The London School of Economics and Political Science

Essays in Financial Econometrics

Christopher Greiner

Thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of Doctor of Philosophy

August 27, 2024

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 39537 words.

Statement of co-authored work

I confirm that Chapter 2 was jointly co-authored with Tommaso Mancini-Griffoli, Christian Julliard and Kathy Yuan. I contributed 30% of this work.

Christopher Greiner

Acknowledgements

I am deeply indebted to my advisors Christian Julliard and Ian Martin. Without their advice, guidance, and encouragement, writing this thesis would not have been possible. Especially Christian's continuous support, patience, and rigour helped me immensely in writing this thesis.

In addition, I would like to thank the faculty at the London School of Economics for the many valuable comments, questions, suggestions, and discussions, both within and outside the seminar room. I am especially grateful to Jean-Pierre Zigrand, Georgy Chabakauri, Walker Ray, Dong Lou, Vicente Cuñat, Daniel Ferreira, Ulf Axelson, Dimitri Vayanos, Liliana Varela, Michela Verardo, Kathy Yuan and Linyan Zhu.

I also want to thank several of my fellow PhD colleagues who have impacted my time and work substantially: Mark Morley, Nicolas Garrido Sureda, Magnus Hjortfors Irie, Oliver Ashtari Tafti, Bijan Aghdasi, Vahid Rostamkhani, Ali Choubdaran Varnosfaderani, Marcus Gomes de Castro, Roberto Maura-Rivero, Alberto Pellicioli and Niels Wagner. Finally, I want to thank several teachers and colleagues for helping me throughout the years at LSE: Thummim Cho, Alexey Onatskiy, Mary Comben, Morag Brocklehurst, Rhys Cadman, Mike Burkart, Martin Oehmke, Daniel Paravisini, Cameron Peng and Igor Makarov. I acknowledge financial support from the ESRC and the Finance Department at the London School of Economics.

Finally, I am immensely grateful for the support I have received from my close family, friends, and my girlfriend, Mariana. Without them, this thesis would not have been possible.

Abstract

This thesis contains three chapters developing and applying novel techniques in financial econometrics.

In the first chapter, I show that to construct factor models for the cross-section of expected returns, principal component factors should be selected based on risk premia or Sharpe ratios – rather than eigenvalues, as is predominantly done in the literature. This is because principal component factors' prices of risk are given by their risk premia divided by eigenvalues. I show that selection based on risk premia (Sharpe ratios) minimizes the sum of squared pricing errors (Hansen-Jagannathan distance) for a factor model and demonstrate empirically that the proposed selection methods lead to substantial in- and out-of-sample improvements. Further, I devise a test to determine the number of factors to approximate the stochastic discount factor.

In the second chapter, co-authored with Tommaso Mancini-Griffoli, Christian Julliard, and Kathy Yuan, using an equilibrium network model and a large international panel of cross-border trade, we analyse empirically the drivers of foreign currency invoicing. First, we find evidence of strategic complementarity in currency invoicing across countries. Second, key players for a given currency are also countries that are central to the international trade network. Third, we find evidence of natural hedging, between the choices of export and import currencies. Fourth, in counterfactual analysis, we find that the position of the USD is inherently fragile.

In the third chapter, I estimate risk preferences through nonparametric methods from option data. The proposed estimator is shown to be consistent and asymptotically normally distributed. The estimated risk preferences are more in line with preferences implied by classical utility functions than other studies suggest. Specifically, formal statistical tests suggest that there is no statistically significant evidence supporting the pricing kernel puzzle. In contrast, constraining estimated risk preferences to be monotonically decreasing improves the associated beliefs' forecasting performance substantially.

Contents

List of Figures

List of Tables

Chapter 1.
Constructing Factor Models that Approximate the Stochastic Discount Factor

I show that to construct factor models for the cross-section of expected returns, principal component factors should be selected based on risk premia or Sharpe ratios – rather than eigenvalues, as is predominantly done in the literature. This is because prices of risk for principal component factors are given by their risk premia divided by eigenvalues. Hence, selection based on eigenvalues does not guarantee selection of economically relevant factors. Formally, I show that selection based on risk premia (Sharpe ratios) minimizes the sum of squared pricing errors (Hansen-Jagannathan distance) for a factor model and demonstrate empirically that the proposed selection methods lead to substantial in- and out-of-sample improvements. For example, for a five-factor model selection based on Sharpe ratios increases the associated maximum Sharpe ratio by 68% out-of-sample. Further, I devise a test to determine the number of factors and find that relatively few principal components approximate the stochastic discount factor.

1.1. Introduction. A central prediction of asset pricing theory is that assets have different exposures to systematic sources of risk which carry risk premia as compensation for bearing such risk. Identifying the correct factors associated with these sources of risk has become a central question of asset pricing as it is the crucial first step for several practical applications such as capital budgeting, performance evaluation of asset managers, estimating risk premia of non-tradable factors, and so on.

Economic theory offers some guidance on the nature of these factors. However, the search for factors has produced hundreds of potential candidates, leaving applied researchers with ambiguity around the relevance as well as significance of different factors and on how to construct a model from them. Due to this proliferation of factors, machine learning and statistical methods for factor model construction have re-emerged in empirical asset pricing with the prevalent approach being principal component analysis (for early examples see Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986); for recent examples see Kelly, Pruitt, and Su (2019) and Giglio and Xiu (2021)).

An attractive feature of this approach is that the use of principal component factors can be motivated by the law of one price. Kozak, Nagel, and Santosh (2018) demonstrate that the projection of the stochastic discount factor (SDF) on excess returns, as derived in L. Hansen and Jagannathan (1991), can always be expressed as a linear function of the N principal component factors, where N denotes the number of assets. This implies that, for any test asset set considered, a correct factor model is the one containing the N principal component factors. Applied researchers use only a subset $K < N$ principal component factors, which consequently gives rise to an approximation of the valid model. This raises the question, what the consequences of this approximation are, and more specifically, how the principal component factors should be selected in light of the SDF.

The predominant approach is to rank principal component factors based on their eigenvalues, i.e. their variance, and retain the K largest. The main result of this paper is that if the objective is to select factors that are relevant to pricing the cross-section – or equivalently to approximating the SDF – selection should be based on principal component factor's risk premia or the Sharpe ratios, not based on their eigenvalues. To see the intuition for this result, note that principal component factors are by construction orthogonal to each other. Therefore, *prices of risk*, i.e. the multivariate regression coefficients with which factors enter the SDF, are equal to the principal component factor's risk premia divided by its eigenvalue. Consequently, selection based on eigenvalues gives little guarantee that the selected factors are relevant to the SDF as their prices of risk are potentially small. If, for example, the risk premium is zero, the factor does not matter regardless of its eigenvalue. In contrast, when focusing on the cross-section of expected returns, risk premia of principal component factors inform us not only if a factor *is priced*, but also if it *is pricing* assets.

Formally, I demonstrate the following results. Selecting a subset of $K < N$ principal component factors leads to pricing errors. If the objective is to select K principal component factors such that the associated factor model's sum of squared pricing errors is minimized, the K principal component factors with the largest (squared) risk premia should be selected. If however, the objective is to find factors that approximate the SDF well, i.e. that minimize the L. Hansen and Jagannathan (1997) (HJ-)distance of the associated factor model, the ranking should be based on (squared) Sharpe ratios. The latter approach has the convenient interpretation that selection based on Sharpe ratios leads to the optimal SDF approximation in the least-square sense and maximizes the Sharpe ratio of the tangency portfolio associated with the constructed factor model.

For asset pricing, principal component factor selection based on eigenvalues was motivated by Chamberlain and Rothschild (1983), who demonstrated that doing so bounds the sum of squared pricing errors of the associated factor model. The presented analysis is consistent with this result and specifically can be viewed as a tightening of their bound. Empirically, I document that the proposed selection rules lead to substantial in- and out-of-sample improvements over selection based on eigenvalues, measured in terms of root mean squared error, the HJ-distance, and the Sharpe ratios associated with a factor model's tangency portfolio. For example, on the anomaly portfolios by Kozak, Nagel, and Santosh (2020) constructing a 5-factor model based on factors associated with the largest squared Sharpe ratios doubles the associated tangency portfolio's Sharpe ratio in-sample and increases it by 68% out-of-sample. Additionally, it reduces the HJ-distance by more than 50% in- and-out-of-sample.

Having established how factors should be selected in light of the SDF, I turn to how many factors should be selected. I devise a method to estimate the number of factors required to approximate the SDF[1]. The approach adds principal component factors to a factor model based on their Sharpe ratio and repeatedly tests whether the HJ-distance is significantly different from zero at a specified significance level. This identifies the number of factors required such that the factor model's HJ-distance is no longer significantly different from zero. To mitigate multiple testing problems, the stopping point $\hat{K}$ is based on the approach by G'Sell et al. (2016), which controls the family-wise error rate. Empirically, I document that for several datasets relatively few principal component factors are required to approximate the SDF well, when selected based on their Sharpe ratios.

The estimator requires establishing the asymptotic distribution for quantities associated with principal component factor models. To do so, I draw on results from the large dimensional factor literature (see Bai and Ng (2008) for a survey), which allows me to

[1] Several estimators for the number of principal component factors exist, however, these are designed to determine the number of factors required to capture the variation in data.

establish the properties of risk premia, prices of risk, pricing errors, and the HJ-distance. Relative to existing studies, I examine the necessary assumptions such that the factor model implied by the law of one price connects to the factor model analyzed in the large dimensional factor literature – this shows that the law of one price implies additional structure for the data generating process. Because of this, under similar assumptions as in Giglio and Xiu (2021), I show that principal component factor's risk premia are recovered up to their sign at potentially faster convergence rates and derive the asymptotic distribution of these and of prices of risk. Under slightly strengthened assumptions, I derive the asymptotic distribution of pricing errors, whose asymptotic covariance matrix exhibits an adjustment reminiscent of the Shanken (1992) adjustment, but larger. Using this result the properties of the HJ-distance are established.

The results of this paper are closely related to other approaches recently developed in the literature. Kozak, Nagel, and Santosh (2020) estimate the SDF via elastic net regularization. When their estimator is applied to principal components of the return data the Lasso component (L^1-penalty) of their estimator performs principal component factor selection based on risk premia. The results of this paper formally justify this component. Empirically, I demonstrate that factor models constructed using the presented optimal selection rules outperform the Kozak, Nagel, and Santosh (2020) estimator in- and out-of-sample across multiple datasets. Lettau and Pelger (2020b) design a method to extract factors for the cross-section of expected returns – risk premium principal component analysis (RP-PCA). I empirically document that while the RP-PCA based factor models outperform factor models constructed from standard principal component factors, they do so to a substantially lesser degree once the optimal selection rules are employed. For example, principal component factor based models with factors selected based on Sharpe ratios typically outperform the RP-PCA in terms of the HJ-distance in- and out-of-sample. I further demonstrate that the presented selection rules also apply to the RP-PCA and document empirically that this further boosts its in- and out-of-sample performance.

To see the broader relevance of the results of this paper, consider the estimation of risk premia of non-traded factors. Giglio and Xiu (2021) highlight that classical approaches – two pass cross-sectional regressions, Fama-MacBeth regressions, and factor mimicking portfolios – are susceptible to omitted variable bias. This arises whenever the factor model specified by the econometrician does not fully account for all priced sources of risk in the economy. They suggest constructing factor models via principal component analysis and demonstrate how this resolves issues arising from omitted variable bias. A crucial step in this approach is that principal component analysis recovers priced factors. The presented results aid in the efficient selection of relevant priced sources of risk for any test asset set under consideration.

Section 1.2 gives an example to highlight the connection between the SDF and principal component factors. Section 1.3 derives the general result and discusses the relationship to Chamberlain and Rothschild (1983). Section 1.4 derives and discusses the asymptotics. Section 1.5 discusses the out-of-sample performance of the selection rules, the estimator for the number of factors and highlights complementarities of the results of this paper with Kozak, Nagel, and Santosh (2020) and Lettau and Pelger (2020b). Finally, section 1.6 concludes.

1.2. The SDF and Principal Components. To understand the role of principal components for factor models consider the following stylized example. Let R_t denote a vector of gross returns and let $\mathbb{E}[R_{t+1}R'_{t+1}] = Q\Lambda Q'$ denote the eigendecomposition of the second unconditional moment, where Q is the collection of eigenvectors and Λ is a diagonal matrix containing eigenvalues. Consider the SDF projection onto the gross return space

$$M^*_{t+1} = i'_N \mathbb{E}[R_{t+1}R'_{t+1}]^{-1} R_{t+1} = \sum_{k=1}^{N} \frac{(i'_N q_k)(q'_k R_{t+1})}{\lambda_k}$$

where i_N is a $N \times 1$ vector of ones, q_k is the kth column of Q and λ_k is the kth element of the diagonal of Λ. By substituting the eigendecomposition of $\mathbb{E}[R_{t+1}R'_{t+1}]$ the SDF projection can be expressed as a N factor model with factors being principal components $q'_k R_{t+1}$. Furthermore, they enter the SDF with coefficient $i'_N q_k/\lambda_k$ highlighting that eigenvectors play an important role in whether a principal component factor matters for the SDF or equivalently the cross-section of expected returns.

To demonstrate this link further, assume that the first eigenvector satisfies $q_1 = i_N/\sqrt{N}$. Due to the orthogonality of eigenvectors it immediately follows that $i'_N q_k = 0$ for all $k \geq 2$. Therefore, the SDF projection reduces to

$$M^*_{t+1} = \frac{(i'_N q_1)(q'_1 R_{t+1})}{\lambda_1}$$

Notice that the eigenvector can be viewed as portfolio weights and $q'_1 R_{t+1}$ denotes the portfolio return. Since $q_1 = i_N/\sqrt{N}$ the first principal component is related to the market portfolio. Define

$$R_{M,t+1} = \frac{1}{N} i'_N R_{t+1} = \frac{1}{\sqrt{N}} q'_1 R_{t+1}$$

and note $\mathbb{E}[R^2_{M,t+1}] = \lambda_1/N$. Therefore,

$$M^*_{t+1} = \frac{R_{M,t+1}}{\mathbb{E}[R^2_{M,t+1}]}$$

which shows that the SDF projection is proportional to the market portfolio. Define the zero beta rate $\gamma = 1/\mathbb{E}[M^*_{t+1}]$ and by the law of one price, $\mathbb{E}[M^*_{t+1}R_{i,t+1}] = 1$, it follows that for any i

$$\mathbb{E}[R_{i,t+1}] - \gamma = -\frac{Cov(R_{M,t+1}, R_{i,t+1})}{Var(R_{M,t+1})} \frac{Var(R_{M,t+1})}{\mathbb{E}[R^2_{M,t+1}]} \frac{1}{\mathbb{E}[M^*_{t+1}]}$$

By applying the prior result to the market portfolio itself, the expression simplifies to

$$\mathbb{E}[R_{i,t+1}] - \gamma = \beta_{M,i}(\mathbb{E}[R_{M,t+1}] - \gamma)$$

where $\beta_{M,i} = Cov(R_{M,t+1}, R_{i,t+1})/Var(R_{M,t+1})$. Hence, assuming $q_1 \propto i_N$, or any $q_k \propto i_N$ for that matter, in combination with the law of one price implies the extreme case in which only one factor, i.e. the market factor, matters for the SDF and by extension implies a factor model resembling the CAPM for the cross-section of expected returns.

Assuming $q_1 = i_N/\sqrt{N}$ is crucial to the above result, however, does not hold in practice. Several studies suggest $q_1 \approx i_N/\sqrt{N}$ (see for example Kozak, Nagel, and Santosh (2018) and Pelger (2019)), however, this cannot hold exactly. To see why, note that in the prior example, the expected market risk premium is positive only if the expected market return, and therefore zero beta rate, is negative[2].

The example illustrates the link between principal components, the SDF, and factor models. A *correct* factor model consists of the N principal component factors. Further,

[2] Since $\gamma = 1/\mathbb{E}[M^*_{t+1}] = \mathbb{E}[R^2_{M,T+1}]/\mathbb{E}[R_{M,T+1}]$ it follows that $\mathbb{E}[R_{M,t+1}] - \gamma = -Var(R^*_{Mt+1})/\mathbb{E}[R_{M,t+1}]$. Therefore the expected market risk premium is positive only if the expected market return is negative.

the eigenvectors of principal components provide insights into the relevance of each factor. The next section builds on this intuition and derives more general results with a focus on how principal component factors should be selected when constructing factor models for asset pricing.

1.3. Economically Motivated Selection Rules for Principal Components. The focus of most studies is on excess or long-short returns. As before, it is straightforward to show that for any cross-section of excess returns, the N principal components give a valid factor model. Analyzing how these factors relate to the SDF allows to assess the quality of factor models constructed from a selection of principal component factors. To do so, let R_{t+1} be a $N \times 1$ vector of excess or long-short returns with $\mu = \mathbb{E}[R_{t+1}]$, $\Sigma = Var(R_{t+1})$ and $\Sigma = Q\Lambda Q'$, where $Q = [q_1, ..., q_N]$ is the collection of eigenvectors and Λ a diagonal matrix containing eigenvalues. I focus on unconditional expectations for convenience, however, all results carry over to conditional expectations.

Assume the law of one price holds and consider the SDF projected onto excess returns. Following L. Hansen and Richard (1987) and L. Hansen and Jagannathan (1991), the SDF projection can be expressed as

$$M^*_{t+1} = 1 - \mu'\Sigma^{-1}(R_{t+1} - \mu)$$

where I have normalized $\mathbb{E}[M^*_{t+1}] = 1$ for convenience but without loss of generality. Note $\mathbb{E}[M^*_{t+1}R_{t+1}] = 0$ holds for the set of assets onto which the SDF is projected exactly. As in Kozak, Nagel, and Santosh (2018), substituting the eigendecomposition in the above the SDF becomes

$$M^*_{t+1} = 1 - \sum_{k=1}^{N} \frac{\mu' q_k}{\lambda_k} q_k'(R_{t+1} - \mu) \tag{1}$$

By substituting (1) in $\mathbb{E}[M^*_{t+1}R_{t+1}] = 0$, the above immediately implies a N factor model for returns of form

$$\mathbb{E}[R_{i,t+1}] = \sum_{k=1}^{N} \mu' q_k \beta_{i,k} \tag{2}$$

where $\beta_{i,k} = Cov(q_k' R_{t+1}, R_{i,t+1})/Var(q_k' R_{t+1})$. Note that $\beta_{i,k} = q_{i,k}$ as $Var(q_k' R_{t+1}) = \lambda_k$ and $Cov(q_k' R_{t+1}, R_{i,t+1}) = \lambda_k q_{i,k}$ due to properties of eigenvectors. In spirit of section 1.2, equations (1) and (2) demonstrate that a valid factor model for the SDF and the cross-section of expected returns consists of the N principal components.

Equation (1) shows that the multivariate regression coefficients with which factors enter the SDF, or their *prices of risk*, are $b_k = \mu' q_k / \lambda_k$ and allows to analyze how a selection of principal component factors enters the SDF. In practice, the predominant approach is to rank principal component factors based on their eigenvalues, λ_k, and build a factor model with factors associated with the largest eigenvalues. Equation (1) demonstrates that this approach gives little guarantee that selected factors are relevant to the SDF as their prices of risk, b_k, are potentially small. The problem is that such a selection method ignores the risk premium of factors, $\mu' q_k$. If for example $\mu' q_k = 0$ the factor should be ignored regardless of its eigenvalue.

To gauge the practical relevance of this observation, I examine two datasets commonly used in the literature – the 25 size and value portfolios (*FF25*) and the 57 long-short anomaly portfolios by Kozak, Nagel, and Santosh (2020) (*KNS57*). For each, I calculate principal component factors and rank these by eigenvalues. Table 1 reports the eigenvalues

Table 1. Factor Statistics and Model Evaluation

		FF25							
		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Factor statistics	$\hat{\lambda}_k$	0.123	0.006	0.005	0.004	0.001	0.001	0.001	0.001
	$\widehat{RP}_k$	0.527	0.083	0.026	0.045	0.046	0.012	0.014	0.041
	$\widehat{SR}_k$	0.433	0.298	0.111	0.208	0.371	0.107	0.128	0.456
Model statistics	$\hat{\alpha}'\hat{\alpha}$	0.015	0.008	0.008	0.006	0.004	0.003	0.003	0.002
	$\widehat{HJ}$	0.076	0.068	0.067	0.064	0.052	0.051	0.050	0.033
	$\hat{b}_k$	0.356***	1.070***	0.477	0.960**	3.010***	0.932	1.210	5.081***
		(0.000)	(0.001)	(0.251)	(0.050)	(0.000)	(0.289)	(0.184)	(0.000)

		KNS57							
		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Factor statistics	$\hat{\lambda}_k$	0.030	0.011	0.008	0.006	0.003	0.003	0.002	0.002
	$\widehat{RP}_k$	0.191	0.072	0.084	0.104	0.111	0.207	0.017	0.081
	$\widehat{SR}_k$	0.320	0.197	0.277	0.404	0.561	1.100	0.105	0.565
Model statistics	$\hat{\alpha}'\hat{\alpha}$	0.111	0.106	0.099	0.088	0.076	0.033	0.033	0.026
	$\widehat{HJ}$	0.637	0.634	0.628	0.614	0.588	0.487	0.486	0.460
	$\hat{b}_k$	0.536**	0.537*	0.916*	1.565***	2.848***	5.841***	0.655	3.957***
		(0.028)	(0.088)	(0.055)	(0.006)	(0.000)	(0.000)	(0.516)	(0.000)

Note: The table reports estimated eigenvalues ($\hat{\lambda}_k$), risk premia ($\widehat{RP}_k = \hat{\mu}'\hat{q}_k$), Sharpe ratios ($\widehat{SR}_k = \hat{\mu}'\hat{q}_k/\sqrt{\hat{\lambda}_k}$), sum of squared pricing errors ($\hat{\alpha}'\hat{\alpha}$), HJ-distance ($\widehat{HJ} = \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$) and prices of risk ($\hat{b}_k = \hat{\mu}'\hat{q}_k/\hat{\lambda}_k$) for the first eight principal components extracted from different datasets. Note risk premia, Sharpe ratios and the sum of squared pricing errors are annualized and factors have been normalized to have positive mean. The sum of squared pricing errors and HJ-distance correspond to a factor model containing the first through K^{th} principal component indicated in each column. In brackets below the price of risk estimates the p-value associated with $H_0 : b_k = 0$ is reported based on the asymptotic distribution derived in section 1.4. The Newey and West (1987) covariance estimator is employed with $T^{1/4}$ lags. Estimates significant at the 10%, 5% and 1% are indicated by *, ** and *** respectively. Each dataset consists of monthly excess or long-short returns – for a description of the data see section 1.5.1.

and annualized risk premia and Sharpe ratios for these. From the principal component factors, I construct factor models subsequently increasing the number of factors. To evaluate the model, I report the annualized sum of squared pricing errors, the HJ-distance, and the price of risk for each principal component factor in table 1.

For both datasets selection of principal components based on eigenvalues does not guarantee that factors with the largest and most significant prices of risk are selected early on. Focussing on squared pricing errors, the largest reductions seem to occur when factors with large risk premia are added to the factor model – for FF25 these are PC2, PC4, and PC5; for KNS57 these are PC4, PC5, and PC6. Turning to the HJ-distance, the largest reductions occur when factors with large Sharpe ratios are added to the factor model – for FF25 these are PC2, PC5, and PC8; for KNS57 these are PC5, PC6, and PC8. Overall, this suggests that selection of principal component factors solely based on eigenvalues may be suboptimal if the goal is to select factors relevant to pricing the cross-section.

To formally analyze how principal component selection affects the ability of the associated factor model to price the cross-section consider the SDF approximation containing K principal components

$$\tilde{M}^*_{t+1} = 1 - \sum_{k=1}^{K} \frac{\mu' q_k}{\lambda_k} q_k'(R_{t+1} - \mu)$$

The approximation will lead to violations of the law of one price, that is $\mathbb{E}[\tilde{M}^*_{t+1}R_{i,t+1}] \neq 0$. These violations can be expressed as

$$\begin{aligned} \alpha_i &= \mathbb{E}[\tilde{M}^*_{t+1}R_{i,t+1}] \\ &= \mathbb{E}[R_{i,t+1}] - \sum_{k=1}^{K} \mu' q_k q_{i,k} \\ &= \sum_{k=K+1}^{N} \mu' q_k q_{i,k} \end{aligned} \tag{3}$$

where I used the definition of $\tilde{M}^*_{t+1}$ to obtain the first equality and (2) to obtain the second equality. These α_i correspond to OLS intercepts of a time series regression of $R_{i,t+1}$ on a constant and the K factors $F_{k,t+1} = q_k' R_{t+1}$.

Jointly expression (1) and (3) reveal how a selection method impacts the quality of the SDF approximation and of the associated factor model, measured by mispricing α_i. Selection based solely on λ_k may lead to a poor SDF approximation or a factor model with poor pricing performance. This insight and what it implies for optimal factor selection can be made more precise by considering explicit criteria with which factor models are evaluated.

1.3.1. *Squared Pricing Errors.* Consider the sum of squared pricing errors across all assets associated with a factor model containing K principal component factors. Let $Q_{N-K} = [q_{K+1}, ..., q_N]$ be the set of omitted eigenvectors. Note (3) can be written more compactly as $\alpha = Q_{N-K}Q'_{N-K}\mu$. Therefore,

$$\alpha'\alpha = \sum_{i=1}^{N} \alpha_i^2 = \sum_{k=K+1}^{N} (\mu' q_k)^2 \tag{4}$$

where I used the fact that $Q'_{N-K}Q_{N-K} = I_{N-K}$. Recall that the principal component factor is $F_{k,t+1} = q_k' R_{t+1}$, so the associated risk premium and variance is $\mathbb{E}[F_{k,t+1}] = q_k'\mu$ and $Var(F_{k,t+1}) = \lambda_k$ respectively.

Expression (4) demonstrates that the sum of squared pricing errors is equal to the sum of squared risk premia belonging to the principal component factors omitted from the SDF approximation or factor model. Equation (1) already demonstrated that risk premia matter for whether or not a principal component factor enters the SDF, as there is a one-to-one mapping between risk premia and prices of risk. The surprising insight of (4) is that if the objective is to select factors that minimize pricing errors only risk premia matter and eigenvalues, λ_k, are irrelevant. Put differently, if we wish to construct a factor model from principal components that minimize $\alpha'\alpha$, we should select factors with the largest risk premia and omit factors with small risk premia.

1.3.2. *HJ-Distance.* Albeit interesting, $\alpha'\alpha$ may not be the right criterion to consider. Specifically, a low $\alpha'\alpha$ does not guarantee that the SDF is well approximated and $\alpha'\alpha$ is susceptible to repackaging and scaling of test assets. The L. Hansen and Jagannathan (1997) (HJ-) distance overcomes these issues.

The objective of the HJ-distance is to assess the magnitude of specification error induced by using a proxy SDF from the set of admissible SDFs. As I consider excess returns, the law of one price does not identify the mean of the SDF. I therefore normalize the mean of the proxy and of the set of admissible SDFs to be one. Under this normalization the

HJ-distance is given as[3]

$$HJ = \alpha'\Sigma^{-1}\alpha$$

and measures the least-squares distance between the proxy SDF and the set of admissible SDFs. Substituting (3) and the eigendecomposition of Σ in the expression yields

$$\begin{aligned} HJ &= \mu' Q_{N-K} Q'_{N-K} Q \Lambda^{-1} Q' Q_{N-K} Q'_{N-K} \mu \\ &= \mu' Q_{N-K} \Lambda^{-1}_{N-K} Q'_{N-K} \mu \\ &= \sum_{k=K+1}^{N} \frac{(\mu' q_k)^2}{\lambda_k} \end{aligned} \tag{5}$$

where I used the orthogonality of eigenvectors to arrive at the second equality. Notice that $(\mu' q_k)^2/\lambda_k$ is the squared Sharpe ratio of principal component factor k. From (5) it follows that

$$\begin{aligned} HJ &= \sum_{k=1}^{N} \frac{(\mu' q_k)^2}{\lambda_k} - \sum_{k=1}^{K} \frac{(\mu' q_k)^2}{\lambda_k} \\ &= SR(M^*_{t+1})^2 - SR(\tilde{M}^*_{t+1})^2 \end{aligned} \tag{6}$$

where $SR(M)$ denotes the Sharpe ratio of the tangency portfolio associated with SDF M[4]. Therefore, in the considered setting, the HJ-distance additionally measures how close the Sharpe ratio associated with the proxy SDF, $\tilde{M}^*_{t+1}$, is to the maximum attainable Sharpe ratio on the test assets which is associated with M^*_{t+1}.

Expression (5) demonstrates that the HJ-distance is equal to the sum of squared Sharpe ratios belonging to the principal component factors omitted from the SDF approximation or factor model. This suggests principal component factors with the largest Sharpe ratios should be selected. In turn, this achieves three objectives. First, it leads to a factor model that minimizes weighted pricing errors. Second, it leads to a SDF approximation that is as close as possible, in the least-squares sense, to the set of admissible SDFs. Third, by (6), it ensures that the Sharpe ratio of the tangency portfolio associated with the approximate SDF is as close as possible to the maximum attainable Sharpe ratio. In contrast, classical selection based on λ_k is not guaranteed to achieve either of the aforementioned objectives.

1.3.3. *Relationship to Chamberlain and Rothschild (1983).* Classical selection of principal components is based on retaining the factors associated with the largest eigenvalues. Economically, this selection procedure was formally motivated by Chamberlain and Rothschild (1983) for asset pricing, who demonstrated that such a selection will bound the sum of squared pricing errors of the associated factor model. The result derived from equation (4) seems to be at odds with their result. I here demonstrate that (4) is consistent with their result and specifically can be viewed as tightening their bound.

Chamberlain and Rothschild (1983) showed that under no arbitrage, as $N \to \infty$, assuming the $K + 1^{st}$ largest eigenvalue is finite, that the squared pricing errors associated with a factor model consisting of the K principal components obey

$$\delta^2 \lambda_{K+1} \geq \sum_{i=1}^{N} \alpha_i^2 \tag{7}$$

[3] Let y be a general SDF proxy, $\mathcal{M}$ be the set of admissible SDFs with $m \in \mathcal{M}$ and let $||.||$ be the standard L^2 norm. L. Hansen and Jagannathan (1997) consider the general problem $HJ \equiv \min_{m\in\mathcal{M}} ||y-m||$, which yields $HJ = \alpha'\mathbb{E}[R_{t+1}R'_{t+1}]^{-1}\alpha$. When we constrain $\mathbb{E}[y] = \mathbb{E}[m] = 1$, Kan and Robotti (2008) show that $HJ \equiv \min_{m\in\mathcal{M},\mathbb{E}[m]=1} ||y-m||$ yields $HJ = \alpha'\Sigma^{-1}\alpha$.

[4] Specifically, it corresponds to the Sharpe ratio of the return $R^*_{t+1} = \mu'_F \Sigma_F^{-1} F_{t+1}$.

where δ is the maximum attainable Sharpe ratio in the economy and λ_{K+1} is the $K+1^{st}$ largest eigenvalue eigenvalue[5]. Hence, selection based on the largest eigenvalues ensures that the above bound is as tight as possible and therefore should lead to lower squared pricing errors.

