less well than HLPs for the piecewise linear DGP.

Second, for horizons greater than 0 (4, 8 and 12), all models show comparable accuracy, yet they encounter difficulties in accurately capturing the true signal, particularly as h increases. This is by design expected since additional interaction terms are present in the DGP in (13) for $h > 0$ which are not taken into account by the estimating equations in (2), (14) and (15). In this way, the estimating equations are consistent with empirical practice, where generally only one interaction term is included in the local projection. Relative to the regression benchmarks, HLPs present better coverage while keeping similar accuracy, but present larger confidence intervals.

Third, the relative performance of HLPs does not seem to be significantly affected by changes in sample sizes. As the sample size decreases (in terms of either T or N), uncertainty around HLPs estimates increases, although the concurrent increases in errors and coverage rates are relatively smaller. The same observation follows for the regressions although these have smaller confidence intervals.

Fourth, HLPs appear to slightly overcoverage, but this behaviour tends to disappear asymptotically as the variance decreases. I note that the uncertainty of HLPs' estimates can be somewhat sensitive to the choice of tree depth, governed by the parameter k, the minimum number of observations it at the subsample level. In simulations, I fix k to be 5% of the total number of observations used for tree estimation. As k increases, we exchange precision for a decrease in variance, which in turn tends to decrease the empirical coverage rates.

5 Firms' financial conditions and the transmission of monetary policy

In this section, I revisit the empirical application in Ottonello and Winberry (2020) on the role of financial heterogeneity in the investment response of firms to monetary policy shocks. Specifically, I focus on estimating the dynamic effects of monetary policy shocks on firm investment for firms facing different financial conditions. This can be implemented by estimating heterogeneous local projections as in (2), and then evaluating the impulse responses in (3) at different levels of firms' financial conditions $X_{i,t-1}$. This methodology generalizes equation (4) in Ottonello and Winberry (2020) (page 2480) in the context of HLPs.

As in Ottonello and Winberry (2020), $Y_{i,t+h}$ is set as firm investment h horizons after the shock, measured as the cumulative growth rate of capital stock, or $\Delta \text{log} k_{i,t+h} = \text{log} k_{i,t+h} - \text{log} k_{i,t-1}$ for $h = 0, ..., H$, where k_{it} is capital stock at the end of period t. W_t is a monetary policy shock based on the high-frequency series in Gurkaynak et al. (2005) and Gorodnichenko and Weber (2016), in which shocks are identified by movements in the current-month fed funds futures around monetary policy announcements. The shock is normalized such that positive values represent interest rate decreases (and transformed to decimal points). $X_{i,t-1}$ is a proxy for firms' default risk, and can be either the leverage ratio (total debt to total assets), or a measure of distance-to-default, which estimates the probability of default by comparing the firm's value to its debt (Gilchrist and Zakrajsek, 2012). As in the original paper, $X_{i,t-1}$ is demeaned with respect to the average value of firm i over time, and then standardized over the entire sample. Finally, the vector of controls $\boldsymbol{C_{it}}$ may include, depending on the specification, (i) firm controls, comprising lagged sales growth, total assets, current assets to total assets ratio and default risk, (ii) an interaction of default risk and last quarter GDP growth, (iii) time-sector, quarter-sector and fiscal-quarter dummies, and (iv) macro controls, comprising four lags of GDP growth, inflation and unemployment. Firm-level data are from quarterly Compustat, and covers the universe of U.S. nonfinancial firms. The number of firms in the sample is $N \approx 5000$ and the average number of time periods is $T \approx 25$.

5.1 Baseline estimates suggest heterogeneous responses

Table 3 reports the impact responses of investment to monetary policy, and replicates table III in Ottonello and Winberry (2020) (columns 2-5). Columns 5-8 extend the baseline results to include more lags of firm-level Avariables, which is discussed in Section C in the appendix. The specification includes an interaction of the shock W_t with a (lagged) proxy for default risk $X_{i,t-1}$, either leverage or distance-to-default, to capture heterogeneous effects, and is as follows

$$\Delta \text{log} k_{it} = a\ (X_{i,t-1} \times W_t) + b\ W_t + \sum_{j=1}^{P} \gamma_j\ C_{j,it} + \delta_i + u_{it}. \tag{16}$$

The estimates suggest that more risky firms, as measured by indebtedness and probability of default, tend to respond *less* to monetary policy. Specifically, column 1 reports that following a 100 bps decrease in interest rates, the investment rate is estimated to be 70 bps smaller for firms that are one standard deviation more indebted than the average firm in the sample. Similarly from column 2, we see that firms that are one standard deviation more distant to default than the average firm are estimated to present an investment rate 120 bps bigger following the same 100 bps base rate decrease. Note that the inclusion of both leverage and distance-to-default in the equation renders leverage insignificant (columns 3, 4, 7 and 8). The coefficient of the shock alone in column 4 indicates that the average investment rate is around 2.5 percentage points higher after an expansionary 100 bps monetary policy shock. This is economically significant given the average investment rate in the sample of 0.4%.

Dynamics. Now we extend (16) to horizons greater than 1, and estimate local projections of the form

$$\text{log} k_{i,t+h} - \text{log} k_{i,t-1} = a^h\ (X_{i,t-1} \times W_t) + b^h\ W_t + \sum_{j=1}^{P} \gamma_j^h\ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H. \tag{17}$$

The above specification can be seen as a dynamic extension to the specification in column 4 of Table 3 except that either leverage or distance-to-default enter the equation, but not both (that is, $X_{i,t-1}$ represents either one variable or the other and $\boldsymbol{C_{it}}$ includes either one or the other). The left-hand side variable is also changed to reflect investment rates at longer

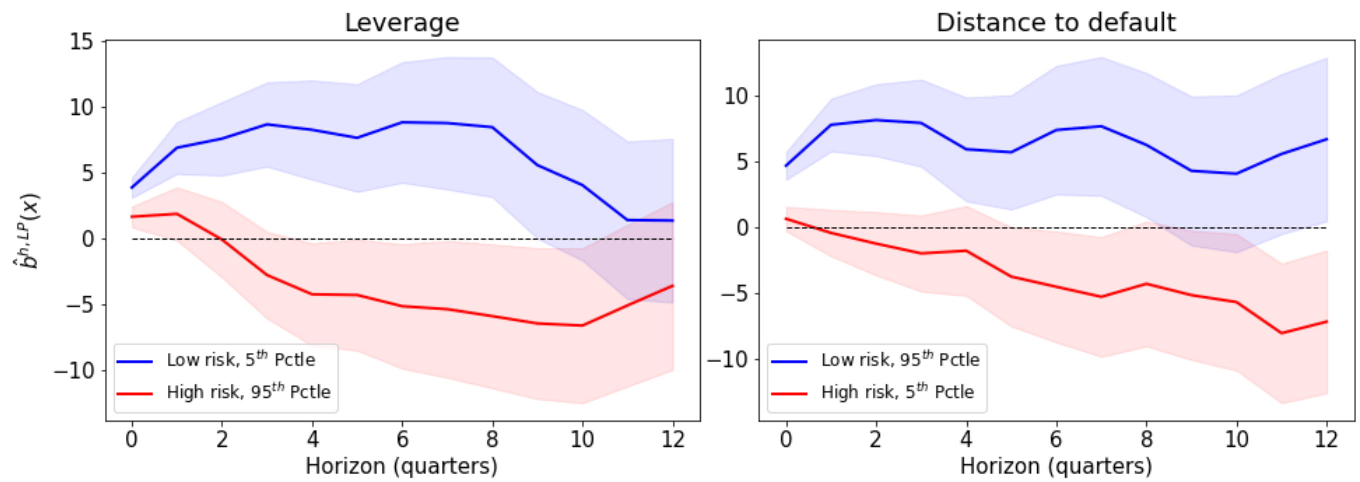
Table 3: Impact response of investment to monetary policy

