Independent and conditionally independent counterfactual distributions

Marcin Wolski*

European Investment Bank
98-100 Boulevard Konrad Adenauer
L-2950, Luxembourg

This version: December 2017

Abstract

This study proposes a novel dependence filtering framework. The filter provides a draw from a counterfactual distribution which is independent from the effects of given covariates. We provide estimation techniques and an inference roadmap for such distributions, and a numerical exercise confirms that the approach performs well in nonlinear environments. Furthermore, we provide bootstrap validity results for the confidence interval of the estimates. We apply the approach to filter out the sovereign risk spill-overs on corporate cost of borrowing in selected euro area countries.

JEL-Classification: C14, E43, E52, G21
Keywords: filtering, nonparametric methods, sovereign risk, cost of borrowing, pass-through

*Electronic address: M.Wolski@eib.org; Corresponding author.

1 Introduction

Counterfactual distributions have become an important tool for policy practicioners to evaluate the performance of proposed strategies, decompose specific policy outcomes, select the most efficient implementation mechanisms, and tailor-make specific policy instruments. In the context of impact assessment, Gertler et al. (2011) point out that counterfactual scenarios constitute an industry benchmark for evidence-driven policy making.

In general terms, one can think of a counterfactual distribution, denoted by Y', as representing the behaviour of distribution Y in the context of information spill-overs from a given set of covariates X. Chernozhukov et al. (2013) point out that distribution Y' can reflect the effects of changes in a covariate distribution X onto the outcome distribution Y, or it can echo a relationship change between X and Y.

Despite the wide array of applications, the vast majority of counterfactual scenarios are user-designed, suffering from a natural over-simplification and potential model misspecification biases. Nevertheless, the recent advances in counterfactual distributions aim at providing possibly assumption-free inference techniques. Chernozhukov et al. (2013) offer a complete toolkit to study counterfactual distributions through a prism of regression methods. Rothe (2010) extends the approach to a fully nonparametric setup and demonstrates that nonparametric estimation has superior Mean Squared Error (MSE) performance in the case of (functional) model misspecification. Rothe (2012) further extends the nonparametric aproach to cover partial distributional effects.

In those techniques, the identification of counterfactual distribution Y' reflects a change in the covariate distribution from X to, say, X', both of which need to be known to the end user. Additionally, for Y' to be well-defined, the support of X' needs to be included in the support of X, $X' \subseteq X$. In this paper we propose a different identification strategy, which follows from the independence and conditional independence principles between random variables.

The framework has its starting point in hypothesis testing literature (Hiemstra and Jones, 1994; Diks and Panchenko, 2006; Diks and Wolski, 2016), and further capitalizes on data sharpening methods proposed by Choi and Hall (1999) and Hall and Minnotte (2002). The goal of the procedure is to provide a sample from counterfactual distribution Y' which is independent from the effects of a given set of covariates X. The method can be viewed as a dependence filtering procedure, with Y' describing a counterfactual distribution of Y which is liberated from the effects of X.

We show a link between the proposed technique and the setups of Chernozhukov et al. (2013) and Rothe (2010).[1] There are, however, three advantages of the proposed method. Firstly, the procedure doesn't require distribution X' to construct counterfactuals of Y. Therefore, it can be applied in studies where counterfactual covariate distributions are not readily available or known to the end user *a priori*, like in the time series setups. Secondly, we provide bootstrap confidence intervals for counterfactual realizations of Y'. We show that in the context of dependence analysis it helps characterizing the evolution of variable co-movement. Thirdly, we utilize a smooth version of kernel density/distribution estimates with process- and data-driven bandwidth selection. In

[1] The method allows also to filter out partial dependence between variables, directly relating to Rothe (2012).

the numerical study we show that such techniques substantially improve the Mean Square Error (MSE) performance of the estimates.

In an empirical application we evaluate to what extent the sovereign risk distorted the pass-through channel in the euro area (EA). To provide liquidity to the troubled EA banking sector and to alleviate the supply-side credit constraints, the European Central Bank (ECB) started a series of interest rates cuts in October 2008. The effects on the retail funding costs heavily differed across the member states revealing the heterogeneity in the ECB monetary policy transmission mechanism and the segmentation of the Euro Area (EA) banking system (EIB, 2016).

A sizeable part of cross-country transmission asymmetries can be attributed to the differences in sovereign risk (Arnold and van Ewijk, 2014). Elevated risk levels raise the borrowing costs of the sovereign, which often serve as a benchmark market rate for loan pricing. Additionally, due to arbitrage relations, increases in sovereign yields are reflected in higher retail rates. Last but not least, weak fiscal positions reduce the ability of a sovereign to support financial sector in times of distress. Disturbances to this implicit sovereign backstop can be perceived as an untight safety net, putting an upward pressure on the required rate of return and increasing consequently the lending rates. The sovereign risk component can be also associated with institutional quality and the rule of law.[2]

There is vast international evidence in favour of nonlinear dependencies in the interest rate transmission mechanisms. Apergis and Cooray (2015) estimate a Nonlinear Auto-Regressive Distributed Lag (NARDL) model and confirm asymmetric interest rate pass-through in the U.S., the U.K. and Australia. In particular, they observe that the lending rates responded more strongly to the hikes in policy rates, rather than to the falls. Using a time-varying parameter state space approach Arnold and van Ewijk (2014) find that heterogeneity in credit risk had a negligible effect on retail lending rates in the EA, in contrast to the sovereign risk. Additionally, they report that the sovereign distortions affect deposit rates to a greater extent than the lending rates. On the example of Japan, Kitamura et al. (2016) find that banks with high shares of relationship lending appear to be more restrained in pass-through of policy rate hikes, compared to the falls. This is somehow in line with a more general finding of Huang et al. (2010) and He and Krishnamurthy (2012) who suggest that banking risk is a nonlinear function of asset exposure.

The responsiveness of bank lending rates can also differ with respect to the instruments used by policy makers. For instance, von Borstel et al. (2016) find that the sovereign debt crisis changed the composition of the pass-through, adjusting for indirect effects from lower sovereign risk premia in the EA. Furthermore, Hristov et al. (2014) suggest that the pass-through distortion can depend on the average size of a structural shock. At a national level, in larger and more diverse economies, the pass-through might be affected by regional differences (Montagnoli et al., 2016) and the characteristics of the ultimate borrower (Stanislawska, 2015).

We focus on four EA countries, i.e. France, Italy, Portugal and Spain, for which we filter out the short- and long-run sovereign risk effect on the retail bank lending rates. Our main findings confirm the pass-through heterogeneity across the EA countries. We also find evidence in favour of

[2] On the sample of 122 countries, Perera and Wickramanayake (2016) conclude that, on top of financial development and market characteristics, institutional and governance factors play a significant role in the pass-through mechanism.

“flight to quality” hypothesis during the financial and sovereign debt crises (Arnold and van Ewijk, 2014). Indeed, the sovereign risk contributions in France and Germany were negative, and in Italy and Spain they were positive, during the crisis periods. Interestingly, as of 2016 the sovereign risk distortions escalated in Germany (although it still remains muted). We speculate that this might come from the fiscal backstop role Germany plays in the EA. We also find that the standard linear models may underestimate the underlying pass-through distortions, suggesting strongly nonlinear sovereign risk spill-overs.

The paper is organized as follows. Section 2 outlines the independent and conditionally independent filtering framework. We provide the estimation techniques of the filtered distribution and we derive its basic properties. We test the methodology numerically in Section 3. We apply the methodology to filter out the sovereign risk effects on the bank lend rates in Section 4. Finally, Section 5 concludes.

2 Methodology

Let us consider a random real-valued outcome variable Y, which is potentially influenced by a d_X-dimensional vector of random real-valued covariates $\boldsymbol{X} = \{X_1, ..., X_{d_X}\}$. The Cumulative Distribution Functions (CDFs) are given by $F_Y(y)$ and $F_{\boldsymbol{X}}(\boldsymbol{x})$, respectively, and assume that the former is invertible around point $y \in \mathbb{R}$. Let us further assume that both Y and $\boldsymbol{X}$ have smooth and well-defined Probability Density Functions (PDFs) over their domains, denoted by $f_Y(y)$ and $f_{\boldsymbol{X}}(\boldsymbol{x})$, respectively. There also exist joint CDFs and PDFs, denoted by $F_{Y,\boldsymbol{X}}(y, \boldsymbol{x})$ and $f_{Y,\boldsymbol{X}}(y, \boldsymbol{x})$. For clarity of exposition and without loss of generality let us imagine that the outcome variable and covariates are not independent, i.e. $F_{Y,\boldsymbol{X}}(y, \boldsymbol{x}) \neq F_Y(y)F_{\boldsymbol{X}}(y, \boldsymbol{x})$, or since we assume the existance of the PDFs, $f_{Y,\boldsymbol{X}}(y, \boldsymbol{x}) \neq f_Y(y)f_{\boldsymbol{X}}(y, \boldsymbol{x})$ for some $y \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^{d_X}$.[3] In principle, if the variables of interest are mutually independent, the following reasoning still holds, and the filtering map is identify by the identity transformation, as suggested later by Lemma 1.

Assume that we have a sample $\{(Y_i, \boldsymbol{X_i}) : i = 1, ..., n\}$, consisting of independent and identically distributed (i.i.d.) draws from $(Y, \boldsymbol{X})$ with properties described above. By filtering out the effects of $\boldsymbol{X}$ onto Y we mean finding a counterfactual distribution of a random variable Y', characterized by a sample $\{Y'_i : i = 1..n\}$, which is independent from $\boldsymbol{X}$. By design, counterfactual distribution Y' is defined over the subset of the support as Y, $Y' \subseteq Y$. In terms of PDFs, one can write the independence condition as

$$f_{Y',\boldsymbol{X}}(y, \boldsymbol{x}) = f_Y(y)f_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

for all y and $\boldsymbol{x}$. Note that in this setup the independence is obtained by adjusting the joint probability space rather than the respective unconditional distributions. We motivate this by the information theory. In case of dependent random variables, variables $\boldsymbol{X}$ provide extra information about variable Y. Consequently, the conditional distribution of Y given $\boldsymbol{X}$ is not equal to the

[3] PDF-representation proved to be computationally more attractive and has been widely used in hypothesis testing (Hiemstra and Jones, 1994; Diks and Panchenko, 2006; Diks and Wolski, 2016). Although, our main results can be also derived from the CDF functionals, the PDF approach allows to link the methodology to the testing environment explicitly.

unconditional distribution of Y. Filtering this extra information should be then naturally defined over the conditional rather than unconditional distribution of Y, which is reflected in Eq. (1). Additionally, such a setting builds an easy link to the literature on counterfactual analysis and inference (see, for instance, Chernozhukov et al. (2013)), which we describe later in the paper.

The sample from a counterfactual distribution Y' is obtained by solving Eq. (1) for realizations $Y_i' = y_i'$. Using the data sharpening framework, one can show that the realizations $Y_i' = y_i'$ are a function of Y_i for given realizations $\boldsymbol{X}_i = \boldsymbol{x}_i$, for each $i = 1, ..., n$, i.e. $Y_i' = \phi(Y_i|\boldsymbol{X}_i = \boldsymbol{x}_i)$. The main result is summarized in Theorem 1.

Theorem 1. *Suppose that we have an i.i.d. sample* $\{(Y_i, \boldsymbol{X_i}) : i = 1, ..., n\}$ *from a continuous distribution with well-defined and sufficiently smooth PDFs. Then, the counterfactual distribution* Y_i'*, satisfying the independence condition given in Eq. (1), follows asymptotically*

$$F_Y(y') = F_{Y|\boldsymbol{X}}(y|\boldsymbol{x}), \tag{2}$$

where $F_{Y|\boldsymbol{X}}$ *is the conditional distribution function of* Y *given* $\boldsymbol{X} = \boldsymbol{x}$.

Proof of Theorem 1 is built upon kernel and data-sharpening methods (see Wand and Jones (1995); Silverman (1998); Hall and Minnotte (2002)), and can be found in Appendix A.1.

Theorem 1 complements and extends the growing body of literature on counterfactual distributions (Chernozhukov et al., 2013; Rothe, 2010, 2012). There are two main novelties in the proposed framework. Firstly, the setup doesn't require a counterfactual covariate distribution $\boldsymbol{X'}$. It makes the procedure an attractive alternative for practical applications, where such distributions are often unknown or unobservable. Secondly, since the RHS of Eq. (2) is identified by the data, and under an additional assumption that F_Y is strictly increasing, Theorem 1 is also identified for any point in the support of (Y, X).

In the context of identification in nonseparable models (Hoderlein and Mammen, 2007; Imbens and Newey, 2009), the framework replaces a potentially strict assumption of (conditional) error exogeneity, by a relatively weaker restriction of strict monotonicity of F_Y. Nevertheless, this raises potential concerns for results' interpretation. To show it formally, it is useful view the counterfactual distribution Y' in relation to the distribution Y, following Lemma 1.