The setup considered thus far is slightly different from Chamberlain and Rothschild (1983). I here work with the weaker assumption that the law of one price holds, make no assumptions on the behaviour of eigenvalues and the results hold for all N. Nevertheless, their bound can also be derived under these assumptions. Notice that under the law of one price, the maximum attainable squared Sharpe ratio is equal to $Var(M^*_{t+1})$ since $\mathbb{E}[M^*_{t+1}] = 1$ (see L. Hansen and Jagannathan (1997)). Therefore,

$$\delta^2 = Var(M^*_{t+1}) = \sum_{k=1}^{K} \frac{(\mu' q_k)^2}{\lambda_k} + \sum_{k=K+1}^{N} \frac{(\mu' q_k)^2}{\lambda_k}$$

which follows by (1) using the properties of principal components. I further split the variance of M^*_{t+1} into two components for exposition. Let λ_{K+1} denote the eigenvalue such that $\lambda_{K+1} \geq \lambda_k$ for all $k \in \{K+1, ..., N\}$. As the first term on the right-hand side is positive and by the definition of λ_{K+1} it follows that

$$\delta^2 \geq \frac{1}{\lambda_{K+1}} \sum_{k=K+1}^{N} (\mu' q_k)^2$$

which using (4) immediately yields the Chamberlain and Rothschild (1983) result in (7).

The difference between the Chamberlain and Rothschild (1983) bound in (7) and the result in (4) is that their bound can be arbitrarily loose. Selecting principal component factors associated with the largest λ_k can ensure that the bound becomes as tight as possible, however, this will not guarantee that it will hold exactly. In contrast, the result in (4) is an identity and therefore selecting principal component factors associated with the largest risk premia directly ensures that $\alpha'\alpha$ is as low as possible. This allows for potentially more efficient selection.

1.3.4. *Comparison of Selection Rules.* To evaluate the relevance of the previously discussed selection rules, I examine the performance of principal component based factor models for three different test asset sets. On each test asset set, I calculate the principal component factors and sort them by eigenvalues, squared risk premia, or squared Sharpe ratios. I then construct pricing errors and the return on the associated tangency portfolio. Based on these I calculate the root mean squared error, the HJ-distance, and the Sharpe ratio. For details on the datasets and estimation see section 1.5.1.

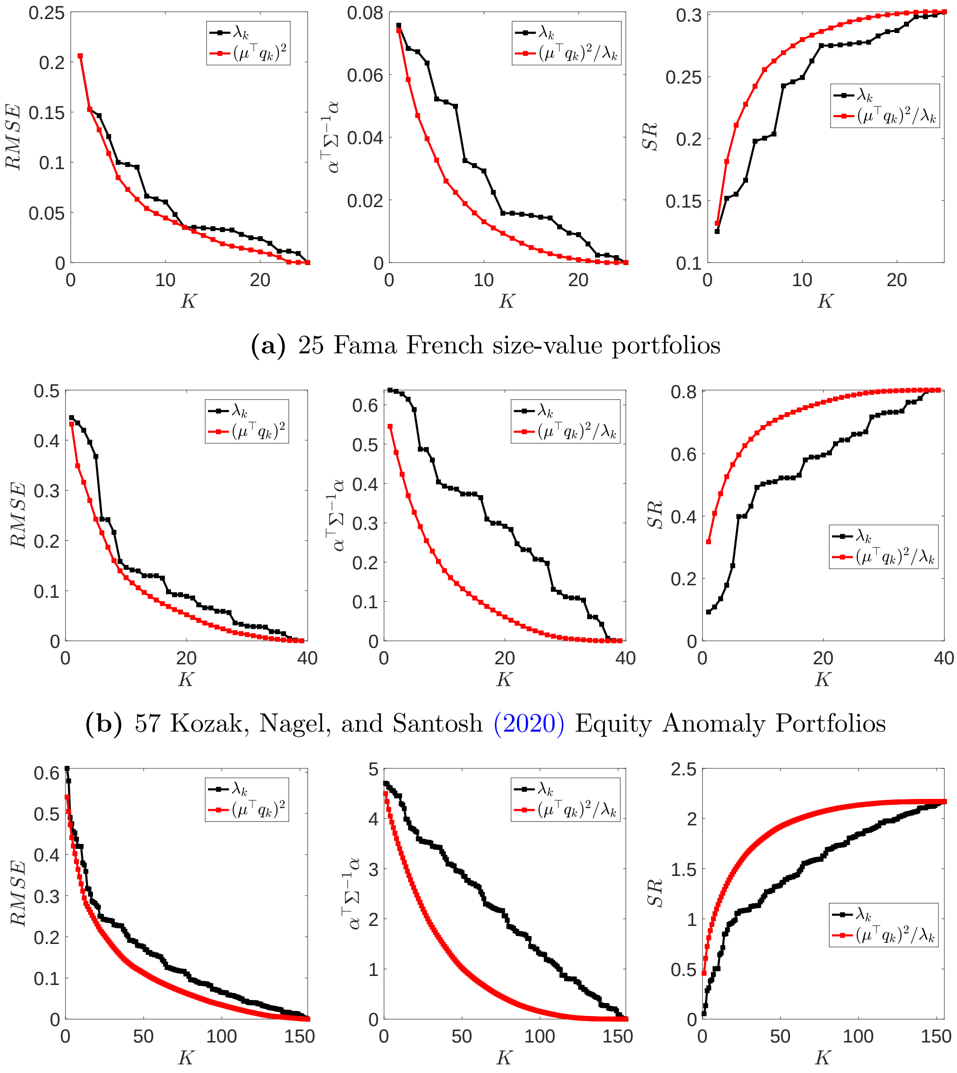
From left to right, the panels in figure 1 report RMSE, the HJ-distance, and the Sharpe ratio for the three different selection methods over the full set of principal components, that is for every approximation degree K. The leftmost figures illustrate that selection based

[5] They show that the pricing errors are bounded by the highest squared Sharpe ratio, δ^2, and the $K+1^{st}$ eigenvalue, that is

$$\sum_{i=1}^{N}(\mu_i - \sum_{k=1}^{K} \tau_k \beta_{i,k})^2 = \sum_{i=1}^{N} \alpha_i^2 \leq \lambda_{K+1}\delta^2$$

where μ_i are mean excess returns and τ_k are coefficients for the betas. The above bound corresponds to their theorem 3' and is for a finite economy with N assets, which they allow to go to infinity (see Chamberlain and Rothschild (1983)). As in the limit λ_{K+1} is finite, they argue principal components corresponding to the largest K eigenvalues can be used to price assets well and the arbitrage pricing theorem holds approximately. Using their lemma 1 it can be shown that their expression for $\tau_k\beta_{i,k}$ in the above is equal to $q_k'\mu q_{i,k}$ using the notation here.

Figure 1. Comparison of PCA Factor Selection Rules



(a) 25 Fama French size-value portfolios

(b) 57 Kozak, Nagel, and Santosh (2020) Equity Anomaly Portfolios

(c) 212 A. Y. Chen and Zimmermann (2022) Equity Anomaly Portfolios

Note: The figure depicts the RMSE, the HJ-distance and Sharpe ratio of R^*_{t+1} for factor models subsequently increasing K, the number of principal components selected. For each test asset set the principal components are extracted. They are then sorted based on the largest λ_k, $(\mu' q_k)^2$ or $(\mu' q_k)^2/\lambda_k$ to construct factor models. All test assets are at monthly frequency and returns correspond to excess returns or long-short returns – for a description of the data see section 1.5.1.

on squared risk premia minimizes RMSE consistently across all approximation degrees and datasets. Trivially, as $K \to N$, a factor model prices assets perfectly – all selection rules lead to zero RMSE. The important feature is that selection based on squared risk premia achieves consistently lower RMSE than the competing methods for all approximation degrees, especially for lower K (see table 11 in the appendix).

The central and rightmost figures illustrate that selection based on squared Sharpe ratios minimizes the HJ-distance and maximizes the Sharpe ratio of the associated tangency portfolio with consistently sizeable gains over selection based on eigenvalues. Again, as all principal components are included the HJ-distance becomes zero for all selection rules, however, selection based on squared Sharpe ratios is able to deliver consistently lower HJ-distances for all approximation degrees. Further, these gains are economically meaningful, even if only few principal components are considered as illustrated in table 11 in the appendix. For $K = 3$, across datasets selection based on principal component Sharpe ratios improves monthly tangency portfolio Sharpe ratios by 0.06-0.44 relative to selection based on eigenvalues. More so, these gains persist across a variety of approximation degrees K until the SDF approximation reaches the limit.

Finally, as discussed and illustrated in table 1 principal component factor selection based on eigenvalues can lead to the inclusion of less important factors in terms of their associated prices of risk. Table 13 in the appendix compares the prices of risk for the first ten factors selected under each selection rule for different datasets. Selection based on squared Sharpe ratios leads to substantial improvements over selection based on eigenvalues. Factors with larger and highly significant prices of risk are selected early on. This is expected as prices of risk, $b_k = \mu' q_k / \lambda_k$, are closely related to Sharpe ratios, $SR_k = \mu' q_k / \sqrt{\lambda_k}$, but not fully mechanical as the selection rule is not directly based on p-values. Selection based on squared risk premia in comparison only leads to minor improvements over selection based on eigenvalues. Therefore, the results in table 13 further strengthen the results obtained for the HJ-distance. If the goal is to construct factor models that approximate the SDF well and include factors relevant to the SDF, selection of principal components should be based on Sharpe ratios.

1.4. Asymptotic Results for Principal Component Factor Models. This section derives asymptotic results to evaluate the quality of a principal component factor model in the context of the SDF. First, I establish asymptotic results to evaluate whether principal component factors enter the SDF with nonzero coefficients, i.e. nonzero prices of risk. Further, to assess the overall quality with which a model approximates the SDF the asymptotic distribution of pricing errors and the HJ-distance is given. All proofs are deferred to the appendix.

1.4.1. *Setup.* Since PCA is typically applied to datasets with large cross-sections the results are derived for both $N, T \to \infty$. To do so, I draw on the *large dimensional factor* literature (see Bai and Ng (2008) for a survey). The starting point in this literature is that data is driven by $K < N$ factors and idiosyncratic errors.

As shown, the law of one price implies that *a* valid factor model consists of the N principal component factors. Writing equation (2) compactly gives

$$\mu = Q_K Q_K' \mu + Q_{N-K} Q_{N-K}' \mu$$

where I separated factors into two groups and used the fact that for each group betas are equal to eigenvectors[6]. It follows that

$$R_t = Q_K Q_K' R_t + Q_{N-K} Q_{N-K}' R_t$$

No error is introduced as the above is an identity since $QQ' = I_N$. To connect the prior to settings considered in the large dimensional factor literature the following assumption is made.

Assumption E.1. *Let* $\mu' q_k = 0 \; \forall \; k \in K+1, \ldots, N$.

Assumption E.1 implies that the second term in the return identity is zero in expectation. For all economic purposes, this component is noise as it does not deliver risk compensation. The assumption therefore separates the process for R_t into a factor and

[6] The N factor representation of μ is valid for any orthonormal matrix or basis. The eigenvector matrix Q of the covariance matrix is a natural choice motivated by the law of one price.

idiosyncratic error component given as

$$R_t = \beta F_t + E_t \tag{8}$$

where $\beta = Q_K$, $F_t = Q'_K R_t$ and $E_t = Q_{N-K} Q'_{N-K} R_t$. It follows that $\mathbb{E}[E_t] = 0$, $\mathbb{E}[F_t E'_t] = 0$, $\beta' E_t = 0$, $Var(F_t) = \Lambda_{T,K}$ which is diagonal and $\beta'\beta = I_K$.

Given (8) assumptions on factors and errors are made to characterize the asymptotic properties of the principal component factors. I make assumptions similar to the strong factor setting by Bai (2003), which establishes asymptotic properties under fairly general conditions. The results can be extended along the line of Bai and Ng (2023) to allow for weaker factors. Relative to existing studies, the setup considered here differs in several aspects. First, coefficients and errors are orthogonal, $\beta' E_t = 0$, which leads to faster convergence rates for coefficients than in e.g. Bai (2003). Second, the restrictions on factors and coefficients imply a specific factor normalization in spirit of Bai and Ng (2013) – theorem 1 demonstrates that risk premia are therefore recovered up to their sign. Finally, as in Lettau and Pelger (2020a) and Giglio and Xiu (2021), returns are allowed to have non-zero means which requires additional assumptions.

1.4.1.1. Assumptions. In what follows, let $\bar{X} = X - i_T \bar{x}'$, where $\bar{x} = 1/T \sum_t X_t$, let $\|.\|$ denote the Frobenius norm and $\|.\|_{sp}$ denote the spectral norm. Also define $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. As limits are taken over both N and T, it is convenient to define the sample covariance matrix $\hat{\Sigma}_T$ and normalized sample covariance matrix $\hat{\Sigma}_{NT} = \hat{\Sigma}_T / N$ – both matrices share the same eigenvectors and their eigenvalues are connected via $\hat{\Lambda}_T = N \hat{\Lambda}_{NT}$.

Assumption A.1. *Let $\gamma_N(s,t) = \mathbb{E}[\frac{1}{N} E'_s E_t]$. For all N and T there is a positive constant $M < \infty$ such that*

(i) $\frac{1}{T} \sum_s \sum_t |\gamma_N(s,t)| \leq M$ *and* $\max_t |\gamma_N(t,t)| \leq M$

(ii) $\frac{1}{T^2} \sum_s \sum_t \mathbb{E}[(\sum_i E_{i,s} E_{i,t} - \mathbb{E}[E_{i,s} E_{i,t}])^2] \leq MN$

Assumption A.2. *For all T and every i there is some positive constant $M < \infty$ such that $\mathbb{E}[\|\sum_t F_t E_{i,t}\|^2] \leq MT$.*

Assumption A.2'. *For all T and every i there is some positive constant $M < \infty$ such that $\mathbb{E}[\|\sum_t \bar{F}_t \bar{E}_{i,t}\|^2] \leq MT$.*

Assumption A.3. *As $N, T \to \infty$, let*

(i) $\frac{1}{\sqrt{N}} \bar{f} \xrightarrow{p} \mu_F$ *where μ_F is finite;*

(ii) $\frac{1}{NT} \bar{F}' \bar{F} \xrightarrow{p} \Lambda_{TN,K}$ *where $\Lambda_{TN,K}$ is diagonal, positive, finite and has distinct elements.*

Assumption A.4. *Define $v_t = \frac{1}{\sqrt{N}} F_t - \mu_F$ so that $\bar{v} = \frac{1}{\sqrt{N}} \bar{f} - \mu_F$. As $T \to \infty$ the following central limit theorem holds*

$$\sqrt{T} \bar{v} \xrightarrow{d} N(0, \Omega_v)$$

where $\Omega_v = \lim_{T \to \infty} T \mathbb{E}[\bar{v}\bar{v}']$.

Assumption A.5. *Define $\bar{e} = \frac{1}{T} \sum_t E_t$. As $T \to \infty$ the following central limit theorem holds*

$$\sqrt{T} \bar{e} \xrightarrow{d} N(0, \Omega_e)$$

where $\Omega_e = \lim_{T \to \infty} T \mathbb{E}[\bar{e}\bar{e}']$.

Assumption A.1 restricts the dependence of E_t. It limits the contribution of errors to the overall return variance, however, allows the largest eigenvalues of the error covariance matrix to grow at rate $\sqrt{N}$ and is therefore more general than the setting in Chamberlain and Rothschild (1983)[7]. Assumption A.2 is as in Bai (2003) and restricts the dependence between errors and factors. Assumption A.2' further constrains the variation of error and factor lead-lag correlation and is used to analyze the properties of $\hat{\alpha}$. Assumption A.3 (ii) corresponds to the standard *pervasiveness condition* and states that factor variances grow with N – it takes a slightly different form than in e.g. Bai (2003) as coefficients in (8) are normalized. Assumption A.3 (i) is necessary to rule out asymptotic arbitrage opportunities or mechanical irrelevance of factors for pricing as together with assumption A.3 (ii) it ensures sample Sharpe ratios converge to a finite limit. Finally assumption A.4 and A.5 state that sample means of factors and errors satisfy central limit theorems.

1.4.2. *Asymptotic Results.* To test if principal component factors are relevant to the SDF, I first establish the asymptotic properties of sample risk premia. Let $\hat{\mu}_F$ denote the sample average of the principal component factors $\hat{F}_t = \hat{Q}'_K R_t$. Theorem 1 presents the asymptotic distribution of $\hat{\mu}_F$ with the proof given in the appendix.

Theorem 1. *Under E.1 and A.1-A.4, as $N, T \to \infty$*

$$\frac{1}{\sqrt{N}}\hat{\mu}_F - S\mu_F = S\bar{v} + O_p(\frac{1}{T} + \frac{1}{N^2})$$

for the $K \times K$ matrix S that is asymptotically diagonal with 1 or -1 on its diagonal. Further, if $\sqrt{T}/N^2 \to 0$

$$\sqrt{T}(\frac{1}{\sqrt{N}}\hat{\mu}_F - S\mu_F) \xrightarrow{d} N(0, \Omega_\mu)$$

where $\Omega_\mu = \Omega_v$ with Ω_v defined in A.4.

Theorem 1 establishes that appropriately scaled risk premia converge to their population counterpart up to sign and are asymptotically normally distributed. The asymptotic covariance matrix does not depend on the covariance matrix of residuals or the estimation error of factor coefficients – in the appendix, I demonstrate that these terms are negligible. The result is similar to Giglio and Xiu (2021), however, the convergence rate is faster. This is because errors are orthogonal to factor coefficients.

The theorem allows to test if factors enter the SDF with nonzero coefficients – or prices of risk – which are given as $b_k = \mu_{F,k}/\lambda_k$. Hence, testing whether b_k is zero is equivalent to testing whether the factor has a zero Sharpe ratio. Corollary C.1 establishes the asymptotic distribution for the respective joint or individual factor test under the null.

Corollary C.1. *Suppose $\hat{\Omega}_\mu \xrightarrow{p} \Omega_\mu$. Under the null $H_0 : \mu'_F \Lambda^{-1}_{K,T} \mu_F = 0$, E.1 and A.1-A.4, as $N, T \to \infty$ with $\sqrt{T}/N^2 \to 0$*

$$T\hat{\mu}'_F \hat{\Lambda}^{-1}_{K,T} \hat{\mu}_F \xrightarrow{d} \chi^2(d, i_K)$$

where $\chi^2(d, i_K)$ denotes a weighted χ^2 distribution[8] with weights d equal to the eigenvalues of $\Lambda^{-1}_{K,NT}\Omega_\mu$ and i_K is a $K \times 1$ vector of ones specifying the degrees of freedom of each chi-square random variable. Further, under $H_0 : \mu^2_{F,k}/\lambda_{T,k} = 0$

$$T\frac{\hat{\mu}^2_{F,k}}{\hat{\lambda}_{T,k}} \xrightarrow{d} \chi^2(d_k, 1)$$

[7] Specifically, A.1 implies $\mathbb{E}[\|E'E\|^2] \leq O(\frac{(TN)^2}{\delta^2_{NT}})$, so $\mathbb{E}[\|\frac{1}{T}E'E\|^2] \leq O(\frac{N^2}{\delta^2_{NT}})$. By Lyapunov's inequality $\mathbb{E}[\|\frac{1}{T}E'E\|] \leq \mathbb{E}[\|\frac{1}{T}E'E\|^2]^{1/2} \leq O(\frac{N}{\delta_{NT}})$. Finally, by the expectation inequality $\|\Sigma_E\| \leq \|\mathbb{E}[\frac{1}{T}E'E]\|$, so by standard norm inequalities $\|\Sigma_E\|_{sp} \leq O(\frac{N}{\delta_{NT}}) = O(\max(\sqrt{N}, N/\sqrt{T}))$.

[8] See B. E. Hansen (2021).

where $d_k = \Omega_{\mu,kk}/\lambda_{NT,k}$ with $\Omega_{\mu,kk}$ denoting the k^{th} element on the diagonal of Ω_μ.

The result in corollary C.1 is helpful to evaluate if factors matter for the SDF, however, often it is more interesting to assess the quality with which a factor model is able to approximate the SDF or price assets. The HJ-distance is the natural metric to consider. To do so, theorem 2 first establishes the asymptotic properties of the pricing errors – note by E.1 these are zero in population.

Theorem 2. *Under E.1, A.1-A.5 and A.2', as $N, T \to \infty$*

$$\hat{\alpha} = \bar{e}(1 + \mu_F' \hat{\Lambda}_{NT,K}^{-1} \mu_F) + O_p(\frac{1}{\sqrt{T}\delta_{NT}}) + O_p(\frac{\sqrt{N}}{\delta_{NT}^3})$$

Further, if $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$

$$\sqrt{T}\hat{\alpha} \xrightarrow{d} N(0, \Omega_\alpha)$$

where $\Omega_\alpha = (1 + \mu_F' \Lambda_{NT,K}^{-1} \mu_F)^2 \Omega_e$ with Ω_e defined in A.5.

Theorem 2 establishes that pricing errors are asymptotically normally distributed around zero. The theorem restricts the rates at which N and T are allowed to grow relative to each other and requires that neither grows too quickly. The conditions are satisfied for example if $T/N \to c < \infty$. Unlike for risk premia, the asymptotic covariance matrix of pricing errors contains an adjustment that resembles the Shanken adjustment (Shanken (1992)) – in contrast to the classical adjustment, the present adjustment is squared and hence larger. The adjustment is higher the larger factor means and the lower factor variances are.

From theorem 2 the asymptotic distribution of the HJ-distance can be established. Note E.1 implies that the HJ-distance is zero in population, so the below result establishes the distribution under the null that the HJ-distance is zero.

Theorem 3. *Define $\Sigma_{\alpha,T}^{-1} = Q_{N-K} \Lambda_{T,N-K}^{-1} Q_{N-K}'$. Let $\hat{d}$ denote the non-zero eigenvalues of $\hat{\Sigma}_{\alpha,T}^{-1} \hat{\Omega}_\alpha$ and suppose $\hat{d} \xrightarrow{p} d$. Under E.1, A.1-A.5 and A.2', as $N, T \to \infty$ with $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$*

$$\frac{T\hat{\alpha}\hat{\Sigma}_T^{-1}\hat{\alpha} - \sum_i^{N-K} \hat{d}_i}{\sqrt{2\sum_i^{N-K} \hat{d}_i}} \xrightarrow{d} N(0,1)$$

Theorem 3 establishes that the scaled and centered HJ-distance is asymptotically standard normally distributed. To see why, note $T\hat{\alpha}\hat{\Sigma}_T^{-1}\hat{\alpha}$ can be written as a weighted sum of asymptotically independent chi-square variables. For finite N, the HJ-distance would converge to a weighted chi-square distribution. However, because $N \to \infty$ this weighted chi-square distribution in turn converges to a normal distribution by a standard central limit argument.

1.5. Empirical Results.

1.5.1. *Data and Implementation.* In what follows I use data on monthly excess or long-short returns for several anomaly portfolios. I consider the 25 size and value portfolios (*FF25*), 57 long-short anomaly portfolios by Kozak, Nagel, and Santosh (2020) (*KNS57*) and the 212 long-short anomaly portfolios by A. Y. Chen and Zimmermann (2022) (*CZ212*). Where necessary, excess returns are calculated using the monthly T-bill rate. To deal with missing observations, I only keep periods for which at least 75% of anomaly portfolio returns are observed and subsequently keep portfolios that are observed for the full period[9]. I focus on these datasets as they are frequently encountered in the literature and results can therefore be compared by the reader easily.

To evaluate the performance of factor models, I construct in-sample and out-of-sample root mean squared error (RMSE), HJ-distances, and the Sharpe ratio of the implied tangency portfolio. This is done for the different principal component factor selection methods whilst varying the number of factors K.

For the in-sample exercises, K principal component factors are selected using the full sample from which I construct pricing errors based on the associated factor model expression in (3) using sample analogs. Pricing errors are hence computed as $\hat{\alpha} = \hat{\mu} - \hat{Q}_K \hat{Q}_K' \hat{\mu}$ from which I calculate $RMSE = \sqrt{\hat{\alpha}'\hat{\alpha}/N}$ and $HJ = \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$. To construct the Sharpe ratio, the return proportional to the associated tangency portfolio is constructed as $R^*_{t+1} = \hat{\mu}'\hat{Q}_K\hat{\Lambda}_K^{-1}\hat{Q}_K' R_{t+1}$ from which the Sharpe ratio is calculated on the full sample.

Out-of-sample quantities are constructed using rolling estimation. Using information over the last 20 years up to time t, K principal component factors are selected. For a set of K factors, denote by $\hat{Q}_{K,t}$, $\hat{\Lambda}_{K,t}$ and $\hat{\mu}_t$ the estimates of factor weights or betas, factor variances and return means using information until t. Let T_{OOS} denote the total number of out-of-sample periods. Out-of-sample pricing errors at $t+1$ are then constructed as $\hat{\epsilon}_{t+1} = R_{t+1} - \hat{Q}_{K,t}\hat{Q}_{K,t}' R_{t+1}$ for all T_{OOS} periods, from which I compute $\hat{\alpha} = 1/T_{OOS} \sum_t \hat{\epsilon}_t$. Using these, RMSE and the HJ-distance are constructed where the HJ-distance uses the sample covariance matrix estimated on the T_{OOS} out-of-sample periods. Out-of-sample returns proportional on the tangency portfolio are constructed as $R^*_{t+1} = \hat{\mu}_t'\hat{Q}_{K,t}\hat{\Lambda}_{K,t}^{-1}\hat{Q}_{K,t}' R_{t+1}$, from which the Sharpe ratios is calculated over the T_{OOS} out-of-sample periods.

1.5.2. *Out-of-Sample Evaluation of Principal Component Factor Selection Rules.* This section evaluates how the selection rules discussed in section 1.3 perform out-of-sample. Figure 2 reports the out-of-sample performance of the different selection methods across K – for clarity, I only compare selection based on eigenvalues with the optimal selection rule for the respective criteria derived in 1.3. Table 12 in the appendix reports the performance for $K = 1, 3, 5$ corresponding to the number of factors often encountered in the literature.

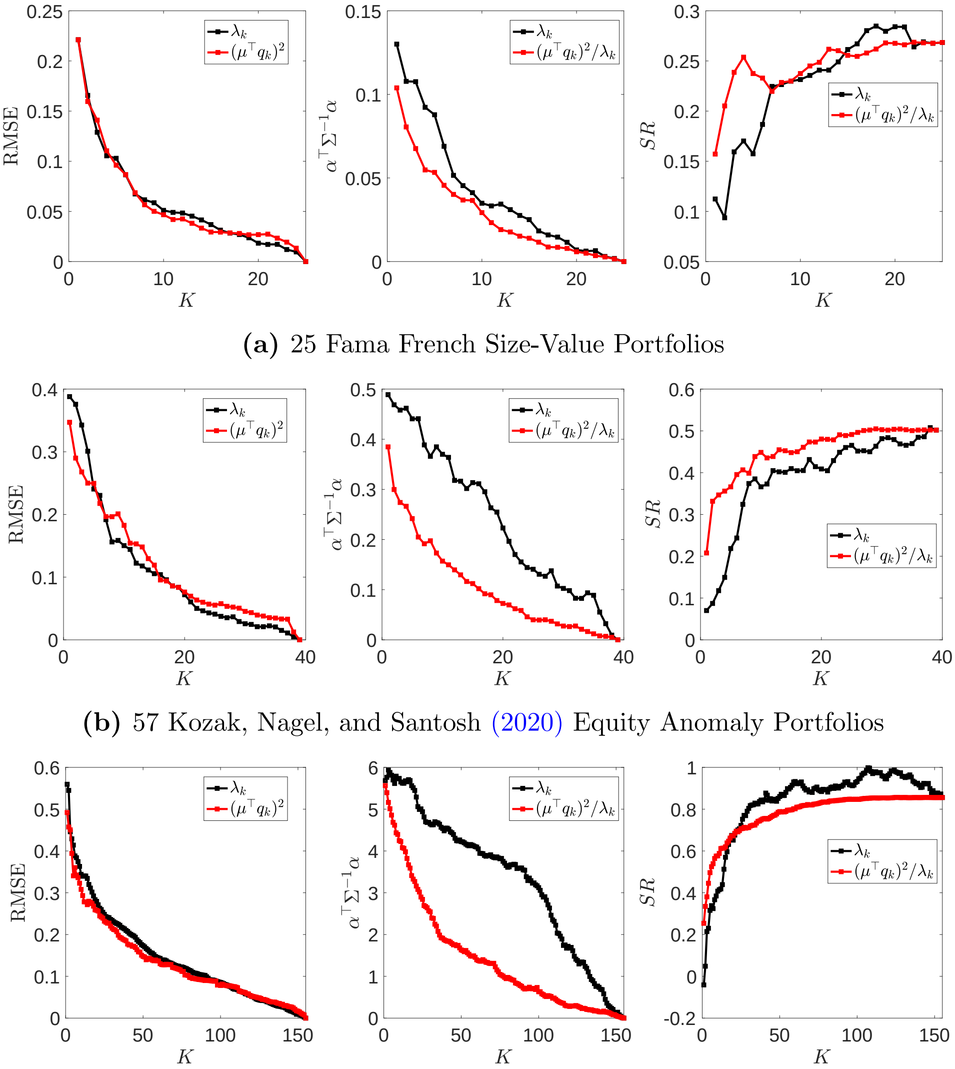
Focussing on RMSE, note that as $K \to N$ the out-of-sample RMSE goes to zero. This is because the rolling eigenvectors, $Q_{K,t}$, form an orthonormal basis[10]. Hence, as before what matters is the speed with which the different selection rules decrease pricing errors relative to each other. For smaller K, selection of principal component factors based on squared risk premia typically leads to improvements over selection based on eigenvalues in terms of out-of-sample RMSE (see table 12). These gains are however not as large or consistent as for the in-sample results and as illustrated in figure 2 do not persist across all K. Overall, selection based on squared risk premia and eigenvalues leads to comparable performance in terms of out-of-sample RMSE.

Turning to the HJ-distance, selection of factors based on squared Sharpe ratios leads to consistent and substantial improvements over selection based on eigenvalues across all K and datasets considered. Comparing the out-of-sample results with their in-sample

[9] For FF25 this leaves 25 portfolios covering July 1926 to November 2023, for KNS57 this leaves 39 portfolios covering July 1963 to December 2019 and for CZ212 this leaves 155 portfolios covering July 1973 to December 2022.

[10] As $K \to N$ it follows that $Q_{K,t}Q_{K,t}' \to I_N$ for every t, so $\epsilon_{t+1} = R_{t+1} - Q_{K,t}Q_{K,t}' R_{t+1} \to 0$ for every t. Therefore, both out-of-sample RMSE and the HJ-distance converge to zero as $K \to N$.

Figure 2. Full OOS Performance



(a) 25 Fama French Size-Value Portfolios

(b) 57 Kozak, Nagel, and Santosh (2020) Equity Anomaly Portfolios

(c) 212 A. Y. Chen and Zimmermann (2022) Equity Anomaly Portfolios

Note: The figure depicts the out-of-sample RMSE, the HJ-distance and Sharpe ratio of R^*_{t+1} for factor models subsequently increasing K, the number of principal components selected. For each test asset set the principal components estimated on the training data are extracted and sorted based on λ_k, $(\mu' q_k)^2$ or $(\mu' q_k)^2/\lambda_k$ to construct the factor models and the tangency portfolio weights. See section 1.5.1 for details on the data and the estimation of out-of-sample statistics.

counterparts in figure 1 shows that the result derived in section 1.3.2 do not fully generalize out-of-sample. For example, on the KNS57 dataset the selection of factors based on conditional squared Sharpe ratios, estimated using 20-year rolling windows, does not ensure that the out-of-sample HJ-distance decreases strictly. Nevertheless, doing so delivers significant out-of-sample improvements over selection based on eigenvalues estimated over the same windows.

In terms of the out-of-sample Sharpe ratio of the tangency portfolio selection of factors based on squared Sharpe ratios leads to sizeable gains over selection based on eigenvalues, especially for smaller K. These gains are economically meaningful. For $K = 3$, across datasets selection based on squared Sharpe ratios improves the monthly out-of-sample Sharpe ratio of the tangency portfolio by 0.08-0.229 relative to selection based on eigenvalues (see table 12). On all except the CZ212 dataset, these gains typically persist. For the latter, selection based on squared Sharpe ratios outperforms selection based on eigenvalues until $K = 22$. As for RMSE and the HJ-distance, this shows that the in-sample results do not fully carry over out-of-sample. To be precise, the selection rules when implemented using 20-year rolling window estimation do not retain the same optimality guarantees out-of-sample as they do in-sample.