	Replication from OW				Lag-augmented controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
leverage × shock	-0.73^{**}		-0.19	-0.23	-0.48		-0.25	-0.17
	(0.29)		(0.39)	(0.61)	(0.35)		(0.56)	(0.58)
dist-to-def × shock		1.20^{***}	1.10^{***}	1.25^{**}		0.84^{**}	0.72^{**}	1.04^{***}
		(0.40)	(0.39)	(0.50)		(0.34)	(0.32)	(0.34)
shock				2.49^{***}				1.40^{***}
				(0.62)				(0.40)
Observations	208,695	143,185	143,185	113,817	161,419	107,178	107,178	107,179
R^2	0.127	0.144	0.145	0.153	0.191	0.196	0.197	0.189
Firm controls	yes	yes	yes	yes	yes	yes	yes	yes
Lag-aug controls	no	no	no	no	yes	yes	yes	yes
Time sector FE	yes	yes	yes	no	yes	yes	yes	no
Macro controls	no	no	no	yes	no	no	no	yes

This table reports estimates of $\Delta \log k_{it} = a\ (X_{i,t-1} \times W_t) + b\ W_t + \sum_{j=1}^{P} \gamma_j\ C_{j,it} + \delta_i + u_{it}$, where k_{it} is capital stock, $X_{i,t-1} = \{\text{leverage, distance-to-default}\}$, W_t is the monetary policy shock normalized such that positive values represent rate decreases (in decimal points). The vector $\boldsymbol{C_{it}}$ includes firm controls (one lag of sales growth, total assets, current assets to total assets ratio and leverage and/or distance-to-default), time-sector dummies if indicated in "Time sector FE", otherwise quarter-sector dummies, fiscal-quarter dummies, an interaction of $X_{i,t-1}$ and previous-quarter GDP, macro controls (four lags of GDP growth, inflation and unemployment), and lag-augmented controls (four lags of firm controls and four lags of the dependent variable). These results replicate table III in Ottonello and Winberry (2020) (OW), in columns (1) to (4), and extend the control set to include more lags of firm-level variables, in columns (5) to (8). Standard errors in parenthesis are two-way clustered by firms and time.

Figure 2: Standard local projections and heterogeneity

The figure presents the heterogeneous impulse responses $\hat{b}^{h,LP}(x)$ of firm investment following a 100bps decrease in the fed funds rate for low and high risk firms according to the standard LPs in (17). Each panel represents a distinct specification in which either leverage or distance-to-default is considered, but not both. Shaded areas are 68% error bands, where standard errors are two-way clustered by firm and time.



horizons.

Figure 2 shows heterogeneous impulse responses according to (17) evaluated at the 5^{th} and 95^{th} percentiles of leverage (left) and distance-to-default (right). The plots confirm the previous findings that low risk firms react more to the shock on impact, and highlight that this heterogeneity is persistent until at least eight quarters after the shock. In fact, over time low risk firms significantly increase their investment after an expansionary shock, while high risk firms significantly decrease investment, especially at larger horizons.

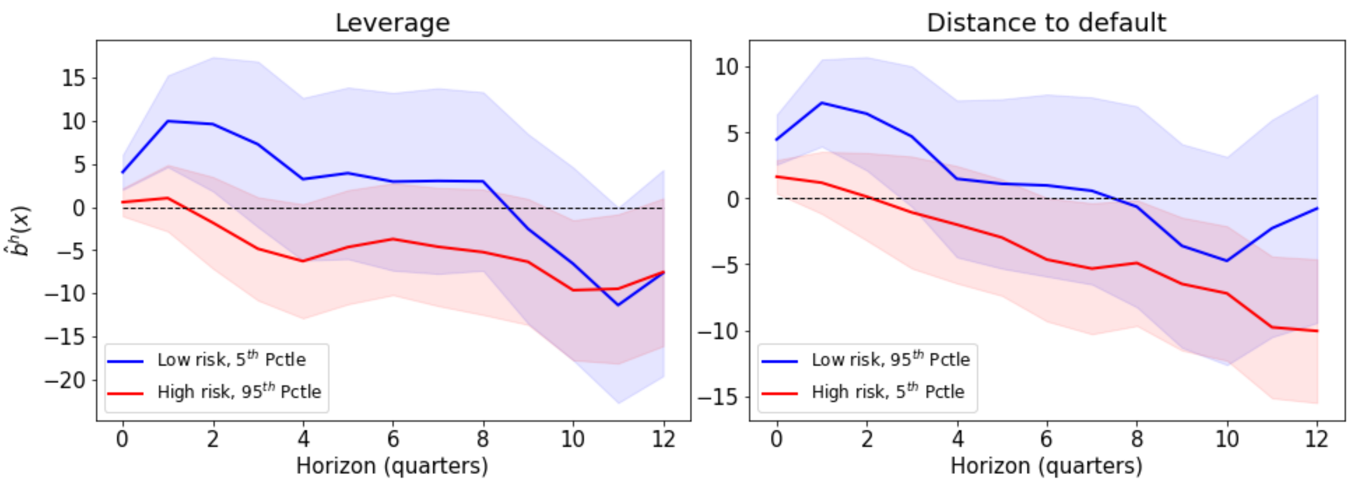
5.2 Heterogeneity from the perspective of HLPs

Heterogeneous local projections estimate the equivalent of equation (2) tailored to the current application,

$$\log k_{i,t+h} - \log k_{i,t-1} = b^h(X_{i,t-1})\ W_t + \sum_{j=1}^{P} \gamma_j^h\ C_{j,it} + \delta_i^h + u_{i,t+h}, \quad h = 0, ..., H. \quad (18)$$

Figure 3: Heterogeneous local projections

The figure presents the heterogeneous impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate for low and high risk firms according to (2). Each panel represents a distinct specification in which either leverage or distance-to-default is considered, but not both. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are one-way clustered by firm.



The variables included in $\boldsymbol{C_{it}}$ are the same as in (17), and here again $X_{i,t-1}$ represents either leverage or distance-to-default, not both.

Figure 3 shows heterogeneous impulse responses from the perspective of HLPs, in a similar vein to Figure 2 for the standard local projection. I note that unlike for the standard case where standard errors are two-way clustered by firm and time, standard errors from HLPs can be interpreted as clustered by firm only, as discussed in Section 3.1. The estimates are consistent with Figure 2 in that low risk firms react more to the shock on impact, but differ on the persistence of the heterogeneity where here differences in responses across firms fade away after about four quarters after the shock. Note in particular that responses according to the standard LP are more dissimilar between low and high risk firms than they are according to HLPs, especially at medium and long horizons. This point is further discussed below. Figure 3 also shows that high risk firms respond insignificantly (leverage) or have responses very close to insignificance (distance-to-default), while low risk firms present a positive significant response to policy on impact, with a semi-elasticity of around 4 for leverage and 4.5 for distance-to-default. As also evident from the standard local projection, note that high risk firms present a negative significant response to policy at long horizons.

The comparison between the standard LP and HLPs can be made more evident with a closer inspection of the impulse responses over a finer grid of percentiles than those available in Figures 2 and 3.[13] These are available in Figures 4 and 5 for HLPs for short $(0, 1, 2, 3)$ and long $(9, 10, 11, 12)$ horizons respectively. Note that for both risk measures the effect of the shock is generally constant (and close to insignificance for $h > 0$) for most of the distribution, but increases for firms at approximately the 25^{th} percentile of leverage or below, or firms at the 75^{th} percentile of distance-to-default or above. This effect is mostly visible up to the second quarter after the shock, and estimates become insignificant for all percentiles at horizon 3 and beyond. One exception is the impact response with respect to leverage, which is approximately linear, although we do observe a more pronounced response for firms below the 25^{th} percentile as well.

Importantly, these differences in responses across quartiles of the data cannot be detected by the standard LP. Figures B.2 and B.3 in the appendix show the results for the standard case according to (17). As expected, note that heterogeneity varies linearly across firms.[14] This linearity restriction implies that whenever heterogeneity is detected on average in the sample, low and high risk firms will tend to exhibit quite distinct responses. This is usually what is implied by a standard regression with an interaction term as in (17), hence implicit from the empirical analysis in Ottonello and Winberry (2020). The main contribution of HLPs in this context is precisely to relax this restriction and allow the response to vary "freely" across the distribution of firms. Specifically in this application, it allows the researcher to detect the presence of the aforementioned threshold in the level of firm risk (approximately the first quartile of leverage and the third quartile of distance-to-default) beyond which monetary policy appears to be less effective.