Lemma 1. *Suppose that* Y' *satisfies the conditions outlined in Theorem 1. Then,*

$$F_{Y'}(y') = \delta(y, \boldsymbol{x}) F_Y(y'), \tag{3}$$

where $\delta(y, \boldsymbol{x}) = F_Y(y) F_{\boldsymbol{X}}(\boldsymbol{x}) / F_{Y,\boldsymbol{X}}(y, \boldsymbol{x})$.

Proof of Lemma1 can be found in Appendix A.2. It confirms that distribution of Y' is directly linked to distribution of Y, with functional $\delta(y, x)$ capturing the dependence between variables Y and $\boldsymbol{X}$ at point $(y, \boldsymbol{x})$.

There are two straightforward implications of Lemma 1. Firstly, it shows that independence between Y and $\boldsymbol{X}$ requires an equivalence between $F_{Y'}(y')$ and $F_Y(y')$ for any point y' in the support of Y'. Secondly, it explicitly defines the sharpening function in terms of dependence structure between model variables. To introduce the concept formally, let us consider the following nonseparable model characteristics

$$Y = m(\boldsymbol{X}, \boldsymbol{\varepsilon}), \tag{4}$$

where $m(.)$ describes the general model dynamics and ε is an unobservable error component. Under exogeneity assumption, i.e. $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$, the model is identified by the data, in the sense that distribution of $Y|\boldsymbol{X} = \boldsymbol{x}$ is equivalent to distribution of $m(\boldsymbol{x}, Q_{\boldsymbol{\varepsilon}}(\boldsymbol{e}))$, where the error component is identified through its quantile function $Q_{\boldsymbol{\varepsilon}}(\boldsymbol{e}) \equiv \inf\{u : F_{\boldsymbol{\varepsilon}}(u) \geq \boldsymbol{e}\}$.[4] If $\boldsymbol{X} \not\perp\!\!\!\perp \boldsymbol{\varepsilon}$ the two distributions differ in an unknown way (Torgovitsky, 2011). Although dependence filtering doesn't focus on properties of functional $m(.)$ directly, which makes the identification less of an issue, violation of the exogeneity condition affects the properties of the filtered distribution Y'.

Functional $m(.)$ characterizes the dependence structure between covariates and the error term through the joint distribution function $F_{\boldsymbol{\varepsilon},\boldsymbol{X}}(u, x)$. Consequently, Lemma 1 allows to represent the dependency structure of model in Eq. 4 as

$$\delta(y, \boldsymbol{x}, \boldsymbol{u})^{-1} = \frac{F_{Y,\boldsymbol{\varepsilon}|\boldsymbol{X}}(y, u|x)}{F_Y(y)F_{\boldsymbol{\varepsilon}|\boldsymbol{X}}(u|x)}. \tag{5}$$

Under exogeneity condition it follows that $F_{\boldsymbol{\varepsilon}|\boldsymbol{X}}(u|x) = F_{\boldsymbol{\varepsilon}}(u)$ and the filtering procedure preserves the original dependency of Y on the unobserved error term $\boldsymbol{\varepsilon}$ for each $\boldsymbol{X} = \boldsymbol{x}$. However, if $\boldsymbol{X} \not\perp\!\!\!\perp \boldsymbol{\varepsilon}$, there is residual dependency on $\boldsymbol{X}$, which is passed on Y' through a cofounding error component. This property can build an interesting case for testing the error exogeneity restriction, however, it is outside the scope of this paper and we leave it for further considerations.

Lemma 1 implies that if the expected value and variance of Y are finite, then the expected value and variance of the filtered variable Y' are also finite. The RHS of Eq. (1) is parallel to the functional of the Probability Integral Transform (PIT) and which follows uniform distribution with well-defined and finite expected value and variance.

We view this result as quite interesting as the filtering procedure doesn't remove the dependence by infinitely increasing the variance, so that the dependence structure dilutes over the sample space.

2.1 Estimation

We propose a coherent, and fully nonparametric, framework to estimate the filtered distribution Y'. To generalize the notation from Section 2, let us consider a multivariate random variable $\boldsymbol{W}$, with a corresponding sample satysfying Assumption 1.

Assumption 1. *Data $\{\boldsymbol{W}_i : i = 1, ..., n\}$, where $\boldsymbol{W}_i = \{W_{1i}, ..., W_{d_W i}\}$, are i.i.d. as a d_W-variate smooth continuous distribution $F_{\boldsymbol{W}}(\boldsymbol{w})$ with well-defined PDF $f_{\boldsymbol{W}}(\boldsymbol{w})$ and respective derivatives, up to a finite order r, being finite, continuous and uniformly bounded on the support.*

Following Wand and Jones (1995), the kernel PDF estimator, around point $\boldsymbol{w}$, is given by

$$\hat{f}_{\boldsymbol{W}}(\boldsymbol{w}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\boldsymbol{w} - \boldsymbol{W_i}), \tag{6}$$

[4]The following reasoning can be naturally extended to a, relatively weaker, conditional exogeneity assumption $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}|\boldsymbol{U}$ where $\boldsymbol{U}$ is an additional conditioning variable.

where $\mathbf{H}$ is a symmetric positive definite $d_W \times d_W$ bandwidth matrix. To simplify the presentation and without the loss of generality, let us further impose that $\mathbf{H}$ is a diagonal with elements $\text{diag}(h_1^2, ..., h_{d_W}^2)$. For clarity of exposition, we set that all elements of $\mathbf{H}_{\boldsymbol{X}}$ are symmetric, i.e. $\mathbf{H}_{\boldsymbol{X}} = \text{diag}(h^2, ..., h^2)$. Expression

$$K_{\mathbf{H}}(\boldsymbol{w}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\boldsymbol{w}), \tag{7}$$

is the scaled kernel with $|.|$ denoting the determinant and $K(.)$ being a d_W-variate kernel function satisfying Assumption 2.

Assumption 2. *Kernel function $K : \mathbb{R}^{d_{\boldsymbol{W}}} \to \mathbb{R}$ behaves as*

$$\begin{aligned} &\int K(\boldsymbol{w})\mathrm{d}\boldsymbol{w} = 1, \\ &\int K(\boldsymbol{w})\boldsymbol{w}^c \mathrm{d}\boldsymbol{w} = 0 \quad \text{for} \quad c = 1, ..., r-1, \\ &\int K(\boldsymbol{w})\boldsymbol{w}^c \mathrm{d}\boldsymbol{w} = \kappa_r I_{d_W} < \infty \quad \text{for} \quad c = r, \end{aligned} \tag{8}$$

and $K(\boldsymbol{w})$ is r-times differentiable, where I_{d_W} is a $d_W \times d_W$ identity matrix.

The corresponding CDF kernel estimator is given by integrating the kernel function

$$\hat{F}_{\boldsymbol{W}}(\boldsymbol{w}) = n^{-1} \sum_{i=1}^{n} \bar{K}_{\mathbf{H}} \left(\boldsymbol{w} - \boldsymbol{W_i}\right), \tag{9}$$

where $\bar{K}_{\mathbf{H}}(\boldsymbol{w}) = \int_{-\infty}^{\boldsymbol{w}} K(\mathbf{H}^{-1/2}\boldsymbol{u})\mathrm{d}\boldsymbol{u}$.

To estimates of the conditional CDF, corresponding to the RHS of Eq. (2), let us introduce variable V for which the joint i.i.d. sample is given by $\{(V_i, \boldsymbol{W}_i) : i = 1, ..., n\}$, satisfying Assumption 1. Following Li and Racine (2008) and Li and Racine (2013), the conditional CDF can be estimated as

$$\hat{F}_{V|\boldsymbol{W}}(v|\boldsymbol{w}) = \frac{n^{-1}\sum_{i=1}^{n} \bar{K}_{\mathbf{H}_V}(v - V_i) K_{\mathbf{H}}\left(\boldsymbol{w} - \boldsymbol{W_i}\right)}{\hat{f}_{\boldsymbol{W}}(\boldsymbol{w})}, \tag{10}$$

where $\mathbf{H}_V$ is the bandwidth matrix smoothing the CDF over marginal V.

For consistency of the estimates we require that elements of $\mathbf{H}$ and $\mathbf{H}_V$ decrease sufficiently slowly with the sample size. However, following Li and Racine (2008), we note that smoothing in the direction of CDF requires sufficiently fast rates of convergence as well, as summarized by Assumption 3.

Assumption 3. *As $n \to \infty$,*

(i) the bandwidth along the cumulative distribution marginal satisfies $n^{1/2}h_V/(\log n)^{1/2} + n^{1/2}h_V^r \to 0$,

(ii) the bandwidth along the density marginals satisfy $n^{1/2}h^{d_{\boldsymbol{W}}}/\log n + n^{1/2}h^r \to 0$.

To guarantee uniform convergence, we require moderately more restrictive bandwidth behaviour than for the point-wise convergence. Assumption 3 suggests that for CDF marginal the sequence of bandwidths of the form $h_V = c_V n^{-\gamma}$, with $c_V > 0$, should converge at the rate

$\gamma \in ((2r)^{-1}, 1)$. For the PDF marginals, setting $h = cn^{-\beta}$, corresponding to $\mathbf{H}$, we have that $\beta \in ((2r)^{-1}, (2d_{\boldsymbol{W}})^{-1})$. It follows that for conditional estimates of $Y|X = x$ we need $r > d_{\boldsymbol{X}}$, which is satisfied for the simplest example with second-order kernels and $d_{\boldsymbol{X}} = 1$. Clearly, for multivariate applications, the kernel smoothing would require a higher-order kernel, to reduce the bias of the estimates, and create a non-empty set of possible β values.

Assumption 3 guarantees that CDF bandwidths do not converge slower than PDF ones as $n \to \infty$. In fact, the dominant terms in the MSE rates of convergence of the smoothed conditional estimators are of the same order of convergence as in the unsmoothed equivalents, as suggested by Theorem 2.2 of Li and Racine (2008). Since the convergence in MSE implies convergence in probability (Li and Racine, 2007), the corresponding uniform rates of convergence should also be similar in their dominant terms. This argument is built along the lines suggested by the uniform rates of convergence for conditional PDFs, as demonstrated by Ferraty et al. (2010), and it opens a door for smooth, instead of step-wise, CDF estimators in the setup.

In practice, Eq. (10) is often expressed as a weighted unconditional CDF of V, with weights determined by

$$\theta(w, W_i) = \frac{K_{\mathbf{H}}\left(\boldsymbol{w} - \boldsymbol{W_i}\right)}{\hat{f}_{\boldsymbol{W}}(\boldsymbol{w})}. \tag{11}$$

This, in fact, reduces computetional complexity of the estimators (Rothe, 2010). Consequently, the plug-in representation of Theorem 1 simplifies to

$$\sum_{i=1}^{n} \bar{K}_{\mathbf{H}_{Y'}}(y' - Y_i) = \sum_{i=1}^{n} \bar{K}_{\mathbf{H}_Y}(y - Y_i)\theta(\boldsymbol{x}, \boldsymbol{X_i}), \tag{12}$$

where we distinguish between three bandwidth matrices: $\mathbf{H}_{Y'} = h^2_{Y'}$ to smooth the uncondional CDF of Y, $\mathbf{H}_Y = h^2_Y$ to smooth the conditional CDF over marginal Y and $\mathbf{H}_{\boldsymbol{X}} = \text{diag}(h^2_{X1}, ..., h^2_{Xd_X})$ to smooth the conditional CDF in $\boldsymbol{X}$ directions.

Filtered sample is obtained by solving Eq. (12) for any point y' from a counterfactual distribution of Y'. In this respect, following Jones (1992), Eq. (12) is equivalent to the plug-in representation of Eq. (2), where one can use a smooth empirical quantile function $\hat{F}_Y^{-1} = \hat{Q}_Y$.[5]

2.2 Asymptotic properties

The (smooth) filtering framework described above can be directly linked to the literature on empirical processes and their convergence. Let us denote the estimator of the independent counterfactual distribution in terms of the quantile transformation as

$$\hat{Y}' \equiv \hat{Y}'(y, \boldsymbol{x}) = \hat{F}_Y^{-1}(\hat{F}_{Y|\boldsymbol{X}}(y|\boldsymbol{x})). \tag{13}$$

There are two main difficulties in deriving convergence properties of $\hat{Y}$. Firstly, the quantile estimator is evaluated at a random element $\hat{F}_{Y|\boldsymbol{X}}$, adding an extra noise to the process. Secondly, the convergence takes place in a smooth probability space. Although, under extra assumptions, smooth empirical processes have been shown to converge in distribution, they require special care.

[5]Because of better computational performance, in this paper we rely on numerical methods rather than on a quantile function transformation. Jones (1992) shows that both methods have the same asymptotic properties.