Still, selection of principal component factors based on squared Sharpe ratios leads to substantial and consistent improvements in the out-of-sample HJ-distance for all K and to economically significant improvements in terms of the out-of-sample Sharpe ratio of the tangency portfolio for most K relative to selection based on eigenvalues. Figure 14 in the appendix shows that these results are robust when decreasing the frequency with which parameters are estimated from monthly to once every 2 years. In practice, especially if the goal is to construct a low-dimensional factor models from principal component factors, the results here further support selection of factors based on squared Sharpe ratios.

1.5.3. *Number of Factors.* The previous sections evaluated factor model performance for various numbers of factors K, however, in practice an estimator for the number of factors to include in a model is required. Several estimators for the number of factors have been suggested in the large dimensional factor literature[11]. These focus on determining the number of factors such that *variation* in the return data is captured. I here present an estimator that focuses on determining the number of factors required to approximate the SDF.

For every factor model including K factors the associated HJ-distance can be calculated, which measures how close the factor model is to the set of admissible SDFs. Using the results established in theorem 3, I can test if the HJ-distance is statistically significantly different from zero. Ideally, factors are added until the test can no longer be rejected. To achieve the maximum reduction in the HJ-distance possible, section 1.3 demonstrated that factors should be sorted based on their squared Sharpe ratios. Intuitively, this examines how many factors are required to price the cross-section and approximate the SDF sufficiently well.

The estimator proceeds as follows. Sort principal component factors based on their squared Sharpe ratio and construct factor models subsequently increasing the number of factors until $N-1$ – the case of N factors is excluded as it will trivially price all assets perfectly. For each factor model test whether the HJ-distance is zero and compute the respective p-value (see theorem 3) which leaves a list of p-values, p_k. The *simple stop* estimator is then defined as

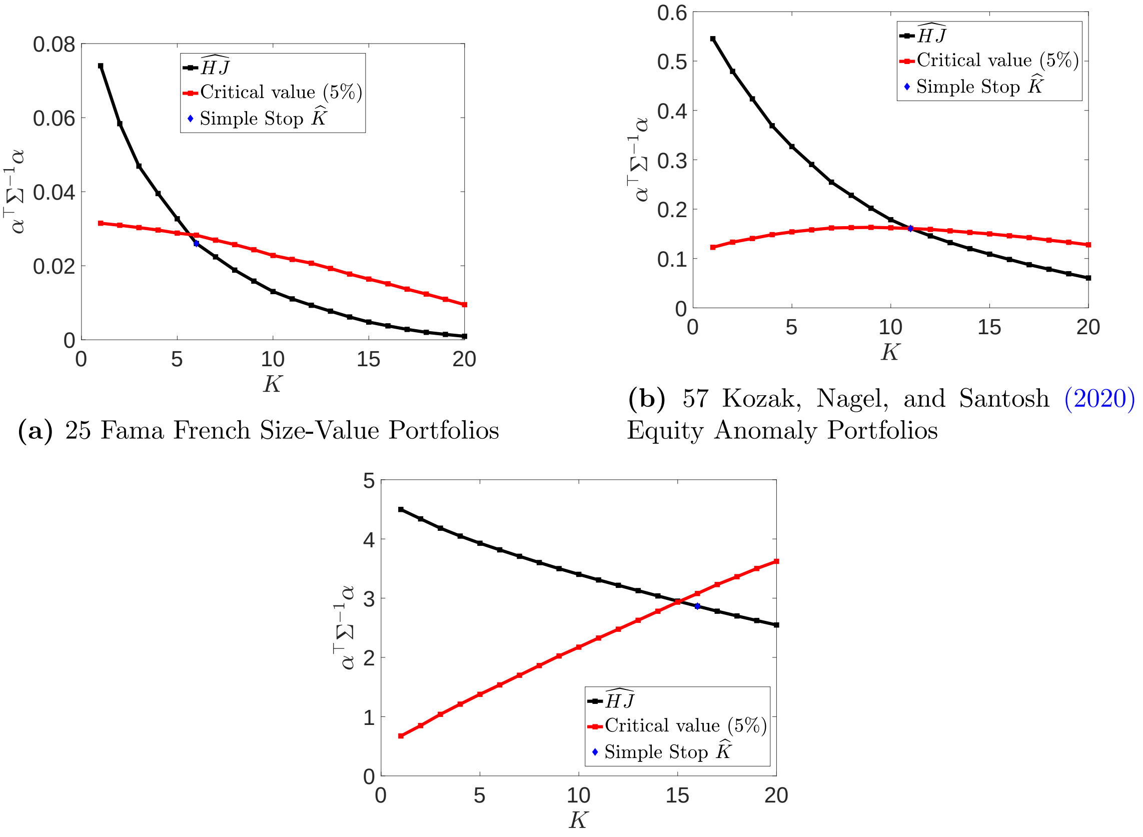
$$\hat{K} = \max\{k \in \{1, ..., N-1\} : p_k \leq a\} + 1 \tag{9}$$

and corresponds to the first k such that the HJ-distance is no longer significantly different from zero at level a, i.e. it simply stops the first time $p_k > a$. Figure 3 illustrates the estimator by plotting the HJ-distance and associated critical values for various K and further marking the simple stop estimator $\hat{K}$ for the 5% significance level across different datasets.

The proposed estimator repeatedly tests the hypothesis that the HJ-distance is zero and only tests a K factor model if the factor models with 1 through $K-1$ factors have been sequentially rejected. Hence, the estimator is a multiple-testing problem in which

[11] See for example Bai and Ng (2002), Onatski (2010), Ahn and Horenstein (2013) and Freyaldenhoven (2022).

Figure 3. Number of Factors



(a) 25 Fama French Size-Value Portfolios

(b) 57 Kozak, Nagel, and Santosh (2020) Equity Anomaly Portfolios

(c) 212 A. Y. Chen and Zimmermann (2022) Equity Anomaly Portfolios

Note: The figure shows the HJ-distance and corresponding critical values for factor models subsequently increasing K, the number of principal components selected sorted by their squared Sharpe ratio. Critical values are calculated based on theorem 3 where covariance matrices are estimated using Newey and West (1987) considering up to $T^{1/4}$ lags. In blue the *simple stop* estimator $\hat{K}$ in equation (9) is highlighted. See section 1.5.1 for details on the data.

hypotheses are rejected in an ordered fashion. It therefore is prone to a false discovery problem. Restricting each individual test's type-I error rate to be a does not control the overall probability of false discoveries. A solution is to instead control the probability of one or more false rejections – the *family-wise error rate* (FWER). The *strong stop* procedure by G'Sell et al. (2016) is designed to control the FWER at level a for settings in which hypotheses are rejected sequentially. The procedure takes the sequential p-values as its input and the estimator is defined as

$$\hat{K}_S = \max\left\{k \in \{1, ..., N-1\} : \exp\left(\sum_{j=k}^{N-1} \frac{\log(p_j)}{j}\right) \leq \frac{ak}{N-1}\right\} + 1 \tag{10}$$

The estimator ensures that $P(\hat{K}_S > K) \leq a$, i.e. ensures the probability of one or more false rejections is below a (see G'Sell et al. (2016))[12]. In the context at hand, the strong stop estimator therefore ensures that the probability of the estimated number of factors being above the true number of factors is below a.

[12] When the p-values under the null are drawn from a uniform distribution, independent to each other, the strong stop procedure controls FWER for a given a. This is the case if the last $N - 1 - K$ p-values are null.

Table 2. Number of Factors

	Dataset		
	FF25	KNS57	CZ212
Simple Stop	6	11	16
Strong Stop	4	8	10

Note: The table depicts the simple stop estimator given in (9) and the strong stop estimator given in (10) for the number of factors at the five percent level for different datasets.

Table 2 depicts results for the simple and strong stop estimator across the different datasets. The simple stop estimator suggests that across datasets 10-28% of total principal component factors are required to approximate the SDF sufficiently well. Comparing the results of the simple with the strong stop estimator, the multiple testing adjustment can be substantial. Focussing on the strong stop estimator, 6-21% of total principal component factors are required to approximate the SDF sufficiently well. Overall, this suggests that once principal component factors are sorted by squared Sharpe ratios relatively few factors are required to approximate the SDF for the datasets considered.

1.5.4. *Comparison to Other Methods.* I here benchmark the results of this paper against other approaches in the literature and highlight complementarities. Specifically, I focus on the methods proposed in Kozak, Nagel, and Santosh (2020) and Lettau and Pelger (2020b). For exposition, I focus on the KNS57 dataset as it is the dataset considered in the two studies. Results for the other datasets are reported in the appendix.

1.5.4.1. Comparison with Kozak, Nagel, and Santosh (2020). The objective of this paper is related to Kozak, Nagel, and Santosh (2020). They focus on constructing a robust SDF, which they define as a SDF that prices assets well out-of-sample in a high-dimensional setting. They achieve this by estimating the prices of risk of the SDF projected onto the test data under L^1- and L^2-penalization, i.e. elastic net estimation. If the SDF is projected onto the principal components of the data their estimator has a closed-form solution, which assuming $\mu' q_k > 0$ for all k is given by

$$M^*_{t+1} = 1 - b'_{KNS} Q'(R_{t+1} - \mu)$$
$$b_{k,KNS} = \begin{cases} \frac{q'_k \mu - \gamma_1}{\lambda_k + \gamma_2} & \text{if } q'_k \mu \geq \gamma_1 \\ 0 & \text{if } q'_k \mu < \gamma_1 \end{cases} \tag{11}$$

where γ_1 and γ_2 control the degree of L^1- and L^2-penalties respectively. The L^2-penalty shrinks prices of risk towards zero. The L^1-penalty performs selection of principal component factors based on risk premia and further shrinks prices of risk.

Kozak, Nagel, and Santosh (2020) motivate the L^2-penalization through priors on the maximum attainable Sharpe ratio in the economy, however, give little justification for the L^1-penalization. The results of this paper are complementary. The law of one price implies that principal component factors should be selected based on risk premia to ensure the factor model and associated SDF approximation minimize squared pricing errors (see equation (4))[13].

[13] There is a subtle difference between the L^1-penalty and the derived selection rule from (4). The law of one price would imply that to minimize squared pricing errors principal component factors with the largest (squared) risk premia should be selected. The L^1-penalty leads to an identical selection rule but in addition subtracts γ_1 after the selection, which further shrinks prices of risk towards zero. The selection

To compare the performance of the factor models constructed from principal component factors selected based on squared risk premia or Sharpe ratios with the Kozak, Nagel, and Santosh (2020) estimator, I implement the elastic net estimator for principal components given in (11). I vary their L^1-penalty so that K factors are selected and determine their L^2-penalty via 3-fold cross-validation on either the full sample (for in-sample results) or 20 years of rolling training data (for out-of-sample results). Results are compared across varying K. As the elastic net estimator is focussed on estimating regularized prices of risk, $\hat{b}_{k,KNS}$, it is straightforward to construct the return on the associated tangency portfolio as $R^*_{t+1} = \hat{b}'_{KNS}\hat{Q}'R_{t+1}$. The estimator however does not map into an associated factor model. Pricing errors therefore are calculated based on the SDF moment condition $\hat{\alpha} = 1/T\sum_t \hat{M}^*_{t+1}R_{t+1}$. For out-of-sample exercises all parameters in (11) – b_{KNS}, Q, and μ – are estimated on the 20 years of rolling training data.

Figure 4a compares the Kozak, Nagel, and Santosh (2020) estimator in-sample against principal component based factor models with factors selected based on either squared risk premia or Sharpe ratios for the previously considered evaluation criteria for varying K. Principal component factor models with factors selected based on squared risk premia lead to substantially lower RMSE. Similarly, factor models with factors selected based on squared Sharpe ratios lead to consistent improvements over the Kozak, Nagel, and Santosh (2020) estimator in terms of HJ-distance and the Sharpe ratio of the associated tangency portfolio across all K. Table 11 illustrates that these gains are economically sizeable even for small K and figure 15 demonstrates that these results hold for the other considered datasets.

This is expected as the elastic net estimator by Kozak, Nagel, and Santosh (2020) shrinks the prices of risk and therefore produces a biased estimate of the SDF. The bias becomes visible in figure 4a: even as $K \to N$ the elastic net estimator does not reduce pricing errors to zero. This bias leads to a deterioration of the in-sample performance in exchange for potentially boosting out-of-sample performance.

Figure 4b presents the same statistics out-of-sample. Focussing on RMSE, principal component factor models with factors selected based on squared risk premia outperform the Kozak, Nagel, and Santosh (2020) estimator across all K out-of-sample. Similarly, selection based on squared Sharpe ratios outperforms in terms of out-of-sample HJ-distances across all datasets and approximation degrees. This is accompanied by consistent and substantial increases in the out-of-sample Sharpe ratio for the tangency portfolio, even for small K (see table 12). Other datasets give rise to similar results (see figure 15).

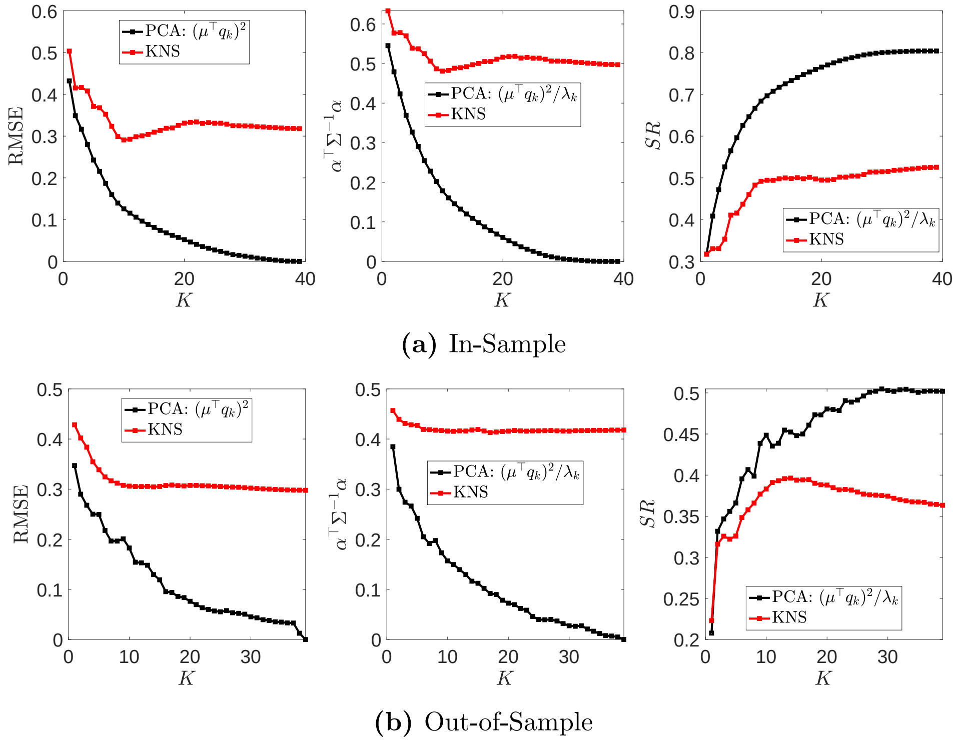
To conclude, the results in section 1.3.1 are complementary to the Kozak, Nagel, and Santosh (2020) estimator as they can formally motivate the use of their L^1-penalty. Having said so, simple factor models constructed from principal component factors selected based on squared risk premia or Sharpe ratios outperform the Kozak, Nagel, and Santosh (2020) estimator consistently across approximation degrees K in-and out-of-sample.

1.5.4.2. Comparison with Lettau and Pelger (2020b). The risk premium PCA (*RP-PCA*) by Lettau and Pelger (2020b) estimates factors by solving a constrained PCA problem that simultaneously minimizes the squared pricing errors. To construct their factors, the eigendecomposition is applied to a modified covariance matrix

$$\Sigma_{RP} = \Sigma + (1+\gamma)\mu\mu' \tag{12}$$

rule derived in section 1.3 is a type of hard-thresholding, while the Kozak, Nagel, and Santosh (2020) L^1-penalty is a type of soft-thresholding.

Figure 4. Performance Comparison with Robust SDF Estimator



(a) In-Sample

(b) Out-of-Sample

Note: The figure depicts the in-sample and out-of-sample RMSE, HJ-distances and Sharpe ratios of the tangency portfolio for factor models subsequently increasing K, the number of factors, constructed from principal component factors selected by squared risk premia $(\mu' q_k)^2$ or squared Sharpe ratios $(\mu' q_k)^2/\lambda_k$ and the Kozak, Nagel, and Santosh (2020) estimator. In-sample results are constructed using the full dataset, whereas out-of-sample results are constructed using a 20-year rolling window. Data corresponds to the Kozak, Nagel, and Santosh (2020) 57 anomaly long-short portfolio returns which are at monthly frequency.

where γ is a hyperparameter penalizing pricing errors (see Lettau and Pelger (2020b)). Let $\Sigma_{RP} = \tilde{Q}\tilde{\Lambda}\tilde{Q}'$ denote the eigendecomposition, where $\tilde{Q}$ is the collection of eigenvectors and $\tilde{\Lambda}$ is a diagonal matrix containing the eigenvalues.

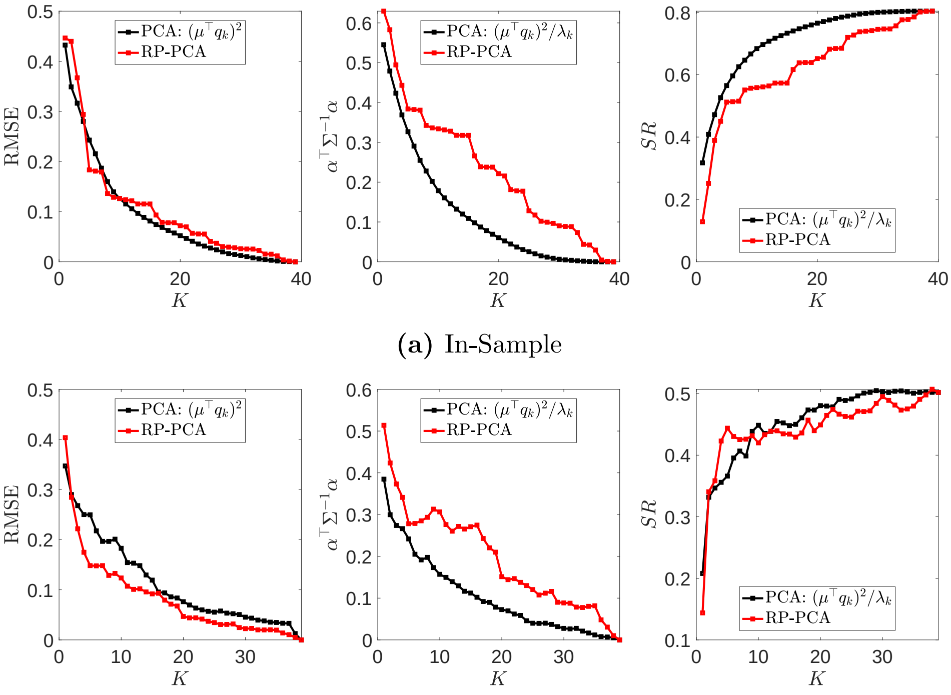
The RP-PCA estimator is concerned with the construction of factors well suited for cross-sectional asset pricing. This paper is concerned with optimal factor selection and it turns out the results are complementary. In appendix 4.3 I demonstrate that similar selection rules as for the standard principal component factors hold for the RP-principal component factors. Specifically, I show that factor selection based on squared Sharpe ratios, $(\mu'\tilde{q}_k)^2/\tilde{\lambda}_k$, is optimal for the HJ-distance and the Sharpe ratio of the tangency portfolio associated with a factor model constructed from RP-principal component factors[14]. A similar result suggests that selection of RP-principal component factors based on squared risk premia, $(\mu'\tilde{q}_k)^2$, is approximately optimal for squared pricing errors.

To compare the performance of the factor models constructed from standard principal component factors with the RP-principal component factors and across different factor selection methods, I implement the RP-PCA following Lettau and Pelger (2020b). I fix $\gamma = 10$ and unless specified differently select RP-principal component factors based on eigenvalues, which corresponds to their baseline specification. Factors are extracted as $\tilde{F}_{k,t} = \tilde{q}_k' R_t$ where $\tilde{q}_k$ is an eigenvector obtained from the eigendecomposition of (12).

[14] Note that for the RP-PCA $(\mu'\tilde{q}_k)^2/\tilde{\lambda}_k$ no longer corresponds to the squared Sharpe ratio as $\tilde{\lambda}_k \neq Var(\tilde{F}_{k,t})$ where $\tilde{F}_{k,t} = \tilde{q}_k' R_t$. However, for $\gamma = 10$ the difference is negligible for most factors, so $(\mu'\tilde{q}_k)^2/\tilde{\lambda}_k \approx \tilde{SR}_k^2$.

Using the factors, betas are estimated through regression. In-sample pricing errors and the return on the tangency portfolio are constructed in line with section 1.5.1. To construct out-of-sample quantities, the relevant parameters are estimated over a 20-year rolling window, and pricing errors and the return on the tangency portfolio are then constructed in line with section 1.5.1.

Figure 5. Performance Comparison with Baseline RP-PCA



(a) In-Sample

(b) Out-of-Sample

Note: The figure depicts the in-sample and out-of-sample RMSE, HJ-distances, and Sharpe ratios of the tangency portfolio for factor models subsequently increasing K, the number of factors, constructed from principal component factors selected by squared risk premia $(\mu' q_k)^2$ or squared Sharpe ratios $(\mu' q_k)^2/\lambda_k$ and the baseline RP-PCA factor model. In-sample results are constructed using the full dataset, whereas out-of-sample results are constructed using a 20-year rolling window. Data corresponds to the Kozak, Nagel, and Santosh (2020) 57 anomaly long-short portfolio returns which are at monthly frequency.

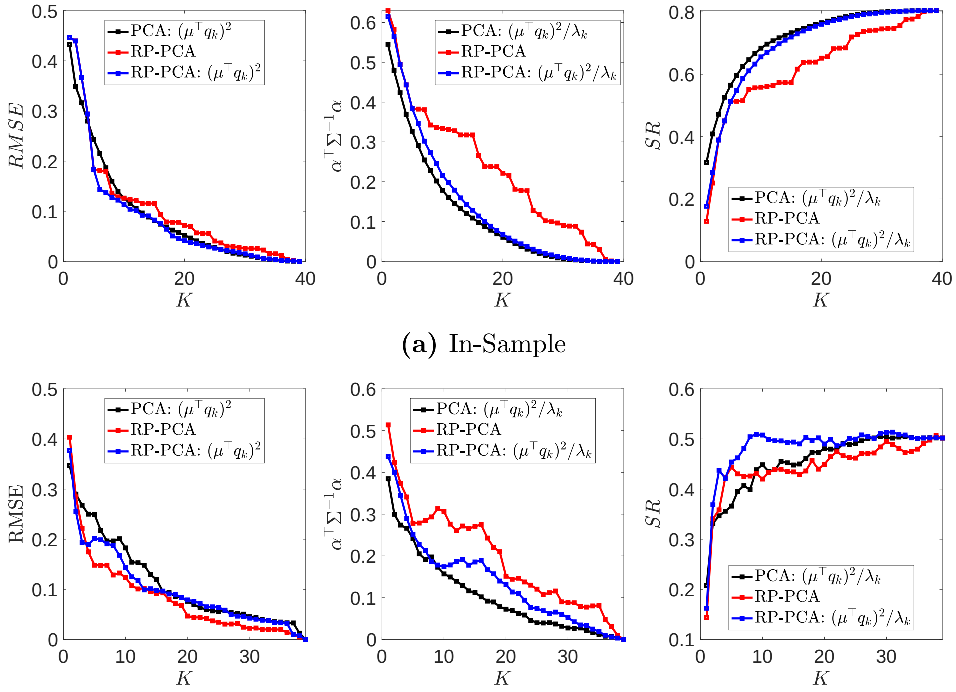
Figure 5a compares the baseline RP-principal component factor model with factor models constructed from standard PCA and factors selected based on either their squared risk premia or Sharpe ratios in-sample. Focussing on RMSE, the RP-PCA approach performs comparable to the standard PCA based model. Principal component factor models with factors selected based on their squared Sharpe ratio outperform the RP-PCA approach in terms of HJ-distance and Sharpe ratio of the tangency portfolio. These improvements are economically sizeable for small K as illustrated in table 11. Other datasets give rise to similar results (see figure 16).

Figure 5b presents the results out-of-sample. The baseline RP-PCA improves over simple principal component factor based models with factors selected based on squared risk premia in terms of out-of-sample RMSE. Principal component factor models with factors selected based on their squared Sharpe ratio lead to substantial and consistent improvements in terms of out-of-sample HJ-distance over the baseline RP-PCA. For all except the CZ212 dataset, the out-of-sample Sharpe ratio of the tangency portfolio is comparable (see figure 16). For the latter, the RP-PCA based models lead to substantial gains in the out-of-sample Sharpe ratio of the tangency portfolio, particularly for small

K (see table 12). Overall the baseline RP-PCA appears to improve the factor models' performance in terms of RMSE, whilst standard principal component factor models with factors selected based on squared Sharpe ratios demonstrate better performance in terms of HJ-distances and mostly comparable performance in terms of the Sharpe ratio of the tangency portfolio.

The performance of the RP-PCA can be further boosted by adopting the discussed factor selection rules. Figure 6a demonstrates this in-sample. For comparability, the performance of the standard PCA with factors selected based on either their squared risk premia or Sharpe ratios are included. Selecting RP-principal component factors based on their squared risk premium further boosts the performance of the factor model in terms of in-sample RMSE. The figure also illustrates that selection based on squared risk premia is only approximately optimal as the largest decreases in RMSE do not one-to-one line up with the ranking based on squared risk premia (see appendix 4.3 for further details). Selecting RP-principal component factors based on squared Sharpe ratios leads to an improvement of the HJ-distance and the Sharpe ratio of the associated tangency portfolio with performance similar to the factor models constructed from standard principal component factors. Other datasets give rise to similar results (see figure 17).

Figure 6. Performance Comparison with Baseline and Optimally Selected RP-PCA



(a) In-Sample

(b) Out-of-Sample

Note: The figure depicts the in-sample and out-of-sample RMSE, HJ-distances and Sharpe ratios of the tangency portfolio for factor models subsequently increasing K, the number factors, constructed from standard or RP-principal component factors selected by squared risk premia $(\mu' q_k)^2$ or squared Sharpe ratios $(\mu' q_k)^2/\lambda_k$ and the baseline RP-PCA factor model. RP-PCA methods set $\gamma = 10$. In-sample results are constructed using the full dataset, whereas out-of-sample results are constructed using a 20-year rolling window. Data corresponds to the Kozak, Nagel, and Santosh (2020) 57 anomaly long-short portfolio returns which are at monthly frequency.

Figure 6b presents the out-of-sample results. Focusing on RMSE, selecting RP-principal component factors based on risk premia does not seem to improve performance for the

KNS57 dataset. Figure 17 illustrates that on the other datasets selection based on squared risk premia leads to marginal improvements (for FF25 performance is unaffected, whilst for CZ212 performance slightly improves). For the HJ-distance, selecting RP-principal component factors based on Sharpe ratios improves the performance of the RP-principal component factor model, however, the factor models based on standard principal component factors selected based on squared Sharpe ratios continue to outperform. Finally, the out-of-sample Sharpe ratio of the RP-principal component factor model generally improves by selecting factors based on squared Sharpe ratios and consistently outperforms the Sharpe ratio of the tangency portfolio associated with a factor model based on standard principal components selected by squared Sharpe ratios. The other datasets give rise to similar observations (see figure 17).

The results of this paper are therefore complementary to the RP-PCA by Lettau and Pelger (2020b). The selection rules derived in section 1.3 are optimal for the RP-PCA (see section 4.3 for details) and lead to improvements in its in- and out-of-sample performance. Interestingly, standard principal component based factor models with factors selected by squared Sharpe ratios generally outperformed the RP-PCA in terms of the HJ-distance in- and out-of-sample.

1.6. Conclusion. This paper examines the selection of principal component factors for the construction of the SDF and by extension factor models describing the cross-section of expected returns. Starting from the law of one price, I show that factor selection based on eigenvalues is typically suboptimal for asset pricing. Specifically, if the goal is to select factors such that the associated factor model's sum of squared pricing errors (HJ-distance) is minimized, factors associated with the largest squared risk premia (Sharpe ratios) should be selected. The intuition for this result is that the latter criteria are linked to principal component factor's prices of risk and hence inform us whether a given factor is useful in pricing other assets. Empirically, this leads to substantial improvements in the factor model's performance in- and out-of-sample and once ranked by Sharpe ratios it appears few principal components are sufficient to approximate the SDF adequately.

Having said so, factor selection will always depend on the objective of the empirical researcher. If the goal is to obtain a lower-dimensional representation capturing the variation of the data, selection should be carried out based on eigenvalues. The contribution of this paper is that if the objective is to identify factors relevant for pricing the cross-section of expected returns other criteria are optimal. Specifically, if the goal is to approximate the SDF well, minimize the weighted sum of squared pricing errors, or maximize the Sharpe ratio of a factor model's associated tangency portfolio, selection based on Sharpe ratios is recommended.

Chapter 2.
The Network Drivers of Trade Currency Invoicing

Co-authored with Tommaso Mancini-Griffoli, Christian Julliard, and Kathy Yuan

Using an equilibrium network model and a large international panel of cross-border trade, we analyse empirically the drivers of foreign currency invoicing. First, we find strong evidence of strategic complementarity in currency invoicing across countries: Exporting countries tend to invoice more in a given currency when their main trade partners invoice in that same currency. This in turn leads to an amplification of domestic shocks through the trade network. Second, key players for a given currency are not only countries that invoice most of their exports in that foreign currency (e.g., China, South Korea and Russia), but also countries that are central in the international trade network (e.g., Japan, Germany and Canada). Third, at the country-level, we find evidence of strategic complementarity, or natural hedging, between the choices of export and import currencies. Fourth, in counterfactual analysis, we find that, due to the large network externalities that we identify, the position of the USD as dominant trade currency is inherently fragile with respect to the currency invoicing choices of EU and BRICS countries.

2.1. Introduction. It is well documented that the vast majority of international trade is invoiced in a small number of dominant currencies, with the USD playing an outsized role (Goldberg and Tille (2016); Gopinath (2015)). Dominant currency pricing has significant implications for monetary policy spillovers, transaction costs, and financial market development Gopinath, Boz, et al. (2020). In this paper, in addition to the factors proposed by the literature on dominant currency, we examine whether the currency invoicing decisions of the firms in a country are affected by those of its trading partners.

Specifically, in the theoretical model, a representative firm in a country chooses the size of invoicing in a dominant currency for international trade transactions based on a cost-benefit analysis. The cost and benefit, as identified by the literature, can be due to a multitude of factors. For example, the interaction of nominal price stickiness with pricing complementarities and input-output linkages across firms generates complementarities in currency choice (Gopinath (2015); Doepke and Schneider (2017); Mukhin (2022); Eren and Malamud (2022)). That is, exporters coordinate on the same currency of invoicing for the following two reasons: to be competitive in output pricing; and to be able to hedge their balance sheet against exchange rate shocks with the denominated currency of imported intermediate (real and financial) inputs. This indicates that the size of the market is an important driver of the dominant currency Mukhin (2022), as well as various price index levels and other macroeconomic and financial variables at the country-level. Additionally, a large body of literature on international finance emphasizes the safety feature of the dominant currencies. Differences in financial development, and hence the differences in access to safe assets Maggiori (2017), or risk aversion of participants Gourinchas and Hélène Rey (2022) may drive the demand for an international safe asset. The dominant currency preserves value added in exchange transactions, leading to its wide use in the global financial market.[15] This indicates that safety/volatility of the currency is potentially a major determinant of its dominant status. Importantly, the network of international trade underlies how these factors affect each country's invoicing decisions. The network of international trade not only captures the potential network effects of neighbouring countries in choosing a dominant currency for pairwise transactions (for stable transactions or cheaper access to working capital or financial borrowing) but also reflects the input-output linkages across countries. Furthermore, the trade network might also give rise to the need for balance sheet currency hedging: A representative firm in a country that is invoiced in a certain dominant currency in its imports has more incentive to invoice their exports in the same currency to hedge its currency risk exposure. In our paper, we use a network model to capture these trade-offs.