At longer horizons, shown in Figure 5, a different nonlinearity appears; responses are more muted for medium-level firms than for firms that lie either in low or high percentiles of the financial position. Note the inverse U-shape of the responses with respect to both measures. This nonlinearity, again, cannot be detected by the standard LP (see Figure B.3

[13] Figure B.1 in the appendix presents HLPs point estimates at four different percentiles of leverage and distance-to-default.

[14] I note that in Figures B.2 and B.3 the graphs do not appear strictly linear because the conditioning variables, leverage and distance-to-default, are not uniformly distributed and can be more closely represented as Gaussian variables.

Figure 4: HLPs and the cross-section variation, Short horizons

The figure presents the impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate for firms at a fine grid of percentiles of leverage and distance-to-default, at quarters $h = 0, 1, 2$ and 3 after the shock. These are two distinct specifications in which either leverage or distance-to-default is considered. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are one-way clustered by firm.

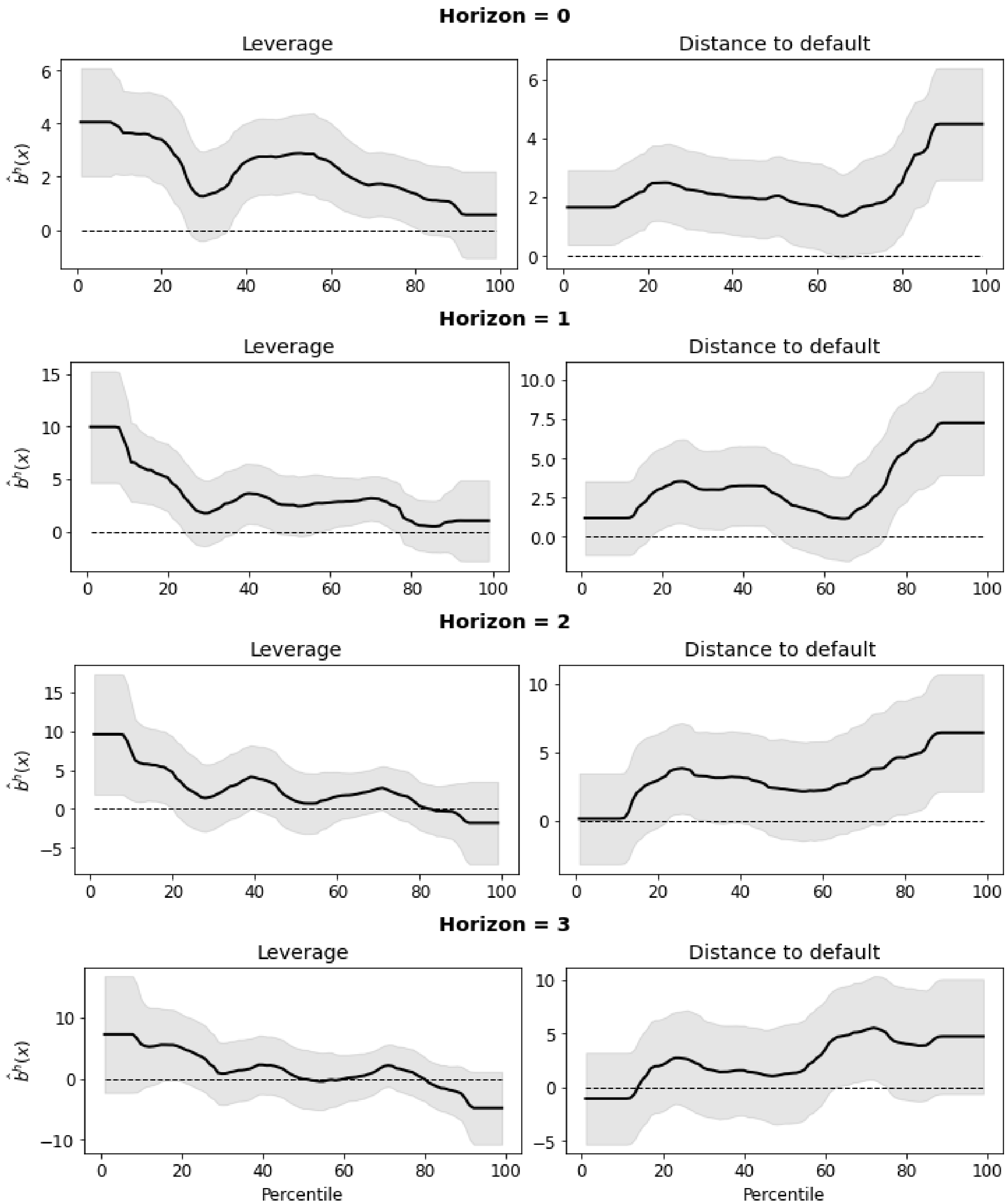
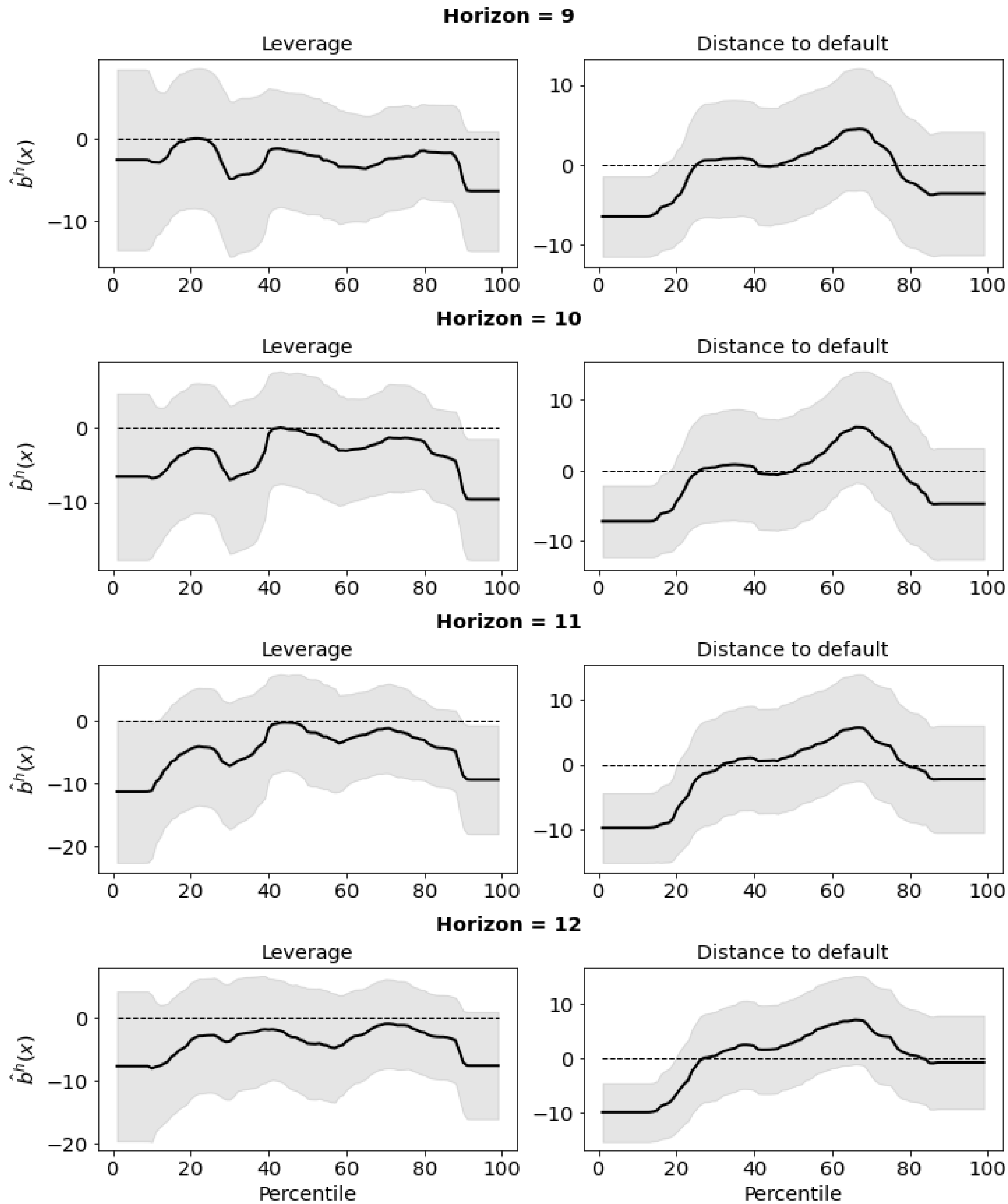