Conequently, we firstly derive convergence properties for the standard, i.e. stepwise, estimator of the empirical process characterized in Eq. (13), and then, following Giné and Nickl (2008), we extend the results to smooth estimators. For the Reader's convenience, the main findings are presented jointly below.

Convergence of empirical processes requires certain additional regularity conditions which are outlined as Assumptions 4. In particular, we need to impose bounds on the quantile function as otherwise the process would be non-convergent in the tails.

Assumption 4. *We assume that (i) distribution functions F_Y and $F_{Y|\boldsymbol{X}}$ are Hadamard differentiable, (ii) F_y^{-1} is uniformly Lipschitz and bounded by $[a,b] \in \mathbb{R}$, (iii) the Y is supported by a compact interval on $J \in \mathbb{R}$ for which $F_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ is uniformly bounded by $[p_1, p_2] \in (0,1)$.*

The assumption on Hadamard differentiability allows to linearize the estimate functionals in converging sequences by the functionals delta method (van der Vaart, 2000). As a result, the estimates can be shown to converge to a Gaussian process, as demonstrated by Theorem 2.

Theorem 2. *Suppose that Assumptions 1-4 hold. Then*

$$\sqrt{n}\left(\hat{Y}' - Y'\right) \xrightarrow{d} N(0, \sigma^2), \tag{14}$$

conditional on the data, where σ^2 is given by

$$\sigma^2 = \frac{F_{Y|\boldsymbol{X}}(y|x)(1 - F_{Y|\boldsymbol{X}}(y|x)) + \hat{F}_{Y|\boldsymbol{X}}(y|x)(1 - \hat{F}_{Y|\boldsymbol{X}}(y|x)))}{f_Y\left(F_Y^{-1}(F_{Y|\boldsymbol{X}}(y|x))\right)}. \tag{15}$$

The proof of Theorem 2 is given in Appendix A.3. This result supports the use of standard inference techniques on $\hat{Y}'$ estimates, as the distribution of the difference $\hat{Y}' - Y'$ contains all the information necessary to assess the precision of $\hat{Y}'$. Following Theorem 2, under the assumption that σ^2 is estimated consistently, the sequence $(\hat{Y}' - Y')/\hat{\sigma}$ tends to a standard Gaussian curve. It is then straightforward to calculate the confidence bounds on standard normal quantiles $z_\alpha = Q_N(\alpha)$, as $[\hat{Y}' + z_{\alpha/2}\hat{\sigma}, \hat{Y}' + z_{1-\alpha/2}\hat{\sigma}]$, where α is the statistical significance level.

In practical applications it may be difficult to estimate $\hat{\sigma}^2$ consistently, however. Therefore, we follow the approach of (Chernozhukov et al., 2013) and Rothe (2010), a propose a naive bootstrap method as an alternative to approximate the true distribution of $\hat{Y}' - Y'$. Let us generate a random sample by resampling (i.e. sampling with replacement) from the original distribution $\{(Y_i, \boldsymbol{X}_i) : i = 1, ..., n\}$. For the specific realization i under the bootsrap draw $b = 1, ..., B$, we then estimate the $\hat{Y}'_{i,b}$ on the sample $\{(Y_{i,b}, \boldsymbol{X}_{i,b}) : i = 1, ..., n\}$ according to Theorem 2. This gives the bootstrap distribution $\hat{Y}'_b = \{(\hat{Y}_{i,b}) : i = 1, ..., n\}$. Theorem 3 verifies that the bootstrap procedure is asymptotically valid, i.e. the distribution of the filtered realizations under the bootstrap $\hat{Y}'_b$ mimics well the distribution filtered of values under the true measure Y' as $n \to \infty$.

Theorem 3. *Suppose that Assumptions 1-4 hold. Then*

$$\sqrt{n}\left(\hat{Y}'_b - \hat{Y}'\right) \xrightarrow{d} N(0, \sigma^2), \tag{16}$$

conditional on the data, where σ^2 is the same as in Theorem 2 but taken over the bootstrap sample.

Proof of is given in Appendix A.4. Lemma 23.3 of van der Vaart (2000) implies that under Theorems 2 and 3, bootstrap confidence intervals, given by $[\hat{Y}' + \hat{\zeta}_{\alpha/2}\hat{\sigma}, \hat{Y}' + \hat{\zeta}_{1-\alpha/2}\hat{\sigma}]$, are asymptotically consistent at the $1-\alpha$ level, where ζ_α is taken as α-quantile from the bootstrap distribution. Formally, confidence bounds can be expressed as

$$\lim_{n\to\infty} \inf P\left(\hat{Y}'(y,\boldsymbol{x}) + \hat{\zeta}_{\alpha/2}\hat{\sigma} \leq Y'(y,\boldsymbol{x}) \leq \hat{Y}'(y,\boldsymbol{x}) + \hat{\zeta}_{1-\alpha/2}\hat{\sigma}\right) \geq 1-\alpha. \tag{17}$$

In practice, the unknown quantities $\hat{\zeta}$ and $\hat{\sigma}^2$ can be obtained from the bootstrap distribution directly, simplifying the derivations.

There are two interesting Corollaries to be pointed out, which may be useful in practical applications and build a foundation for the numerical example below. Firstly, we note that the density estimated over the joint $(Y', \boldsymbol{X})$ distribution should demonstrate independence between Y and $\boldsymbol{X}$.

Corollary 1. *Given consistent estimate of Eq. (1) for which $Y' = \phi(Y)$, the estimate of $f_{Y',\boldsymbol{X}}(y,\boldsymbol{x})$, given by*

$$\hat{f}_{Y',\boldsymbol{X}}(y,\boldsymbol{x}) = n^{-1}\sum_{i=1}^{n} K_{\mathbf{H}_Y}(y-\phi(Y_i))K_{\mathbf{H}_{\boldsymbol{X}}}(\boldsymbol{x}-\boldsymbol{X_i}), \tag{18}$$

tends to $f_Y(y)f_{\boldsymbol{X}}(\boldsymbol{x})$ as $n\to\infty$.

The proof of Corolloary 1 is an implication of Theorem 1. The joint density estimate evaluated over the filtered marginal Y' has to be equal the product of the marginals Y and $\boldsymbol{X}$ at any given point $y\in\mathbb{R}$ and $\boldsymbol{x}\in\mathbb{R}^{d_X}$, as required by the filtering procedure. This verifies that the filtering procedure removes any dependence between Y and $\boldsymbol{X}$.

Secondly, we highlight the asymptotic properties of the density estimates over the counterfactual joint distribution.

Corollary 2. *Given consistent estimate of Eq. (1), the estimate $\hat{f}_{Y',\boldsymbol{X}}(y,\boldsymbol{x})$ has bias and variance of the same order as $\hat{f}_Y(y)\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})$ under the same bandwidth rates of convergence.*

Proof of Corollary 2 can be found in Appendix A.5. In fact, Corollary 2 can be useful for practical inference on distribution Y', including hypothesis testing and asymptotic approximations. In particular, given the order of bias and variance, one can readily observe that the rate of convergence of the bandwidth minimizing the Asymptotic Mean Integrated Squared Error (AMISE) of $\hat{f}_{Y',\boldsymbol{X}}(y,\boldsymbol{x})$ is the same as the one of $\hat{f}_Y(y)\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})$. This is again quite an important result as the kernel procedures derived for a generic estimator $\hat{f}_{\boldsymbol{W}}(\boldsymbol{w})$ can be applied to the filtered sample.

2.3 Optimal bandwidths

The fact that that the smoothing parameter along the Y marginal has a not slower rate of convergence, ensures that the uniform rate of convergence of conditional distribution functions is driven by smoothing along $\boldsymbol{X}$ marginals. Without a loss of generality, let us assume for now that by symmetry, all elements of $\mathbf{H}$ are of the same order of convergence. After Ferraty et al. (2010), one finds that the uniform rate of convergence of $\hat{F}_{Y|\boldsymbol{X}}$ is of order $(nh^{d_W}/\log(n))^{-1/2} + h^r + h_V^r$. As a result, Assumption 3 guarantees the conditional distribution estimates converge uniformly at a sufficiently fast rate across the support.

Our current estimation setup requires two sets of bandwidth matrices, i.e. to smooth the unconditional and conditional distributions of Y. Let us denote the former by $\mathbf{H}_Y = h_Y^2$ and the latter as a block diagonal matrix

$$\mathbf{H}_{Y|X} = \begin{bmatrix} \mathbf{H}_Y & 0 \\ 0 & \mathbf{H}_X \end{bmatrix} = \begin{bmatrix} h_0^2 & 0 & \dots & 0 \\ 0 & h_1^2 & 0 & \\ \vdots & 0 & \ddots & 0 \\ 0 & & 0 & h_{d_{\boldsymbol{X}}}^2 \end{bmatrix}. \tag{19}$$

Under Assumption 3, the MSE-optimal rate of convergence of the unconditional distribution estimate satisfies the uniform convergency condition. As a result we take that $h_Y \sim n^{-1/3}$. Provided that the order of the smoothing kernel is sufficiently high, our numerical experiments suggest that taking the middle point of the allowed interval $((2r)^{-1}, (2d_{\boldsymbol{X}})^{-1})$ delivers good size/power performance. In the simplest case with $r = 2$ and $d_{\boldsymbol{X}} = 1$ we have that $h_i \sim n^{-1/3}$ for $i = \{0, 1\}$.

We extract the constants by rescaling the MSE-optimal bandwidths to match the rates of convergence, in line with Li and Racine (2008). We describe several possible bandwidth selection methods and evaluate their performance in the numerical experiment below.

2.4 Conditionally independent counterfactuals

Theorem 1 can be further extended to filter the dependence of Y on one over many possible covariates. Let us assume that in additional to variable $\boldsymbol{X}$, variable Y depends also on a d_Q-dimensional vector of covariates $\boldsymbol{Q} = \{Q_1, ..., Q_{d_Q}\}$.

Following Hiemstra and Jones (1994) and Diks and Panchenko (2006), any additional information about the conditional distribution $Y|\boldsymbol{Q} = \boldsymbol{q}$ which is contained in variable $\boldsymbol{X}$ can be reflected as a distance between the respective conditional distributions. Consequently, the independence condition, for which the filtering procedure removes the effects of $\boldsymbol{X}$ but preserves the dependence on $\boldsymbol{Q}$, can be written as

$$f_{Y'|\boldsymbol{Q},\boldsymbol{X}}(y|\boldsymbol{q}, \boldsymbol{x}) = f_{Y|\boldsymbol{Q}}(y|\boldsymbol{q}). \tag{20}$$

As a result we arrive at the following Theorem.

Theorem 4. *Suppose that we have an i.i.d. sample* $\{(Y_i, \boldsymbol{Q_i}, \boldsymbol{X_i}) : i = 1, ..., n\}$ *from a continuous distribution with well-defined and sufficiently smooth PDFs. Then, the filtered realizations* Y_i' *satisfying the independence condition given in Eq. (20), follow asymptotically*

$$F_{Y|\boldsymbol{Q}}(y'|\boldsymbol{q}) = F_{Y|\boldsymbol{Q},\boldsymbol{X}}(y|\boldsymbol{q}, \boldsymbol{x}). \tag{21}$$

The proof of Theorem 4 follows directly from the steps behind the proof of Theorem 1 described in the Appendix A.1. Similarly, assuming that Eq. (21) is estimated consistently under Assumptions 1-4, Theorems 2 and 3, together with Corollaries 1 and 2, still hold for a generalized setting with variable $\boldsymbol{Q}$.

3 Numerical performance

The goal of the numerical exercise is twofold. Firstly, we assess the advantages of smoothing techniques in constructing independent counterfactual distributions. Secondly, we compare the MSE performance of various bandwidth selectors.

Let us consider a time-series setting with a bivariate ARCH process given by

$$\begin{aligned} X_i &\sim \mathrm{N}(0,1) \\ Y_i &\sim \mathrm{N}\left(0, d + aX_i^2\right), \end{aligned} \tag{22}$$

The process is set up such that $\{X_i\}$ is affecting the second moment of $\{Y_i\}$ and is similar to the one considered by Diks and Wolski (2016). To satisfy the criteria set by Assumption 1, we require that $d > 0$ and $0 < a < 1$. The setting builds an interesting testing environment as it is allow for nonlinear effects, which standard linear models may find difficult to capture. For the simulations we choose $d = 1$ and $a = 0.4$.