In the model, the currency invoicing decision of a representative firm in a country is affected not only by its own economic conditions, such as the size of the economy, inflation, financial market conditions, and other relevant economic variables, but also the invoicing decision of its trading partners and its trading partner country's economic variables. Similarly, the currency invoicing decisions of its trading partners are affected by those of their own trading partners, as well as their economic variables, and so forth. In equilibrium, we show that this network dependency is captured by a network attenuation factor ϕ, the key parameter whose sign determines whether the Nash equilibrium features strategic substitution ($\phi < 0$) or complementarity ($\phi > 0$) in agents' invoicing decisions. Each agent's equilibrium invoicing amount in the dominant currency depends not only on the network attenuation factor ϕ but also on its network centrality measure.

We estimate the equilibrium based on four different sets of restrictions on the model parameters, that is, four models – Panel, Spatial Error (SEM), Spatial Lag (SLM) and Spatial Durbin (SDM) – using Bayesian methods. For example, setting $\phi = 0$ in the equilibrium condition yields a simple panel structure for the data. We use a Bayesian procedure for model specification and assess whether the data support the presence of network externalities and if so, which spatial specification. Our empirical analysis focusses on excessive USD or EUR invoicing for each country to assess the use of the USD or EUR as a vehicle currency in international trade.[16] The measure is constructed at monthly frequency based on the payment share dataset by Boz et al. (2022) and the Direction of Trade Statistics database by the International Monetary Fund. We further augment the payment share dataset with a proprietary dataset obtained from the Society for Worldwide Interbank Financial Telecommunications (SWIFT) to increase the cross-sectional coverage. The final dataset contains 84 countries from January 2004 to December 2019 and covers on average 91% (93%) of worldwide exports (imports).

Our analysis shows that there is overwhelming evidence of network spillovers: The panel specification with no spatial dependency is never preferred by the data. There is strong evidence of strategic complementarity in currency invoicing across countries: Exporting countries tend to invoice more in a given currency when their main trade partners invoice in that same currency. This in turn leads to an amplification of domestic shocks through the trade network. In fact, the SDM model – the specification of our theoretical formulation

[15] Furthermore, Gopinath and Stein (2021) argue that assets denominated in the dominant currency can be used as a savings device for export producers to hedge against invoicing risk. Chahrour and Valchev (2022) additionally suggest that safe assets are used as collateral to overcome contractual frictions in cross-border transactions.

[16] For robustness, we also use aggregate currency invoicing amounts in USD or EUR. The results are similar and reported in the appendix.

– is always strongly preferred by the data. We conduct analysis based on the SDM specification and include among the control variables the lagged values of the dependent variables to capture time series autocorrelation. Therefore, the model not only captures the contemporaneous, or short-term, impact of a shock originating from any of the independent variables but also their long-term effects. Due to the network specification we are further able to decompose these effects into direct and indirect effects, the latter being impacts originating from trade partners propagated through the trade network. The network attenuation factor for USD excessive invoicing is around 0.24, while that of the EUR is 0.16. This indicates that USD excessing invoicing is inherently less stable: A small negative shock might lead to a substantial reduction in the use of the USD as a vehicle currency.

We find that key players for a given currency are not only countries that invoice most of their exports in that foreign currency (e.g., China, South Korea, and Russia) but also countries that are central in the international trade network (e.g., Japan, Germany, and Canada). The driver of the former set of countries is based on their large direct impacts, while that of the latter set of countries is due to the network amplification and to the central position that these countries have in the trade network.

Furthermore, at the country-level, we find evidence of *strategic complementarity* between exports and imports in a given currency, lending support to the natural hedge hypothesis proposed in the literature (Doepke and Schneider (2017); Amiti, Itskhoki, and Konings (2022)). This investigation is based on a reduced-form Vector Autoregression (VAR) in our SDM specification with four dependent variables (and their respective controls): Excessive currency invoicing in EUR and USD of *both* export and import. We employ identification via cross-sectional heteroskedasticity. That is, we exploit heterogeneity across residuals' covariances of our panel dataset to pin down structural parameters. The key identification assumption this requires is that all countries have identical contemporaneous reactions to USD or EUR import- or export-based shocks.

Finally, we conduct a counterfactual analysis to examine the impact of a set of countries choosing to abandon the USD for excessive invoicing. We conduct this exercise for Russia, Brazil, India and China, the EU block, and the members of BRICS in our sample, i.e. Brazil, Russia, India and China jointly. The estimated effects are quantitatively large, with the effects of the BRIC(S) block (EU) abandoning the USD for excess invoicing resulting in a 42% (11%) reduction in the overall use of this currency. More so, the channels through which these reductions arise are quite different. For the BRIC(S) block, most of the effect is driven by the direct reduction in the use of the USD by these countries. For the EU, almost half of the effect is driven by indirect network externalities. This underlines the fragility of a dominant trade currency such as the USD – coordinated abandonment can have substantial impacts on the overall use of a dominant currency for trade invoicing, specifically as the network externalities highlighted in this paper lead to an amplification.

The remainder of this paper is organized as follows. In Section 2.2, we review the related literature. In Section 2.3, we present a network model to guide our analysis of currency choice for invoicing. In Section 2.4 we present our estimation methodology. In Section 2.5, we describe the data and variable construction. In Section 2.6 we present and discuss the estimation results and conduct the counterfactual analysis. Section 2.7 concludes.

2.2. Related Literature. The current international macro literature has shown that the choice of currency in trade invoicing is an active firm-level decision with some degree of persistence over time Amiti, Itskhoki, and Konings (2022). This evidence is contrary to the conventional international literature, which assumes exogenous producer currency pricing (PCP) or local currency pricing (LCP), that is, trades are denominated either in the producer's currency or the importer's currency. Instead, recent studies postulate the existence of dominant currencies for international trade that can endogenously emerge – dominant currency pricing (DCP) – and focus on their determinants and implications for optimal monetary policies with different spillover dynamics.

The DCP literature on international trade proposes that the interaction of nominal price stickiness with pricing complementarities and input-output linkages across firms generate complementarities in currency choice (Gopinath (2015); Doepke and Schneider (2017); Gopinath, Boz, et al. (2020); Mukhin (2022); Eren and Malamud (2022)). That is, exporters coordinate on the same currency of invoicing to be competitive in output pricing and to be able to hedge their balance sheet against exchange rate shocks with the denominated currency of imported intermediate (real and financial) inputs. Financial intermediate inputs can be thought of as working capital, trade credit, or any form of financial borrowing. According to this line of research, important determinants for dominant currencies include the competitiveness or importance of the destination market and the currencies that denominate the intermediate inputs. Exporting firms that import a lot of intermediate products, or borrow capitals in the international markets, are more likely to choose DCP. The theoretical result in Mukhin (2022) also indicates that the market size is important in determining the dominant currency. Our model of the network effects of currency invoicing choices is motivated by this line of research. The complementarities in invoicing currency choice we identify are via the import/export network channel.

Additionally, there are many other determinants for currency choices in trade invoicing. Existing empirical work using transactions of firm-level import or export data shed light on these determinants. Gopinath, Itskhoki, and Rigobon (2010) and Goldberg and Tille (2016) analyze transaction-level data on currency invoicing for, respectively, US and Canadian imports, and find that USD pricing is more common in sectors classified as producing homogeneous goods and hence likely substitutes. Chung (2016) finds that a 1% decrease in the share of imported inputs priced in sterling decreases the probability that UK exporters invoice in sterling by about 18% using UK trade transaction data with non-EU countries. Studying the currency invoicing choices of Belgian exporting firms, Amiti, Itskhoki, and Konings (2022) find that large and import-intensive firms tend to invoice their exports in USD. There is also evidence supporting firms choosing their invoicing currency to hedge their financial input risk. BIS (2014) document that traded financial contracts are mostly USD denominated even though they are sourced through local banks, indicating most trades are financed in USD. Bahaj and Reis (2020) find that when the cost of financing working capital in Renminbi (RMB) is lower due to swap arrangements by central banks, trades are more likely to be denominated in RMB. Furthermore, although invoicing and settlement currencies do not necessarily coincide, in most transaction they are the same currency Gopinath, Boz, et al. (2020). Therefore, the choice of an invoicing currency also depends on its liquidity level. In our estimation, we control for these determinants identified by the trade literature and examine whether these determinants might have additional impact propagated through the trade network.

Finally, recent work in the DCP literature also emphasizes financial frictions in cross-border transactions as an important factor in currency choices. A dominant currency such as the USD preserves its value during global market crises, and thus is widely used as an international reserve or safe asset. This safety feature offered by assets denominated in dominant currency means that the dominant currency preserves value added in exchange transactions, leading to its wide use in global financial market. Differences in financial development or the differences in access to safe asset Maggiori (2017) or risk aversion of participants Gourinchas and Hélène Rey (2022) may drive the demand for an international safe asset. Chahrour and Valchev (2022) propose that safe assets are used as collateral to overcome contractual frictions in cross-border transactions. Gopinath and Stein (2021) argue that assets denominated in the dominant currency can be used as saving devices for export producers to hedge against invoicing risk. In our analysis, we also examine the impact of these financial frictions for dominant currencies at the country-level in relation to the trade network.

2.3. The Network Model. In this section, we construct a network model of currency invoicing decisions for international trade transactions. This framework directly guides our empirical estimation of network effects. In this model, there is a representative firm in each country. They simultaneously decide in which currency to invoice their trades given the import and export trade network structure. For the ease of exposition, the optimizing agent in the model is described as a country (rather than a representation of firms in a country). A country's optimal currency invoicing decision not only depends on its own characteristics but also responds to all other countries' invoicing decisions (and potentially characteristics) through the given trade network. In Section 2.4, we lay out the steps for structurally estimating the model.

The Network. There are n countries. The time-t trade network is predetermined, characterized by an n-square adjacency matrix $\mathbf{G}_t$. If its element $g_{ij\neq i,t}$ are not null, country i and j are connected. To construct $\mathbf{G}_t$ in the structural estimation, we use country-level bilateral trade data. Specifically, we focus on the case where $g_{ij\neq i,t}$ is the fraction of imports or exports by country i from or to country j in the previous month.

The matrix $\mathbf{G}_t$ keeps track of all direct connections – links of order one – between any pair of countries in the network. Similarly, the matrix $\mathbf{G}_t^h$, for any positive integer h, encodes all links of order h between countries, that is, the paths of length h between any pair of countries. The coefficient in the (i, j)th cell of $\mathbf{G}_t^h$, i.e., $\left\{\mathbf{G}_t^h\right\}_{ij}$, gives the exposure of country i to country j in h steps. Since $\mathbf{G}_t$ is a right stochastic matrix (with each row summing up to one), it can be interpreted as a Markov chain transition kernel, and $\mathbf{G}_t^h$ as the h-step transition matrix.

Countries and their Invoice Currency Preference. We study the amount of *excessive* currency countries choose to invoice their trade partners at the beginning of each period t. The amount of excessive currency invoiced for a given currency is defined as any amount above the corresponding bilateral trade volume. For example, it is natural for a country to invoice their trade counterparts located in the US in USD. However, if this country invoices trades with other countries in USD, we define these transactions as excessive USD invoicing. We aim to explain this decision at the country-level.

Let $y_{i,k,t}$ denote the total excessive currency k invoiced by country i. We model the impact of the trade network on country i's choice of $y_{i,k,t}$. In the rest of the paper, we drop the index k for expositional simplicity and often refer to k as the dominant currency (DC).

Given the predetermined network $\mathbf{G}_t$ measured by trade links, each country chooses excessive DC invoicing simultaneously to maximize its own payoff. In our model, we specify the per-unit value $\tilde{\mu}_{i,t}$ of excess DC for country i in network $\mathbf{G}_t$ as

$$\tilde{\mu}_{i,t} := \mu_{i,t} + x_{i,t}\delta + x_{p,t}\rho + \phi \sum_j g_{ij,t} y_{j,t} + \sum_j g_{ij,t} x_{i,j,t} \theta \tag{13}$$

where $x_{i,t}$ is a row vector of country-specific characteristics, $x_{p,t}$ denotes aggregate common controls, $x_{i,j,t}$ captures pair-specific covariates, $\mu_{i,t} := \bar{\mu}_i + \epsilon_{i,t}$ contains a countries specific fixed effect ($\bar{\mu}_i$), and shock ($\epsilon_{i,t}$), δ, ρ and θ are conformable column vectors and ϕ is a scalar coefficient. The first and second set of variables capture the country-level variables that directly affect the value of a unit of DC such as bilateral (local versus the invoicing currency) exchange rate volatility, inflation, and other macro variables. The third set of variables capture aggregate factors that might affect all countries' choices.

The fourth term in equation (13) is a network-dependent component of valuation of extra unit of DC invoicing: $\phi \sum_j g_{ij,t} y_{j,t}$. It is motivated by the DCP hypothesis. According to DCP, firms in country i have a preference for invoicing their exports in the same currency as their imports so to minimize the currency mismatch on their assets and liabilities. Network effects lead to certain currencies such as the USD emerging as the dominant currency for countries to coordinate their invoices.

The last term in equation (13) highlights that the network-dependent mechanism may also operate via the characteristics of the trading partners. Variable $x_{i,j,t}$ denotes match-specific control variables and the characteristics of other countries, and θ is a vector of suitable dimension. That is, in addition to the aggregate information embedded in the neighbouring countries' level of excessive DC invoicing, macro variables and the neighbouring countries' characteristics affect the per-unit valuation of the excessive DC invoicing. For example, suppose that country j has a well-developed financial market denominated in EUR instead of USD. This means that trade financing, work capital financing, and currency risk hedging in EUR would be cheaper and readily available. Firms in country j would have more incentives to invoice their trades in EUR, which can impact the invoicing decision of country i. The DCP empirical literature has identified several such determinants for $x_{i,j,t}$, such as the size of the economy, financial market development, foreign direct investment, foreign debt, and inflation.

Next, assuming a quadratic cost for holding $y_{i,t}$ amounts of DC, we specify country i's utility from DC invoicing as

$$u_i(y_t|g_t) \quad = \quad \tilde{\mu}_{i,t} y_{i,t} - \frac{1}{2} y_{i,t}^2. \tag{14}$$

The bilateral network influences in our model are captured by the following cross-derivatives for $i \neq j$:

$$\frac{\partial^2 u_i(y_{i,t}, \{y_{j,t}\}_{j\neq i} | \mathbf{G}_t)}{\partial y_{i,t} \partial y_{j,t}} = \phi g_{ij,t},$$

where ϕ is the network attenuation factor, the key parameter whose sign determines whether the Nash equilibrium features strategic substitution ($\phi < 0$) or complementarity ($\phi > 0$). We are agnostic about the sign of ϕ, and we instead estimate it empirically.

We solve countries' optimal excessive DC invoicing decision in the Nash equilibrium of simultaneous action. The optimal response function for each country is then

$$y_{i,t}^* = \mu_{i,t} + x_{i,t}\delta + x_{p,t}\rho + \phi \sum_j g_{ij,t} y_{j,t} + \sum_j g_{ij,t} x_{i,j,t} \theta. \tag{15}$$

Note that the empirical counterpart of the above best response is the spatial Durbin model.

Let us denote $\mu_{i,t}+x_{i,t}\delta+x_{p,t}\rho+\sum_j g_{ij,t}x_{i,j,t}\theta$ by $\breve{\mu}_t$. The following result is immediate.

Proposition 1. *Suppose that $|\phi| < 1$. Then, there is a unique interior solution for the individual equilibrium outcome given by*

$$y^*_{i,t}(\phi, g) = \{\mathbf{M}(\phi, \mathbf{G}_t)\}_{i.}\, \breve{\mu}_t, \tag{16}$$

where $\{\}_{i.}$ is the operator that returns the i-th row of its argument, $\breve{\mu}_t := [\breve{\mu}_{1,t}, ..., \breve{\mu}_{n,t}]^\top$, $y_{i,t}$ denotes the total excessive DC invoicing by country i, and

$$\mathbf{M}(\phi, \mathbf{G}_t) \equiv \mathbf{I} + \phi\mathbf{G}_t + \phi^2\mathbf{G}_t^2 + \phi^3\mathbf{G}_t^3 + ... = \sum_{k=0}^{\infty} \phi^k \mathbf{G}_t^k = (\mathbf{I} - \phi\mathbf{G}_t)^{-1}, \tag{17}$$

where I is the $N \times N$ identity matrix.

Proof. The first-order condition identifies the individual country's optimal response. Applying Theorem 1(b) in Calvo-Armengol, Patacchini, and Zenou (2009), we know that the necessary equilibrium condition is $|\phi\lambda^{max}(\mathbf{G}_t)| < 1$, where the function $\lambda^{max}(\cdot)$ returns the largest eigenvalue. Since $\mathbf{G}_t$ is a right stochastic matrix, its largest eigenvalue is 1. Hence, the condition requires $|\phi| < 1$, and if so, the infinite sum in equation (17) is finite and equal to the stated result (Debreu and Herstein (1953)). □

The condition $|\phi| < 1$ states that network externalities must be small enough in order to prevent the feedback triggered by such externalities to escalate without bounds. In vector form, $y_t \equiv [y_{1,t}, ..., y_{N,t}]^\top$, and in equilibrium,

$$y^*_t = \mathbf{M}(\phi, \mathbf{G}_t)\, \breve{\mu}_t \tag{18}$$

Network Propagation. In equilibrium, the matrix $\mathbf{M}(\phi, \mathbf{G}_t)$ contains information about the *centrality* of network players.[17] Multiplying the rows (columns) of $\mathbf{M}(\phi, \mathbf{G}_t)$ by a unit vector of conformable dimensions, we recover the indegree (outdegree) Katz–Bonacich centrality measure. The indegree centrality measure provides the weighted count of the number of ties directed to each node (i.e., inward paths), while the outdegree centrality measure provides the weighted count of ties that each node directs to the other nodes (i.e., outward paths). That is, the i-th row of $\mathbf{M}(\phi, \mathbf{G}_t)$ captures how country i loads on the trade network as whole, while the i-th column of $\mathbf{M}(\phi, \mathbf{G}_t)$ captures how the trade network as a whole loads on country i. Therefore, the trade network centrality of a country affects the dominance status of its currency, potentially also through its characteristics (captured by variables x_i).

2.4. Estimation Method. Making explicit the role of the shocks, $\epsilon_{i,t}$, and country fixed effects, $\bar{\mu}_i$, in the first order condition (15), yields the empirical representation

$$y_{i,t} = \bar{\mu}_i + x_{i,t}\delta + x_{p,t}\rho + \phi\sum_j g_{ij,t}y_{j,t} + \sum_j g_{ij,t}x_{i,j,t}\theta + \epsilon_{i,t} \tag{19}$$

where the covariates, $x_{i,t}$, are contemporaneously independent from the time t shock. Hence, we can also accommodate, among other controls, the lagged value of the both import and export excess currency invoicing. The above formulation is the so-called

[17] This centrality measure takes into account the number of both direct and indirect connections in a network. For more on the Bonacich centrality measure, see Bonacich (1987) and Matthew O. Jackson (2010). For other economic applications, see Ballester, Calvo-Armengol, and Zenou (2006) and Acemoglu et al. (2012). For an excellent review of the literature, see Matthew O Jackson and Zenou (2012).

spatial Durbin model (SDM – see, e.g., LeSage and Pace (2009)). We estimate the model using monthly data, and we include year fixed effects to control for unobserved macro factors. At time t, the network is predetermined and $g_{ij,t}$ is measured by the fraction of country i's imports or exports from or to country j. We include a very broad set of country-, and pair-specific, characteristics suggested in the previous literature (such as existence of swap line, bilateral exchange rate volatility, consumer price index volatility, financial market development, denomination of corporate sector FX liability), and lagged values of the excess currency invoicing of both exports and imports. All control variables are lagged by one period for predeterminancy.

The general formulation in equation (19) nests several more restrictive models considered in the previous literature on network spillovers. For instance, setting the vector θ to zero, i.e. shutting down the direct dependency of country i's outcome variable on the covariates of all other countries, we have a simple spatial lag (SLM) as in Ozdagli and Weber (2023). Furthermore, restricting $x_{i,j,t} = [x_{j,t}, x_{p,t}]$ and $\theta = -\phi[\delta^\top, \rho^\top]^\top$, we have a spatial error model (SEM) as in Denbee et al. (2021). As shown in Bramoullé, Djebbari, and Fortin (2009), the identification conditions for SDM and SLM boils down to the requirement of linear independence of the identity matrix, the adjacency matrix containing the network weights ($g_{ij,t}$), and the square of this matrix, while in SEM identification arises from the implied restriction on the covariance matrix of the error terms.[18]

Note that in all three formulations for the spatial dependency (SDM, SLM, and SEM), assessing the presence of network externalities boils down to testing whether the coefficient ϕ is different from zero, and setting $\phi = 0$ yields a simple panel structure for the data. Frequentist estimation of these models is possible via e.g., (quasi) maximum likelihood and the Generalized Method of Moments (see, e.g. Anselin (1988)). Nevertheless, we opt for a Bayesian procedure, since we aim to select a specification, and assess whether the data support the presence of network externalities (i.e., a $\phi \neq 0$), with a procedure that is robust to model misspecification in that it does not require testing under the null of a correctly specified model. Nevertheless, since we employ flat priors for the parameters,[19] and we assume Gaussianity for the error terms, the posterior modes coincide with the quasi maximum likelihood estimates (a consistent estimator in this setting). When comparing models (the SDM, SLM, SEM, and panel specifications), we assign equal prior probability to each formulation, and posteriors are sampled via the Gibbs sampling procedure detailed in Appendix 5.2.

Furthermore, the estimate of ϕ also reveals the type of equilibrium on the network, i.e., strategic substitution (when $\phi < 0$) or complementarity (when $\phi > 0$). Note also that from equation (15) we have that the conditional covariance of y_t is

$$Var_{t-1}(y_t) = Var_{t-1}\left(\mathbf{M}\left(\phi, \mathbf{G}_t\right)\epsilon_t\right) = \mathbf{M}\left(\phi, \mathbf{G}_t\right)\Sigma_\epsilon \mathbf{M}\left(\phi, \mathbf{G}_t\right)^\top \tag{20}$$

since $\mathbf{G}_t$ is predetermined at time t and $\Sigma_\epsilon \equiv Var(\epsilon_t)$. Hence, the variance is increasing in ϕ: The stronger the degree of strategic complementarity, the larger is the endogenous amplification of shocks to the system, and the higher is the volatility of total excess invoicing in the network. To see this, note that the variance of total excess invoicing is

[18] See Denbee et al. (2021) for a detailed discussion.

[19] We use improper flat priors for δ, ρ, and θ, since these parameters are common across specifications, and consequently the improper prior does not invalidate the posterior model probabilities. For ϕ instead we employ a Gaussian prior and modify the acceptance rate to ensure proper support.

$Var_{t-1}(\mathbf{1}^\top y_t)$, hence, a unit shock equally spread among all N countries has a contemporaneous impact on total excess invoicing equal to $\mathbf{1}_N^\top \mathbf{M}(\phi, \mathbf{G}_t)\,\mathbf{1}_N/N = 1/(1-\phi)$, where $\mathbf{1}_N$ denotes a vector of ones with length N.[20]

2.5. Variable Construction and Data Description. The main focus of our analysis is excessive currency invoicing, $y^x_{i,k,t}$, where i denotes a country, t the time period, k the currency (USD or EUR), and x the trade direction (export or import). To construct the variable we rely on the dataset by Boz et al. (2022) and the Direction of Trade Statistics database by the International Monetary Fund. The former is augmented with data from SWIFT to increase cross-sectional coverage[21] and provides data on the shares of aggregate exports or imports invoiced in USD and EUR by country over time, which we denote as $PS^x_{i,k,t}$. The latter provides data on the value of merchandise exports or imports disaggregated according to a country's trading partners over time, which we denote as $T^x_{i,j,t}$.

It is natural to assume that merchants in country i when exporting to the United States (Euro Area) invoice these exports in USD (EUR). However, if they also invoice in USD (EUR) when exporting to other destination countries these are not the local official currency, we refer to these transactions as excessive USD (EUR) invoicing.

To calculate our variable of interest from the data, let j_k be the set of countries j with home currency k, e.g., if $k = EUR$, then j_k denotes all Euro Area countries. Using the previously defined variables, we then have

$$y^x_{i,k,t} = PS^x_{i,k,t} \sum_{j \in J_{i,t}} T^x_{i,j,t} - T^x_{i,j_k,t} \tag{21}$$

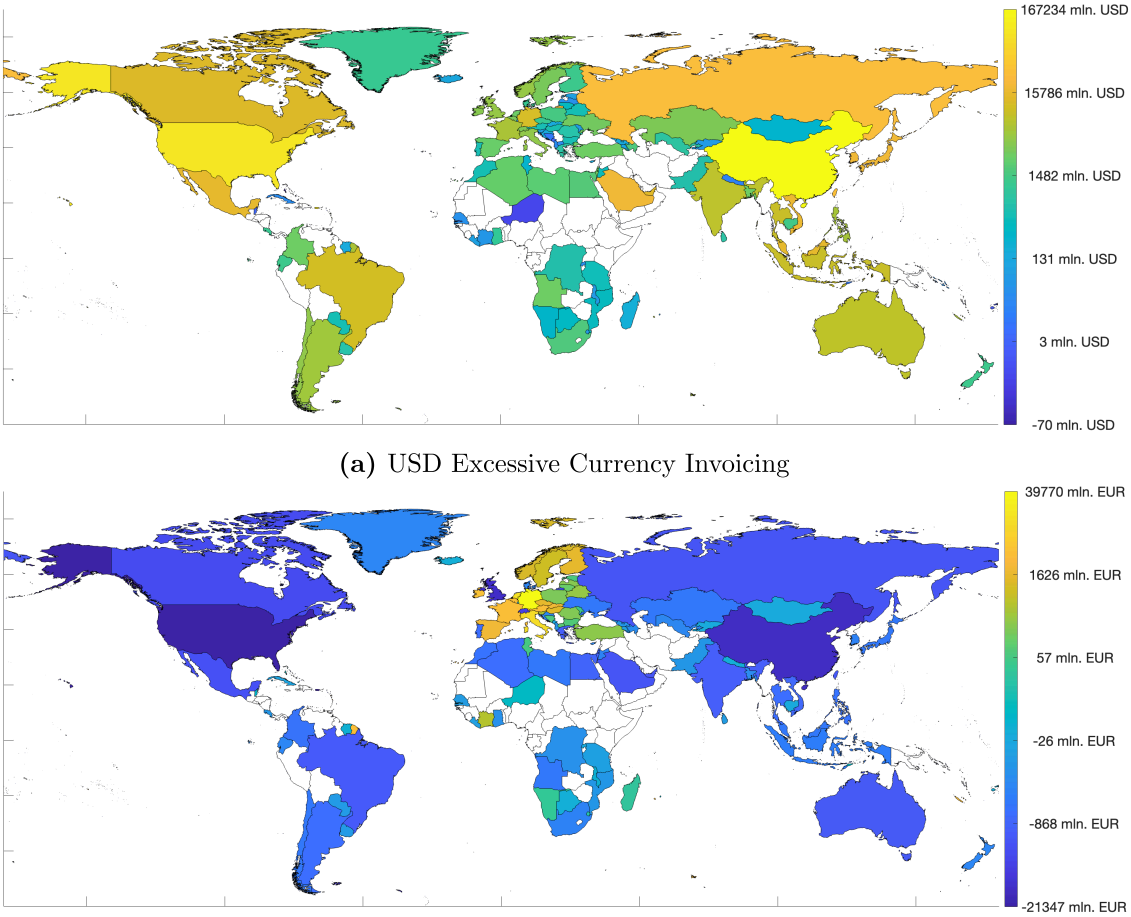
where $J_{i,t}$ denotes the set of trade counterparties of country i at time t. The first term captures the aggregate currency invoicing of country i in currency k with direction x. The second term deducts the trade conducted with the countries that have currency k as their home currency, thereby isolating the excessive amount of currency k invoiced. More details on the construction of $y^x_{i,k,t}$ based on the raw datasets is given in appendix 5.1.1.

Figure 7 depicts the geographic distribution of the average export-based excessive currency invoicing for the USD and the EUR in our sample. Figure 18 in appendix 5.3 depicts the distribution for import-based excessive currency invoicing. In total, we cover 119 countries in our dataset. Focussing on the USD export-based excessive currency invoicing, all countries, except the Bahamas, Niger, and the Republic of Fiji, use the USD in excess of their trade with the United States on average. That is, we find substantial use of the USD as a vehicle currency to conduct export-based trade. Particularly Asian and some Latin American countries have large positive export-based excessive USD invoicing positions on average. Generally, European countries have positive, albeit relatively lower, positions. The USD import-based measure in Figure 18 panel (a) shows similar patterns. Only one country, the Bahamas, has a negative USD import-based excessive currency invoicing position on average. Judging from the magnitude of positions, there is comparable usage of the USD across exports and imports as a vehicle currency. On the import side, Asian and European countries have particularly large positive import-based excessive USD invoicing positions on average.

[20] Since $\mathbf{1}_N = (I_N - \phi\mathbf{G}_t)^{-1}(I_N - \phi\mathbf{G}_t)\mathbf{1}_N = (I_N - \phi\mathbf{G}_t)^{-1}\mathbf{1}_N(1-\phi)$ due to $\mathbf{G}_t$ being a right stochastic matrix. Hence, $\mathbf{M}(\phi, \mathbf{G}_t)\,\mathbf{1}_N = (1-\phi)^{-1}\mathbf{1}_N$.

[21] Crucially, the SWIFT dataset allows us to cover China, Hong Kong, Mexico, Canada, the United Arab Emirates, Singapore, Vietnam and Sri Lanka. For details on the augmentation see appendix section 5.1.1.

Figure 7. Export-Based Excessive Currency Invoicing across Countries



(a) USD Excessive Currency Invoicing

(b) EUR Excessive Currency Invoicing

The figure depicts the average monthly excessive currency invoicing across countries over our sample. All amounts are in USD equivalents. The countries marked in white are not included in our sample due to missing observations. The top ten countries by export-based excessive USD invoicing positions in our sample are: China, the United States, Taiwan, Russia, South Korea, Saudi Arabia, Japan, Vietnam, Singapore, and Mexico. The top ten countries by export-based excessive EUR invoicing positions in our sample are: Germany, the Netherlands, Italy, Ireland, France, Belgium, Austria, Spain, the Slovak Republic, and the Czech Republic. Panel (a): USD excessive currency invoicing. Panel (b): EUR excessive currency invoicing.

Focussing on the EUR export-based excessive currency invoicing, the majority of countries use less EUR relative to their trade with the Euro Area countries on average, leading to negative excessive currency invoicing positions. Mostly European and their immediate neighbouring countries have large positive export-based excessive EUR invoicing positions. The fact that some European countries on average have large positive EUR export-based excessive currency invoicing positions indicates a form of producer currency pricing. The EUR import-based measure in figure 18 panel (b) again shows similar patterns, in that mostly European and neighbouring countries have large positive import-based excessive EUR invoicing positions. Interestingly, we again observe that European countries on average have a large positive EUR import-based excessive currency invoicing position. This indicates a form of local (destination) currency pricing. Together, these patterns lend support to claims that the EUR is less of a globally, but more of a regionally dominant currency.