Figure 5: HLPs and the cross-section variation, Long horizons

The figure presents the impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate for firms at a fine grid of percentiles of leverage and distance-to-default, at quarters $h = 9, 10, 11$ and 12 after the shock. These are two distinct specifications in which either leverage or distance-to-default is considered. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are one-way clustered by firm.



in the appendix), where the effect of the shock increases monotonically from high to low risk firms. We do observe however consistency between LPs and HLPs' estimates in that high risk firms tend to present significantly negative responses of investment at long horizons, which is mostly visible from the distance-to-default specification.

Relation to theory. These results highlight an important nonlinearity in the transmission of monetary policy with respect to default risk: short term effects on investment are mostly significant for firms below a certain threshold in the level of leverage or above a certain threshold in the level of distance-to-default, while beyond this threshold the effect of monetary policy is essentially insensitive to the level of default risk. According to the estimations, this threshold is situated around the best quartile of the distribution of default risk (the first quartile of leverage or the third quartile of distance-to-default).

This observed nonlinear effect is consistent with models of heterogeneous firms that incorporate constraints to capital reallocation (Khan and Thomas, 2013) and borrowing issuance costs (Jeenas, 2019). Firms with low enough net worth do not find it optimal to issue new debt given the initial cost or tighter credit constraints, which implies that a change in borrowing costs - triggered by a monetary policy shock for example - will not have a significant effect on the investment decisions of these firms. Alternatively, firms that actively participate in the credit market (and consequently do not face a binding borrowing constraint) are more responsive to changes in market interest rates and borrowing spreads. According to theoretical predictions in Ottonello and Winberry (2020), this later group of firms face smaller credit spreads following a decrease in interest rates, which in turn flattens the slope of their marginal cost curve and induces a change in investment. As the level of net worth increases, credit spreads decrease further and the induced response from policy is larger. This behaviour is consistent with the estimates in Figure 4 for firms with high enough net worth, where here we regard leverage and distance-to-default as proxies for net worth.

Perhaps the most striking result is the estimated threshold itself, which implies that for approximately three quarters of firms in the sample the semi-elasticity of investment is insensitive to financial conditions in the short term. This result is consistent with theoretical predictions from Khan and Thomas (2013) in that frictions to capital reallocation create an heterogeneous investment behaviour with respect to the strength of credit constraints. According to their model, the majority of firms may not engage in new borrowing due to binding constraints or even the prospect that these constraints might bind in the future. In the same line, Leary and Roberts (2005) estimate that 72% of the time firms do not adjust their capital structure due to fixed costs. This suggests that at least for a significant period of time firms tend to be inactive in reballancing their portfolios and consequently in engaging in new investment, which corroborates to the economically relevant threshold estimated in this paper.

6 Concluding remarks

This paper introduces heterogeneous local projections (HLPs), a non-parametric method for the estimation of impulse responses based on random forests, to estimate the transmission of monetary policy shocks to firm investment. The method is useful to uncover possible nonlinearities in the transmission of monetary policy, since it does not impose any assumptions on how the transmission mechanism varies across firms. Using data on US non-financial firms until the Great Recession, my estimates suggest that there exist a threshold in the level of firm risk above which monetary policy is much less effective, particularly for middle-risk firms.

HLPs can be thought as a non-parametric generalization of local projections that nests several linear and nonlinear local projection specifications, and can be used with several common identification schemes in macroeconomics. Unlike other non-parametric techniques, HLPs can be estimated in high dimensions as well. In the context of the investment channel of monetary policy discussed in this paper, this would allow conditioning the impulse responses to monetary shocks on a large pool of firms' characteristics, possibly shedding light on the relevant transmission channels. I do not however explore the high dimension capability of HLPs in this paper and leave it for future work.

Acknowledgements

I would like to thank the invaluable guidance and supervision of Ana Galvao and Giovanni Ricco. I also would like to thank Miguel Bandeira, Luis Candelaria, Danilo Cascaldi-Garcia, Mingli Chen, Roger Farmer, James Fenske, Haroon Mumtaz, Dennis Novy, Roberto Pancrazi, Thijs Van Rens, and Marija Vukotic for helpful comments, as well as all the participants of the Warwick Macro & International workshop.

Funding

This work was supported by the Department of Economics at The University of Warwick.

References

Angrist, J. D., O. Jorda, and G. M. Kuersteiner. (2018). "Semiparametric estimates of monetary policy effects: String theory revisited." Journal of Business and Economic Statistics 36, 371–387.

Arellano, M.. (1987). "Computing robust standard errors for within-groups estimators." Oxford Bulletin of Economics and Statistics 49, 431–434.

Athey, S., and G. Imbens. (2016). "Recursive Partitioning for Heterogeneous Causal Effects." Proceedings of the National Academy of Sciences 113, 7353–7360.

Athey, S., J. Tibshirani, and S. Wager. (2019). "Generalized Random Forests." The Annals of Statistics 47, 1148–1178.

Athey, S., and S. Wager. (2018). "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." Journal of the American Statistical Association 113, 1228–1242.

Auerbach, A.J., and Y. Gorodnichencko. (2013). "Fiscal multipliers in recession and expansion." Alesina, A., Giavazzi, F. (eds.): Fiscal Policy after the Financial Crisis, 19–55, University of Chicago Press.

Bernanke, B. S., M. Gertler, and S. Gilchrist. (1999). "Chapter 21: The financial accelerator in a quantitative business cycle framework." volume 1 of Handbook of Macroeconomics, 1341–1393, Elsevier.

Biau, G.. (2012). "Analysis of a Random Forest Model." Journal of Machine Learning Research 13, 1063–1095.

Breiman, L.. (2001). "Random Forests." Machine Learning 45, 5–32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. G. Stone (1984). Classification and regression trees. Wadsworth International Group, Belmont, California, USA.

Chipman, H.A., E. I. George, and R. E. McCulloch. (2010). "BART: Bayesian additive regression trees." The Annals of Applied Statistics 4, 266–298.

Cloyne, J., C. Ferreira, M. Froemel, and Surico P.. (2018). "Monetary Policy, Corporate Finance and Investment." NBER Working Papers, no. 25366 .

Cloyne, J. S., O. Jorda, and A. M. Taylor. (2023). "State-Dependent local projections: Understanding impulse response heterogeneity." Working Paper Series2023-05, Federal Reserve Bank of San Francisco .

Davis, A. R., and M. S. Nielsen. (2020). "Modeling of time series using random forests: Theoretical developments." Electronic Journal of Statistics 14, 2, 3644–3671.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Efron, B.. (2014). "Estimation and Accuracy after Model Selection." Journal of the American Statistical Association 109, 991–1007.

Efron, B., T. Hastie, and S. Wager. (2014). "Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife." Journal of Machine Learing Research 15, 1625–1651.

Fortin, N., T. Lemieux, and S. Firpo. (2011). "Chapter 1 - Decomposition Methods in Economics." volume 4 of Handbook of Labor Economics, 1–102, Elsevier.

Friedberg, R., J. Tibshirani, S. Athey, and S. Wager. (2021). “Local Linear Forests.” Journal of Computational and Graphical Statistics 30, 503–517.