Corollary 1 offers a natural performance indicator. In line with Li and Racine (2013), we test the sample MSE of the estimator of $f_{Y',X}(y,x)$ and product of true marginal densities $f_Y(y)f_X(y)$ given by

$$\mathrm{MSE}(\hat{f}_{Y',X}) = n^{-1}\sum_{i=1}^{n}\left(\hat{f}_{Y',X}^{-i}(y_i,x_i) - f_Y(y_i)f_X(x_i)\right)^2. \tag{23}$$

In the filtering procedure as well as to calculate the MSE estimates we use leave-one-out kernel estimators given by $\hat{f}_{\boldsymbol{W}}^{-i}(\boldsymbol{w_i}) = (n-1)^{-1}\sum_{i\neq j} K_{\mathbf{H}}(w_i - w_j)$ for PDFs and by $\hat{F}_{V|\boldsymbol{W}}^{-i}(v_i|\boldsymbol{w_i}) = (n-1)^{-1}\sum_{i\neq j} \bar{K}_{\mathbf{H}_V}(v_i - v_j)K_{\mathbf{H}}(w_i - w_j)/\hat{f}_{\boldsymbol{W}}^{-i}(\boldsymbol{w_i})$ for conditional CDFs. We apply the standard Gaussian kernel of order $r = 2$. In line with Assumption 4, we restrict the support of Y' to wall within $[-3.7\hat{\sigma}_y, 3.7\hat{\sigma}_y]$. In case a filtered realization $\hat{y}'$ cannot be found numerically for some $i = 1,...,n$, we set $\hat{y}_i' \equiv y_i$.

We compare four estimators. The first one is the empirical conditional CDF, with smoothing only along X marginals. As demonstrated in the proof of Theorem 1, it is a first-line testing benchmark, which relates to the estimators used by Rothe (2010).

Secondly, we test two process-driven bandwidth selectors. After Silverman (1998), a simple rule of thumb, which assumes normality and mean-dependence of the underlying data generating processes and normal kernel, suggests that along marginals Y one should use $c_Y = c_0 = 1.59\hat{\sigma}_y$, for the unconditional and conditional distributions, respectively. At the same time, marginals X should use $c_i = 1.06\hat{\sigma}_{xi}$ for $i = \{1,...,d_{\boldsymbol{X}}\}$, where $\hat{\sigma}_x$ corresponds to an empirical dispersion measure of the relevant marginal.

A natural extension of the normal-reference rule, is to calculate the MSE-optimal constants for the process given in Eq. 22. It is readily seen that the constant for the uncodnitional CDF coincides with the normal-reference value. Li and Racine (2013) suggest that the integrated MSE of the smooth conditional CDF estimators in the bivariate setting can be approximated as

$$\mathrm{IMSE}(\hat{F}_{Y|X}) = A_0h_0^4 + A_1h_0^2h_1^2 + A_2h_1^4 + A_3(nh_1)^{-1}, \tag{24}$$

where except for A_1, all A_j's are positive constants, given by $A_0 = \int B_0(y|x)^2p(y,x)\mathrm{d}y\mathrm{d}x$, $A_1 = 2\int B_0(y|x)B_1(y|x)p(y,x)\mathrm{d}y\mathrm{d}x$, $A_2 = \int B_1(y|x)^2p(y,x)\mathrm{d}y\mathrm{d}x$ and $A_3 = \int Vr(y|x)p(y,x)\mathrm{d}y\mathrm{d}x$.

Conditional on data, terms B_0 and B_1 correspond to the estimate bias along Y and X marginals, respectively, Vr is the estimator variance given by $\kappa F_{Y|X}(1 - F_{Y|X})/f_X$, with $\kappa = \int K(w)^2 dw$. Additionally, $p(x, y)$ is the weight function which guarantees non-explosiveness of the IMSE in the tails. More specifically, we take $p(x, y)$ as trimming function taking values of 1 if argument falls within g standard deviations of the mean, and 0 otherwise. Substituting for Gaussian kernel and process properties from Eq. 22, one finds that

$$A_0 = \frac{g}{8\sqrt{\pi} d \sqrt{ag^2 + d}}, \tag{25}$$

$$A_1 = \frac{a\left(\sqrt{a}g\left(ag^2(a+4d) + 2d(a+2d)\right) + 2d\left(ag^2+d\right)^{3/2}\left(\log(ad) - 2\log\left(\sqrt{a\left(ag^2+d\right)} + ag\right)\right)\right)}{16\sqrt{\pi}d\left(a\left(ag^2+d\right)\right)^{3/2}}, \tag{26}$$

$$A_2 = \frac{ag\left(3a^4g^4 + 4a^3dg^2\left(2g^4 - 4g^2 + 1\right) + 4a^2d^2\left(10g^4 - 6g^2 + 1\right) + 8ad^3\left(7g^2 - 1\right) + 24d^4\right)}{32\sqrt{\pi}ad\left(ag^2+d\right)^{5/2}} + \frac{4\sqrt{a}d(3d-a)\left(ag^2+d\right)^{5/2}\left(\log(ad) - 2\log\left(\sqrt{a\left(ag^2+d\right)} + ag\right)\right)}{32\sqrt{\pi}ad\left(ag^2+d\right)^{5/2}}, \tag{27}$$

$$A_3 = \frac{e^{\frac{g^2}{2}}\left(\sqrt{2}ag - 2(a-2d)D_+\left(\frac{g}{\sqrt{2}}\right)\right)}{2\sqrt{\pi}\sqrt{d}}, \tag{28}$$

where D_+ is a Dawson function. Solving for h_0 and h_1 which minimize Eq. 24, one finds that the bandwidth constants for the process in Eq. 22, with $d = 1$, $a = 0.4$ and $g = 3$, are given by $c_0 = 2.16$ and $c_1 = 2.26$.

The above bandwidth selectors rely on regularity and availability of the data generating process. As the fourth alternative we propose a data-driven bandwidth selector, following closely Li and Racine (2013). In particular, Eq. 24 builds a natural benchmark for cross-validation as

$$CV = \int \text{IMSE}(\hat{F}_{Y|X}) f_X(x) \mathrm{d}y \mathrm{d}x. \tag{29}$$

Li and Racine (2013) propose choosing bandwidths by minimizing the CV objective function using the least-square method. The constants are then extracted by rescaling the CV-optimal bandwidths, denoted by c_0^{CV} and c_0^{CV}, by their respective rates of convergence, i.e.

$$\begin{aligned} c_0 &= c_0^{CV}/n^{-1/(4+d_{\boldsymbol{x}})}, \\ c_1 &= c_1^{CV}/n^{-1/(4+d_{\boldsymbol{x}})}. \end{aligned} \tag{30}$$

A similar approach is taken to calculate optimal bandwidths for the unconditional CDF, however, since the CV-optimal rates of converge coincide with uniform rates, there is no need to rescale the bandwidths.

We run 1000 Monte Carlo replications of process in Eq. (22) for different sample sizes. True density $f_Y(y)$ and filtered realizations y' are calculated using numerical methods. The medians and standard deviation intervals, aggregated over all the runs, are presented in Tables 1 and 2.

It can be readily observed that smoothing along the CDF marginal results in instant MSE gains, both in terms of accuracy and dispersion. Process-driven bandwidths demonstrate still better

Table 1: Median MSE of joint density estimated over independent counterfactual distributions.

Bandwidth selector	n=30	n=50	n=100	n=200	n=500	n=1000
no smoothing	9.462	6,424	3,819	2,252	1.17	0.688
rule of thumb	6.917	5,023	3,205	1,963	1,081	0.649
process-driven	5.341	3.851	2.407	1.511	0.859	0.527
LS-CV	4,829	3.590	2.139	1.217	0.676	0.393

Notes: Medians taken over 1000 Monte Carlo results of Eq. (23) for the ARCH process given in Eq. (22). For transparency, the values are scaled by a factor 10^4.

Table 2: MSE dispersion of joint density estimated over independent counterfactual distributions.

Bandwidth selector	n=30	n=50	n=100	n=200	n=500	n=1000
no smoothing	3,280	2,541	1,586	1,086	0.556	0.342
rule of thumb	3,054	2,277	1,467	1,022	0.533	0.331
process-driven	3,354	1,905	1,186	0.831	0.447	0.285
LS-CV	4,929	2,087	1,085	0.705	0.337	0.208

Notes: Standard deviations taken over 1000 Monte Carlo results of Eq. (23) for the ARCH process given in Eq. (22). For transparency, the values are scaled by a factor 10^4.

performance than the simple heuristics suggested by Silverman (1998), however, the best efficiency gains, in terms of MSE accuracy and dispersion, can be observed for data-driven bandwidths, confirming the common findings in the kernel density and distribution estimation (Li and Racine, 2008).

4 Empirical application

In order to show the practical application of the filtering procedure we choose the interest rate dynamics in the EA. In normal times, the changes in the bank funding rates are well represented in the retail lending rates as a consequence of arbitrage opportunities and market competition. In abnormal times, as a result of heterogeneous risk components and market segmentation, the changes in bank lending conditions do not have to correspond to the changes in the market funding rates. In the largely banking-dependent EA this can have a detrimental influence on the real economy as the ECB rate cuts are not fully passed to the bank lending rates in several member states and regions where accommodative monetary policy would be very welcome.

The factors which can hamper the effectiveness of monetary policy transmission to the bank lending rates include, *inter alia*, high level of sovereign debt, sluggish economic activity, insufficient banks' capital positions and high economic uncertainty (EIB, 2016). High levels of sovereign indebtedness, and therefore higher sovereign funding costs, are a sign of unsustainable fiscal policy. In case of a sovereign default the debtors who are government-dependent would find it more difficult to finance their obligations which could have a direct effect on banks. Similarly, in a slow-growing or stagnating economy the incomes and investments are stalled, making the debt more burdensome than it had been assumed when applying for a loan. Weak banking positions, i.e. low capital ratios or high riskiness of their balance sheets, limit the incentives to take additional risk by granting new loans as well as that could increase the counterparty risk in the interbank market. Those two effects put an upward pressure on the retail rates. Similarly, high economic uncertainty increases the risk that the debtor will not be able to repay its obligations which eventually makes banks demand higher yields on the loans they give.

In the EA the cross-country differences in the above-mentioned factors make monetary policy particularly difficult as the ECB controls only the area-wide instruments. Nevertheless, in the recent years the ECB took non-standard measures trying to alleviate the bank funding constraints in the countries which were did not fully admire the comfort of low-interest-rate environment and restore the market confidence in the whole euro area. Those included the creation of temporary rescues programs (in the forms of the European Financial Stabilization Mechanism, European Financial Stability Facility and European Stability Mechanism), the bond-purchase and long-term refinancing programs, creation of the Single Supervisory Mechanism and most recently the Quantitative Easing (QE) program.

In this study we focus to what extent the interest rate pass-through was distorted by country-specific sovereign characteristics. We choose four largest EA economies, i.e. Germany, France, Italy and Spain, as being representative of the EA's dynamics during the financial and sovereign debt crises. In particular, it has been observed that two latter countries suffered more during the crisis and consequently the monetary policy transmission channel was there less effective, than in the core EA countries (Darracq Paries et al., 2014).

We consider the Error Correction Model (ECM), which has been widely applied to evaluate the interest rate pass-through dynamics (see Darracq Paries et al. (2014)). We adapt the framework to match the filtering procedure described in Section 2. In particular, the goal of the exercise is to provide counterfactual realizations of country-specific bank lending rates as if there was no short-term effect from sovereign risk. By comparing counterfactuals with actual realizations one can approximate by how much sovereign risk escalated, or alleviated, the the pass-through rigidities over the short term in a given country.

We focus on the short-term effects for three main reasons. Firstly, temporary pass-through shocks, resulting, for instance, from a EA-wide financial disturbance or country-specific turbulence, should be, by definition, represented in the short-term dynamics rather than long-run run equilibrium. Consequently, short-term co-movement between the variables serves as a vital source of information for central bankers and practitioners. Secondly, short-run dynamics is of particular interest for policy makers, not only as a warning flag but it could also be directly influenced by policy actions. Additionally, since the filtering conditions require stationarity of the underlying processes, the ECM equation cannot be directly filtered out.

4.1 Estimation strategy

We begin by adjusting the parametric setup to our filtering procedure. The standard pass-through model, adjusted for the sovereign risk factors, can be written as

$$\Delta B_t = \sum_{k=0}^{lR} \beta_{kR} \Delta R_{t-k} + \sum_{j=1}^{lB} \beta_{jB} \Delta B_{t-k} + \sum_{m=1}^{lS} \beta_{mS} \Delta S_{t-m} + \alpha \nu_{t-1} + \varepsilon_t, \tag{31}$$

where B_t is a bank lending rate, R_t is the monetary policy reference rate, S_t is the sovereign risk component, ν_t is the error correction factor and ε_t is the standard error term. In such a specification coefficients β reflect the short-run pass-through effects, whereas α represents the speed of adjustment to the long-run equilibrium path given by

$$B_t = \mu_0 + \mu_R R_t + \mu_S S_t + \nu_t, \tag{32}$$

with μ_R and μ_S reflecting the long-run elasticity of bank lending rates to the reference rate and sovereign risk factors, respectively. Finally, subscript $t = 1, ..., T$ denotes the time dimension and Δ is the first difference operator.