2.5.1. *Dataset Construction.* The focus of the empirical study is to investigate the drivers of USD and EUR excessive currency invoicing, and in particular to asses whether network externalities affect the currency invoicing decision. We construct our dependent variable at monthly frequency based in exports and imports and add a large set of country-specific variables, suggested in the previous literature as potential drivers of the currency invoicing decisions, to our dataset.

In particular, the literature has found that exporters might coordinate on a certain invoicing currency to improve their pricing competitiveness in a certain market or hedge against exchange rate shocks to their inputs (e.g., labor, capital, or intermediate goods) (see Gopinath (2015), Doepke and Schneider (2017), Mukhin (2022) and Eren and Malamud (2022)). It is important to control for price volatilities and the size of the market. Hence, among the covariates, we include consumer price index-based inflation and inflation volatility, the change in domestic exchange rates and exchange rate volatility with the USD or EUR and the share of total aggregate imports or exports. Bahaj and Reis (2020) find that currency invoicing decisions of exporters depend on the level of financial services provided to exporting and importing firms denominated in certain currencies. For example, suppose a large share of counterparties of country i have a well-developed financial market denominated in EUR instead of USD. Trade and working capital financing as well as currency risk hedging in EUR would be cheaper and readily available. Then, firms in country i would have more incentive to invoice their trades in EUR. Similar ideas are also found in Maggiori (2017), Gourinchas, Helene Rey, and Sauzet (2019), and Gopinath and Stein (2021). In these papers, it is the characteristics of a country's trading partner countries, such as whether these partner countries have a well-developed credit or debt market denominated in USD or EUR, that determine whether USD, or EUR, or any other currency is used for invoicing. Motivated by these findings, we also include the aggregate level of firm-level debt denominated in the USD or EUR, dummy variables indicating whether a country has swap lines with the United States Federal Reserve or the European Central Bank, the financial development index, and the foreign direct investment inflows or outflows of a country. For detailed variable definitions, data sources and data-cleaning steps, see Appendix 5.1.

To construct the final dataset we lag all independent variables with respect to their original frequency. We then add lagged dependent variables, that is, lags of import- or export-based USD or EUR excessive currency invoicing and time- and country-fixed effects. The final dataset, including explanatory variables for our baseline specification, covers 84 countries from January 2004 to December 2019. These countries cover on average 91% (93%) of worldwide exports (imports) reported in the Direction of Trade Statistics database during the sample period.

Before estimation we standardise our data in two ways. First, we divide our dependent variable by lagged nominal gross domestic product. We also divide our foreign direct investment and aggregate level of firm-level debt variables by contemporaneous nominal gross domestic product. Second, we normalise all variables (independent and covariates), except the swap line dummy variable, by their sample standard deviation.

2.6. Network Analysis. In this section, our focus is to explore the influence of network externalities on the currency invoicing choices for a broad set of countries in the context of trade-induced transactions. We incorporate a substantial set of covariates, previously suggested in relevant literature, as potential factors driving currency invoicing decisions, and we examine whether the structure of the trade network itself acts as an additional driving force.

First, we start by conducting a comparative analysis between dynamic panel specifications for the invoicing decision and alternative specifications (SLM, SEM, and SDM) that account for potential network spillover effects. The data overwhelmingly indicate that a country's currency invoicing decision is significantly influenced by the currency invoicing choices made by its trade partners. In essence, network externalities emerge as one of the significant drivers impacting the currency invoicing choices. Second, we revisit the evidence on the determinants of the currency invoicing choice through the lenses of spatial dependency across countries. In doing so, we discover that certain findings from previous analyses become more nuanced when we consider both direct and indirect (i.e., effects through the network) effects. Third, employing our estimated structural model, we identify the key players in currency invoicing – countries whose decisions wield the most significant influence on total currency invoicing. Remarkably, countries and regions exhibiting substantial trade network centrality emerge as critical actors in this process. Fourth, we analyse the spillovers between export and import invoicing, to shed light on the *natural hedge* (Doepke and Schneider (2017); Amiti, Itskhoki, and Konings (2022)) channel of currency invoicing determination, and whether USD and EUR are complementary or substitute in the invoicing decisions. Fifth, we perform counterfactual analysis to assess the potential fragility of the current dominant currency equilibrium.

2.6.1. *Are there Trade-Network Spillovers in Currency Invoicing?* The first question we ask the data is whether there are indeed network spillovers driving the currency invoicing decision. We do so by computing the posterior probability of the spatial Durbin model (SDM) implied by our currency invoicing model in equation (15) and the same quantity for the panel specification obtainable by shutting down the trade-network channel (i.e., setting ϕ and the vector θ equal to zero). Furthermore, we consider alternative sources of spatial dependence. In particular, we consider two alternative canonical cases. First, the case in which the invoicing decision of country i depends on the invoicing of other countries in the network, but not directly on the other countries' covariates (i.e., $\phi \neq 0$ but the vector θ equal to zero in equation (15)). This is the so-called spatial lag or spatial autocorrelation model (SLM). Second, we also consider network spillovers purely driven by network propagation of the shocks (as, e.g., in Denbee et al. (2021)). That is, the invoicing decision of each country does not depend directly on the invoicing decision of any other countries or on other countries' covariates (i.e., ϕ and the vector θ equal to zero as in a panel specification), but the shocks in each country are linked via the network, i.e. $\mu_{i,t} = \mu_i + z_{i,t}$, where $z_{i,t} = \phi \sum_j g_{i,j,t} z_{j,t} + \nu_{i,t}$, where $\nu_{i,t}$ denotes cross-sectionally uncorrelated shocks.

Posterior model probabilities, that is the likelihood of the various models being the true data generating process, are computed assuming equal prior probabilities for all the models (i.e., assuming that the various specification are ex ante equally likely). That is, the posterior probability of the m-th model is $prob_m = \frac{p_m}{\sum_m p_m}$, where p_m denotes the so-called marginal likelihood of model m (the value of the integrated unnormalized posterior, i.e. likelihood times the prior, over the parameter space).

Log marginal likelihood values and posterior specification probabilities are reported in table 3. Each column considers a different dependent variable: excessive currency invoicing of exports in USD (column 1) and EUR (column 2), and excess currency invoicing of imports in the same two currencies (respectively, columns 3 and 4). Several observations are in order. First, there is overwhelming evidence of network spillovers: The panel

Table 3. The Posterior Likelihood of Trade-Network Spillovers

Specification:		ECI_{USD}^{Ex}	ECI_{EUR}^{Ex}	ECI_{USD}^{Im}	ECI_{EUR}^{Im}
Panel	$\ln p_m$	203.701	-1999.739	721.933	-2653.269
	$prob_m$	0.000	0.000	0.000	0.000
SEM	$\ln p_m$	178.036	-1973.738	710.634	-2634.372
	$prob_m$	0.000	0.000	0.000	0.000
SLM	$\ln p_m$	227.448	-2001.040	732.393	-2570.273
	$prob_m$	0.000	0.000	0.000	0.000
SDM	$\ln p_m$	274.748	-1863.344	897.106	-2479.632
	$prob_m$	1.000	1.000	1.000	1.000

The table reports the logarithm of the marginal likelihood ($\ln p_m$) of the data, given the model and the posterior model probabilities ($prob_m$). Note that the marginal likelihoods are adjusted by subtracting the logarithm of the number of observations. The models are separately estimated on each dataset using our baseline specification. Depending on the dataset, the baseline specification uses, respectively, USD or EUR export- or import-based excessive currency invoicing as the dependent variable. As independent variables, we include lags of inward foreign direct investments, a USD SWAP line dummy, exchange rate changes with the USD and EUR, realized exchange rate volatility with the USD and EUR, the share of aggregate exports, CPI-based inflation and CPI-based inflation volatility, USD export-, USD import-, EUR export-, and EUR import-based excessive currency invoicing, and country- and time-fixed effects.

specification with no spatial dependency is never preferred by the data. Second, the SDM model – the specification of our theoretical formulation – is always strongly preferred by the data, with posterior probability approaching 1 in all cases considered. Third, even alternative spatial formulations generally dominate the specification with no network dependency with the SLM formulation being almost always strongly preferred to the panel one. Fourth, the SEM model is typically the worst performing among the spatial specifications considered, and it is a less likely data generating process (DGP) than the dynamic panel in all cases: This emphasizes that the measured network spillovers are driven by the effect of a country's invoicing decision on its traded partner's invoicing decisions, rather than being merely the result of common shocks propagated via the trade network. Furthermore, as shown in table 5.4 of the appendix, the above findings hold if we use actual currency invoicing, or aggregate currency invoicing, instead of our preferred measure of *excessive* currency invoicing as the dependent variable.

Overall, table 3 emphasizes the need to account for network spillovers when analyzing the currency invoicing choice and provides strong support for the formulation (the SDM) adopted in our model.

2.6.2. *The Drivers of Excessive Currency Invoicing.* Having established the presence of network spillovers and empirical support for our SDM formulation, we now turn to analyse the implications of our model for the determinants of currency invoicing suggested in the previous literature.

Direct interpretation of coefficients for spatial models is difficult, as they often do not represent the marginal effects of the explanatory variables. This is because marginal effects in spatial models depend on potentially non-zero cross-derivatives. Intuitively, this is because the change in an explanatory variable for an individual country can potentially affect the dependent variable in all other countries, through, for example, feedback loops. Hence, covariates have both a direct and indirect (through the network dependency) effect on the outcome variables. Furthermore, since we also include among the controls the lagged values of the dependent variables to capture time series autocorrelation, perturbations of any of the covariates have both short- and long-run effects.

Hence, we report direct and total effects – the difference between the two indicating the indirect effect – in both the short-term (i.e., contemporaneously) and the long-term. The total effect, which we define as in LeSage and Pace (2009), is the time series average of the average row sum of partial derivatives (since, due to time variation in $\mathbf{G_t}$, the partial derivatives are time varying). This corresponds to the average impact on the individual dependent variable $y_{i,t}$ resulting from changing a given explanatory variable by the same amount across all individual countries. It is important to account for changes across multiple countries, as this allows us to trace out the spatial impact. The direct effect, which we also define similarly to LeSage and Pace (2009), is the time series average of the diagonal partial derivatives. This corresponds to the average impact on individual observation $y_{i,t}$ by changing its own i^{th} observation of a given explanatory variable. This statistic is closely related to the marginal effect in a standard regression model. To see this, suppose ϕ and θ are all zero. We would then find that the direct effect is exactly the β coefficient associated with the given covariate. Finally, the indirect effect is defined as the difference between total and direct effect.

We report these estimated effects in tables 4 and 5 for excessive and, respectively, aggregate currency invoicing. In each table panel A and B focus, respectively, on USD and EUR denominated invoicing. We add different groups of regressors, while always including country- and time-fixed effects, and lagged values of inward foreign direct investments, outward foreign direct investments, USD export-, USD import-, EUR export-, and EUR import-based excessive currency invoicing, as control variables. In both tables, long-term and short-term effects are similar in sign and significance. We report both for completeness.

Several observations are in order. First, for both USD and EUR denominated excess currency invoicing, we observe large and highly statistically significant network effects: The ϕ coefficient ranges from 0.245 to 0.296 for USD denominated invoicing and from 0.144 to 0.188 for EUR denominated invoicing. These estimated coefficients are not only statistically significant at any customary confidence level but also very stable across specifications.

Second, being positive, the estimates of ϕ imply strategic complementarity in the currency invoicing decision: If a country increases its invoicing in a given currency, its trade partners are also likely to do so. In particular, the coefficients imply an average amplification of the shocks to currency invoicing of about 17%–42% relative to a world with $\phi = 0$.

Moreover, as shown in table 5, this strong evidence of network-induced strategic complementarity in the currency invoicing choice is supported by the data even if we use aggregate invoicing values rather than our preferred excessive currency invoicing measure. If anything, in this robustness check, the measured network spillovers are even stronger. This stability of the estimated network effect is extremely reassuring.

Third, albeit most estimates of the direct effects of covariates conform with previous findings in the literature, these are much less stable across specifications and currency denomination, and, most importantly, it is not uncommon to find significant direct effects – akin to those estimated with a panel specification – while the total effects are not statistically significant, and vice versa. This is not too surprising given the strong evidence in Table 3 in favour of spatial modeling, which indicates that evidence produced ignoring

Table 4. The Drivers of Excess Currency Invoicing

		Independent Variables													
		(1)		(2)				(3)	(4)				(5)		(6)
		CPI	$CPIVol$	$FXChng_{USD}$	$FXChng_{EUR}$	$FXVol_{USD}$	$FXVol_{EUR}$	TS^{Ex}	FD_{USD}	$FDChng_{USD}$	FD_{EUR}	$FDChng_{EUR}$	$SWAP_{USD}$	$SWAP_{EUR}$	FMI
		Panel A. Dependent Variable: Export Excess Currency Invoicing in USD (ECI^{Ex}_{USD})													
Short-term:	Direct Effect	-0.011**	0.028***	-0.017	-0.004	-0.014	-0.004	-0.097***	0.042***	0.019**	0.012	-0.011	0.048	-0.045	0.006
		(0.092)	(0.000)	(0.182)	(0.668)	(0.210)	(0.687)	(0.000)	(0.000)	(0.017)	(0.106)	(0.148)	(0.129)	(0.208)	(0.317)
	Total Effect	-0.093***	0.019**	-0.083	-0.074***	0.070	-0.072**	-0.136***	0.104***	0.017	0.042	-0.037	0.324***	-0.453***	0.007
		(0.000)	(0.080)	(0.113)	(0.009)	(0.203)	(0.014)	(0.000)	(0.004)	(0.460)	(0.195)	(0.264)	(0.001)	(0.000)	(0.828)
Long-term:	Direct Effect	-0.044**	0.108***	-0.067	-0.017	-0.055	-0.016	-0.434***	0.159***	0.074**	0.044	-0.043	0.192	-0.178	0.025
		(0.086)	(0.000)	(0.176)	(0.651)	(0.219)	(0.668)	(0.000)	(0.000)	(0.017)	(0.105)	(0.147)	(0.123)	(0.199)	(0.319)
	Total Effect	-0.399***	0.084**	-0.363	-0.325**	0.304	-0.317**	-0.749***	0.426***	0.073	0.172	-0.150	1.370***	-1.915***	0.031
		(0.001)	(0.086)	(0.116)	(0.010)	(0.206)	(0.018)	(0.000)	(0.005)	(0.458)	(0.201)	(0.268)	(0.001)	(0.000)	(0.829)
	ϕ	0.252***		0.245***				0.248***	0.296***				0.259***		0.248***
		(0.000)		(0.000)				(0.000)	(0.000)				(0.000)		(0.000)
	R^2	0.948		0.948				0.949	0.950				0.948		0.948
	NObs	10812		10835				10835	9303				10835		10835
	log marginal	194.757		198.141				291.295	-75.518				206.735		182.530
		Panel B. Dependent Variable: Export Excess Currency Invoicing in EUR (ECI^{Ex}_{EUR})													
Short-term:	Direct Effect	-0.006	-0.009	0.033**	-0.006	-0.005	0.023**	0.046***	-0.005	0.037***	-0.012	0.009	-0.198***	-0.106**	0.000
		(0.459)	(0.252)	(0.028)	(0.609)	(0.740)	(0.044)	(0.000)	(0.637)	(0.000)	(0.165)	(0.367)	(0.000)	(0.013)	(0.962)
	Total Effect	0.079***	-0.042***	-0.025	-0.035	-0.017	0.044	0.060***	0.103***	-0.046**	0.051	-0.103***	-0.047	-0.516***	0.009
		(0.005)	(0.000)	(0.653)	(0.250)	(0.771)	(0.161)	(0.000)	(0.008)	(0.064)	(0.138)	(0.003)	(0.638)	(0.000)	(0.798)
Long-term:	Direct Effect	-0.017	-0.029	0.105**	-0.019	-0.015	0.074**	0.141***	-0.014	0.105***	-0.034	0.024	-0.606***	-0.324**	0.001
		(0.494)	(0.244)	(0.028)	(0.607)	(0.738)	(0.044)	(0.000)	(0.638)	(0.000)	(0.165)	(0.367)	(0.000)	(0.013)	(0.961)
	Total Effect	0.286***	-0.154***	-0.081	-0.112	-0.055	0.142	0.196***	0.296**	-0.131**	0.146	-0.295***	-0.139	-1.528***	0.029
		(0.006)	(0.001)	(0.655)	(0.251)	(0.769)	(0.163)	(0.000)	(0.010)	(0.066)	(0.144)	(0.003)	(0.637)	(0.000)	(0.799)
	ϕ	0.161***		0.144***				0.144***	0.188***				0.144***		0.147***
		(0.000)		(0.000)				(0.000)	(0.000)				(0.000)		(0.000)
	R^2	0.963		0.963				0.963	0.966				0.963		0.963
	NObs	10813		10836				10836	9304				10836		10836
	$\ln p_m$	-1864.490		-1872.266				-1867.117	-1910.905				-1845.850		-1884.534

The table reports the posterior means of the estimated effects and their respective p-values in brackets. Coefficient estimates significant at the 10%, 5% and 1% levels are indicated by *, **, and *** respectively. Estimation is carried out separately for the two different datasets. In addition to the listed independent variables, we always include country- and time-fixed effects, lags of inward foreign direct investments, outward foreign direct investments, USD export-, USD import-, EUR export-, and EUR import-based excessive currency invoicing as control variables.

Table 5. The Drivers of Aggregate Currency Invoicing

		Independent Variables													
		(1)		(2)				(3)	(4)				(5)		(6)
		CPI	$CPIVol$	$FXChng_{USD}$	$FXChng_{EUR}$	$FXVol_{USD}$	$FXVol_{EUR}$	TS^{Ex}	FD_{USD}	$FDChng_{USD}$	FD_{EUR}	$FDChng_{EUR}$	$SWAP_{USD}$	$SWAP_{EUR}$	FMI
Panel A. Dependent Variable: Export Aggregate Currency Invoicing in USD (ACI^{Ex}_{USD})															
Short-term	Direct Effect	-0.001	0.021***	-0.021**	0.006	-0.014	-0.008	-0.121***	0.029***	0.020***	0.005	-0.016**	0.039	-0.043	0.007
		(0.858)	(0.000)	(0.049)	(0.447)	(0.163)	(0.367)	(0.000)	(0.000)	(0.004)	(0.368)	(0.012)	(0.160)	(0.175)	(0.201)
	Total Effect	-0.112***	0.017	-0.102**	-0.091***	0.079	-0.093***	-0.194***	0.060	0.012	0.014	-0.128***	0.096	-0.612***	0.003
		(0.000)	(0.149)	(0.074)	(0.005)	(0.183)	(0.004)	(0.000)	(0.171)	(0.674)	(0.700)	(0.001)	(0.403)	(0.000)	(0.932)
Long-term	Direct Effect	-0.001	0.101***	-0.098**	0.033	-0.067	-0.033	-0.651***	0.129***	0.090***	0.024	-0.069**	0.182	-0.182	0.033
		(0.957)	(0.000)	(0.055)	(0.412)	(0.149)	(0.399)	(0.000)	(0.000)	(0.004)	(0.368)	(0.018)	(0.164)	(0.218)	(0.195)
	Total Effect	-0.416***	0.063	-0.388**	-0.348***	0.302	-0.358***	-0.998***	0.172	0.035	0.041	-0.370***	0.349	-2.218***	0.011
		(0.000)	(0.154)	(0.079)	(0.008)	(0.188)	(0.007)	(0.000)	(0.179)	(0.666)	(0.704)	(0.002)	(0.406)	(0.000)	(0.937)
	ϕ	0.387***		0.391***				0.369***	0.450***				0.393***		0.393***
		(0.000)		(0.000)				(0.000)	(0.000)				(0.000)		(0.000)
	R^2	0.963		0.963				0.964	0.963				0.963		0.963
	NObs	10812		10835				10835	9271				10835		10835
	$\ln p_m$	1663.108		1663.017				1842.718	1180.256				1677.898		1644.081
Panel B. Dependent Variable: Export Aggregate Currency Invoicing in EUR (ACI^{Ex}_{EUR})															
Short-term	Direct Effect	0.008	0.018***	-0.027***	-0.004	-0.005	0.004	-0.101***	0.016**	0.028***	0.002	-0.014**	-0.015	-0.027	0.001
		(0.133)	(0.001)	(0.007)	(0.634)	(0.567)	(0.581)	(0.000)	(0.018)	(0.000)	(0.741)	(0.015)	(0.560)	(0.357)	(0.854)
	Total Effect	-0.046	0.014	-0.125**	-0.109***	0.026	-0.056	-0.229***	-0.028	0.017	0.027	-0.198***	0.085	-0.792***	-0.042
		(0.143)	(0.282)	(0.048)	(0.002)	(0.695)	(0.123)	(0.000)	(0.555)	(0.568)	(0.493)	(0.000)	(0.505)	(0.000)	(0.285)
Long-term	Direct Effect	0.042	0.094***	-0.141***	-0.018	-0.028	0.023	-0.563***	0.083**	0.142***	0.009	-0.071**	-0.079	-0.109	0.006
		(0.130)	(0.001)	(0.008)	(0.669)	(0.559)	(0.563)	(0.000)	(0.016)	(0.000)	(0.751)	(0.020)	(0.544)	(0.457)	(0.835)
	Total Effect	-0.231	0.069	-0.565**	-0.495***	0.117	-0.254	-1.121***	-0.116	0.077	0.119	-0.860***	0.328	-2.973***	-0.200
		(0.153)	(0.289)	(0.055)	(0.004)	(0.697)	(0.133)	(0.000)	(0.574)	(0.569)	(0.500)	(0.000)	(0.504)	(0.000)	(0.293)
	ϕ	0.499***		0.491***				0.488***	0.543***				0.496***		0.493***
		(0.000)		(0.000)				(0.000)	(0.000)				(0.000)		(0.000)
	R^2	0.964		0.965				0.965	0.967				0.964		0.964
	NObs	10813		10836				10836	9272				10836		10836
	$\ln p_m$	2491.015		2493.079				2638.641	2047.666				2515.026		2473.841

The table reports the posterior means of the estimated effects and their respective p-values in brackets. Coefficient estimates significant at the 10%, 5% and 1% levels are indicated by *, **, and *** respectively. Estimation is carried out separately for the two different datasets. In addition to the listed independent variables, we always include country- and time-fixed effects, lags of inward foreign direct investments, outward foreign direct investments, USD export-, USD import-, EUR export-, and EUR import-based aggregate currency invoicing as control variables.

the network spillovers is affected by a large degree of misspecification. Hence, reduced-form evidence that ignores the spatial dependency should be taken with a substantial grain of salt.

Tables 4 and 5 further highlight that impacts for USD and EUR invoicing are noticeably different in terms of sign, magnitude, and significance. To a large degree the differences in results across USD and EUR, as well as across excess and aggregate currency invoicing, can be reconciled with explanations treating the USD as a globally dominant currency and the EUR as a regionally dominant currency, i.e. a pecking order of dominant currencies.

For example, we find that the larger a country's share of worldwide exports (TS^{Ex}) is, the smaller is its amount of USD excess and aggregate invoicing. This finding is in line with the theoretical results in Mukhin (2022), predicting that the larger a country's market size the lower is its reliance on vehicle currencies for invoicing. For EUR aggregate invoicing, similar patterns arise, however, for EUR excess invoicing we find that a country's trade size leads to higher EUR excess invoicing. EUR excess currency invoicing proxies for a country's EUR trade conducted with non-Euro Area countries. This indicates that larger countries tend to be more likely to invoice in EUR with non-Euro Area countries. Finally, judging by the marginal likelihoods ($\ln p_m$), a country's trade size is one of the most important determinants across the considered covariates for currency invoicing.

Next, we find that swap lines with the US federal reserve ($SWAP_{USD}$) lead to an increase in USD excess and aggregate invoicing. This corroborates the findings by Bahaj and Reis (2020) for Chinese RMB trade invoicing. The coefficients for aggregate invoicing are overall insignificant, indicating that swap lines typically impact vehicle currency trade invoicing. Further, for USD denominated excess invoicing, only the total effect is significant, while the direct effect remains insignificant. This illustrates that swap lines lead to an increase in USD excess invoicing predominantly through indirect network effects. The significantly negative effect on EUR excess invoicing suggests substitution between EUR and USD when the swap line is activated, consistent with a pecking order and the dominance of the USD. Turning to swap lines with the European Central Bank ($SWAP_{EUR}$), we find only negative effects. Closer examination of the data showed that in total only ten non-Euro Area countries had a swap line with the European Central Bank. These became effective during the great financial crisis or European debt crisis, explaining the estimated negative effects of EUR swap lines for both USD and EUR invoicing[22].

Turning to exchange rate related variables, we document that when a country's currency depreciates with respect to the USD or EUR ($FXChng_{USD}$ and $FXChng_{EUR}$ respectively), in general excess and aggregate invoicing in both USD and EUR decrease. The only exception is that for depreciation with respect to the USD ($FXChng_{USD}$), we find a positive statistically significant direct effect for EUR excess invoicing. Note however, the total effect is negative albeit insignificant and for aggregate EUR invoicing the effects are negative and significant. Overall, the estimated effects are more significant for aggregate than for excess currency invoicing, indicating that exchange rate depreciation is relatively less relevant for vehicle currency invoicing. Theoretically, exporting firms invoice in vehicle currency to hedge against exchange rate volatility. We find weak evidence to support that exchange rate volatility increases dominant currency invoicing. In Table 4, we observe that the direct effect of higher EUR exchange rate volatility ($FXVol_{EUR}$)

[22] The non-Euro Area countries are the United States, the United Kingdom, Denmark, Sweden, Switzerland, Canada, Japan, China, Hungary, and Poland. Swap lines were activated in 2007-2011 and kept in place throughout the sample for all countries except Poland and Hungary.

is positive and significant on excess EUR invoicing while the total effect is positive but not significant. By comparison, the direct effect of higher USD exchange rate volatility ($FXVol_{USD}$) on USD excess invoicing is actually negative, whereas the total effect after accounting for network impact is positive as expected, although neither estimate is statistically significant. In table 5, when we study the driver of aggregate currency invoicing, the impacts of exchange rate volatility are statistically insignificant. We also observe a cross-currency impact: a larger EUR exchange rate volatility lowers both excess and aggregate USD invoicing, and the total effects are statistically significant. This, again, might reflect a substitution effect between the two currencies when exporting firms hedge against relevant exchange rate volatility.

Inflation and inflation volatility have statistically significant but different impacts on excessive currency invoicing decisions for both USD and EUR. For aggregate USD currency invoicing the effects appear similar, however, significance levels change somewhat. The level of domestic inflation (CPI) has negative direct and total effects. This might reflect that domestic inflation makes exporters more competitive due to reduced local cost and less concerned about price competition in target markets. Interestingly, domestic inflation volatility ($CPIVol$) has positive direct and total effects. Therefore, inflation risk appears to be a more important driver than the level of inflation for dominant currency invoicing. This is intuitive because exporters use currency invoicing to manage balance sheet risks caused by inflation volatility. For excess EUR currency invoicing the signs of the effects are the opposite, however, they change across the two tables. This highlights that the effects are sensitive with respect to EUR invoiced trade conducted with Euro Area countries. The negative effect of inflation volatility for excess rather than aggregate EUR invoicing means that a higher CPI volatility leads to a lower amount of EUR invoicing to non-Euro Area countries, suggesting that the EUR is used as a regional currency.

We also find that foreign debt, which is measured as amount of corporate debt in USD (FD_{USD}) relative to domestic gross domestic product, is positively linked with excess and aggregate USD and EUR invoicing. The variable can be viewed as proxying whether a country has access to international capital markets and hence suggests that countries with more access to international capital markets tend to invoice more of their trade using vehicle currencies. We find that corporate debt in EUR (FD_{EUR}) is not associated with a significant effect. This can be viewed as further evidence of the USD dominance relative to that of the EUR. We further examine the change in corporate debt in USD ($FDChng_{USD}$) or EUR ($FDChng_{EUR}$), which typically lead to effects with plausible signs. Note however that overall the model incorporating data on corporate debt performs worst in terms of its associated marginal likelihood ($\ln p_m$), indicating that corporate debt is a less relevant determinant relative to the other variables considered.

Finally, we do not find that the financial market index (FMI), an index aiming to summarise the broad financial development of a country, has any significant impact on trade invoicing. Together with the previous results, this suggests that only certain financial frictions such as liquidity (captured by swap lines or the level of foreign firm debt relative to gross domestic product) affect trade invoicing decisions.

Tables 15 and 16 in the appendix depict our model baseline specification underlying the subsequent analysis. We included all of the discussed variables except the financial market index and the foreign debt variables. The former was excluded as the overall impact was insignificant. The latter were excluded as judging by the marginal likelihood other model specifications were preferred. Finally, we excluded the swap line with the European Central Bank due to the aforementioned issues. We further dropped outward foreign direct investments, which improved the marginal likelihood of the resulting models.

2.6.3. *The Network Key Players.* With the estimated spatial Durbin we can evaluate which country's shocks are expected to have the largest impact on the overall network – that is, we can identify the key players in the trade network. To see this, note that the SDM, singling out the role of the lagged dependent variables and merging the x covariates into a more compact notation, can be rewritten in vector form as

$$y_t = \alpha y_{t-1} + \eta G_t y_{t-1} + \phi G_t y_t + X_t \beta + G_t X_t \theta + \epsilon_t \tag{22}$$

where $\epsilon_t \sim N(0, \sigma^2 I)$. Define $M_t = (I - \phi G_t)^{-1}$ and $A_t = (\alpha I + \eta G_t)$. Suppose that y_t, as in our empirical implementation, is scaled by the GDP levels for stationarity and normalized to have unit variance. Let $D_t = diag(GDP_{i,t}^{-1})$ and $\Lambda = diag(\sigma_i^{-1})$, where $\sigma_i = Var(y_{i,t}/GDP_{i,t})^{1/2}$. Hence, we have $y_t = \Lambda D_t y_t^{\$}$, where $y_t^{\$}$ is our independent variable in USD units. The above immediately implies that the standard deviation of the errors of the USD unit dependent variable, $\epsilon_t^{\$}$, are heteroskedastic and vary over time. Specifically, $\sigma_{i,t}^{\$} = Var(\epsilon_{i,t}^{\$})^{1/2} = \sigma \sigma_i GDP_{i,t}$. We are mostly interested in computing statistics in terms of USD units from here onward. In what follows, let e be a column vector of ones of size $N \times 1$. Hence, $Y_t = e' y_t^{\$}$ denotes the total USD excess currency invoicing at time t.

Given the presence of lagged independent variables on the right-hand side of equation (22), a country-specific shock affects all other countries both contemporaneously and over time. That is, impulse-response functions (IRFs) of this model have both a spatial (across countries) and a temporal (across time) dimension. We define the USD unit spatiotemporal impulse-response function (STIRF) of $Y_t = e' y_t^{\$}$, to a one standard deviation shock to country i, as

$$STIRF_{i,t,\tau} = \frac{\partial Y_{t+\tau}}{\partial \epsilon_{i,t}^{\$}} \sigma_{i,t}^{\$} = \begin{cases} e'\{D_t^{-1} \Lambda^{-1} M_t\}_{\cdot,i} \sigma \text{ for } \tau = 0 \\ e'\{D_{t+\tau}^{-1} \Lambda^{-1} \prod_{j=0}^{\tau-1} M_{t+\tau-j} A_{t+\tau-j} M_t\}_{\cdot,i} \sigma \text{ for } \tau \geq 1 \end{cases}$$

where $\{\}_{\cdot,i}$ is the operator returning the ith column of a matrix.