Gallant, A. R., P. E. Rossi, and G. Tauchen. (1993). “Nonlinear Dynamic Structures.” Econometrica 61, 4, 871–907.

Gertler, M., and S. Gilchrist. (1994). “Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms.” The Quarterly Journal of Economics 109, 2, 309–340.

Gilchrist, S., and E. Zakrajsek. (2012). “Credit Spreads and Business Cycle Fluctuations.” American Economic Review 102, 4, 1692–1720.

Gonçalves, S., A. M. Herrera, L. Kilian, and E. Pesavento. (2024). “State-dependent local projections.” Journal of Econometrics 105702.

Gorodnichenko, Y., and M. Weber. (2016). “Are Sticky Prices Costly? Evidence from the Stock Market.” American Economic Review 106, 1, 165–99.

Gourieroux, C., and J. Jasiak. (2005). “Nonlinear innovations and impulse responses with applications to VaR sensitivity.” Annales d’Economie et de Statistique 78, 1–31.

Green, D. P., and H. L. Kern. (2012). “Modelling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees.” Public Opinion Quarterly 76, 491–511.

Gurkaynak, R. S., B. Sack, and E. T. Swanson. (2005). “Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements.” International Journal of Central Banking 1, 55–93.

Hansen, B. (2022). Econometrics. Princeton University Press.

Hastie, T., R. Tibshirani, and J. Friedman (2001). The Elements of Statistical Learning. Springer Series in Statistics, New York, NY, USA: Springer New York Inc.

Hill, J., and Y. S. Su. (2013). “Assessing Lack of Common Support in Causal Inference using Bayesian Nonparametrics: Implications for Evaluating the Effect of Breastfeeding on Childrens Cognitive Outcomes.” Annals of Applied Statistics 7, 1386–1420.

Jeenas, P.. (2019). "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics." Working Paper .

Jorda, O.. (2005). "Estimation and Inference of Impulse Responses by Local Projections." American Economic Review 95, 161–182.

Kapetanios, G.. (2008). "A Bootstrap Procedure for Panel Data Sets with Many Cross-sectional Units." The Econometrics Journal 11, 2, 377–395.

Khan, A., and J. K. Thomas. (2008). "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics." Econometrica 76, 2, 395–436.

Khan, A., and J. K. Thomas. (2013). "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity." Journal of Political Economy 121, 6, 1055–1107.

Koby, Y., and C. K. Wolf. (2020). "Aggregation in Heterogeneous-Firm Models: Theory and Measurement." Unpublished Working Paper .

Koop, G., M. H. Pesaran, and S. M. Potter. (1996). "Impulse response analysis in nonlinear multivariate models." Journal of Econometrics 74, 1, 119–147.

Leary, M. T., and M. R. Roberts. (2005). "Do Firms Rebalance Their Capital Structures?" The Journal of Finance 60, 6, 2575–2619.

MacKinnon, J.. (2006). "Bootstrap Methods in Econometrics." The Economic Record 82, S2–S18.

MacKinnon, J., and H. White. (1985). "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." Journal of Econometrics 29, 3, 305–325.

Meinshausen, N.. (2006). "Quantile Regression Forests." The Journal of Machine Learning Research 7, 983–999.

Mumtaz, H., and M. Piffer. (2022). "Impulse response estimation via flexible local projections." arXiv .

Olea, J. L. M., and M. Plagborg-Moller. (2021). “Local projection inference is simpler and more robust than you think.” Econometrica 89, 4, 1789–1823.

Ottonello, P., and T. Winberry. (2020). “Financial Heterogeneity and the Investment Channel of Monetary Policy.” Econometrica 88, 6, 2473–2502.

Plagborg-Moller, M., and C. K. Wolf. (2021). “Local Projections and VARs Estimate the Same Impulse Responses.” Econometrica 89, 955–980.

Powers, S., J. Qian, K. Jung, Schuler A., N.H. Shah, Hastie T., and Tibshirani R.. (2018). “Some methods for heterogeneous treatment effect estimation in high dimensions.” Stat Med. 37, 1767–1787.

Ramey, V. A., and S. Zubairy. (2018). “Government spending multipliers in good times and in bad: Evidence from U.S. historical data.” Journal of Political Economy 126, 850–901.

Ramey, V.A.. (2016). “Chapter 2 - Macroeconomic Shocks and Their Propagation.” volume 2 of Handbook of Macroeconomics, 71–162, Elsevier.

Scornet, E., G. Biau, and J.P. Vert. (2015). “Consistency of Random Forests.” The Annals of Statistics 43, 1716–1741.

Stock, J., and M. Watson. (2018). “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” The Economic Journal 128, 917–948.

Tenreyro, S., and G. Thwaites. (2016). “Pushing on a string: U.S. monetary policy is less powerful in recessions.” American Economic Journal: Macroeconomics 8, 43–74.

Wager, S., and G. Walther. (2015). “Adaptive Concentration of Regression Trees, with Application to Random Forests.” arXiv .

Willke, R. J., Z. Zheng, P. Subedi, R. Althin, and C. D. Mullins. (2012). “From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer.” BMC Medical Research Methodology 12.

Zetlin-Jones, A., and A. Shourideh. (2017). "External financing and the role of financial frictions over the business cycle: Measurement and theory." Journal of Monetary Economics 92, 1–15.

A Derivation of equation 9

Suppose we grow $|\mathcal{T}|$ trees according to the methodology described in Section 3. Define $L_{\mathcal{T}}(\boldsymbol{x})$ as the set of observations it that are in the same subsample of $\boldsymbol{x}$ according to tree $\mathcal{T}$, and let $\widehat{b}_{\mathcal{T}}(\boldsymbol{x})$ denote the prediction of tree $\mathcal{T}$ at $\boldsymbol{x}$. (In this section I omit the h superscript to simplify the notation). As in Breiman (2001), the forest estimator is defined as the average of all tree predicitons in the ensemble,

$$\widehat{b}(\boldsymbol{x}) = \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \widehat{b}_{\mathcal{T}}(\boldsymbol{x}), \tag{A.1}$$

where the prediction of tree $\mathcal{T}$ at $\boldsymbol{x}$ is the impulse response from (1) in the subsample associated with $\boldsymbol{x}$, denoted $\{it : \boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\}$. The prediction of tree $\mathcal{T}$ can then be written as

$$\begin{aligned} \widehat{b}_{\mathcal{T}}(\boldsymbol{x}) &= \frac{\frac{1}{|L_{\mathcal{T}}(\boldsymbol{x})|} \sum_{\{it:\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\}} \left(Y_{i,t+h} - \bar{Y}_i\right)\left(W_{it} - \bar{W}_i\right)}{\frac{1}{|L_{\mathcal{T}}(\boldsymbol{x})|} \sum_{\{it:\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\}} \left(W_{it} - \bar{W}_i\right)^2} \\ &= \frac{\sum_{it} \frac{\mathbf{1}\left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right]}{|L_{\mathcal{T}}(\boldsymbol{x})|} \left(Y_{i,t+h} - \bar{Y}_i\right)\left(W_{it} - \bar{W}_i\right)}{\sum_{it} \frac{\mathbf{1}\left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right]}{|L_{\mathcal{T}}(\boldsymbol{x})|} \left(W_{it} - \bar{W}_i\right)^2} \end{aligned} \tag{A.2}$$

where $\mathbf{1}$ is the indicator function and $|L_{\mathcal{T}}(\boldsymbol{x})|$ denotes the number of observations in partition $L_{\mathcal{T}}(\boldsymbol{x})$. In (A.2), I also rely on the prior orthogonalization of both the dependent and shock variables with respect to the set of controls as described in the main text.