To match the notation described in Section 2, let R_t^{lR} describe a $(lR+1)$-dimensional vector of lags given by $R_t^{lR} = \{R_{t-lR}, ..., R_t\}$ and similarly $B_t^{lB} = \{B_{t-lB}, ..., B_t\}$ and $S_t^{lS} = \{S_{t-lS}, ..., S_t\}$. Following Eq. (20), the counterfactual distribution of ΔB, denoted by $\Delta B'$, which is free from the short-term effects of ΔS, can be calculated from

$$f_{\Delta B'|.}(\Delta B_t | \Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \Delta S_{t-1}^{lS}, \nu_{t-1}) = f_{\Delta B|.}(\Delta B_t | \Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \nu_{t-1}). \tag{33}$$

Consequently, by applying Theorem 4, the counterfactual realizations $\Delta B_t'$ can be estimated by solving for each t

$$\frac{\int_{-\infty}^{\Delta B_t'} \hat{f}_{\Delta B,...}(u, \Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \nu_{t-1}) \mathrm{d}u}{\hat{f}_{\Delta R^{lR},...}(\Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \nu_{t-1})} = \frac{\int_{-\infty}^{\Delta B_t} \hat{f}_{\Delta B,...}(u, \Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \Delta S_{t-1}^{lS}, \nu_{t-1}) \mathrm{d}u}{\hat{f}_{\Delta R^{lR},...}(\Delta R_t^{lR}, \Delta B_{t-1}^{lB}, \Delta S_{t-1}^{lS}, \nu_{t-1})}. \tag{34}$$

4.2 Data description and results

The baseline relation, given in Eq. (34), focusses on the responsiveness of the bank lending rate to the policy reference rate and sovereign risk factors. In this study we focus on smaller loans, as being mostly associated with Small and Medium-size Enterprises (SMEs). Since SMEs find it more difficult to raise funds through capital markets, their dependence on the banking sector, and consequently on the cost of bank financing, is of particular importance for the recovery of the EA (EIB, 2016). Having pointed this out, we focus on the average interest rate on loans of up to and including EUR 1m of all maturities. As the reference policy rate we take 1-year EURIBOR rate, however, the results are largely robust to different lending and reference rates' maturities. Both time series come from the ECB. As a proxy for the sovereign risk we take 1-year Probabilities of Default (PDs) of a sovereign, calculated by Kamakura corporation. PDs are a composite measure of riskiness which include the dynamics of price- and quantity-based metrics, including fiscal positions and macroeconomic dynamics. They are further calibrated to match the empirically-observed default rates. Consequently, we consider PDs to better reflect the riskiness of a given sovereign, rather than purely price-based indices. The data sample is at a monthly frequency and covers the period from January 2000 until December 2016. The basic summary statistics, together with the stationarity results and optimal model lag selection, are depicted in Table 3.

Germany								
Variable	Obs.	Mean	Std. Dev.	Min	Max	ADF	ADF (Δ)	Lags
Bank lending rate	204	4.509	1.482	2.360	7.670	0.809	0.000	3
Sovereign risk	204	0.062	0.088	0.012	0.671	0.022	0.000	1
EURIBOR (1 year)	204	2.308	1.611	-0.080	5.393	0.923	0.000	3

France								
Variable	Obs.	Mean	Std. Dev.	Min	Max	ADF	ADF (Δ)	Lags
Bank lending rate	204	3.994	1.165	1.690	6.100	0.978	0.000	1
Sovereign risk	204	0.091	0.048	0.039	0.309	0.044	0.000	1
EURIBOR (1 year)	204	2.308	1.611	-0.080	5.393	0.923	0.000	3

Italy								
Variable	Obs.	Mean	Std. Dev.	Min	Max	ADF	ADF (Δ)	Lags
Bank lending rate	204	4.393	1.087	2.230	6.680	0.941	0.000	3
Sovereign risk	204	0.250	0.159	0.106	0.995	0.000	0.000	1
EURIBOR (1 year)	204	2.308	1.611	-0.080	5.393	0.923	0.000	1

Spain								
Variable	Obs.	Mean	Std. Dev.	Min	Max	ADF	ADF (Δ)	Lags
Bank lending rate	204	4.587	0.979	2.440	6.570	0.967	0.000	3
Sovereign risk	204	0.248	0.153	0.071	0.722	0.063	0.000	1
EURIBOR (1 year)	204	2.308	1.611	-0.080	5.393	0.923	0.000	1

Table 3: Summary statistics of the data used in the empirical analysis. Bank lending rate is the average rate given on bank loans (new businesses) of up to and including EUR 1m, sovereign risk is approximated by 1-year probabilities of default. ADF and ADF (Δ) denote the p-values from the Augmented Dickey-Fuller test on levels and first differences, respectively. Optimal number of lags is chosen by the Bayesian information criterion. Source: ECB and Kamakura.

It can be readily observed that the average lending rates were the highest in Spain, followed by Germany, Italy and France. Possibly surprising levels in Germany are mostly driven by early-2000 period. In fact, excluding pre-2004 period the lending rates were around 3.91 pp. and were much closer to the French levels. Interestingly, the volatility of the bank lending rates was lower in Italy and Spain than in Germany or France. This result, in fact, builds an avenue which we further explore in the filtering exercise as it confirms that the responsiveness of the French and German rates to the reference rate was indeed larger than in the case of Italy or Spain. After the interest rate cut in 2008 and further non-conventional monetary measures introduced by the ECB, the lending rates in the core Member States (MS) adjusted much faster than in the vulnerable MS.

In fact, only recently the lending rates on small loans in Italy and Spain started converging to the core MS levels.

The sovereign PDs depict that the sovereign risk in France and Germany was at negligible level of below 10 bps. To the contrary, the risk in Italy and Spain was nearly three times higher, on average, suggesting substantial tensions in the sovereign sectors.

We test the stationarity of the time series with the Augmented Dickey-Fuller test (Fuller, 1995). The results suggest that both interest rates are $I(1)$, whereas the PDs are stationary in levels at conventional significance levels. This indeed can create identification problems as there is weak evidence for cointegration between the variables. Looking at the results in Table 4 there appears to be no cointegration between model variables in France, Italy and Spain.

	Germany	France	Italy	Spain
ADF	0.006	0.290	0.261	0.108

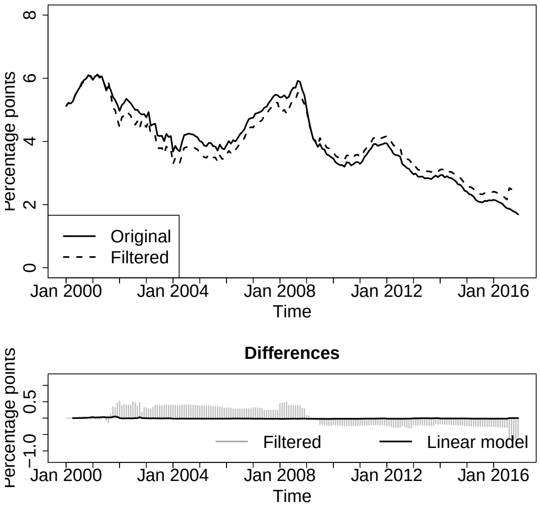
Table 4: Cointegration between the model variables. ADF denotes the p-values from the Augmented Dickey-Fuller test on the residuals from the cointegrating relation given by Eq. (32).

There seems to be no need to estimate the ECM form of the pass-through model as the standard Auto-Regressive Distributed Lag (ARDL) model is correctly specified. Therefore, to guarantee the robustness of the results, we study both specifications, i.e. with and without the ECM component. However, since the results for both are very close to each other, for transparency reasons, the results for the ARDL model are given in Appendix B.

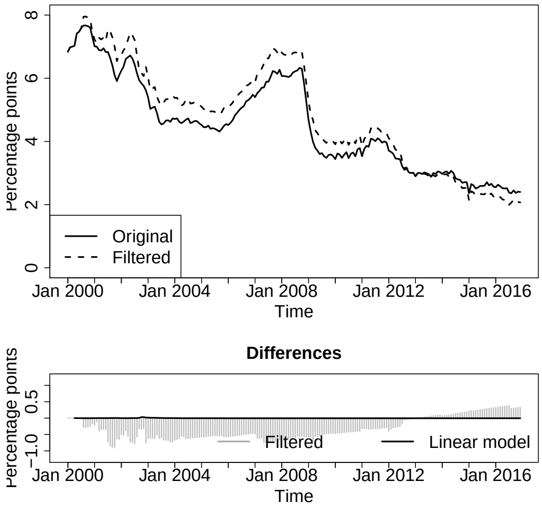
In the filtering procedure we use the multivariate Gaussian kernel. Optimal bandwidths are calculated using the least squares cross validation technique, as indicated in Section 2. We estimate Eq. (34) on the time series standardized by a normal transformation and we scale the filtered values by the moments of the original distribution. We then convert the filtered difference estimates into levels. The effects of the sovereign risk factors on the bank lending rates is taken as a difference between the observed and filtered, i.e. counterfactual, bank lending rates. We additionally compare the sovereign risk contributions with their linear equivalents, calculated from Eq. (31) by setting $\beta_{mS} = 0$ for $m = 1, ..., lS$. The results are presented in Figure 1.

Several interesting features can be readily observed. Firstly, the sovereign risk substantially disturbed the interest rate pass-through in Italy and Spain, compared to France and Germany. On average, the retail lending rates would have been lower by 3 bp in France and higher by 32bp in Germany if we filter out the sovereign risk transmission effects. To the contrary, in Italy they could have been lower by as much as 61bp, and by around 31bp in Spain. This is in fact a confirmation of asymmetric interest rate transmission in various EA jurisdictions (Arnold and van Ewijk, 2014; Darracq Paries et al., 2014).

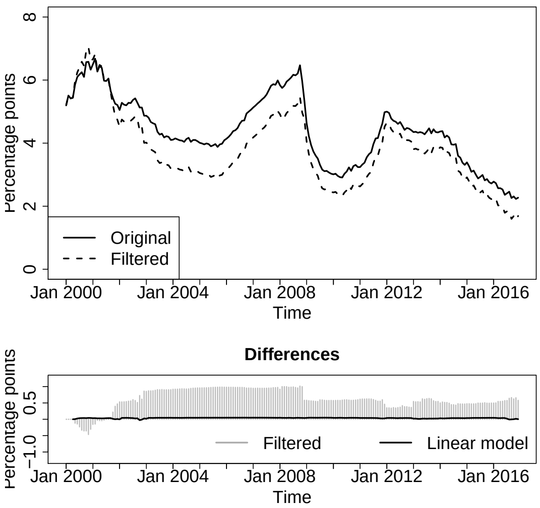
Secondly, we observe a change in sovereign risk contributions around the period of the global financial crisis and sovereign debt crisis, and in the most recent years. In 2008 there was a visible negative effect around 2008-2009 in France, Germany and Italy. In Spain one could observe a modest upward dynamics around the same period. These findings are in line with the findings presented by EIB (2016). Between 2012 and 2013 one can observe a positive sovereign risk transmission in Italy and Spain, and evaporation of the effects in Germany. French transmission channel



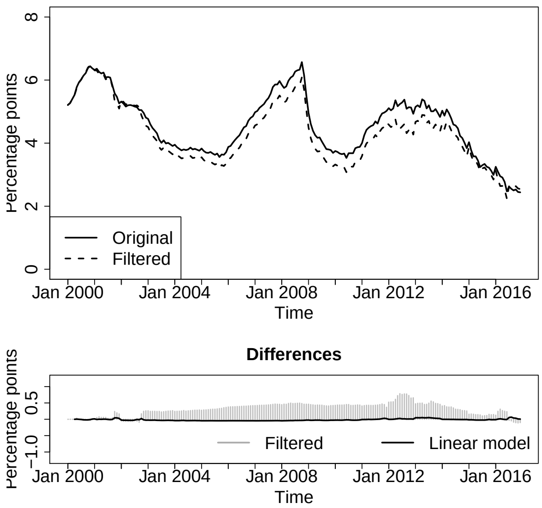
(a) France



(b) Germany



(c) Italy



(d) Spain

Figure 1: Short-term sovereign risk transmission into bank lending rates in (a) France, (b) Germany, (c) Italy and (d) Spain. Filtering equation includes the error correction term. Top figures in each panel compare the realized against the filtered bank lending rates. Bottom figures show the differences between the realized and filtered rates and compare them against the linear model.

was roughly stable around that period. Interestingly, in 2016 one observe that the sovereign risk transmission decreased in Spain but remained elevated in Italy. There was also a disconnect between French and German risk effects, with the former going into more negative territories of as much as 65bp and the latter gradually increasing and as of the end of 2016 reaching around 30bp. This indeed signals the heterogeneity among the sovereign tensions across the MS (von Borstel et al., 2016) and a shift of risk transmission to Germany which might be viewed as a fiscal backstop of the EA.