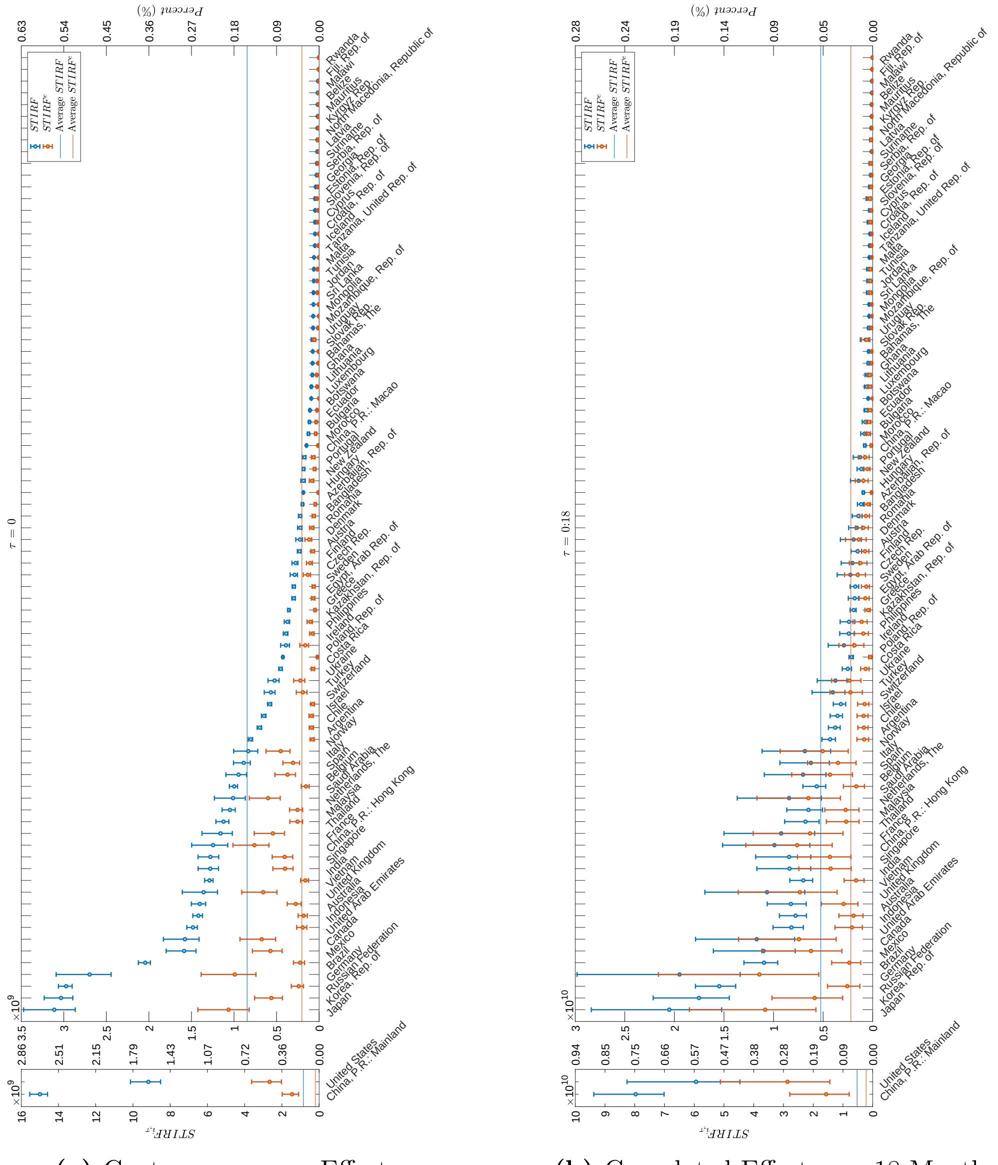
We can also isolate the purely network-driven part of the STIRF – which is the effect in excess of the original shock

$$STIRF_{i,t,\tau}^{e} = \begin{cases} e'\{D_t^{-1} \Lambda^{-1} M_t\}_{\cdot,i} \sigma - \sigma \sigma_i GDP_{i,t} \text{ for } \tau = 0 \\ e'\{D_{t+\tau}^{-1} \Lambda^{-1} \prod_{j=0}^{\tau-1} M_{t+\tau-j} A_{t+\tau-j} M_t\}_{\cdot,i} \sigma - \alpha^{\tau} \sigma \sigma_i GDP_{i,t} \text{ for } \tau \geq 1 \end{cases}$$

Figures 8 and 9 report $STIRF_{i,\tau}$ and $STIRF_{i,\tau}^{e}$ of, respectively, USD and EUR excess currency invoicing evaluated at the average adjacency matrix during the sample and average gross domestic product of each individual country.

Focusing on the USD denominated currency invoicing in Figure 8, China is the key player for USD invoicing: A one standard deviation shock generates a contemporaneous change in total ECI of 15 billion USD with about 10% of the effect driven by trade externalities. The United States follows in second place – this is mostly because China exports almost 1.5 times as much as the United States and invoices 92% of its exports in USD on average. Four other countries stand out: Japan, South Korea, Russia, and Germany. A one standard deviation shock to the ECI of these countries would result in a contemporaneous change (panel a) in total ECI of about 2.7-3.1 billion USD invoicing, i.e. about 0.5% of the total ECI in USD and a total cumulative effect (panel b) after 18 months of 15-21 billions – about 0.2% of total USD ECI over the same period. However, the drivers of these large effects are very different in nature, as outlined by the STIRFs in excess of the original shocks. For Russia, and to a lesser extent South Korea, the effects are almost entirely driven by the direct effect of a change in their USD invoicing, whereas for Germany, and to a lesser extent Japan, more than a third of the effect of a domestic shock is due to the network amplification and the central position that these countries have in the trade network. Furthermore, for countries such as Canada, the UK, Hong Kong, France, and the Netherlands, we also observe that a large share of the total effect is generated by network externalities.

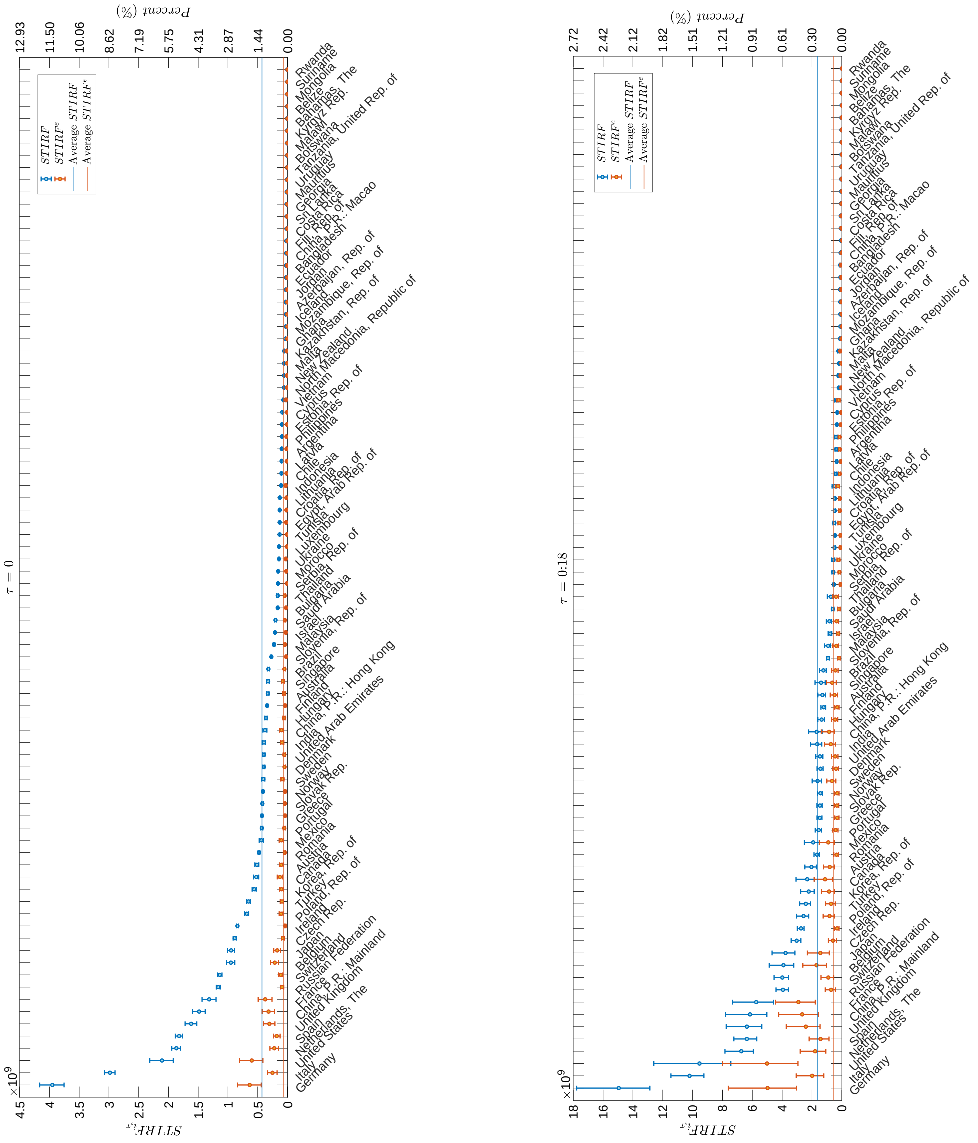
Figure 8. Impulse-Response Functions of USD Excess Currency Invoicing



(a) Contemporaneous Effect **(b)** Cumulated Effect over 18 Months

Spatiotemporal impulse-response functions to a domestic one standard deviation shock. Left axis = USD. Right axis = percentage of monthly total excess currency invoicing in USD over the same horizon. Panel (a): Contemporaneous effect. Panel (b): Cumulative effect after 18 months. Box-plots report posterior means and centered 95% posterior coverage.

Figure 9. Impulse-Response Functions of EUR Excess Currency Invoicing



(a) Contemporaneous Effect **(b)** Cumulated Effect over 18 Months

Spatiotemporal impulse-response functions to a domestic one standard deviation shock. Left axis = EUR. Right axis = percentage of monthly total excess currency invoicing in EUR the over same horizon. Panel (a): Contemporaneous effect. Panel (b): Cumulative effect after 18 months. Box-plots report posterior means and centered 95% posterior coverage.

Focusing on the EUR denominated currency invoicing in Figure 9, Germany is a clear outlier: a one standard deviation shock in this country would imply a contemporaneous change (panel a) of about 4 billion EUR (about 11.5% of total EUR ECI), with around 16% of the effect due to the network externalities generated by this country. Similarly, Germany generates the largest cumulative impulse-response after 18 months, with a total effect of about 15 billion EUR (or around 2.3% of EUR ECI over the same period). The second and third largest STIRFs are generated, in order of magnitude, by Italy and the United States (with STIRFs, respectively, about 75% and 50% of the German ones). Interestingly, albeit the effect of a shock to Italian EUR ECI is larger than that of the United States, the shocks arising in the latter are characterised by a larger degree of network amplification. It is worth noticing that Russia, with the ninth largest STIRFs, seems to play an important role also for the EUR denominated ECI but this effect, as in the USD case, is almost entirely direct in nature, rather than being amplified through the network.

In figures 19 and 20 of the appendix we report the network impulse-response functions for actual currency invoicing, rather than our excess invoicing measure, and find extremely similar results. Overall, the stability of the estimated effects is extremely reassuring.

2.6.4. *Cross-Currency and Export-Import Spillovers.* So far we have considered the determination of excess currency invoicing of exports in USD and EUR separately, while including as controls the lagged values of ECI in the currencies for both exports and imports. This allows us to consistently estimate the spatial spillovers in a specific currency export or import pair, but it does not provide an estimate of the *contemporaneous* links *within* a country of the ECI in different currencies and of imports and exports.

Nevertheless, our SDM specification implies that, after accounting for the spatial dependency, the estimation equations for the ECI in EUR and USD have the same structure as the corresponding equations of a reduced-form Vector Autoregression (VAR) system with four dependent variables (in addition to contemporaneously independent covariates and fixed effects): ECI in EUR and USD for *both* exports and imports. This observation implies that, by also estimating our SDM specification for the ECI of imports, we have a complete reduced-form VAR (with spatial dependencies in the mean processes) for these four variables. Hence, as in the Structural-VAR literature (see, e.g., Sims and Zha (1999)), one can recover the contemporaneous relationship between ECI of imports and exports and of different currencies, that is the matrix Γ, from the covariance matrix of the reduced-form VAR residuals.

Recall from (19) that $\epsilon_{i,t}$ correspond to country-specific shocks, which in our model in (13) is interpreted as shocks to a country's value of excess invoicing. To emphasise, $\epsilon_{i,t}$ measures the shock to a country without propagating it through the network system. It measures an idiosyncratic country-specific shock to the value of a country's excess invoicing. Using our estimates, we can calculate the $N_t \times 1$ vectors $\hat{\epsilon}^x_{k,t,b} = (I - \hat{\phi}^x_{k,b} G^x_t) y^x_{k,t} - (X_t \hat{\beta}^x_{k,b} + G^x_t X_t \hat{\theta}^x_{k,b})$, where k denotes the currency, t denotes the time period, b denotes the posterior sample draw, and x denotes the trade direction. Note that parameter estimates have subscript k and superscript x to emphasise that estimation is carried out separately on the four types of ECI (USD export, EUR export, USD import and EUR import).

For the sake of exposition, let us suppress the dependence on b, the posterior sample draw. We then organise a country's residuals as a 4×1 vector $\hat{\epsilon}_{i,t} = [\hat{\epsilon}^{Im}_{\$,i,t}, \hat{\epsilon}^{Im}_{\text{€},i,t}, \hat{\epsilon}^{Ex}_{\$,i,t}, \hat{\epsilon}^{Ex}_{\text{€},i,t}]^\top$. Let Σ_i denote the corresponding covariance matrix of dimension 4×4. The covariance matrices Σ_i can be used to recover the matrix of contemporaneous linkages Γ_i, since $\Sigma_i \equiv \Gamma_i^{-1} \Lambda_i (\Gamma_i^\top)^{-1}$, where Λ_i is a diagonal matrix with entries equal to the country-specific variance of the structural shocks.

Since Σ_i is symmetric, it has only $\frac{4\times(4+1)}{2}$ distinct entries, while the matrix of contemporaneous linkages Γ_i has 4×4 free entries and the matrix of structural variances has four free entries, Γ_i and Λ_i cannot be recovered without imposing additional restrictions. To achieve identification, we assume that Γ_i is identical across countries, i.e. $\Gamma_i = \Gamma$. Put differently, we assume that all countries USD or EUR export or import ECI react the same way to structural shocks. Notice that this still allows for cross-sectional heterogeneity in the structural variances, Λ_i. This implies that $\Sigma_i = \Gamma^{-1} \Lambda_i (\Gamma^{-1})^\top \ \forall \ i$. To emphasize, Γ

encodes the contemporaneous relationships between ECI of import and export for different currencies for a specific country.

The above identification strategy mimics ideas from identification via time series heteroskedasticity. In these models, time variation in Σ_t is used to identify the structural parameters (see Brunnermeier et al. (2021) for a recent example). Instead of time series variation, we utilize cross-sectional heteroskedasticity. That is, we use the variation in Σ_i to identify the structural parameters. The reason we opt for identification via cross-sectional heteroskedasticity is twofold. First, our dataset covers only a relatively short time period from 2004 to 2019. Second, by employing identification via heteroskedasticity, we are not required to take a stance on short-run, long-run, or sign restrictions, allowing us to identify effects under fairly general conditions. As long as the contemporaneous relationship matrix Γ is shared across the countries, identification via cross-sectional heteroskedasticity is possible.

Note that the previous assumptions achieve identification for the different products of Γ and Λ_i, however, similar to Brunnermeier et al. (2021), we cannot separate Γ from Λ_i[23]. Therefore, we need to impose one more restriction

$$\frac{1}{N}\sum_{i=1}^{N}\lambda^{x}_{k,i} = 1$$

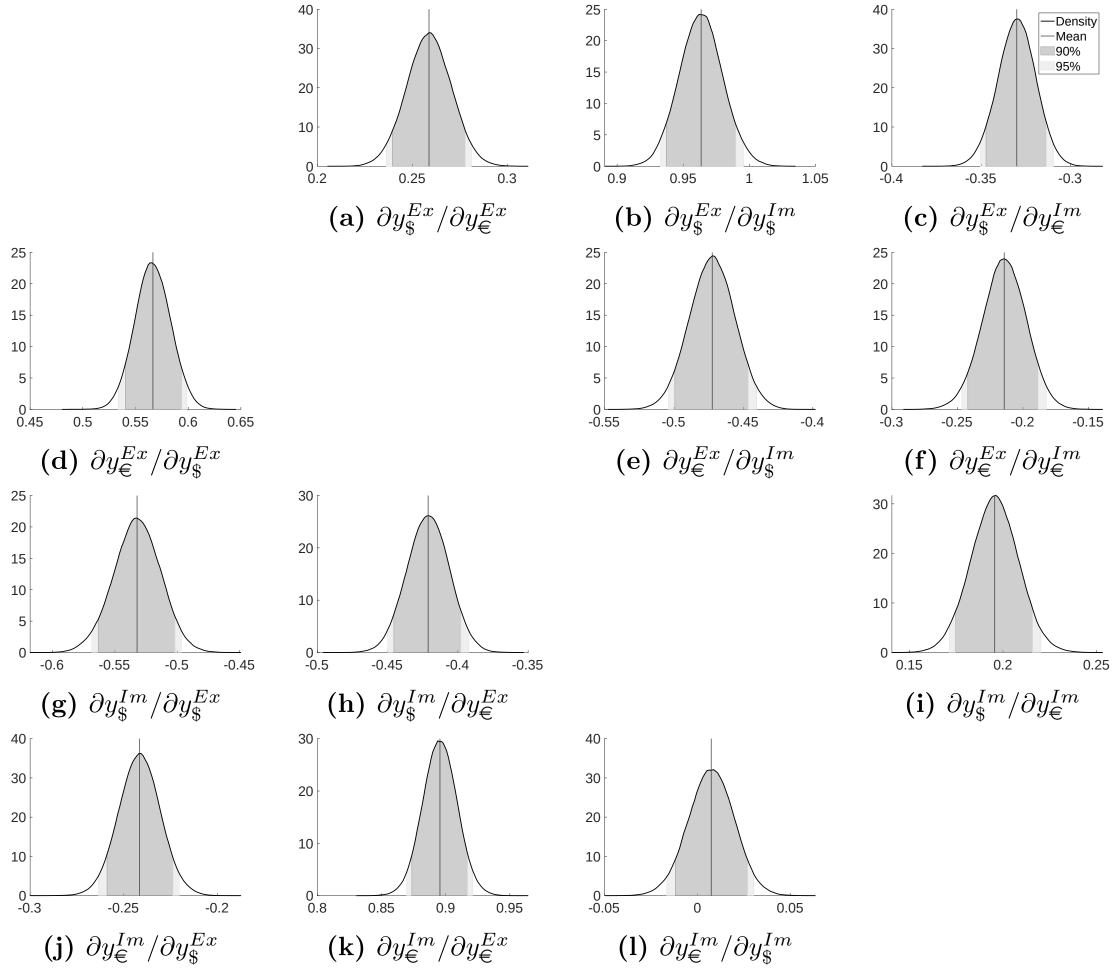
where $\lambda^{x}_{k,i}$ is one of the diagonal elements of Λ_i. The interpretation of this normalization is that we force the cross-country average structural variance to be one in each equation. This ensures identification of Γ up to flipping the sign of a row and identification of Γ and the set of Λ_i up to permuting the order of rows. We rule out former permutations by requiring Γ to have a positive sign on the diagonal and the latter permutations by selecting the final rows to ensure that Γ has its largest element on the diagonal. We estimate Γ and Λ_i using Bayesian techniques. The estimation procedure yields posterior samples of the 4×4 matrix Γ_b, and details on the estimation are given in appendix 5.2.4.

The posterior distribution of the off-diagonal elements of Γ – the contemporaneous effects matrix – is reported in Figure 10. Coefficients have been normalized such that contemporaneous partial derivatives between variables can be identified immediately. Several observations are in order.

First, we find some evidence of *natural hedging*, i.e. countries limiting their currency mismatch between imports and exports. Consider an increase in excessive USD denominated imports and focus on panels b and e. As the excessive USD denominated imports increase, USD denominated exports increase and EUR denominated exports decrease. This suggests that countries actively rebalance the currency denomination of their exports as they are faced with higher USD denominated imports. A similar pattern is observed for EUR denominated exports. Panels h and k illustrate that when countries excessive EUR denominated exports increase, in response USD denominated imports decrease and EUR denominated imports increase. These findings are in line with natural hedging and are fairly robust when considering our alternative aggregate currency invoicing measure (see figure 21 in appendix 5.3). Further, the coefficients in b and k are close to one, indicating almost perfect natural hedging for excessive currency invoicing – when considering aggregate currency invoicing these same coefficients are substantially below one. This indicates

[23] Multiplying the rows of Γ and the set of Λ_i by scale factors leaves the likelihood function unchanged (see appendix 5.2.4 equation (41)).

Figure 10. Cross-Currency and Export-Import Spillovers



(a) $\partial y_{\$}^{Ex}/\partial y_{€}^{Ex}$ **(b)** $\partial y_{\$}^{Ex}/\partial y_{\$}^{Im}$ **(c)** $\partial y_{\$}^{Ex}/\partial y_{€}^{Im}$

(d) $\partial y_{€}^{Ex}/\partial y_{\$}^{Ex}$ **(e)** $\partial y_{€}^{Ex}/\partial y_{\$}^{Im}$ **(f)** $\partial y_{€}^{Ex}/\partial y_{€}^{Im}$

(g) $\partial y_{\$}^{Im}/\partial y_{\$}^{Ex}$ **(h)** $\partial y_{\$}^{Im}/\partial y_{€}^{Ex}$ **(i)** $\partial y_{\$}^{Im}/\partial y_{€}^{Im}$

(j) $\partial y_{€}^{Im}/\partial y_{\$}^{Ex}$ **(k)** $\partial y_{€}^{Im}/\partial y_{€}^{Ex}$ **(l)** $\partial y_{€}^{Im}/\partial y_{\$}^{Im}$

The figures depict the posterior distribution of the elements of Γ, identified via cross-sectional heteroskedasticity. For the sake of interpretation, we have scaled the draws of Γ_b such that the diagonal only contains ones and then multiplied each row by negative one. Additionally, the figures depict the posterior mean, as well as 90% and 95% confidence intervals.

that the hedging motive is particularly prevalent when counterparties use a vehicle currency to conduct trade, i.e. when the currency does not correspond to the home currency of either involved country.

The responses associated with increases in excessive USD denominated exports and EUR denominated imports are less in line with natural hedging. Specifically, the posterior mean of the relevant coefficients in panel g and, respectively, panel f have the wrong sign. However, these coefficient estimates seem to be less robust as when considering our alternative aggregate currency invoicing measure the coefficient in g becomes insignificant and in f flips sign, i.e. starts to support natural hedging (see figure 21). Furthermore, evaluating the joint responses associated with increases in excessive USD denominated exports or excessive EUR denominated imports (panels d, g, j and, respectively, c, f, i), could indicate that other currencies play a role not considered in the analysis here. For example, panels d, g, and j suggest that as excessive USD denominated exports increase, so do excessive EUR denominated exports, but excessive USD and EUR denominated imports decrease. The increase in exports can be reconciled either with additional domestic production or imports denominated in another currency. However, with the data available to us this is difficult to distinguish.

Second, we find evidence of complementarity across currencies used for export and import invoicing. Panels a and d illustrate that generally as a country increases its excessive currency invoicing for exports in either currency, excessive invoicing of exports in the other currency also increases contemporaneously. This suggests that as a country increases its international exports, it tends to do so both in USD and EUR. Panels i and l suggest a similar pattern for imports. However, the response of excessive EUR denominated imports to excessive USD denominated imports is insignificant.

When examining aggregate currency invoicing (see figure 21) the coefficients for d and l flip signs. This indicates that as a country increases its aggregate USD denominated exports (imports), their EUR denominated exports (imports) decrease. As aggregate EUR denominated exports or imports increase, their USD counterparties still increase. These patterns can be interpreted as confirming the dominant status of the USD over the EUR, at least for exports and imports evaluated in aggregate currency invoicing. The latter indicate that as a country exports or imports more, it does so in either currency, whereas the former stress the special role of the USD: EUR denominated trade is substituted for USD denominated trade.

2.6.5. *Counter-Factual Analysis.* Suppose a country decides to permanently stop using a vehicle currency, and let us focus on the USD for exposition. This would lead to a reduction of $y^x_{i,USD,t+\tau}$ to zero for all τ. Notice that since we focus on excessive currency invoicing, this allows for countries to continue using the USD to trade with the United States but requires countries to use another currency as a vehicle currency, such as the EUR or Renminbi, when trading with other counterparties.

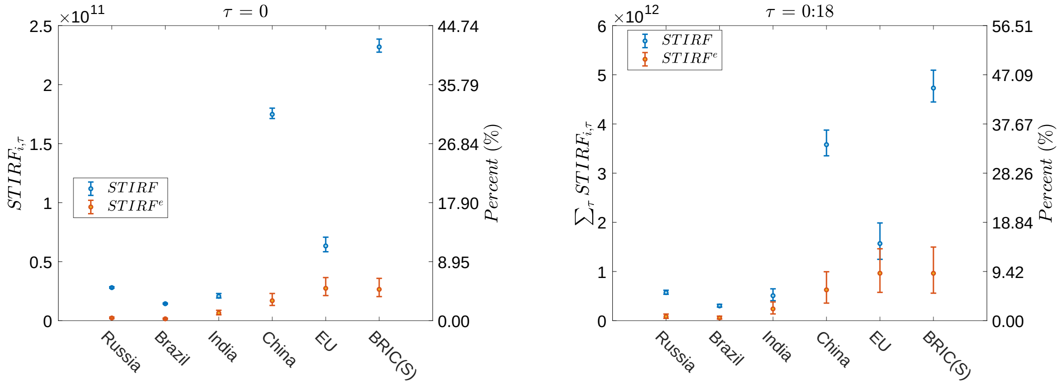
Within our framework, using our previous notation, we can view this as a shock $\epsilon^{\$}_{i,t}$, such that $y^{\$}_{i,t} = 0$ in period t. In period $t+1$, we then seek a shock $\epsilon^{\$}_{i,t+1}$, such that $y^{\$}_{i,t+1} = 0$, taking into account the previous shock $\epsilon^{\$}_{i,t}$. Due to scaling and the network effects, $\epsilon^{\$}_{i,t+\tau}$ is multiplied by a matrix $(D^{-1}_{t+\tau}\Lambda^{-1}M_{t+\tau}\Lambda D_{t+\tau})$, and it is therefore not immediately clear what size the shock must take.

Let S denote a set of countries that we want to shock such that for $i \in S$, we require $y^{\$}_{i,t+\tau} = 0$. Let $|S|$ be the cardinality of set S and let $y^{\$}_{S,t+\tau}$ and $\epsilon^{\$}_{S,t+\tau}$ for $\tau \geq 0$ be the sequence of vectors of size $|S| \times 1$ containing ECI and shocks of countries within the set. Let $\{D^{-1}_{t+\tau}\Lambda^{-1}M_{t+\tau}\Lambda D_{t+\tau}\}_{S,S}$ denote the submatrix corresponding to the countries within the set. Using this notation, we can show that the sequence of shocks $\epsilon^{\$}_{S,t+\tau}$ needs to satisfy

$$0 = y^{\$}_{S,t+\tau} + \underbrace{\{D^{-1}_{t+\tau}\Lambda^{-1}M_{t+\tau}\Lambda D_{t+\tau}\}_{S,S}\epsilon^{\$}_{S,t+\tau}}_{\text{Impact of shock at } t+\tau}$$
$$+ \underbrace{\sum_{j=1}^{\tau}\{D^{-1}_{t+\tau}\Lambda^{-1}\prod_{i=0}^{j-1} M_{t+\tau-i}A_{t+\tau-i}M_{t+\tau-j}\Lambda D_{t+\tau-j}\}_{S,S}\epsilon^{\$}_{S,t+\tau-j}}_{\text{Impact of shocks up until } t+\tau-1}$$

Assuming that $\{D^{-1}_{t+\tau}\Lambda^{-1}M_{t+\tau}\Lambda D_{t+\tau}\}_{S,S}$ is invertible, the above has a unique solution and allows us to solve for $\epsilon^{\$}_{S,t+\tau}$ sequentially. That is, given $\epsilon^{\$}_{S,t}$, we can determine $\epsilon^{\$}_{S,t+1}$. Then given $\epsilon^{\$}_{S,t}$ and $\epsilon^{\$}_{S,t+1}$, we can determine $\epsilon^{\$}_{S,t+2}$, and so forth. Once these shocks are obtained, we can calculate the STIRFs of each shock and aggregate them to assess the impact on total excessive currency invoicing over time.

Figure 11. Counterfactual: Abandonment of USD as Vehicle Currency



(a) Contemporaneous Effect **(b)** Cumulative Effect over 18 Months

Spatiotemporal impulse-response functions to a shock sequence that sets the excessive currency invoicing of the specified countries to zero permanently. EU contains all 19 EUR-Area countries while BRIC(S) contain the BRICS countries excluding South Africa, due to missing observations. Left axis = USD. Right axis = percentage of monthly total excess currency invoicing in USD. Panel (a): Contemporaneous effect. Panel (b): Cumulative effect after 18 month. Box-plots report posterior means and centered 95% posterior coverage.

Figure 11 depicts this counterfactual exercise for Russia, Brazil, India, and China individually, the EU block, and the members of BRICS in our sample (Brazil, Russia, India and China jointly). The calculations are done using average values. The estimated effects are quantitatively large, with the effects of the BRIC(S) block (EU) abandoning the USD for excess invoicing resulting in a 42% (11%) reduction in the overall use of this currency. But the channels through which these large effects arise are quite different. In the case of the BRIC(S) countries, most of the effect is driven by the direct reduction in the use of this currency by these countries, while in the EU case almost half of the effect is due to network externalities. That is, due to the EU countries central role in the trade network, and the strategic complementarity in the invoicing currency choice, if the block were to abandon the USD, the consequent reduction in the usage of this currency would almost double the direct effect.

To emphasize, the figures depict the impact on total excessive currency invoicing if the aforementioned sets of countries stop using the USD as vehicle currency permanently. The exercise still allows for countries to continue trading in USD or EUR with the United States or the EU, respectively. In figures 22 of the appendix we report results on the same exercise using aggregate currency invoicing. The findings are similar.

The above counterfactual stresses the inherent fragility of the dominant currency equilibrium we uncover in the data: In the presence of strategic complementarity in currency choice, the abandonment of the dominant currency by players that are large or central to the trade network can have dramatic effects.

2.7. Conclusion. In this paper, we examine the drivers of dominant currency invoicing for cross-border trades by constructing and estimating an equilibrium network model. The network externality arises because when choosing in which foreign currency to invoice their trades, besides being affected by its own macro and microeconomic conditions, agents are impacted by their trading partners' invoicing decisions either due to balance sheet hedging, competition, financing considerations or various other reasons.

Our estimation results show strong evidence of strategic complementarity in currency invoicing across countries: Exporting countries tend to invoice more in a given currency when their main trade partners invoice in that same currency. We find that the USD as a dominant currency is less stable than the EUR. The estimated network attenuation factor, ϕ, for the USD export trade is 0.241 and for the EUR export trade is 0.159. This suggests that there is a higher degree of complementarity for USD invoicing than for EUR invoicing. Conversely, this underscores that the USD as a dominant currency is more fragile than for example the EUR – the ϕ coefficients imply an average shock amplification of about 32% for the USD export ECI compared to 19% for the EUR export ECI.

We also identify key players in a given dominant currency, that is, countries that would have a sizeable impact if they were to abandon a certain dominant currency. Some of these key player countries are those that invoice most of their exports in that foreign currency (e.g., China, South Korea, and Russia). Some are countries that are central in the international trade network (e.g., Japan, Germany, and Canada). Furthermore, we find evidence for strategic complementarity between the choices of export and import currencies supporting the balance sheet hedging hypotheses for currency invoicing.