Replacing (A.2) in (A.1), we have

$$\widehat{b}(\boldsymbol{x}) = \frac{\sum_{it} \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \frac{\mathbf{1}\left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right]}{|L_{\mathcal{T}}(\boldsymbol{x})|} \left(Y_{i,t+h} - \bar{Y}_i\right)\left(W_{it} - \bar{W}_i\right)}{\sum_{it} \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \frac{\mathbf{1}\left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right]}{|L_{\mathcal{T}}(\boldsymbol{x})|} \left(W_{it} - \bar{W}_i\right)^2} = \frac{\sum_{it} \alpha_{it}(\boldsymbol{x}) \left(Y_{i,t+h} - \bar{Y}_i\right)\left(W_{it} - \bar{W}_i\right)}{\sum_{it} \alpha_{it}(\boldsymbol{x}) \left(W_{it} - \bar{W}_i\right)^2} \tag{A.3}$$

where we define $\alpha_{it}(\boldsymbol{x}) \equiv \frac{1}{|\mathcal{T}|} \sum_{\mathcal{T}} \mathbf{1}\left[\boldsymbol{X_{i,t-1}} \in L_{\mathcal{T}}(\boldsymbol{x})\right] / |L_{\mathcal{T}}(\boldsymbol{x})|$ as the weights (i.e. kernel) implied by the forest model.

B Additional results for Section 5

Figure B.1: HLPs and the distribution of impulse responses

The figure presents the heterogeneous impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate at different levels of firm risk according to leverage and distance-to-default. Each panel represents a distinct specification in which either leverage or distance-to-default is considered, but not both. No uncertainty bands are plotted to facilitate visualization. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$.

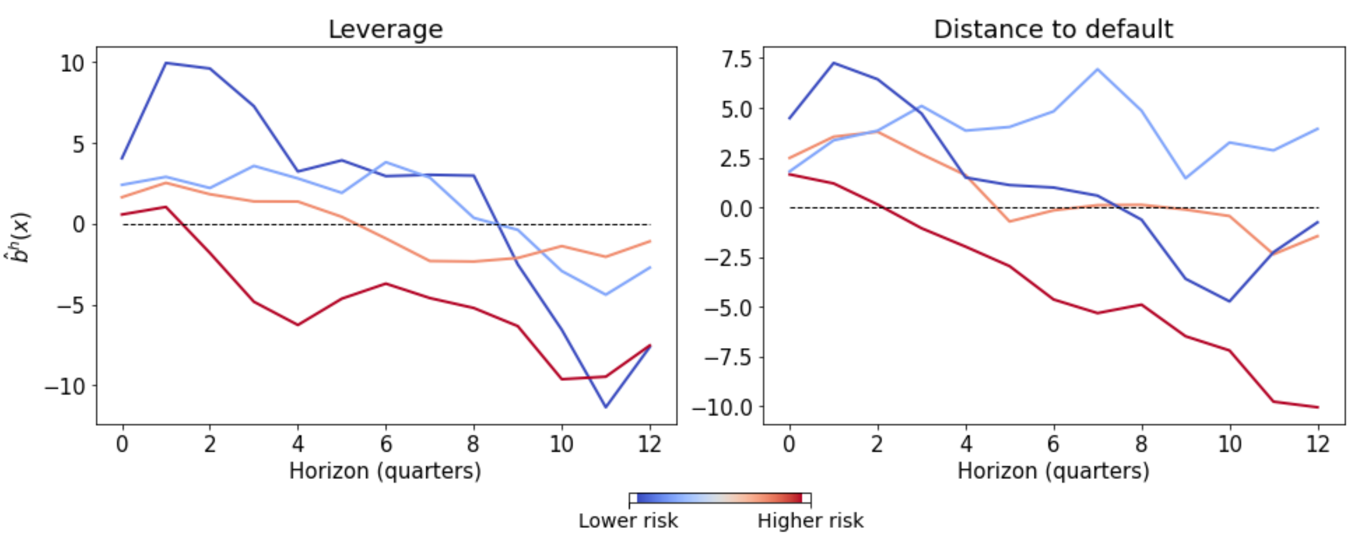


Figure B.2: Standard LPs and the cross-section variation, Short horizons

The figure presents the impulse responses according to standard LPs $\hat{b}^{h,LP}(x)$ of firm investment following a 100bps decrease in the fed funds rate for firms at a fine grid of percentiles of leverage and distance-to-default, at quarters $h = 0, 1, 2$ and 3 after the shock. Left and right panels represent distinct specifications in which either leverage or distance-to-default is considered, but not both. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are two-way clustered by firm and time.

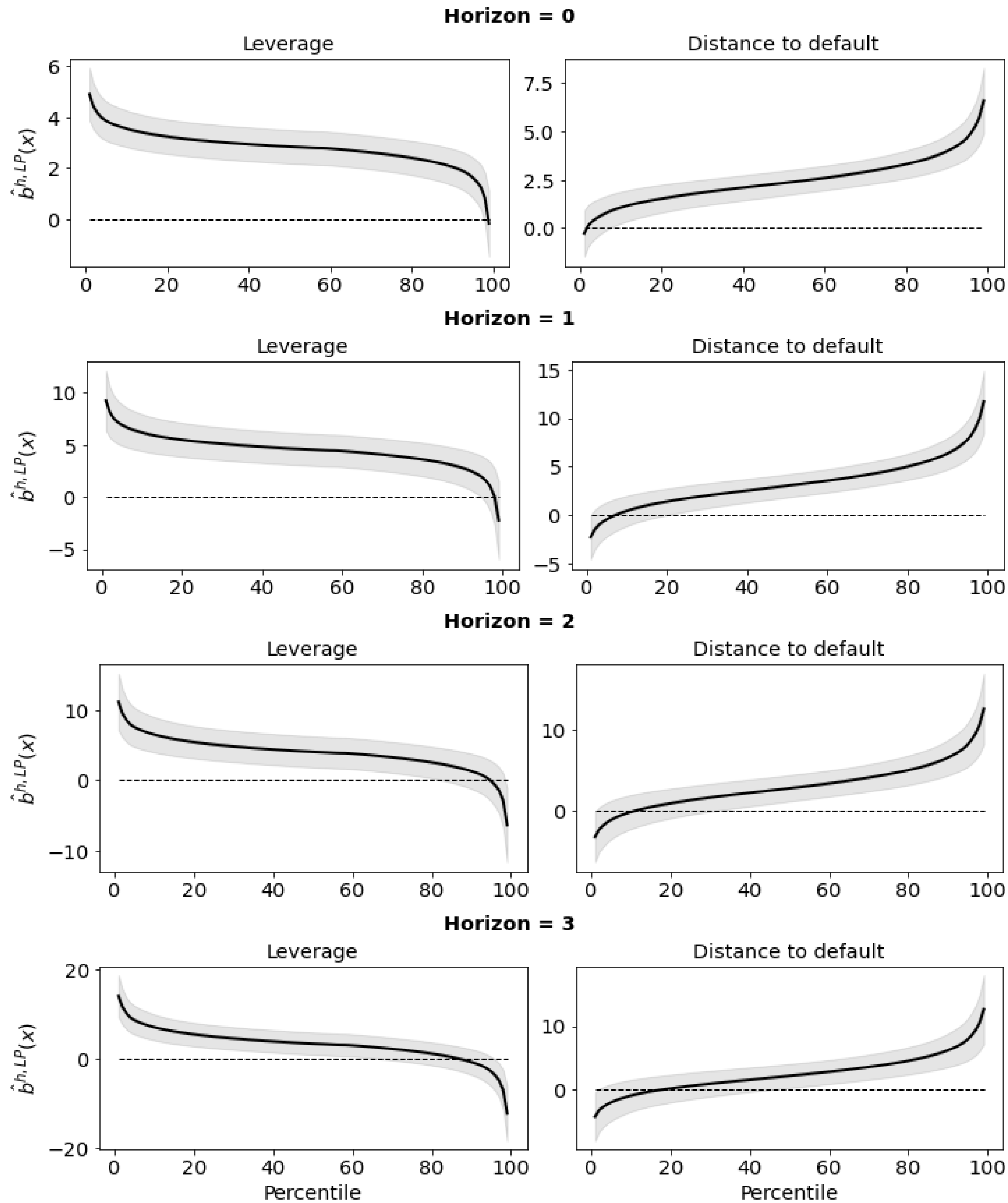


Figure B.3: Standard LPs and the cross-section variation, Long horizons

The figure presents the impulse responses according to standard LPs $\hat{b}^{h,LP}(x)$ of firm investment following a 100bps decrease in the fed funds rate for firms at a fine grid of percentiles of leverage and distance-to-default, at quarters $h = 9, 10, 11$ and 12 after the shock. Left and right panels represent distinct specifications in which either leverage or distance-to-default is considered, but not both. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are two-way clustered by firm and time.