Thirdly, one can clearly observe a difference between the dependence filtering approach and linear specification. The latter appears to substantially underestimate the short-term sovereign risk transmission mechanisms. Although in the standard ECM pass-through model the risk effects seem to affect the lending rates mostly through the long-term channel, the results for the ARDL model confirm negligible linear effects (see Appendix B). Since the filtering approach doesn't assume any underlying model parametrization, it seems that the sovereign risk transmission is highly nonlinear in nature.

5 Conclusions

This study develops a novel dependence filtering framework. Under mild regularity conditions, and without assuming any specific parametric structure, we derive independent and conditionally independent distributions, filtering out the effects of given covariates. We show that the filtering methodology has desired asymptotic properties and can be easily extended to a multivariate setting. We additionally confirm the accuracy of the method numerically, on an example of an ARCH process.

In an empirical application, we demonstrate that the proposed framework can be easily adjusted to match standard parametric regressions. In particular, we look at the standard interest rate pass-through model, widely used by central bankers (Darracq Paries et al., 2014), and we demonstrate that the filtering approach can be used to calculate counterfactual distribution of bank lending rates which is free from the sovereign risk distortions.

The main findings confirm the heterogeneity in the ECB interest rate pass-through between France, Germany, Italy and Spain. Interestingly, we find that the standard linear models can underestimate the short-term sovereign risk distortions, especially for Germany in 2016. We speculate that the elevated risk levels of Germany can coincide with the implicit fiscal backstop role it plays in the EU. We additionally confirm elevated sovereign risk pass-through in Italy, with negligible effects in France and convergence in Spain.

The results build an interesting avenue for policy makers and practitioners. With non-responsive retail rates the standard monetary policy actions are largely ineffective. The fact that the ECB operates near to the zero-lower bound environment makes the bank lending only more difficult to stimulate as the conventional tools run out of scope. Strategies which could improve the effectiveness of the transmission channels should gain more attention. These could include policies aiming at restoring market confidence, cross-border risk-sharing and fostering financial integration in the EA.

Furthermore, this study highlights the nonlinear structure of the sovereign risk transmission to the bank lending rates. It therefore contributes to the ongoing discussion on the nature of risk spill-overs between financial variables (Apergis and Cooray, 2015; Huang et al., 2010; He and Krishnamurthy, 2012). Although the sovereign distortions substantially elevated the bank lending rates in Spain and Italy during times of distress, they affected the lending rates in France and Germany to a different extent and in a way which was not captured by linear specification.

The filtering approach we develop in this study shows only a top of an iceberg in terms of the nature of sovereign risk spill-overs. To get a better understanding of the influence of risk factors on bank-specific characteristics and lending, further research is therefore desired. In particular, one could extrapolate the filtering methodology to a panel framework or fit the pass-through model to a different set of risk proxies, controlling for, for instance, macroeconomic and financial risks. Furthermore, the filtering framework can be supported by an accuracy score or confidence bounds, which would make it a more attractive approach in practice.

References

N. Apergis and A. Cooray. Asymmetric interest rate pass-through in the u.s., the u.k. and australia: New evidence from selected individual banks. *Journal of Macroeconomics*, 45:155 – 172, 2015. doi: 10.1016/j.jmacro.2015.04.010.

I. J. Arnold and S. E. van Ewijk. A state space approach to measuring the impact of sovereign and credit risk on interest rate convergence in the euro area. *Journal of International Money and Finance*, 49, Part B:340 – 357, 2014. doi: 10.1016/j.jimonfin.2014.04.005.

V. Chernozhukov, I. Fernndez-Val, and B. Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013. doi: 10.3982/ECTA10582.

E. Choi and P. Hall. Data sharpening as a prelude to density estimation. *Biometrika*, 86(4): 941–947, 1999. doi: 10.1093/biomet/86.4.941.

M. Darracq Paries, D. Moccero, E. Krylova, and C. Marchini. The retail bank interest rate pass-through: The case of the euro area during the financial and sovereign debt crisis. Technical Report no 154, European Central Bank, 2014.

C. Diks and V. Panchenko. A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30(9–10):1647–1669, 2006. doi: 10.1016/j.jedc.2005.08.008.

C. Diks and M. Wolski. Nonlinear Granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics*, 31(7):1333–1351, 2016. doi: 10.1002/jae.2495.

EIB. *Investment and Investment Finance in Europe.* European Investment Bank, Luxembourg, 2016.

F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu. Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, 140(2):335 – 352, 2010. doi: 10.1016/j.jspi.2009.07.019.

W. A. Fuller. *Introduction to Statistical Time Series (Wiley Series in Probability and Mathematical Statistics)*. Wiley-interscience: New York, 1995. doi: 10.1002/9780470316917.

P. J. Gertler, S. Martinez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch. *Impact Evaluation in Practice*. World Bank, Washington, D.C., 2011.

E. Giné and R. Nickl. Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields*, 141(3):333–387, 2008. doi: 10.1007/s00440-007-0087-9.

P. Hall and M. C. Minnotte. High order data sharpening for density estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(1):141–157, 2002. doi: 10.1111/1467-9868.00329.

Z. He and A. Krishnamurthy. A macroeconomic framework for quantifying systemic risk. Working Paper 233, National Bank of Belgium, 2012.

C. Hiemstra and J. D. Jones. Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, 49(5):1639–1664, 1994. doi: 10.2307/2329266.

S. Hoderlein and E. Mammen. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, 75(5):1513–1518, 2007. doi: 10.1111/j.1468-0262.2007.00801.x.

N. Hristov, O. Hulsewig, and T. Wollmershauser. The interest rate pass-through in the euro area during the global financial crisis. *Journal of Banking & Finance*, 48:104 – 119, 2014. doi: 10.1016/j.jbankfin.2014.08.004.

X. Huang, H. Zhou, and H. Zhu. Systemic risk contributions. Working Paper 60, Bank for International Settlements, 2010.

G. W. Imbens and W. K. Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009. doi: 10.3982/ECTA7108.

M. C. Jones. Estimating densities, quantiles, quantile densities and density quantiles. *Ann Inst Stat Math*, 44(4):721–727, 1992. doi: 10.1007/BF00053400.

T. Kitamura, I. Muto, and I. Takei. Loan interest rate pass-through and changes after the financial crisis: Japans evidence. *Journal of the Japanese and International Economies*, 42:10 – 30, 2016. doi: 10.1016/j.jjie.2016.10.002.

Q. Li and J. S. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton and Oxford, 2007.

Q. Li and J. S. Racine. Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26(4):423–434, 2008. doi: 10.1198/073500107000000250.

Q. Li and J. S. Racine. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics*, 31(1):57–65, 2013. doi: 10.1080/07350015.2012.738955.

A. Montagnoli, O. Napolitano, and B. Siliverstovs. Regional interest rate pass-through in italy. *Regional Studies*, 50(8):1404–1419, 2016. doi: 10.1080/00343404.2015.1022311.

A. Perera and J. Wickramanayake. Determinants of commercial bank retail interest rate adjustments: Evidence from a panel data model. *Journal of International Financial Markets, Institutions and Money*, 45:1 – 20, 2016. doi: 10.1016/j.intfin.2016.05.006.

C. Rothe. Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155 (1):56 – 70, 2010. doi: 10.1016/j.jeconom.2009.09.001.

C. Rothe. Partial distributional policy effects. *Econometrica*, 80(5):2269–2301, 2012.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics & Applied Probability, Chapman and Hall/CRC: London – New York, 1998.

E. Stanislawska. Interest rate pass-through in poland: Evidence from individual bank data. *Eastern European Economics*, 53:324, 2015. doi: 10.1080/00128775.2015.1033362.

A. Torgovitsky. Identification of nonseparable models with general instruments. Working paper, Yale University, 2011.

A. W. van der Vaart. Weak convergence of smoothed empirical processes. *Scandinavian Journal of Statistics*, 21(4):501–504, 1994.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 2000.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

A. W. van der Vaart and J. A. Wellner. Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, 55:234–252, 2007.

J. von Borstel, S. Eickmeier, and L. Krippner. The interest rate pass-through in the euro area during the sovereign debt crisis. *Journal of International Money and Finance*, 68:386 – 402, 2016. doi: 10.1016/j.jimonfin.2016.02.014.

M. P. Wand and M. C. Jones. *Kernel Smoothing*. Monographs on Statistics & Applied Probability, Chapman and Hall: New York, 1995.

A Proofs of theorems and lemmas

A.1 Theorem 1

Proof. Let us write that the filtered sample Y_i' is defined as a function of Y_i conditional on realization of $\boldsymbol{X}_i = \boldsymbol{x}_i$, for each $i = 1, ..., n$, such that $Y_i' = \phi(Y_i|\boldsymbol{X}_i = \boldsymbol{x}_i) \equiv \phi(Y_i)$. Let us further assume that ϕ is invertible. Transformation $\phi : \mathbb{R} \to \mathbb{R}$ might be viewed as a perturbation, or following Choi and Hall (1999) and Hall and Minnotte (2002), a sharpening function. Consequently, the plug-in estimator of the Left-Hand Side (LHS) of Eq. (1), which is defined over Y' marginal, can be rewritten as

$$\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}\left(y - \phi(Y_i), \boldsymbol{x} - \boldsymbol{X_i}\right), \tag{A.1.1}$$

where $\mathbf{H}$ is a $(1 + d_X) \times (1 + d_X)$ bandwidth matrix and $K_{\mathbf{H}}$ is a scaled multivariate kernel function satisfying the standard regularity conditions (Wand and Jones, 1995; Silverman, 1998). For presentation, in notation we explicitly distinguish between variables Y and $\boldsymbol{X}$ as sharpening is defined over the Y marginal only. Consequently, let us assume that bandwidth matrix consists of two blocks: one for marginal Y and one for $\boldsymbol{X}$.

By taking the expected value of the estimator from Eq. (A.1.1) with respect to the joint distribution of $(Y, \boldsymbol{X})$, and substituting into Eq. (1) we get

$$\int K_{\mathbf{H}}\left(y - \phi(Y), \boldsymbol{x} - \boldsymbol{X}\right) \mathrm{d}F(Y, \boldsymbol{X}) = f_Y(y) f_{\boldsymbol{X}}(\boldsymbol{x}), \tag{A.1.2}$$

as the sample size $n \to \infty$. By substituting $V = \phi(Y)$

$$\int K_{\mathbf{H}}\left(y - V, \boldsymbol{x} - \boldsymbol{X}\right) \mathrm{d}F(\phi^{-1}(V), \boldsymbol{X}) = f_Y(y) f_{\boldsymbol{X}}(\boldsymbol{x}). \tag{A.1.3}$$

Changing variables we observe that

$$\int K\left(u, \boldsymbol{z}\right) g\left(y - \mathbf{H}_Y^{1/2} u \middle| \boldsymbol{x} - \mathbf{H}_{\boldsymbol{X}}^{1/2} \boldsymbol{z}\right) \mathrm{d}\boldsymbol{z}\mathrm{d}u = f_Y(y) f_{\boldsymbol{X}}(\boldsymbol{x}). \tag{A.1.4}$$

where $\mathbf{H}_Y$ and $\mathbf{H}_X$ are the block matrices associated with marginals Y and $\boldsymbol{X}$, respectively. We also defined $g = \partial G / \partial a$ where

$$G(a|\boldsymbol{b}) = \int_{-\infty}^{a} f_{Y,\boldsymbol{X}}(\phi^{-1}(u), \boldsymbol{b}) \mathrm{d}u. \tag{A.1.5}$$

Assuming g is sufficiently smooth, the LHS of Eq. (A.1.4) is equal $g(y|\boldsymbol{x}) + o(\mathrm{tr}(\mathbf{H}))$. By substituting it back into Eq. (A.1.4) and integrating, we conclude that $G + o(\mathrm{tr}(\mathbf{H})) = D$ where

$$D(y|\boldsymbol{x}) = F_Y(y) f_{\boldsymbol{X}}(\boldsymbol{x}). \tag{A.1.6}$$

Combining it with Eq. (A.1.5), up to the order of $o(\mathrm{tr}(\mathbf{H}))$, we get

$$\int_{-\infty}^{y} f_{Y,\boldsymbol{X}}(\phi^{-1}(u), \boldsymbol{x}) \mathrm{d}u = D(y|\boldsymbol{x}). \tag{A.1.7}$$