Finally, we conduct counterfactual analyses to examine how the use of dominant currency for trade invoicing were impacted if some countries were to abandon the USD in coordination. We find that if the BRIC(S) block (EU) were to bring their excessive invoicing in USD to zero, there would be a 42% (11%) reduction in the usage of USD in international trade with countries other than the United States.

Chapter 3.
Estimating Risk Preferences from Option Data

This paper estimates risk preferences through nonparametric methods from option data, and as a by-product recovers rational beliefs. The proposed estimator is shown to be consistent and asymptotically normally distributed. The estimated risk preferences are much more in line with preferences implied by classical utility functions than other studies suggest. Specifically, formal statistical tests suggest that there is no statistically significant evidence supporting the pricing kernel puzzle. In contrast, by constraining estimated risk preferences to be monotonically decreasing, the associated beliefs' forecasting performance improves substantially. Finally, CRRA risk preferences with a risk aversion coefficient of two are shown to approximate the nonparametric estimator reasonably, implying risk-neutral densities can be translated into physical densities easily.

3.1. Introduction. In this paper, I study the extraction of risk preferences from asset prices. Using non-parametric methods, risk preferences, and as a by-product rational beliefs are recovered from data on realized returns and forward-looking option data.

The starting point of this paper is an identity that relates conditional risk preferences, the stochastic discount factor, to Arrow-Debreu securities and probabilities. In the absence of arbitrage opportunities

$$M_t(\omega) = \frac{a_t(\omega)}{\pi_t(\omega)} \tag{23}$$

where $M_t(\omega)$ is the time t conditional stochastic discount factor, $\pi_t(\omega)$ is the conditional probability (or rational beliefs) and $a_t(\omega)$ is the price of an Arrow-Debreu security, i.e. the price of an asset that pays of 1\$ if state ω realizes.

When specified with respect to general states of the world, equation (23) is of little use for estimating risk preferences as neither of the three variables is observed. Progress can be made by focussing on states of the world summarized by the return R of a specific asset – formally this corresponds to projecting each variable onto the return space. Breeden and Litzenberger (1978) demonstrated that Arrow-Debreu securities over the return space of an asset, $a_t(R)$, can be inferred from an asset's option prices. Further, past return realizations contain information about the probability distribution as they correspond to draws from $\pi_t(R)$. Combining the two then allows to infer risk preferences over the return space, $M_t(R)$.

To highlight the intuition of the estimation approach in this paper, consider a generic candidate, $h_t(R)$, for the stochastic discount factor function. Based on (23) and the observed Arrow-Debreu securities, $a_t(R)$, the associated conditional probability function is pinned down: $\pi_t(R; h) = a_t(R)/h_t(R)$. Evaluating this expression for the historically observed returns, $\pi_t(R_{t+1}; h)$, allows to compute the likelihood of the historical returns implied by the candidate. This allows me to estimate risk preferences that lead to beliefs consistent with the data generating process underlying observed returns.

Formally, this approach searches over a space of functions and estimates the stochastic discount factor function through maximum likelihood. The space of candidate functions is large, hence, some structure is imposed to make estimation feasible. I assume that the true stochastic discount factor is proportional to a time-invariant function. The stochastic discount factor function is then estimated using the method of sieves, a nonparametric estimation technique that approximates the unknown function through a combination of basis functions (see X. Chen (2007)). I demonstrate that the estimator is consistent and asymptotically normally distributed, i.e. comes with the guarantee that asymptotically the correct risk preferences are recovered. A caveat of the method of sieves is that for finite samples there are few results available to guide the choice of basis functions and the degree of sieve approximation K. I develop intuitive criteria for selecting basis functions and the degree of approximation, exploiting the structure by equation (23). Specifically, I demonstrate that using polynomial basis functions up to degree K ensures implied beliefs, $\pi_t(R; h)$, are consistent with sample moments up to the K^{th} power, $\mathbb{E}[R_t^K]$.

To test whether assuming that the stochastic discount factor function is proportional to a time-invariant function is empirically plausible, I employ the procedure by Müller and Petalas (2010). I find that there is no significant evidence of time variation in the parameters describing the shape of the stochastic discount factor function. Together with the asymptotic results, this makes the proposed estimation strategy attractive.

Examining the estimated stochastic discount factor function, I find that risk preferences appear downward sloping over returns. This finding is in line with predictions from standard utility functions, which imply a monotonically decreasing stochastic discount factor over returns, reflecting investor's high marginal utility in low return states and low marginal utility in high return states.

However, the findings of this paper are in contrast with other studies in the literature, which document the *pricing kernel puzzle*, i.e. that risk preferences are generally not monotonically decreasing over returns (see Cuesdeanu and J. Jackwerth (2018) for a survey). To investigate this, I develop a novel formal test of the pricing kernel puzzle. I find that there is no statistically significant evidence against monotonically decreasing risk preferences. Further, I compute the conditional mean from the estimated beliefs, $\mathbb{E}_t[R_{t+1}]$, and evaluate its ability to forecast returns in terms of out-of-sample R^2, constructed as in Welch and Goyal (2008). I find that constraining estimated risk preferences to be monotonically decreasing leads to substantial improvements in the out-of-sample R^2 across all considered forecasting horizons. For example, for forecasts at the monthly horizon the out-of-sample R^2 increases by 67%. Overall, these results call into question the existence of the pricing kernel puzzle.

The results of this paper suggest that risk preferences are much more in line with preferences implied by standard utility functions than previously thought. Specifically, I find that CRRA preferences with a risk aversion coefficient of around two give rise to a reasonable approximation of the nonparametric estimate. For practitioners, this has the advantage that the stochastic discount factor for CRRA preferences is available in closed-form. Therefore, the physical conditional return distribution can be extracted from Arrow-Debreu securities (or the risk-neutral distribution) easily.

3.1.1. *Related Literature.* This paper contributes to a large body of work studying the relationship between risk preferences, the (physical) probability distributions of returns, and option prices both theoretically and empirically. Based on equation (23), the literature can be roughly split into two approaches. The first type estimates the probability distribution of returns and studies the associated risk preferences implied by option data (see e.g. J. C. Jackwerth (2000), Aıt-Sahalia and Lo (2000) and Rosenberg and Engle (2002)). The second type makes explicit or high-level assumptions on risk preferences and studies the implied conditional probability distribution of returns (see e.g. Bliss and Panigirtzoglou (2004), Kostakis, Panigirtzoglou, and Skiadopoulos (2011), S. A. Ross (2015) and Martin (2017)).

The first strand of the literature requires estimation of the conditional return distribution, which is generally non-trivial. Several approaches, such as J. C. Jackwerth (2000) and Aıt-Sahalia and Lo (2000), estimate the conditional return distribution using rolling windows of historical return data. Returns and specifically their distributions are known to fluctuate wildly over time, implying there is little guarantee that a rolling window estimate gives an accurate depiction of the conditional return density[24]. Alternative approaches, such as Rosenberg and Engle (2002), assume a specific form for the conditional return distribution, i.e. for the data generating process (e.g. GARCH), which is prone to suffer from misspecification[25]. This type of literature has documented the pricing kernel puzzle, i.e. they find that risk preferences are not monotonically decreasing over the return space. Crucially, these approaches depend on specifying and estimating $\pi_t(R)$ correctly. Therefore, the pricing kernel puzzle may arise precisely because of the aforementioned drawbacks. The presented approach allows me to revisit the pricing kernel puzzle without estimating or specifying the conditional return distribution explicitly, overcoming challenges in existing approaches of this literature. Doing so, I find that there is no statistically significant evidence supporting the puzzle. These findings corroborate recent results by Linn, Shive, and Shumway (2018).

The second strand of the literature either assumes a specific preference function (e.g. Bliss and Panigirtzoglou (2004)) or makes higher-level assumptions about preferences, such as the negative correlation condition in Martin (2017). The approach taken in this paper is therefore closer to this literature – the key assumption that makes estimation feasible is that the risk preference function is time-invariant. However, due to the non-parametric methods employed, the approach does not overly restrict the functional form of the stochastic discount factor function a priori. Rather the data is allowed to tell us the most likely shape of the unknown risk preferences over returns.

The results of this paper are further related to a large literature examining whether option data or information extracted from option data, has predictive power for realized returns (see Christoffersen, Jacobs, and Chang (2013)). I find that the probability density function extracted through the presented approach from option data has predictive power for returns. Further, imposing the economically motivated restriction that risk preferences are monotonically decreasing improves the predictive power, in spirit of Campbell and Thompson (2008).

Finally, the presented approach is related to an increasing body of studies employing the method of sieves – a popular sieve are neural networks (see X. Chen (2007)). The presented estimator employs a linear sieve paired with a maximum likelihood objective function. Implicitly, the approach models the conditional return distribution as belonging to the family of exponential distributions (see E. Lehmann and Casella (2006)), and therefore

[24] To take the argument to the extreme, suppose the return distribution is based on the full history of returns, i.e. the rolling window is made arbitrarily long. By definition, this is an estimate of the unconditional density. Equation (23) requires the conditional density, meaning this approach is only valid if conditional and unconditional distributions are identical.

[25] A standard GARCH model, implies that $\pi_t(R)$ is a normal distribution. Risk-neutral distributions (Arrow-Debreu securities scaled by the risk-free rate) are typically negatively skewed and leptokurtic. Therefore, assuming normality for $\pi_t(R)$ is bound to create non-monotonicities in the stochastic discount factor function.

the estimator inherits several convenient properties. Specifically, it belongs to the family of concave extended linear models (Huang (2001)).

The paper is structured as follows. Section 3.2 reviews observed Arrow-Debreu securities, $a_t(R)$, and examines the stochastic discount factor and beliefs that can be constructed from these. Section 3.3 presents the nonparametric estimator, discusses its intuition and its asymptotic properties. Further, the selection of basis functions and sieve approximation degree K in finite samples is examined. Section 3.4 describes the data. Section 3.5 presents and discusses the results and section 3.6 concludes.

3.2. Preferences Associated with Observed Arrow-Debreu Securities. Arrow-Debreu securities extracted from option data are defined over the return space of the asset underlying the options (e.g. the S&P 500). By extension, preferences associated with these Arrow-Debreu securities correspond to the projection of the (general) preferences onto the return space of the underlying asset.

To see this formally, consider a setting with discrete time, where $\mathcal{F}_t$ denotes the sigma-algebra used to depict the information available at time t. Let $\omega \in \Omega$ denote the realization of a state of nature. The absence of arbitrage opportunities implies the existence of positive Arrow-Debreu state prices (Dybvig and S. Ross (2003)). For tractability, but without loss of generality, let the relevant state variables be the gross return of an asset next period, R_{t+1}, and a supplementary variable X_{t+1}. Then $a_t(R_{t+1}, X_{t+1})$ is the price of an Arrow-Debreu security that pays 1\$ (or one unit of the consumption good) if R_{t+1} and X_{t+1} realize. Preferences are connected to these securities through the stochastic discount factor by the *pricing rule representation theorem* (Theorem 2, Dybvig and S. Ross (2003)). Formally,

$$M_t(R_{t+1}, X_{t+1}) = \frac{a_t(R_{t+1}, X_{t+1})}{\pi_t(R_{t+1}, X_{t+1})}$$

where $M_t(R_{t+1}, X_{t+1})$ denotes the stochastic discount factor and $\pi_t(R_{t+1}, X_{t+1})$ denotes the conditional probability of realization R_{t+1} and X_{t+1}. Subscripts t indicate that the variables are conditional on $\mathcal{F}_t$.

Breeden and Litzenberger (1978) demonstrated that a type of Arrow-Debreu security is observed through a rich set of option data. Let $C_t(K, t+1)$ be the time t call price with strike K and maturity next period. They showed that the time t price of an Arrow-Debreu security paying 1\$ if return R_{t+1} realizes is equal to the second derivative of the call price curve evaluated at $K = R_{t+1}S_t$, where S_t is the current price of the underlying. Formally,

$$a_t(R_{t+1}) = \frac{\partial^2 C_t(K, t+1)}{\partial K^2}\bigg|_{K=R_{t+1}S_t}$$

Importantly, these pay 1\$ if R_{t+1} realizes regardless of the realization of X_{t+1}. Hence, by no arbitrage $a_t(R_{t+1})$ corresponds to $a_t(R_{t+1}, X_{t+1})$ summed over all potential realizations of X_{t+1}, keeping R_{t+1} fixed. Formally,

$$a_t(R_{t+1}) = \int a_t(R_{t+1}, X) dX \tag{24}$$

Therefore, the stochastic discount factors associated with $a_t(R_{t+1})$ and $a_t(R_{t+1}, X_{t+1})$ are connected as follows

$$\begin{aligned} M_t(R_{t+1}) = \frac{a_t(R_{t+1})}{\pi_t(R_{t+1})} &= \int \pi_t(X|R_{t+1}) \frac{a_t(R_{t+1}, X)}{\pi_t(R_{t+1}, X)} dX \\ &= \mathbb{E}_t[M_t(R_{t+1}, X_{t+1})|R_{t+1}] \end{aligned}$$

where to obtain the second equality I used equation (24) and the fact that $\pi_t(X|R_{t+1}) = \pi_t(R_{t+1}, X)/\pi_t(R_{t+1})$. The final equality follows by definition of $M_t(R_{t+1}, X_{t+1})$. This formally demonstrates that the stochastic discount factor, or more generally preferences, based on the observed Arrow-Debreu securities extracted following Breeden and Litzenberger (1978) corresponds to the projection of the stochastic discount factor defined over multiple state variables onto the return of the option's underlying asset.

It immediately follows that $M_t(R_{t+1})$ inherits several properties of its general counterpart. First, it is positive. Second, by the law of iterated expectations it has the same mean as the actual stochastic discount factor[26]. Third, since $M_t(R_{t+1})$ is a projection it is generally less variable than $M_t(R_{t+1}, X_{t+1})$. For example, by the law of total variance, it follows that $Var_t(M_t(R_{t+1}, X_{t+1})) \geq Var_t(M_t(R_{t+1}))$. A similar result can be obtained in terms of entropy. Hence, the variability of $M_t(R_{t+1})$ gives a lower bound for $M_t(R_{t+1}, X_{t+1})$ and can be contrasted empirically with results from existing bounds such as L. Hansen and Jagannathan (1991) and Alvarez and Jermann (2005). Fourth, unlike $M_t(R_{t+1}, X_{t+1})$ the projection $M_t(R_{t+1})$ generally does not price all assets. However, if the conditional projection error is uncorrelated with an asset, $M_t(R_{t+1})$ will price it correctly[27]. Further, as long as Arrow-Debreu securities are extracted to ensure $\int Ra_t(R)dR = 1$, $M_t(R_{t+1})$ will always price the underlying asset correctly.

Most relevant to the subsequent estimation, this discussion highlights that the stochastic discount factor associated with Arrow-Debreu securities extracted following Breeden and Litzenberger (1978) is a function over the return space of the option's underlying asset, generally with dependence on $\mathcal{F}_t$. Further, any stochastic discount factor extracted from these is a projection of the true stochastic discount factor and inherits several its properties.

3.3. Estimation. The objective of this paper is to estimate the unknown function $M_t(R)$ under the assumption that its general shape does not change over time. For practical reasons, any function of R is assumed to have support over the closed interval $\mathcal{R} \in \mathbb{R}$. This is because Arrow-Debreu security prices, $a_t(R)$, are typically only observed over a closed interval.

Estimating $M_t(R)$ can be thought of as selecting a function h from the space of functions $\mathcal{M}$ with support over $\mathcal{R}$. To see this, note any candidate h needs to give rise to a stochastic discount factor function that is positive and implies proper probabilities given the observed Arrow-Debreu securities. Given that $M_t(R)$ is assumed to be proportional to a time-invariant function, the first constraint is ensured if $\log(M_t(R)) = h(R) + \log(c_t)$. The associated probabilities are $\pi_t(R; h) = a_t(R)/c_t \exp(h(R))$. The second constraint therefore pins down c_t since

$$\begin{aligned} 1 &= \int_{\mathcal{R}} \pi_t(R; h)dR = \int_{\mathcal{R}} \frac{a_t(R)}{c_t \exp(h(R))} dR \\ \Rightarrow c_t &= \int_{\mathcal{R}} \frac{a_t(R)}{\exp(h(R))} dR \end{aligned}$$

Note, c_t varies due to $a_t(R)$. Hence, estimating $M_t(R)$ is equivalent to estimating $h(R)$.

[26] As long as Arrow-Debreu securities are extracted to ensure that $\int a_t(R)dR = 1/R_{f,t+1}$, it follows that $\mathbb{E}_t[M_t(R_{t+1})] = 1/R_{f,t+1}$.

[27] The projection error is defined as $e_{t+1} = M_t(R_{t+1}, X_{t+1}) - \mathbb{E}_t[M_t(R_{t+1}, X_{t+1})|R_{t+1}]$. Consider some asset with return $R_{i,t+1}$ and suppose it is uncorrelated with e_{t+1}. Since $\mathbb{E}_t[e_{t+1}] = 0$, this implies $\mathbb{E}_t[R_{i,t+1}e_{t+1}] = 0$. Therefore, $\mathbb{E}_t[M_t(R_{t+1})R_{i,t+1}] = \mathbb{E}_t[(M_t(R_{t+1}, X_{t+1}) - e_{t+1})R_{i,t+1}] = 1$, since $M_t(R_{t+1}, X_{t+1})$ prices all assets.

Next, a criterion needs to be defined. To do so, I exploit the structure implied by the pricing rule representation theorem. Let $m \in \mathcal{M}$ denote the function corresponding to the true $M_t(R)$. Given a candidate $h \in \mathcal{M}$, the log-likelihood of realized returns is defined as $l_t(h) = \log(\pi_t(R_{t+1}; h))$ by equation (23). Suppose the unconditional expectation taken with respect to the true probability measure induced by m exists. The Kullback-Leibler information inequality in combination with the law of iterated expectations ensures $\mathbb{E}[l_t(m)] \geq \mathbb{E}[l_t(h)]$ for all $h \in \mathcal{M}$. Denote by $L_T(h)$ the sample analog of $\mathbb{E}[l_t(h)]$. Therefore, a natural way of estimating $M_t(R)$ is by maximizing $L_T(h)$ over $\mathcal{M}$, i.e. to estimate $M_t(R)$ by maximum likelihood as

$$\hat{m} = \arg\max_{h \in \mathcal{M}} \frac{1}{T} \sum_{t=1}^{T} \log(\pi_t(R_{t+1}; h))$$

This estimator formalizes the idea of estimating the stochastic discount factor function that makes the realization of historical returns most likely, given the information up to time t, particularly exploiting the information contained in the observed Arrow-Debreu securities. To illustrate how to search over some abstract function space $\mathcal{M}$, consider a simplified example with CRRA preferences.

Example 3.1 (CRRA preferences)**.** *Consider a standard one-period portfolio choice problem with CRRA preferences defined over next period wealth and let R denote the optimal portfolio return. The implied stochastic discount factor is*

$$M_t(R) = \frac{R^{-\gamma}}{\mathbb{E}_t[R_{t+1}^{1-\gamma}]}$$

where γ is the coefficient of relative risk aversion. CRRA preferences can therefore be thought of as restricting the function space $\mathcal{M}$ to only include functions of type $h(R) = -\gamma \log(R)$. Note that, $c_t = \mathbb{E}_t[R_{t+1}^{1-\gamma}]^{-1} = \int a_t(R) R^{\gamma} dR$ to ensure probabilities integrate to one. Selecting a function from this space is equivalent to estimating the parameter γ. The estimation problem therefore becomes the simple parametric problem

$$\hat{\gamma} = \arg\max_{\gamma} \frac{1}{T} \sum_{t=1}^{T} \log(\pi_t(R_{t+1}; \gamma)) \tag{25}$$

with $\pi_t(R_{t+1}; \gamma) = a_t(R_{t+1}) R_{t+1}^{\gamma} (\int a_t(R) R^{\gamma} dR)^{-1}$. This problem is straightforward to solve numerically given data on Arrow-Debreu securities and returns.

3.3.1. *Estimation through Method of Sieves.* Since generally $\mathcal{M}$ is infinite-dimensional, maximizing $L_T(h)$ over $\mathcal{M}$ tends to be infeasible and, if possible, may exhibit undesirable large sample properties. To address this issue, the method of sieves is employed. Instead of estimating over $\mathcal{M}$, estimation is carried out over a sequence of approximating spaces $\mathcal{M}_K$. The sieve spaces are less complex, but their complexity grows with the sample size so as to be dense in the original space. These approximating spaces are constructed through a combination of basis functions, e.g. a linear combination of polynomials up to degree K increasing with T (see X. Chen (2007) for a detailed discussion).

The key ingredients for the method of sieves are the choice of the criterion function and of the basis functions. The former is suggested by the setup and is based on maximum likelihood. The latter is chosen to ensure that the unknown function can be well approximated. Motivated by the *Weierstrass approximation theorem*[28], a linear sieve using

[28] The Weierstrass approximation theorem (Weierstrass (1885)) states that if $f(.)$ is a continuous real-valued function defined on the real interval $[a, b]$, then for every $\epsilon > 0$ there exists a polynomial $p_K(.)$

polynomials is implemented. Further finite-sample arguments for using polynomials are given in section 3.3.2.

Let $\{b_j(R), j = 0, 1, ...\}$ denote a sequence of polynomial basis functions, e.g. standard, Hermite, or Laguerre polynomials. Since $M_t(R)$ is approximated using a linear combination of basis functions, the location normalization $h(1) = 0$ is imposed[29]. The sieve approximation space is $\mathcal{M}_K = \{h(R) = \sum_{j=0}^{K} \theta_j b_j(R), h(1) = 0, R \in \mathcal{R}, \theta_j \in \mathbb{R}\}$ where $\dim(\mathcal{M}_K) = K \to \infty$ as $T \to \infty$ with $K/T \to 0$. Therefore, $\mathcal{M}_K$ becomes dense in $\mathcal{M}$ by the Weierstrass approximation theorem. For $\pi_t(R_{t+1}; h) = a_t(R_{t+1})/c_t \exp(h(R))$ the sieve estimator is defined as

$$\hat{m} = \arg\max_{h \in \mathcal{M}_K} \frac{1}{T} \sum_{t=1}^{T} \log(\pi_t(R_{t+1}; h)) \tag{26}$$

and is a finite-dimensional problem in which K parameters are estimated subject to the location normalization constraint – θ_0 is determined by the constraint. Denote by $\hat{\theta}_T$ the vector of coefficient estimates. The estimator for the stochastic discount factor function then is

$$M_t(R; \hat{\theta}_T) = \underbrace{\exp\Big(\sum_{j=0}^{K} \hat{\theta}_j b_j(R)\Big)}_{\exp(\hat{h}(R))} \underbrace{\int_{\mathcal{R}} \frac{a_t(R)}{\exp\big(\sum_{j=0}^{K} \hat{\theta}_j b_j(R)\big)} dR}_{\hat{c}_t} \tag{27}$$

where the first term captures the general shape of the stochastic discount factor function and the second allows the function to shift over time.

3.3.1.1. Consistency. The sieve estimator defined in equation (26) is consistent under mild regularity conditions. Formally, the following theorem establishes the consistency for the estimator.

Theorem 4. *Let $K, T \to \infty$ and $K/T \to 0$. Under assumptions A.6-A.9 it follows that $||\hat{m} - m||_\infty = o_p(1)$.*

The assumptions and proof can be found in section 6.1 in the appendix. The proof relies on satisfying the conditions of theorem 3.1 in X. Chen (2007). To do so the assumption that the stochastic discount factor is proportional to a time-invariant function is used. Further, the proof relies on a variation of the Weierstrass approximation theorem to argue that the sieve approximation is asymptotically accurate. Finally, I exploit the fact that the probabilities implied by $h \in \mathcal{M}_K$, $\pi_t(R; h)$, belong to the exponential family (for a definition see E. L. Lehmann, Romano, and Casella (2005) and E. Lehmann and Casella (2006)). Hence, the estimator in equation (26) belongs to the family of concave extended linear models (see Huang (2001)), which implies $\log(\pi_t(R; h))$ is strictly concave in h.

3.3.1.2. Asymptotic Distribution. Under additional regularity conditions, the estimator is asymptotically normally distributed. Formally, the following theorem establishes the asymptotic distribution for the estimator.

of some degree K sufficiently large such that for all $x \in [a, b]$ we have $||f(.) - p_K(.)||_\infty < \epsilon$, where $||.||_\infty$ denotes the supremum norm.

[29] Note that for any constant c, it follows that $h(R) + c + \log(\int_{\mathcal{R}} a_t(R) \exp(-(h(R) + c)) dR) = \log(M_t(R))$, meaning $M_t(R)$ is not identified if $h(R)$ is modeled to be a member of some linear space. Common location normalizations are constraints such as $h(R^*) = 0$ for some $R^* \in \mathcal{R}$. This ensures the mapping from $h(R) \mapsto \log(M_t(R))$ is one-to-one, i.e. $M_t(R)$ is identified. See also Stone (1990), Huang (2001) and example 2.6 in X. Chen (2007).

Theorem 5. *Under the assumptions for theorem 4 and assumptions A.10-A.14 it follows that* $\sqrt{T}(\hat{m} - m) \xrightarrow{d} N(0, \frac{b_m}{i_m^2})$.

The assumptions and proof can be found in appendix 6.1. The proof relies on the fact that the scores of the estimator defined over $\mathcal{M}$, i.e. evaluated at m, constitute a martingale difference sequence with respect to the filtration $\mathcal{F}_t$. Therefore, a central limit theorem for martingale difference sequences applies. The additional assumptions are standard and amongst other things ensure that the approximation error is negligible.

Confidence intervals for the estimated SDF function can be constructed from the sample scores and Hessians of the log-likelihood with respect to the parameters θ and by applying the delta method. Note under the assumptions, $b_m = -i_m$ by lemma L.6. In practice when working with overlapping data, e.g. using daily data on returns and options with a maturity of more than one day, scores will be serially correlated and the sandwich estimator with a robust estimator for the covariance matrix of scores is used (e.g. Newey and West (1987)).

3.3.2. *Selection of Basis Function and Approximation Degree in Finite Samples.* In the previous section, the sieve dimension K was allowed to grow with the sample size, and asymptotic results were used to justify the choice of the basis function. However, there is little guidance on how to select K and the basis functions for any finite sample. This section demonstrates that for linear sieves the estimator defined in equation (26) exhibits special features that can guide both the selection of basis functions and of K for finite samples. Specifically, polynomial basis functions ensure moments based on $\pi_t(R; \hat{\theta}_T)$ satisfy the law of iterated expectations up to the K^{th} moment.

To see this, consider the first order conditions of (26) with respect to θ_k, evaluated at $\hat{\theta}_T$

$$0 = \sum_{t=1}^{T} -b_k(R_{t+1}) + \frac{\int_{\mathcal{R}} a_t(R) \exp(-\sum_k \hat{\theta}_k b_k(R)) b_k(R)}{\int_{\mathcal{R}} a_t(R) \exp(-\sum_k \hat{\theta}_k b_k(R))}$$

By the definition of $\pi_t(R; \hat{\theta}_T)$ the above can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\mathbb{E}}_t[b_k(R_{t+1})] = \frac{1}{T}\sum_{t=1}^{T} b_k(R_{t+1}) \tag{28}$$

where $\hat{\mathbb{E}}_t[.]$ denotes the expectation taken with respect to $\pi_t(R; \hat{\theta}_T)$. Equation (28) illustrates that θ_k is selected to ensure that the sample average of the conditional expectation of the k^{th} basis function is equal to its sample analog. Put differently, θ_k is chosen to ensure $\hat{\mathbb{E}}_t[b_k(R_{t+1})]$ satisfies the law of iterated expectations.

Equation (28) can be used to inform both the choice of K and of basis functions in finite samples. The condition holds for any $b_k(R)$, so the choice of a basis function controls what moment conditions the estimated density should match in the data. Relative to other common basis functions, polynomial basis functions have an appealing feature. They ensure conditional moments up to order K implied by the estimated density, $\hat{\mathbb{E}}_t[R_{t+1}^k]$ for $k \leq K$, are consistent with moments calculated from the sample data[30].

[30] Other basis functions give rise to less intuitive moment conditions. Consider for example trigonometric basis functions or basis functions of polynomial splines. Trigonometric basis functions are $b_k^{trig}(R) = \cos(2k\pi R) + \sin(2k\pi R)$. For a polynomial spline of degree K with J knot points, the basis functions are $b_k^{spl}(R) = R^k$ for $k = 0, 1, ..., K$ and $b_{K+j}^{spl}(R) = \max(R - t_j, 0)^K$ for $j = 1, ..., J$, where t_j denotes knot j. The moment conditions associated with $b_k^{trig}(R)$ and $b_{K+j}^{spl}(R)$ seem less desirable.

Hence, for polynomial basis functions, K controls both the approximation flexibility and the maximum number of sample moments matched by the estimated conditional density. This introduces a trade-off since typically, the higher a moment the more samples are required to estimate it reliably. For the purposes of this paper, it appears sensible to set $K = 4$. Therefore, the estimated density will match the historical sample mean, variance, skewness, and kurtosis.

3.4. Data. To implement the estimator, I construct a dataset based on raw option price data obtained from IvyDB US Optionmetrics. The dataset spans 4-Jan-1996 to 31-Dec-2020 and contains daily data on the continuously compounded dividend yield of the S&P 500, continuously compounded zero coupon interest rates for several maturities, and the expiration date, strike price, highest closing bid and lowest closing ask of all call and put options on the S&P 500. Furthermore, data on the S&P 500 closing price is obtained from CRSP. Several filters, cleaning, and processing steps are implemented and detailed in section 6.2.1 in the appendix.

For the empirical exercise, I require an option dataset for identical maturities at each point in time and a set of standardized strikes that are identical over time. Therefore, inter- and extrapolation of the raw data is necessary. Several approaches have been developed. The overall goal is to extract representative Arrow-Debreu security prices. A survey discussing the practical details of extracting these Arrow-Debreu securities is given in Figlewski (2018).

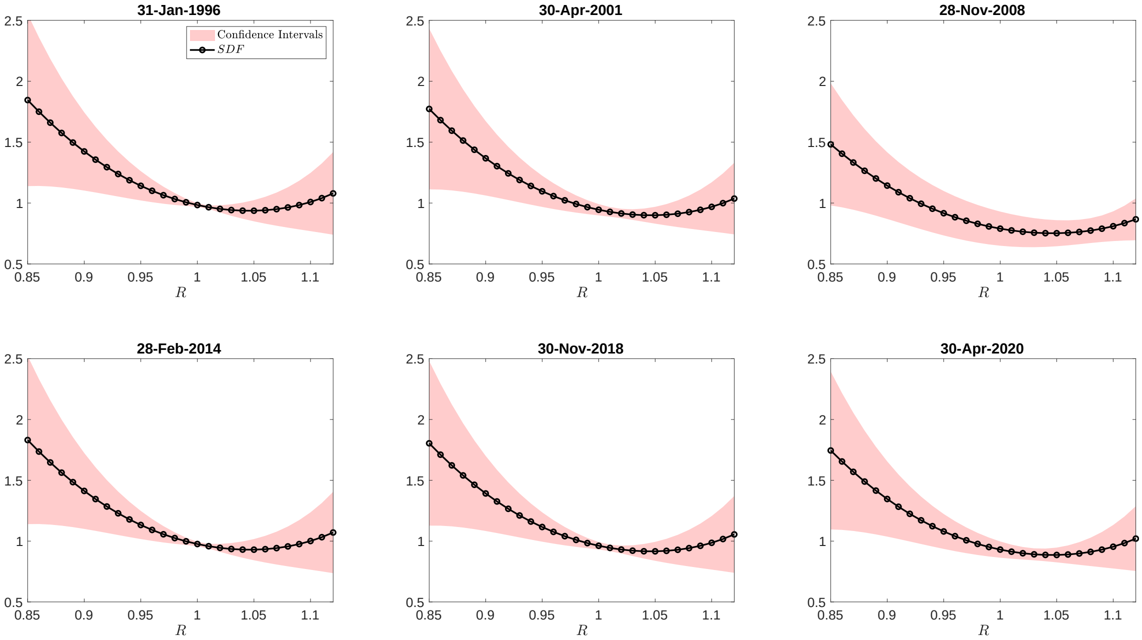
The baseline dataset follows Gatheral and Jacquier (2014) and allows for inter-/ extrapolation taking into account the full option dataset on each date across maturities and strikes while enforcing several no-arbitrage constraints. For robustness, I implement three additional approaches, following Figlewski (2008), Ulrich and Walther (2020) and J. Jackwerth and Menner (2020). See section 6.2.2 in the appendix for details. The baseline dataset has daily Arrow-Debreu security price curves for a maturity of one, two, three, six, and twelve months over a discrete grid of strikes. The grid corresponds to gross returns with 1% increments related to strikes via $K_i = R_i S_t$, where S_t is the time t price of the S&P 500. The grid range is chosen to capture the minimum and maximum observed return of the S&P 500 over the sample period for each maturity horizon, e.g. for the 1-month horizon, it ranges from 66% to 125%.