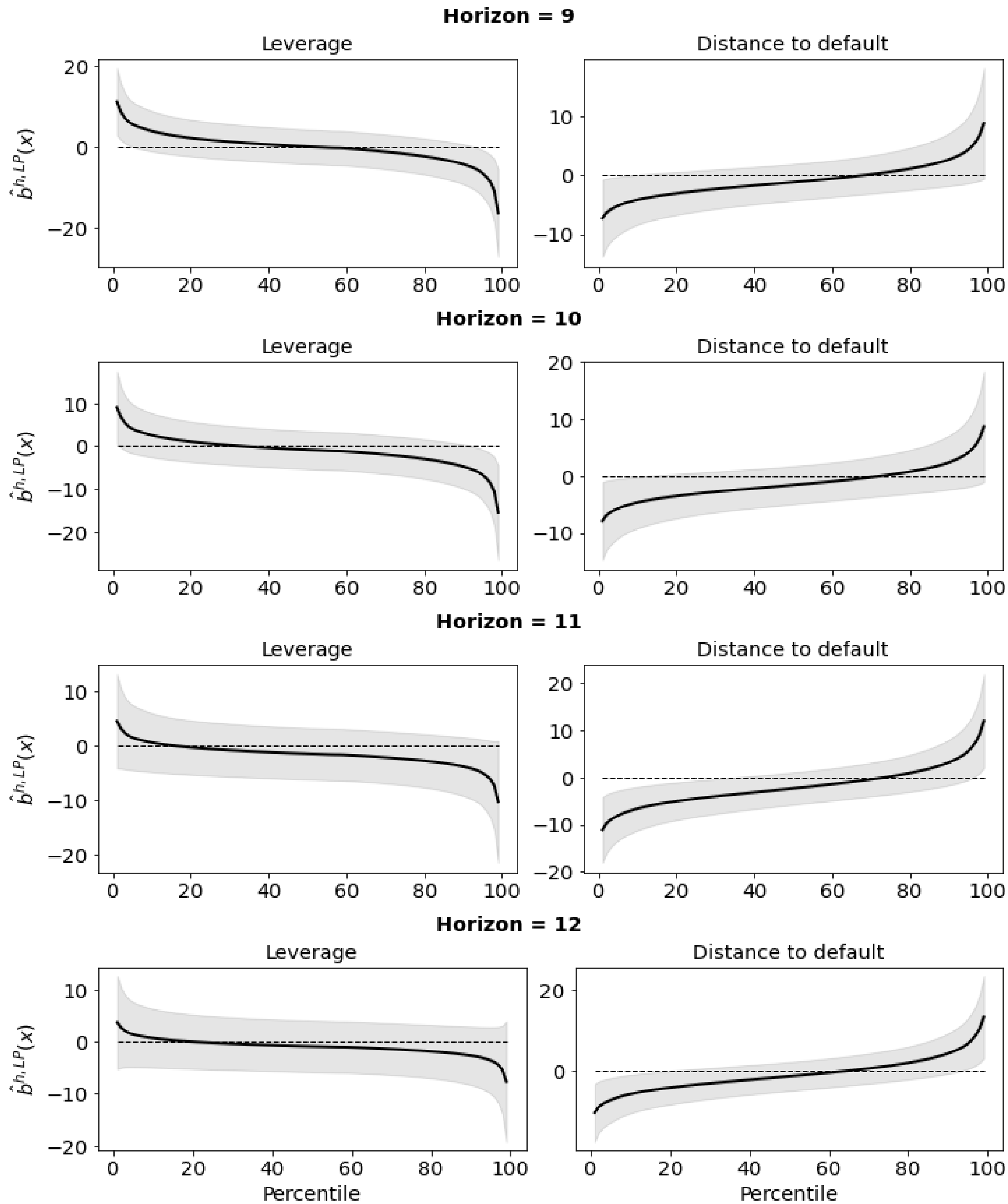
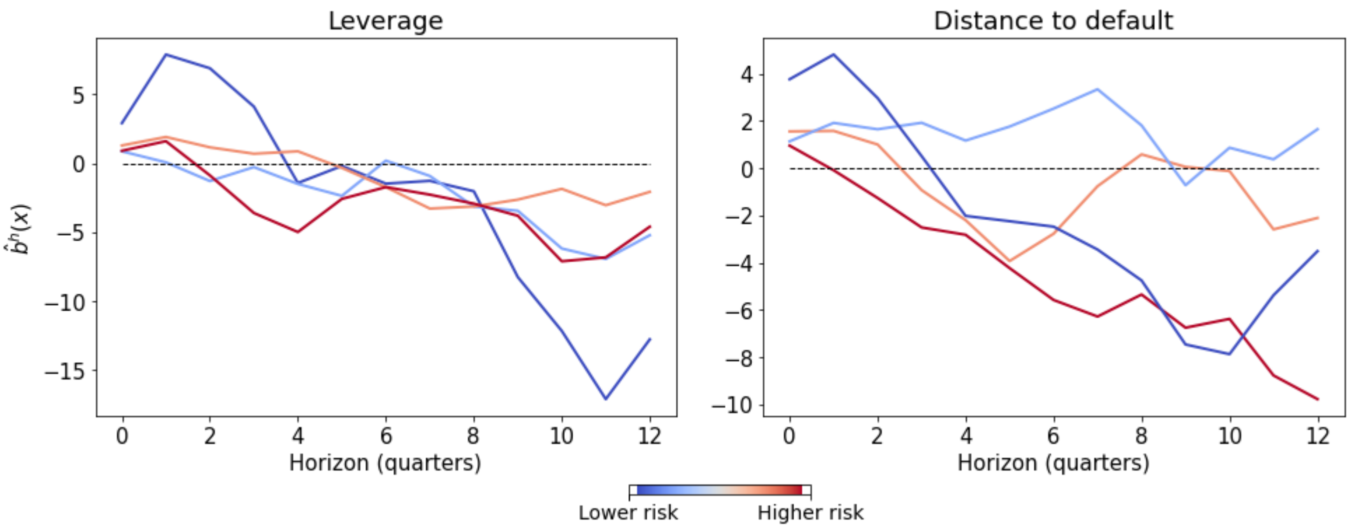


Figure B.4: HLPs and the distribution of impulse responses, Lag-augmented version

The figure presents the heterogeneous impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate at different levels of firm risk according to leverage and distance-to-default. The set of controls is similar to that in specification from column 8 in Table 3. Each panel represents a distinct specification in which either leverage or distance-to-default is considered, but not both. No uncertainty bands are plotted to facilitate visualization. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$.