Denote LHS of Eq. (A.1.7) as $J(\phi^{-1}(y)|\boldsymbol{x})$. Consequently, one finds that $\phi(y) = D^{-1}(J(y|\boldsymbol{x})|\boldsymbol{x})$ or

$$\begin{aligned} \phi(y) &= D^{-1}\left(\int_{-\infty}^{y} f_{Y,\boldsymbol{X}}(u, \boldsymbol{x}) \mathrm{d}u\right) \\ &= F_Y^{-1}\left(\frac{\int_{-\infty}^{y} f_{Y,\boldsymbol{X}}(u, \boldsymbol{x}) \mathrm{d}u)}{f_{\boldsymbol{X}}(\boldsymbol{x})}\right), \end{aligned} \tag{A.1.8}$$

or alternatively, by substituting $y' \equiv \phi(y)$,

$$F_Y(y') = F_{Y|\boldsymbol{X}}(y|\boldsymbol{x}). \tag{A.1.9}$$

□

A.2 Lemma 1

Proof. It can be observerd that the sharpening transformation is defined over a random variable Y and let us define the functional representation of dependency between Y and $\boldsymbol{X}$ in terms of exponent of the Shannon's entropy around points $(Y, \boldsymbol{x})$, as suggested by $\delta(Y, \boldsymbol{x})$. We have

$$\begin{aligned} F_{Y'}(y') &= \mathrm{P}(F_Y^{-1}(F_{Y|\boldsymbol{X}}(Y|\boldsymbol{x}) \leq y') \\ &= \mathrm{P}(F_{Y,\boldsymbol{X}}(Y, \boldsymbol{x}) \leq F_Y(y')F_{\boldsymbol{X}}(\boldsymbol{x})) \\ &= \mathrm{P}\left(F_Y(Y) \leq F_Y(y')\delta(Y, \boldsymbol{x})\right) \\ &= \mathrm{P}\left(Y \leq F_Y^{-1}F_Y(y')\delta(Y, \boldsymbol{x})\right) \\ &= F_Y(y')\delta(Y, \boldsymbol{x}), \end{aligned} \tag{A.2.1}$$

where the last equivalence follows from the involutory nature of inverse functions. □

A.3 Theorem 2

Proof. To derive the assymptotic properties of the process indexed by Eq. (13), let us introduce the basic concepts from the (smooth) empirical process literature. Therefore, let $Z_1, ..., Z_n$ be an i.i.d. random elements in a measureable space $(\mathcal{Z}, \mathcal{A})$ with law (distribution) P, and for a measureable function $g : \mathcal{Z} \to \mathbb{R}$ let the expectation, the empirical measure and empirical process at g be denoted by

$$Pg = \int g \mathrm{d}P, \qquad \mathbb{P}_n g = \frac{1}{n}\sum_{i=1}^{n} g(Z_i), \qquad \mathbb{G}_n g = \sqrt{n}(\mathbb{P}_n - P)g \tag{A.3.1}$$

where μ_n is a sequence of probability measures converging weakly to zero. Following van der Vaart (1994), in our setup μ_n is random and satisfies $\mathrm{d}\mu_n = K_h$.

A.3.1 Empirical measures

Let us firstly consider empirical measures, i.e. not smoothed, of F_Y and $F_{Y|\boldsymbol{X}}$. Consequently, in the notation in this section we treat the estimators as defined over non-convoluted space. Let us write

$$\sqrt{n}(\hat{Y}' - Y') = \sqrt{n}(\hat{F}_Y^{-1} \circ F_{Y|\boldsymbol{X}} - \hat{F}_Y^{-1} \circ F_{Y|\boldsymbol{X}}). \tag{A.3.2}$$

We observe that $\hat{F}_{Y|\boldsymbol{X}}$ can be viewed as a pseudo-observation, i.e. an estimate rather than point realization of the function. After van der Vaart and Wellner (2007), we note that $F_Y^{-1} \circ F_{Y|\boldsymbol{X}}$ is Donsker under Assumption 3, and we can decompose Eq. (A.3.2)

$$\begin{aligned} \sqrt{n}(\hat{F}_Y^{-1} \circ \hat{F}_{Y|\boldsymbol{X}} - F_Y^{-1} \circ F_{Y|\boldsymbol{X}}) &= \mathbb{G}_n(F_Y^{-1} \circ \hat{F}_{Y|\boldsymbol{X}} - F_Y^{-1} \circ F_{Y|\boldsymbol{X}}) + \mathbb{G}_n F_Y^{-1} \circ F_{Y|\boldsymbol{X}} \\ &+ \sqrt{n}P(F_Y^{-1} \circ \hat{F}_{Y|\boldsymbol{X}} - F_Y^{-1} \circ F_{Y|\boldsymbol{X}}) \\ &= A1 + A2 + A3 \end{aligned} \tag{A.3.3}$$

Before handling Eq. (A.3.3) it is useful to remind that under Assumptions 1-4, both empirical quantile function and empirical conditional distribution functions are uniformly consistent, i.e.

$$\sup_{y\in\mathbb{R}}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\hat{F}_{Y|\boldsymbol{X}}(y|\boldsymbol{x})-F_{Y|\boldsymbol{X}}(y|\boldsymbol{x})|=O_P\left(\left(\frac{\log n}{nh^{d_{\boldsymbol{X}}}}\right)^{1/2}+h^r\right), \tag{A.3.4}$$

$$\sup_{w\in J}|\hat{F}_Y^{-1}(w)-F_Y^{-1}(w)|=O_P\left(\left(\frac{\log n}{n}\right)^{1/2}\right), \tag{A.3.5}$$

where J is a compact subset of $\mathbb{R}$ (Li and Racine, 2007).

To deal with term $A1$ we observe that, by Assumption 4 (ii), if F_Y^{-1} is uniformly bounded by $[a,b]=[F_Y^{-1}(p_1)-\epsilon, F_Y^{-1}(p_2)+\epsilon]$ for some $\epsilon>0$ where $0<p_1<p_2<1$, and that F_Y^{-1} is uniformly Lipschitz, i.e.

$$|F_Y^{-1}(r_1)-F_Y^{-1}(r_2)|\leq C|r_1-r_2|, \tag{A.3.6}$$

for every pair of points $r_1,r_2\in[p_1,p_2]$ and a constant $C>0$. Following van der Vaart and Wellner (1996) (page 157), functional F_Y^{-1} has finite uniform entropy integral and by Theorem 3.1 of van der Vaart and Wellner (2007)

$$\sup_{y\in J}|\mathbb{G}_n(F_Y^{-1}\left(\hat{F}_{Y|\boldsymbol{X}}\right)-F_Y^{-1}\left(F_{Y|\boldsymbol{X}}\right))|\xrightarrow{p}0, \tag{A.3.7}$$

in $\ell^\infty(J,\mathbb{R})$ where $J\subset\mathbb{R}$ for which $F_{Y|\boldsymbol{X}}$ is uniformly bounded by $[p_1,p_2]$.

To handle terms $A2$ and $A3$ we note that the quantile function is defined over distribution F_Y, i.e. for $\Gamma_Y\equiv F_Y^{-1}$, the empirical quantile function is given by

$$\Gamma_Y(\hat{F}_Y(\tau))=\inf\{y:\hat{F}_Y(y)\geq\tau\}, \tag{A.3.8}$$

with $\Gamma:J\to[p_1,p_2]$. Term $A2$ is the usual empirical process for i.i.d. one-dimensional transformation of a random variable $F_{Y|\boldsymbol{X}}(Y_1|\boldsymbol{x}),...,F_{Y|\boldsymbol{X}}(Y_n|\boldsymbol{x})$ and it converges weakly by Eq. (A.3.5). Since by Assumption 4 Γ_Y is Hadamard differentiable at F_Y tangentially to $S[a,b]$, where $S[a,b]$ is the set of continuous functions on $[a,b]$, by the distributive property of considered functions, functional delta method and continuous mapping theorem, respectively, we have that

$$\begin{aligned}\mathbb{G}_nF_Y^{-1}\circ F_{Y|\boldsymbol{X}}&=\sqrt{n}(\hat{F}_Y^{-1}-F_Y^{-1})\circ F_{Y|\boldsymbol{X}}\\&=\sqrt{n}(\Gamma_Y(\hat{F}_Y)-\Gamma_Y(F_Y))\circ F_{Y|\boldsymbol{X}}\\&=\Gamma'_Y(\sqrt{n}(\hat{F}_Y-F_Y))\circ F_{Y|\boldsymbol{X}}+o_P(1)\\&=\Gamma'_Y(\mathbb{G}_{n,F_Y})\circ F_{Y|\boldsymbol{X}}+o_P(1)\\&\xrightarrow{d}\Gamma'_Y(\mathbb{G}_{F_Y})\circ F_{Y|\boldsymbol{X}},\end{aligned} \tag{A.3.9}$$

where $\mathbb{G}_{F_Y}$ is the F_Y-Brownian bridge, defined as $\mathbb{G}_{F_Y}=\mathbb{G}_\lambda\circ F_Y$ where $\mathbb{G}_\lambda$ is the standard Brownian bridge process, and convergence is in $\ell^\infty(J)$. As a result we have that

$$A2\xrightarrow{d}N\left(0,\frac{F_{Y|\boldsymbol{X}}(1-F_{Y|\boldsymbol{X}})}{f_Y^2(F_Y^{-1}(F_{Y|\boldsymbol{X}}))}\right), \tag{A.3.10}$$

and convergence is in $\ell^\infty(J,\mathbb{R})$.

Following similar reasoning for the term $A3$, we have

$$\begin{aligned}\sqrt{n}P(F_Y^{-1}(\hat{F}_{Y|\boldsymbol{X}}) - F_Y^{-1}(F_{Y|\boldsymbol{X}})) &= P\left[\sqrt{n}\left(\Gamma_Y\left(F_Y(\hat{F}_{Y|\boldsymbol{X}})\right) - \Gamma_Y\left(F_Y(F_{Y|\boldsymbol{X}})\right)\right)\right] + o_P(1)\\ &= P\left[-\Gamma'_Y\left(\sqrt{n}(\hat{F}_{Y|\boldsymbol{X}} - F_{Y|\boldsymbol{X}})\right)\right] + o_P(1)\\ &= P\left[-\Gamma'_Y\left(\mathbb{G}_{n,F_{Y|\boldsymbol{X}}}\right)\right] + o_P(1)\end{aligned} \tag{A.3.11}$$

where we exploited the fact that $\mathrm{d}/\mathrm{d}F_{Y|\boldsymbol{X}}\Gamma_Y(F_Y) = -\Gamma'_Y$. We further build an argument along the lines of Corollaries 5.1 and 5.3 of van der Vaart and Wellner (2007). By Hadamard differentiability assumption we remind that

$$P\left[-\Gamma'_Y\left(\mathbb{G}_{F_{Y|\boldsymbol{X}}}\right)\right] = L(\sqrt{n}(\eta_n - \eta)), \tag{A.3.12}$$

where L is continuous and linear map and

$$\eta_n - \eta \equiv \eta_n(y, \boldsymbol{x}) - \eta(y, \boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{I}\{Y_i \leq y\}\hat{\theta}(x, X_i) - F_{Y|\boldsymbol{X}}(y|x)\right). \tag{A.3.13}$$

In line with Lemma 4.2 from van der Vaart and Wellner (2007), we have that

$$\begin{aligned}L(\sqrt{n}(\eta_n - \eta)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L\left(\mathbb{I}\{Y_i \leq y\}\hat{\theta}(x, X_i) - F_{Y|\boldsymbol{X}}(y|x)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L\left(\mathbb{I}\{Y_i \leq y\}\hat{\theta}(x, X_i)\right) - \sqrt{n}L(\eta)\\ &= \mathbb{G}_\eta L\left(\mathbb{I}\{Y_i \leq y\}\hat{\theta}(x, X_i)\right)\end{aligned} \tag{A.3.14}$$

Substituting back to the original notation, and expanding function Γ'_Y, we receive

$$P\left[-\Gamma'_Y\left(\mathbb{G}_{F_{Y|\boldsymbol{X}}}\right)\right] = \frac{\mathbb{G}_{F_{Y|\boldsymbol{X}}} E\left[\mathbb{I}\{Y \leq y\}\hat{\theta}(x, X)\right]}{f_Y\left(F_Y^{-1}\left(F_{Y|\boldsymbol{X}}(y|x)\right)\right)} \tag{A.3.15}$$

and convergence is in $\ell^\infty(J, \mathbb{R})$. By the properties of a Brownian bridge it follows that

$$A3 \xrightarrow{d} N\left(0, \frac{\hat{F}_{Y|\boldsymbol{X}}(1 - \hat{F}_{Y|\boldsymbol{X}})}{f_Y^2(F_Y^{-1}(F_{Y|\boldsymbol{X}}))}\right). \tag{A.3.16}$$

Putting the results together we find that $\sqrt{n}(\hat{Y}' - Y)$ is asymptotically tight for the empirical (non-smoothed) estimators.

A.3.2 Smooth empirical measures

To extend the results to the convoluted probability spaces, let us firstly remind that, in the current setup, the probability law $\mathbb{P}$ has a density with respect to the Lebesgue measure. It follows that the smooth empirical measure and smooth empirical process indexed at g as

$$\tilde{\mathbb{P}}_n g = \mathbb{P}_n * \mu_n(g), \qquad \tilde{\mathbb{G}}_{\mu_n} g = \sqrt{n}(\mathbb{P}_n - P) * \mu_n(g), \tag{A.3.17}$$

where μ_n is a sequence of probability measures converging weakly to zero. Following van der Vaart (1994), in our setup μ_n is random and satisfies $\mathrm{d}\mu_n(\boldsymbol{w}) = K_{\mathbf{H}}(\boldsymbol{w})\mathrm{d}\boldsymbol{w}$. Since convoluted estimates converge in MSE, conditions in Eqs. (A.3.4) and (A.3.5) are also satisfied for smooth estimates.

It is useful to restate the uniform convergence rates for the smooth empirical measures, as

$$\sup_{y\in\mathbb{R}}\sup_{\boldsymbol{x}\in\mathbb{R}^d} |\hat{F}_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - F_{Y|\boldsymbol{X}}(y|\boldsymbol{x})| = O_P\left(\left(\frac{\log n}{nh^{d_{\boldsymbol{X}}}}\right)^{1/2} + h^r + h_Y^r\right), \tag{A.3.18}$$

$$\sup_{w\in J} |\hat{F}_Y^{-1}(w) - F_Y^{-1}(w)| = O_P\left(\left(\frac{\log n}{n}\right)^{1/2} + h_Y^r\right). \tag{A.3.19}$$

Following the argument for Theorem 1 of Rothe (2010), we observe that the classes of functions $\mathcal{F}_1$ and $\mathcal{F}_2$, given by $\mathcal{F}_1 = \{F_Y(y) : y \in \mathbb{R}\}$ and $\mathcal{F}_2 = \{F_{Y|\boldsymbol{X}}(y, .) : y \in \mathbb{R}\}$ are r-times differentiable by Asssumption 4. Hence, $\mathcal{F}_1 \in C^r(J)$ and $\mathcal{F}_2 \in C^r(J, \mathbb{R})$ with bounded derivatives up to order r. Class $\mathcal{F}$ is Donsker by definition since $d_Y = 1$ in our setting, and class $\mathcal{F}_2$ is Donsker for $r > (d_{\boldsymbol{X}})/2$, which is satisfied as Assumption 3 implies that $r > d_{\boldsymbol{X}}$, as $\mathbf{H}_Y$ vanishes faster than $\mathbf{H}_{\boldsymbol{X}}$. In fact, this condition allows for application of the usual second order Kernels in the bi-variate settings with $d_Y = d_{\boldsymbol{X}} = 1$.

Since the function $F_Y^{-1} \circ F_{Y|\boldsymbol{X}}$ is Donsker, and together with Assumptions 1-3, by Theorem 5 of Giné and Nickl (2008) we have that $\tilde{\mathbb{G}}_{\mu_n} F_{Y|\boldsymbol{X}}$ converges in probability in $\ell^\infty(J, \mathbb{R})$ to the $F_{Y|\boldsymbol{X}}$-Brownian bridge. As a consequence, all the implications from Eqs. (A.3.7), (A.3.9) and (A.3.15) hold for convoluted probability spaces. In particular we have for the term $A3$ we have

$$A3 = \frac{\mathbb{G}_{F_{Y|\boldsymbol{X}}} E\left[\bar{K}_{H_y}(y - Y)\hat{\theta}(x, X)\right]}{f_Y\left(F_Y^{-1}\left(F_{Y|\boldsymbol{X}}(y|x)\right)\right)}. \tag{A.3.20}$$

As a result, we confirm that also for smooth estimates $\sqrt{n}(\hat{Y}' - Y)$ converges in distrbution to a zero-mean Gaussian variable with variance as stated in Theorem 2. □

A.4 Theorem 3

Proof. Proof follows directly from the delta method for bootstrap. We observe that both F_Y^{-1} and $F_{Y|\boldsymbol{X}}$ are Donsker and that $\hat{F}_Y^{-1}$ and $\hat{F}_{Y|\boldsymbol{X}}$ converge uniformly to the true functionals. Then in line with Theorem 23.5 of van der Vaart (2000), the proof follows the stages outlined for Theorem 2. □

A.5 Corollary 2

Proof. Let us assume that $\hat{f}_Y(y)$, $\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})$ and $\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x})$ are estimated using the same bandwidth matrices $\mathbf{H}_Y$ and $\mathbf{H}_{\boldsymbol{X}}$. Then, following Wand and Jones (1995) and Silverman (1998), it might be shown that

$$E[\hat{f}_Y(y)] - f_Y(y) = \frac{1}{2}\kappa_2 \mathbf{H}_Y f_Y''(y) + o(\mathbf{H}_Y), \tag{A.5.1}$$

$$\mathrm{var}\{\hat{f}_Y(y)\} = n^{-1}\mathbf{H}_Y^{-1/2} R(K) f_Y(y) + o\left(n^{-1}\mathbf{H}_Y^{-\frac{1}{2}}\right), \tag{A.5.2}$$

$$E[\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})] - f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2}\kappa_2 \mathrm{tr}\left(\mathbf{H}_{\boldsymbol{X}}\mathcal{H}(\boldsymbol{x})\right) + o\left(\mathrm{tr}(\mathbf{H}_{\boldsymbol{X}})\right), \tag{A.5.3}$$

$$\text{var}\{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})\} = n^{-1}|\mathbf{H}_{\boldsymbol{X}}|^{-1/2}R(K)f_{\boldsymbol{X}}(x) + o\left(n^{-1}|\mathbf{H}_{\boldsymbol{X}}|^{-\frac{1}{2}}\right), \tag{A.5.4}$$

where $\mathcal{H}(\boldsymbol{x})$ is the Hessian matrix of $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $R(K) = \int K(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$.

The expected value of $\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x})$ can be written as

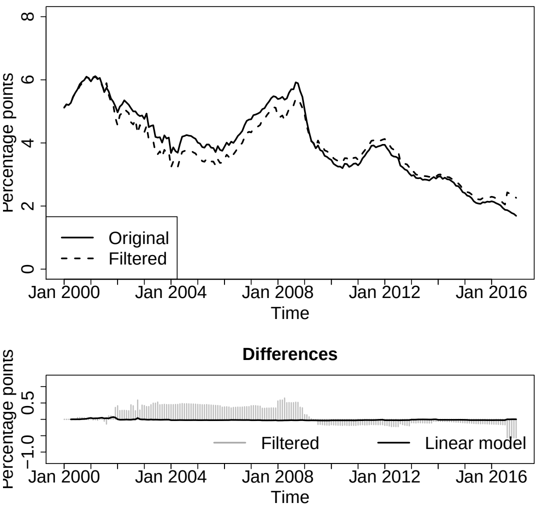
$$\begin{aligned}
E[\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x})] &= \int K_{\mathbf{H}}\left(y - \phi(Y), \boldsymbol{x} - \boldsymbol{X}\right)\mathrm{d}F(Y, \boldsymbol{X}) \\
&= \int K_{\mathbf{H}}\left(y - V, \boldsymbol{x} - \boldsymbol{X}\right)\mathrm{d}F(\phi^{-1}(V), \boldsymbol{X}) \\
&= \int K\left(u, \boldsymbol{z}\right) g\left(y - \mathbf{H}_Y^{1/2}u \middle| \boldsymbol{x} - \mathbf{H}_{\boldsymbol{X}}^{1/2}\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\mathrm{d}u \\
&= \int K\left(u, \boldsymbol{z}\right) f_Y(y - \mathbf{H}_Y^{1/2}u) f_{\boldsymbol{X}}\left(\boldsymbol{x} - \mathbf{H}_{\boldsymbol{X}}^{1/2}\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\mathrm{d}u \\
&= E[\hat{f}_Y(y)]E[\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})].
\end{aligned} \tag{A.5.5}$$

One can also rewrite the variance of $\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x})$ as

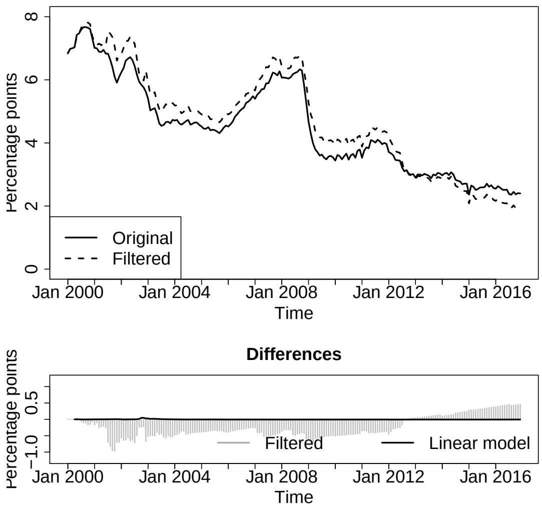
$$\begin{aligned}
\text{var}\{\hat{f}_{Y',\boldsymbol{X}}(y, \boldsymbol{x})\} &= n^{-1}\Bigg[|\mathbf{H}_Y|^{-1/2}|\mathbf{H}_{\boldsymbol{X}}|^{-1/2}\int K\left(u, \boldsymbol{z}\right)^2 f_Y(y - \mathbf{H}_Y^{1/2}u) f_{\boldsymbol{X}}\left(\boldsymbol{x} - \mathbf{H}_{\boldsymbol{X}}^{1/2}\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\mathrm{d}u \\
&\quad - \left\{\int K\left(u, \boldsymbol{z}\right) f_Y(y - \mathbf{H}_Y^{1/2}u) f_{\boldsymbol{X}}\left(\boldsymbol{x} - \mathbf{H}_{\boldsymbol{X}}^{1/2}\boldsymbol{z}\right)\mathrm{d}\boldsymbol{z}\mathrm{d}u\right\}^2\Bigg] \\
&= n^{-1}|\mathbf{H}_Y|^{-1/2}|\mathbf{H}_{\boldsymbol{X}}|^{-1/2}R(K)f_Y(y)f_{\boldsymbol{X}}(\boldsymbol{x}) + o\left(n^{-1}|\mathbf{H}_Y|^{-1/2}|\mathbf{H}_{\boldsymbol{X}}|^{-1/2}\right) \\
&= \text{var}\{\hat{f}_Y(y)\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})\}
\end{aligned} \tag{A.5.6}$$

□

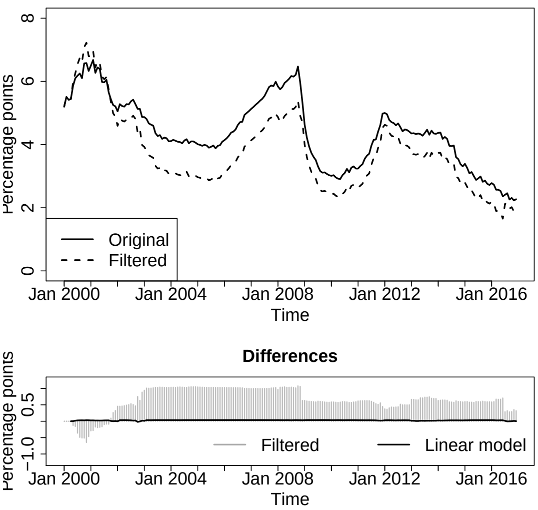
B Additional results



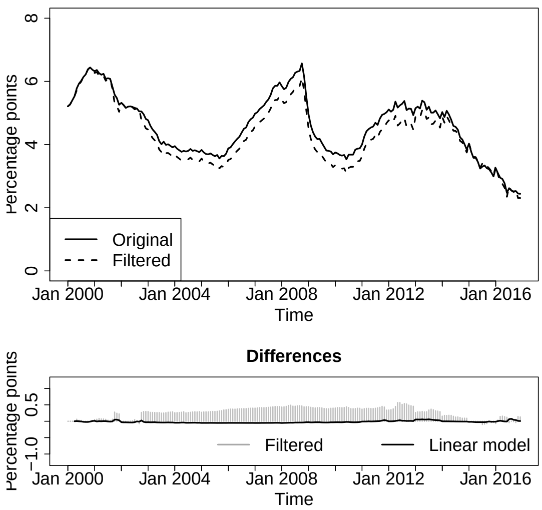
(a) France



(b) Germany



(c) Italy



(d) Spain

Figure 2: Short-term sovereign risk transmission into bank lending rates in (a) France, (b) Germany, (c) Italy and (d) Spain. Filtering equation does not include the error correction term. Top figures in each panel compare the realized against the filtered bank lending rates. Bottom figures show the differences between the realized and filtered rates and compare them against the linear model.