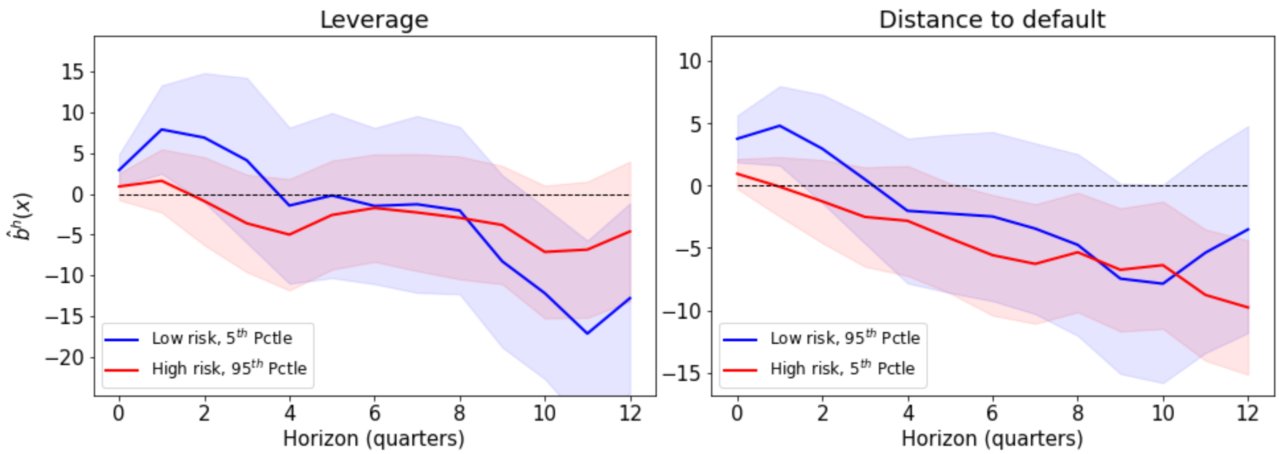
C Robustness

Lag-augmented HLPs. It is common practice in local projection estimation to compute autocorrelation-robust standard errors due to the presence of serial correlation in the residuals. Although the method of HLPs does not account directly for this adjustment, it allows for a very simple recipe that obliviates the need for this correction. Olea and Plagborg-Moller (2021) show that one can simply augment the regression of interest with lags of all the variables in the system and disregard the adjustment for serial correlation. The results in the previous section do not involve lag-augmented regressions to allow for the comparison with results in the literature. Here I show that the same qualitative results from HLPs can be obtained with lag-augmented regressions.

I augment the vector of controls to include four lags of firm-level controls (instead of only one lag as before), as well as four lags of the dependent variable. It is interesting to first analyse the baseline results on impact for this lag-augmented case, which are displayed

Figure C.1: Heterogeneous local projections, Lag-augmented version

The figure presents the heterogeneous impulse responses $\hat{b}^h(x)$ of firm investment following a 100bps decrease in the fed funds rate for low and high risk firms according to (2). The set of controls is similar to that in specification from column 8 in Table 3. Each panel represents a distinct specification in which either leverage or distance-to-default is considered, but not both. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$, and $k = 2500$. Shaded areas are 68% error bands, where standard errors are one-way clustered by firm.



in columns 5-8 of Table 3. A few points to note. When controlling for more lags, there is less evidence of heterogeneity. Specifically, column 5 shows that there is no evidence of heterogeneous responses with respect to leverage, while in column 6 we see that there is less evidence of heterogeneity with respect to distance-to-default, although the estimate is still positive and significant. The coefficient of the shock alone (column 8), which represents the average impact of the shock on investment, is positive and significant, but here again smaller than in the baseline case.

Figure C.1 displays HLPs' impulse responses for leverage and distance-to-default for the lag-augmented version. We see that HLPs' responses on impact indicate less heterogeneity compared to the non lag-augmented case (Figure 3), in particular for the leverage specification, consistent with the results from Table 3. Figure B.4 shows a more granular distribution of responses across percentiles, where it is more visible that the lag-augmented specification entails less heterogeneity, especially for the leverage version (note the comparison with the original specification in Figure B.1). Nonetheless, we do observe differences between low and high risk firms as before. For example, firms at the 95^{th} percentile of distance-to-default are estimated to have a semi-elasticity of investment of around 3.7, while this number decreases to roughly 1 for those at the 5^{th} percentile.

Sensitivity to $\boldsymbol{k}$**.** The parameter k governing tree depth is usually important in determining the properties of the forest estimator. Several works interested in the consistency of random forests assume $k \to \infty$ as the total number of observations increases, or equivalently, a lower bound for k seems to be necessary to achieve consistency in practice (e.g. Biau, 2012; Scornet et al., 2015; Wager and Walther, 2015; Davis and Nielsen, 2020). Although the honesty property (described in Section 3.1) in principle permits trees to grow deep (it allows for low values of k) while maintaining consistency, it could be informative to analyse the sensitivity of the impulse response estimates for different values of k.

Figure C.2: Sensitivity to tree depth k

HLPs impulse responses for different values of k, the minimum number of observations *it* imposed at the subsample level. The specification conditions the impulse responses to depend on distance-to-default only, and are with respect to low (95^{th} percentile) and high ((5^{th} percentile) risk firms. For the estimations, the number of trees in the forest is $|\mathcal{T}| = N \approx 5000$. Shaded areas are 68% error bands.

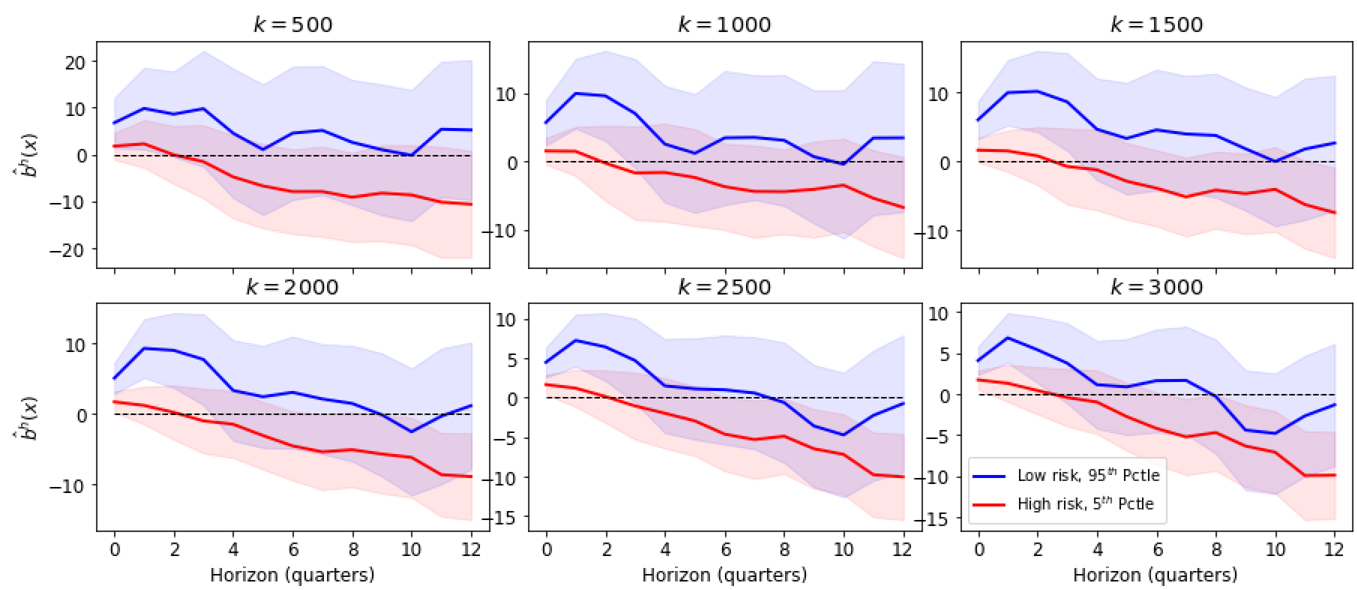


Figure C.2 shows HLPs estimates with respect to distance-to-default for $k = \{500, 1000, 1500, 2000, 2500, 3000\}$. These roughly correspond to assuming that k is $\{2.5\%, 5\%, 7.5\%, 10\%, 12.5\%, 15\%\}$ of the total number of observations used for tree estimation respectively. I note that results reported previously use $k = 2500$. Overall, it is reassuring to confirm that the estimated responses are not significantly different when tree size changes. In particular, especially for $k \geq 1000$, point estimates at short horizons are very similar and surprisingly stable as we vary k. However, for very small values of k (for very large trees) the estimates become more unstable while the uncertainty bands increase as a result. What threshold to use is in fact an empirical matter. On the one hand, bigger trees are less stable, and on the other, smaller trees are less able to properly identify the underlying heterogeneity, where in the limit the estimate becomes the average impulse response (case of one single sample with no splits). In this application, it seems safe to use any $k \in [1000, 3000]$ as the respective point estimates are stable in this interval, while maintaining relatively big trees to pick up the desired heterogeneity in the responses.