3.5. Empirical Results. This section presents results on the estimated risk preferences. First, in comparison with other approaches, I find estimated risk preferences are more in line with traditional utility functions, i.e. generally decreasing over the return space. Second, I find empirical support for the assumption that the stochastic discount factor function is proportional to a time-invariant function through parameter instability tests. Third, I document that there is little evidence of the pricing kernel puzzle. Specifically, I find no statistically significant evidence against monotonically decreasing risk preferences. On the contrary, when imposing $M_t(R)$ to be monotonically decreasing, the associated conditional return distributions can better predict return realizations. Fourth, I demonstrate that CRRA preferences with a risk aversion coefficient of two approximate the nonparametric estimates of $M_t(R)$ well.

3.5.1. *Estimated Risk Preferences and Parameter Stability.* The estimator is implemented for monthly returns, sampled at daily frequency. Based on the discussion in section 3.3, regular polynomial basis functions, $b_k(R) = R^k$, are employed with $K = 4$. The sieve approximation degree is set slightly higher than what is suggested by classical model selection criteria, such as the Akaike or the Bayesian information criterion (see table 17 in the appendix).

Estimated risk preferences, $M_t(R;\hat{\theta}_T)$, are available at each point in time. Figure 12 depicts these for several snapshot dates.

Figure 12. Estimated Risk Preferences



Note: The figure depicts the estimated stochastic discount factor function given in equation (27) on several dates using regular polynomial basis functions with $K = 4$. Further, 95% confidence intervals obtained via the delta method are reported. I employ a Newey and West (1987) type estimator with lag length 21 for the covariance matrix of the scores.

The estimated risk preference function appears generally decreasing over the return space. The point estimates exhibit a slight *u-shape*, which taking into account the confidence intervals does not appear to be substantial. Confidence intervals are relatively tight for gross returns between -5% to 5% and widen exponentially towards the left and right ends of the grid. For the sample, roughly 80% of monthly S&P 500 returns fall into this range, giving rise to more precise estimates in this area. Shifts of the risk preference function due to c_t become apparent when comparing the different snapshots, however, are overall minor.

The estimator assumes that the true stochastic discount factor function is proportional to a time-invariant function. The shape of the estimated function is characterised by the parameters θ. A way of testing whether the stochastic discount factor function is indeed proportional to a time-invariant function is to examine whether the shape parameters, θ, vary over time.

Several parameter instability tests exist and I here implement the procedure by Müller and Petalas (2010), which is designed for testing parameter stability in nonlinear models. Their procedure is convenient, as it only requires the scores, Hessian, and (robust) covariance matrix of the scores, which are readily available from (27). Further, their test statistic is sufficiently general. It is specified against a wide range of alternative parameter dynamics that lead to persistent time-varying parameter paths of relatively small variability. Naturally, as with most tests of parameter instability, it cannot test against all types of alternative time-variation[31]. However, it encompasses a variety of alternatives of interest such as random walks or breaks at unknown dates (see Elliott and Müller (2006) and Müller and Petalas (2010) for a detailed discussion).

Table 6 depicts the results for the Müller and Petalas (2010) qLL test statistic. Note the test statistic rejects the null hypothesis of parameter stability for small (negative) values.

Table 6. Test of Parameter Stability

K	2	3	4	5	6
qLL	-15.036	-15.598	-15.652	-24.912	-24.835
p-value	(0.034)	(0.245)	(0.705)	(0.267)	(0.644)

Note: The table repots the estimated qLL test statistic following Müller and Petalas (2010) for the estimator in (27) using non-overlapping observations and regular polynomial basis functions. The different columns correspond to the number of parameters, or the approximation degree, K. The qLL test is calculated following Müller and Petalas (2010) using $c = 10$ for *robustified* scores. The reported p-values are calculated by simulating the distribution of the qLL test statistic with $c = 10$ from the random variable in Elliott and Müller (2006) lemma 2.

The results indicate that the test fails to reject the null of parameter stability across the considered approximation degrees, except for $K = 2$. The results in tables 18 and 19 of the appendix, which report the qLL test statistic for other values of the hyperparameter c (see Müller and Petalas (2010) for details), further corroborate this finding. Overall, these findings are encouraging. They suggest it is reasonable to assume that the stochastic discount factor function is proportional to a time-invariant function.

3.5.2. *Revisiting the Pricing Kernel Puzzle.* Several studies that have estimated risk preferences from observed Arrow-Debreu securities have found that the stochastic discount factor is not monotonically decreasing over the return space – the so-called pricing kernel puzzle. This is surprising since we generally expect marginal utility to decrease over wealth or returns, i.e. we would expect a monotonically decreasing stochastic discount factor function. Figure 13 contrasts the estimator developed in this paper with two representative alternatives based on J. C. Jackwerth (2000) and Rosenberg and Engle (2002).

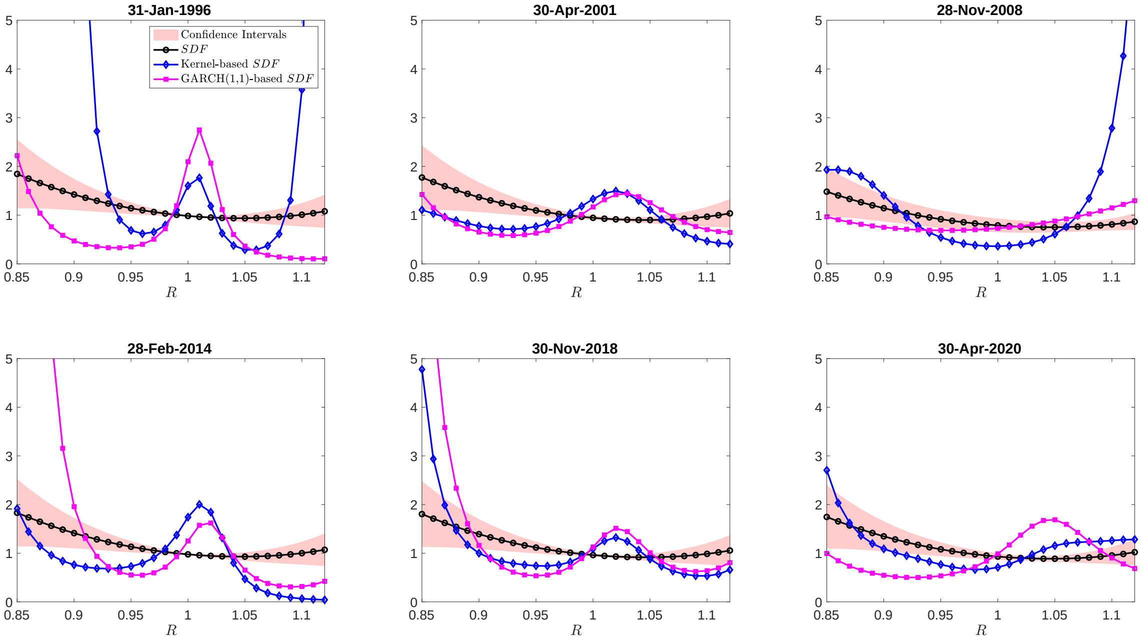
The alternative methods give rise to a stochastic discount factor function that fluctuates wildly over time. Further, they exhibit features indicative of the pricing kernel puzzle, i.e. the stochastic discount factor functions are generally not monotonically decreasing over the return space. In contrast, the estimator developed in this paper appears to be consistent with a monotonically decreasing stochastic discount factor function.

Section 3.3 demonstrated that if the stochastic discount factor function is proportional to a time-invariant function, the developed estimator is consistent. The alternative approaches achieve this only under fairly restrictive assumptions. The J. C. Jackwerth (2000) approach estimates $\pi_t(R)$ from returns over the last four years. This recovers the conditional return distribution with respect to information set $\mathcal{F}_t$ only under fairly restrictive assumptions[32]. The Rosenberg and Engle (2002) approach overcomes this, by employing

[31] For example, the Müller and Petalas (2010) test may fail to detect parameter instability that is less persistent and displays substantial mean reversion (see Calvori et al. (2017)).

[32] To illustrate the shortcomings, consider the extreme example in which $\pi_t(R)$ is estimated on the full history of returns up to t. By definition, this corresponds to an estimate of the unconditional return distribution. Hence, the estimation approach would only be valid if the conditional return distribution is equal to the unconditional return distribution.

Figure 13. Comparison of Estimated Risk Preferences



Note: The figure depicts the estimated stochastic discount factor function on several dates for the estimator proposed in this paper (black), the J. C. Jackworth (2000) estimator (blue) and the Rosenberg and Engle (2002) estimator (purple). The J. C. Jackworth (2000) estimator divides Arrow-Debreu securities by the return distribution estimated from 4 years of prior non-overlapping monthly return data using a kernel estimator. The Rosenberg and Engle (2002) estimator divides Arrow-Debreu securities by the conditional distribution of a GARCH(1,1) model fit to non-overlapping monthly data using the full sample.

a GARCH model. However, this requires specifying the data generating process. The estimator in figure 13 assumes that $\pi_t(R)$ is a normal distribution. Since risk-neutral distributions (Arrow-Debreu securities scaled by the risk-free rate) are typically leptokurtic and negatively skewed, misspecification of the conditional return distribution may lead to a biased estimator of the stochastic discount factor function.

In contrast, the estimator developed in this paper overcomes several of these issues as it does not require assumptions regarding the data generating process of returns, while simultaneously allowing risk preferences to be estimated flexibly. To formally evaluate the pricing kernel puzzle, I conduct two tests. First, I develop a statistical test examining whether a monotonically decreasing pricing kernel can be rejected. Second, I examine the ability of $\pi_t(R;h)$ to forecast returns when the pricing kernel function is constrained to be monotonically decreasing. Both tests suggest that there is no support for the pricing kernel puzzle.

3.5.2.1. Statistical Test of the Pricing Kernel Puzzle. The pricing kernel puzzle states that the stochastic discount factor projection is not monotonically decreasing. To evaluate the puzzle, the estimator in (27) can be implemented subject to the constraint that the function is monotonically decreasing. In the context of the suggested linear sieve, the constraint takes the form

$$\sum_{k=0}^{K} \theta_k \frac{\partial}{\partial R_i} b_k(R_i) \leq 0 \tag{29}$$

Denote by $\tilde{\theta}$ the coefficients estimated from (27) subject to the constraint in (29). By comparing the constrained and unconstrained parameter estimates, the pricing kernel puzzle can be evaluated formally. Specifically, I will evaluate the hypothesis that the constrained and unconstrained estimates are equal, i.e. the null hypothesis $H_0 : \theta = \tilde{\theta}$.

To test the hypothesis a Lagrange multiplier test is implemented as it has a particularly convenient interpretation. Generally, the Lagrange multiplier test examines if the average scores under the constraint are significantly different from zero. A constraint sets $\tilde{\theta}_k$ such that the average score is no longer exactly equal to zero. Based on equation (28), the moment condition for the k^{th} basis function will therefore be violated. The Lagrange multiplier test therefore examines whether the constraint leads to a conditional distribution that is no longer consistent with the sample moments.

Given the hypothesis, the Lagrange multiplier statistic takes the form

$$LM = \frac{1}{T}\left(\sum_{t=1}^{T} u_t(\tilde{\theta}_T)\right)' B_{\tilde{\theta}}^{-1}\left(\sum_{t=1}^{T} u_t(\tilde{\theta}_T)\right) \tag{30}$$

where $B_{\tilde{\theta}}$ is the asymptotic covariance matrix of the scores evaluated at the estimate for $\tilde{\theta}$ and can be estimated e.g. via a Newey and West (1987) estimator. Note the LM statistic is asymptotically χ^2 distributed with K degrees of freedom due to the location normalization constraint. For large values, the Lagrange multiplier test will reject the monotonically decreasing stochastic discount factor function, i.e. favour the pricing kernel puzzle.

Table 7 reports the Lagrange multiplier test statistic. Figure 23 in the appendix further compares the unconstrained with the constrained estimator.

Table 7. Pricing Kernel Puzzle Tests

K	2	3	4	5	6
LM	3.400	0.883	3.611	3.298	3.482
p-value	(0.183)	(0.830)	(0.461)	(0.654)	(0.746)

Note: The table presents the Lagrange multiplier test statistic for the estimator in (27) subject to the constraint in (29) using regular polynomial basis functions for various K as indicated in the table columns. The p-values are based on a χ^2 distribution with K degrees of freedom.

For all values of K, the Lagrange multiplier test fails to reject the constrained estimator in favour of the unconstrained estimator. Put differently, there is no statistically significant evidence against a monotonically decreasing stochastic discount factor function. Figure 23 further illustrates these findings. The constrained estimate of the stochastic discount factor is fairly close to the unconstrained estimate and falls well within the confidence interval surrounding the unconstrained estimator. The appendix reports several robustness checks for the above results. First, table 20 illustrates that the results are robust if Arrow-Debreu securities are extracted through alternative methods than Gatheral and Jacquier (2014). Second, table 21 illustrates that the results are robust if Arrow-Debreu securities are extracted through mid, bid, or ask prices. Third, table 22 illustrates that the results are robust if other basis functions such as Legendre, Laguerre, or Hermite polynomials are used. Fourth, table 23 illustrates that the results hold for other maturity horizons, i.e. when considering returns and Arrow-Debreu securities corresponding to two-, three-, six- or twelve-month horizons.

3.5.2.2. Out-of-sample Performance of a Monotonically Decreasing Pricing Kernel. To evaluate the pricing kernel puzzle along a different dimension, I examine the ability of $\pi_t(R;\theta)$ to predict return realizations. Specifically, I compute the conditional mean, $\hat{\mathbb{E}}_t[R_{t+1}]$, implied by the constrained and unconstrained estimator. To evaluate its predictive performance, I compute the associated out-of-sample R^2 following the procedure by Welch and Goyal (2008). Table 8 depicts the out-of-sample R^2 associated with the estimated $\pi_t(R;\theta)$ and $\pi_t(R;\tilde{\theta})$ for several forecasting horizons. A positive value indicates that $\hat{\mathbb{E}}_t[R_{t+1}]$ outperforms the expanding window sample mean benchmark and vice versa.

Table 8. Out-of-Sample Performance across Maturities

	τ	30	60	90	180	360
R^2_{OOS}	Unconstrained	0.030	0.030	0.029	-0.040	-0.423
	Constrained	0.050	0.074	0.076	0.005	-0.412

Note: The table depicts the forecasting performance as measured through the out-of-sample R^2 (Welch and Goyal (2008)) across multiple horizons of the conditional mean implied by the estimator in (27) with and without the constraint in (29). Regular polynomial basis functions with $K = 4$ are employed. See section 6.3 for details on the computation of the out-of-sample R^2.

Table 8 illustrates that the conditional mean based on the constrained estimator outperforms its unconstrained counterpart. Across all maturities, imposing the constraint that the stochastic discount factor function is monotonically decreasing increases the ability of the conditional mean associated with $\pi_t(R;\theta)$ to predict returns. At the one-month horizon, the out-of-sample R^2 increases by 67%, and at the two- and three-month horizon it more than doubles. For larger horizons, the out-of-sample R^2 is generally negative, however, the constraint still leads to improvements. In the spirit of Campbell and Thompson (2008), this demonstrates the added value of imposing economically motivated constraints.

3.5.3. *CRRA Approximation.* The results presented thus far suggest that risk preferences estimated from option data are much more in line with preferences implied by standard utility functions than previously thought. Stochastic discount factor functions implied by standard utility functions are convenient, as they are available closed-form (see e.g. example 3.1) and allow practitioners to translate Arrow-Debreu securities (or risk-neutral distributions) into physical return distributions easily. To what extent can risk preferences associated with standard utility functions approximate the nonparametric estimate?

To make progress on this question, I consider the stochastic discount factor function implied by CRRA preferences[33]. As discussed in example 3.1, this requires estimation

[33] Alternative preferences that often appear in the literature are CARA preferences, i.e. exponential utility (see Bliss and Panigirtzoglou (2004)). I do not consider these here for the following reasons. First, CARA preferences have several questionable implications, e.g. for portfolio choice and risk premia in a growing economy (see Campbell (2017) section 2.1.3 for a discussion). Second, portfolio choice problems with CARA preferences imply a stochastic discount factor over next period wealth, not return as considered here. Third, if initial wealth is normalized to be one, CARA preferences are nested in the nonparametric estimator employing regular polynomial basis functions in equation (27) by setting $K = 1$. Based on section 3.3.2, this implies beliefs implied by CARA preferences only match the first sample moment of returns. Further, the model selection criteria in table 17 suggest K to be at least two, providing suggestive evidence against CARA preferences.

of the risk aversion coefficient as in (25). Table 9 below reports the results for various maturity horizons.

Table 9. CRRA Coefficient Estimates

τ	30	60	90	180	360
$\hat{\gamma}$	2.073	1.832	1.914	1.897	1.862
$se(\hat{\gamma})$	(0.798)	(0.792)	(0.798)	(0.922)	(1.153)

Note: The table depicts the point estimates of the CRRA coefficients across different maturities. Below the point estimates robust standard errors are reported, employing a Newey and West (1987) estimator with lag length corresponding to the respective horizon in trading days.

The estimated risk aversion coefficient, $\hat{\gamma}$, is close to two and stable across maturities. To examine whether the associated stochastic discount factor functions approximate the nonparametric estimate well, I conduct a Lagrange multiplier test. To do so, I project the logarithm of the CRRA implied stochastic discount factor function with $\hat{\gamma}$ onto the basis function space, subject to the location normalization[34] and calculate the associated Lagrange multiplier statistic given in equation (30). Formally, this examines whether the CRRA implied stochastic discount factor function is rejected in favour of the nonparametric estimate. Table 10 reports the test results and figure 24 in the appendix depicts the nonparametric and CRRA implied stochastic discount factor functions.

Table 10. Are CRRA Preferences Rejected?

	τ	30	60	90	180	360
CRRA	LM	7.378	8.593	8.195	3.813	0.922
	p-value	(0.117)	(0.072)	(0.085)	(0.432)	(0.921)

Note: The table depicts the test results that examine whether we can reject the CRRA implied stochastic discount factor (constrained estimator) in favour of the nonparametric estimate across different maturities. To do so the logarithm of the CRRA implied stochastic discount factor function evaluated at $\hat{\gamma}$ reported in table 9 is projected onto the basis function space. Regular polynomial basis functions with $K = 4$ are employed. The p-values are based on a χ^2 distribution with K degrees of freedom.

Table 10 illustrates that for the one-month horizon the CRRA implied stochastic discount factor function cannot be rejected at the conventional 10% level. Further, across frequencies, there is no rejection at the 5% level. By the properties of the Lagrange multiplier test, this implies that the unconditional moments derived from the beliefs implied by the CRRA preferences are approximately consistent with the sample moments observed in the data. Further, this suggests that the CRRA implied stochastic discount factor function is close to its nonparametric counterpart. This is further corroborated by the estimates depicted in figure 24 in the appendix. The CRRA implied stochastic discount factor function lies within the confidence interval around the nonparametric estimate.

Overall, this suggests that CRRA preferences with a risk aversion coefficient of two approximate the nonparametric estimate of $M_t(R)$ well, at least for the S&P 500 at the

[34] Ignoring the location normalization and letting B denote the basis function matrix over the return grid, this amounts to determining the regression coefficients $\tilde{\theta} = (B'B)^{-1}B'(-\hat{\gamma}\log(R))$. The log-likelihood, scores, and Hessian of (26) can then be evaluated at $\tilde{\theta}$ as usual.

one-month horizon. For practitioners, this has the advantage that CRRA preferences with $\gamma = 2$ are easily implemented as they are available in closed-form. This is convenient, since Arrow-Debreu securities (or risk-neutral distributions) can then easily be translated to physical conditional distributions.

3.6. Conclusion. This paper develops a nonparametric estimator of risk preferences from option data and observed returns as long as the true stochastic discount factor is proportional to a time-invariant function. Relative to other approaches in the literature, the estimator is attractive as it is consistent and asymptotically normal.

The estimated risk preferences appear in line with preferences implied by standard utility functions. Specifically, formal tests suggest that there is no evidence in support of the pricing kernel puzzle. In contrast, by constraining estimated risk preferences to be monotonically decreasing – as is suggested by classical utility functions – the associated beliefs are found to better forecast returns.

Since the estimated risk preferences are much more in line with standard utility functions than previously thought, it is reasonable to expect that risk preferences implied by some classical utility function approximate the nonparametric estimate well. The stochastic discount factor implied by CRRA preferences with a risk aversion coefficient of two is found to approximate the nonparametric estimate well. As these are available in closed-form, practitioners can easily translate Arrow-Debreu securities (or risk-neutral distributions) into physical conditional return distributions.

References

Acemoglu, Daron et al. (2012). “The network origins of aggregate fluctuations”. In: *Econometrica* 80.5, pp. 1977–2016.

Ahn, Seung C and Alex R Horenstein (2013). “Eigenvalue ratio test for the number of factors”. In: *Econometrica* 81.3, pp. 1203–1227.

Ait-Sahalia, Yacine and Andrew W Lo (1998). “Nonparametric estimation of state-price densities implicit in financial asset prices”. In: *The Journal of Finance* 53.2, pp. 499–547.

Aıt-Sahalia, Yacine and Andrew W Lo (2000). “Nonparametric risk management and implied risk aversion”. In: *Journal of econometrics* 94.1-2, pp. 9–51.

Alvarez, Fernando and Urban Jermann (2005). “Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth”. In: *Econometrica* 73.6, pp. 1977–2016.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings (July 2022). “Dominant Currencies: How Firms Choose Currency Invoicing and Why it Matters”. In: *The Quarterly Journal of Economics* 137.3, pp. 1435–1493. ISSN: 0033-5533. DOI: 10.1093/qje/qjac004.

Anselin, L (1988). *Spatial econometrics: methods and models*. Studies in operational regional science. Kluwer Academic Publishers. ISBN: 9789024737352.

Atmaz, Adem and Suleyman Basak (2019). “Option prices and costly short-selling”. In: *Journal of Financial Economics* 134.1, pp. 1–28.

Bahaj, Saleem and Ricardo Ferreira Reis (2020). “Jumpstarting an International Currency”.

Bai, Jushan (2003). “Inferential theory for factor models of large dimensions”. In: *Econometrica* 71.1, pp. 135–171.

Bai, Jushan and Serena Ng (2002). “Determining the number of factors in approximate factor models”. In: *Econometrica* 70.1, pp. 191–221.

– (2008). “Large dimensional factor analysis”. In: *Foundations and Trends® in Econometrics* 3.2, pp. 89–163.

– (2013). “Principal components estimation and identification of static factors”. In: *Journal of Econometrics* 176.1, pp. 18–29.

– (2023). “Approximate factor models with weaker loadings”. In: *Journal of Econometrics* 235.2, pp. 1893–1916.

Ballester, Coralio, Antoni Calvo-Armengol, and Yves Zenou (2006). “Who’s who in networks. Wanted: The key player”. In: *Econometrica* 74, pp. 1403–1417.

Bergman, Yaacov Z (1995). “Option pricing with differential interest rates”. In: *The Review of Financial Studies* 8.2, pp. 475–500.

BIS (2014). “Trade finance: developments and issues”. URL: www.bis.org.

Black, Fischer and Myron Scholes (1973). “The Pricing of Options and Corporate Liabilities”. In: *The Journal of Political Economy* 81.3, pp. 637–654.

Bliss, Robert R and Nikolaos Panigirtzoglou (2004). “Option-implied risk aversion estimates”. In: *The Journal of Finance* 59.1, pp. 407–446.

Bonacich, Phillip (Mar. 1987). “Power and centrality: A family of measures”. In: *American Journal of Sociology* 92.5, pp. 1170–1182. ISSN: 0002-9602. DOI: 10.1086/228631.

Boz, Emine et al. (May 2022). “Patterns of invoicing currency in global trade: New evidence”. In: *Journal of International Economics* 136, p. 103604. ISSN: 18730353. DOI: 10.1016/j.jinteco.2022.103604.

Bramoullé, Y, H Djebbari, and B Fortin (2009). “Identification of Peer Effects through Social Networks”. In: *Journal of Econometrics* 150.1, pp. 41–55.

Breeden, Douglas and Robert Litzenberger (1978). “Prices of State-Contingent Claims Implicit in Option Prices”. In: *Journal of Business*, pp. 621–651.

Brunnermeier, Markus et al. (2021). “Feedbacks: financial markets and economic activity”. In: *American Economic Review* 111.6, pp. 1845–1879.

Calvo-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009). “Peer effects and social networks in education”. In: *Review of Economic Studies* 76, pp. 1239–1267.

Calvori, Francesco et al. (2017). “Testing for parameter instability across different modeling frameworks”. In: *Journal of Financial Econometrics* 15.2, pp. 223–246.

Campbell, John Y (2017). *Financial decisions and markets: a course in asset pricing.* Princeton University Press.

Campbell, John Y and Samuel B Thompson (2008). “Predicting excess stock returns out of sample: Can anything beat the historical average?” In: *The Review of Financial Studies* 21.4, pp. 1509–1531.

Chahrour, Ryan and Rosen Valchev (July 2022). “Trade Finance and the Durability of the Dollar”. In: *The Review of Economic Studies* 89.4, pp. 1873–1910. ISSN: 0034-6527. DOI: 10.1093/restud/rdab072.

Chamberlain, Gary and Michael Rothschild (1983). “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets”. In: *Econometrica: Journal of the Econometric Society*, pp. 1281–1304.

Chen, Andrew Y and Tom Zimmermann (2022). “Open Source Cross-Sectional Asset Pricing”. In: *Critical Finance Review* 11.2, pp. 207–265.

Chen, Xiaohong (2007). “Large sample sieve estimation of semi-nonparametric models”. In: *Handbook of Econometrics* 6, pp. 5549–5632.

Christensen, Timothy M (2017). “Nonparametric stochastic discount factor decomposition”. In: *Econometrica* 85.5, pp. 1501–1536.

Christoffersen, Peter, Kris Jacobs, and Bo Young Chang (2013). “Forecasting with option-implied information”. In: *Handbook of economic forecasting* 2, pp. 581–656.

Chung, Wanyu (2016). “Imported inputs and invoicing currency choice: Theory and evidence from UK transaction data”. In: *Journal of International Economics* 99.C, pp. 237–250. DOI: 10.1016/j.jinteco.2015.11.

Connor, Gregory and Robert A Korajczyk (1986). “Performance measurement with the arbitrage pricing theory: A new framework for analysis”. In: *Journal of Financial Economics* 15.3, pp. 373–394.

Cuesdeanu, Horatio and Jens Jackwerth (2018). “The pricing kernel puzzle: Survey and outlook”. In: *Annals of Finance* 14.3, pp. 289–329.

Debreu, G and I N Herstein (1953). “Nonnegative Square Matrices”. In: *Econometrica* 21, pp. 597–607.

Denbee, Edward et al. (2021). “Network risk and key players: A structural analysis of interbank liquidity”. In: *Journal of Financial Economics* 141.3, pp. 831–859. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2021.05.010.

Doepke, Matthias and Martin Schneider (Sept. 2017). “Money as a Unit of Account”. In: *Econometrica* 85.5, pp. 1537–1574. ISSN: 1468-0262. DOI: 10.3982/ECTA11963.

Dybvig, Philip and Stephen Ross (2003). “Arbitrage, state prices and portfolio theory”. In: *Handbook of the Economics of Finance* 1, pp. 605–637.

Elliott, Graham and Ulrich K Müller (2006). "Efficient tests for general persistent time variation in regression coefficients". In: *The Review of Economic Studies* 73.4, pp. 907–940.

Eren, Egemen and Semyon Malamud (May 2022). "Dominant currency debt". In: *Journal of Financial Economics* 144.2, pp. 571–589. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2021.06.023.

Figlewski, Stephen (2008). "Estimating the implied risk neutral density". In.

– (2018). "Risk-neutral densities: A review". In: *Annual Review of Financial Economics* 10, pp. 329–359.

Freyaldenhoven, Simon (2022). "Factor models with local factors—Determining the number of relevant factors". In: *Journal of Econometrics* 229.1, pp. 80–102.

G'Sell, Max Grazier et al. (2016). "Sequential selection procedures and false discovery rate control". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78.2, pp. 423–444.

Gatheral, Jim and Antoine Jacquier (2014). "Arbitrage-Free SVI Volatility Surfaces". In: *Quantitative Finance* 14.1, pp. 59–71.

Giglio, Stefano and Dacheng Xiu (2021). "Asset pricing with omitted factors". In: *Journal of Political Economy* 129.7, pp. 1947–1990.

Goldberg, Linda S. and Cédric Tille (Sept. 2016). "Micro, macro, and strategic forces in international trade invoicing: Synthesis and novel patterns". In: *Journal of International Economics* 102, pp. 173–187. ISSN: 00221996. DOI: 10.1016/j.jinteco.2016.07.004.

Gopinath, Gita (2015). "The International price system". In: *Jackson Hole Symposium* 1, pp. 1–7.

Gopinath, Gita, Emine Boz, et al. (Mar. 2020). "Dominant currency paradigm". In: *American Economic Review* 110.3, pp. 677–719. ISSN: 19447981. DOI: 10.1257/aer.20171201.

Gopinath, Gita, Oleg Itskhoki, and Roberto Rigobon (Mar. 2010). "Currency Choice and Exchange Rate Pass-Through". In: *American Economic Review* 100.1, pp. 304–336. ISSN: 0002-8282. DOI: 10.1257/aer.100.1.304.

Gopinath, Gita and Jeremy C. Stein (Mar. 2021). "Banking, Trade, and the Making of a Dominant Currency". In: *The Quarterly Journal of Economics* 136.2, pp. 783–830. ISSN: 0033-5533. DOI: 10.1093/qje/qjaa036.

Gourinchas, Pierre-Olivier and Hélène Rey (Jan. 2022). "Exorbitant Privilege and Exorbitant Duty".

Gourinchas, Pierre-Olivier, Helene Rey, and Maxime Sauzet (2019). "The International Monetary and Financial System". In: *Annual Review of Economics* 11.1, pp. 859–893.

Hamilton, James Douglas (2020). *Time series analysis*. Princeton university press.

Hansen, Bruce E. (2021). "Criterion-Based Inference Without the Information Equality: The Weighted Chi-Square Distribution". In.

Hansen, Lars and Ravi Jagannathan (1991). "Implications of Security Market Data for Models of Dynamic Economies". In: *Journal of Political Economy* 99.2, pp. 225–262.

– (1997). "Assessing specification errors in stochastic discount factor models". In: *The Journal of Finance* 52.2, pp. 557–590.

Hansen, Lars and Scott F Richard (1987). "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models". In: *Econometrica: Journal of the Econometric Society*, pp. 587–613.

Huang, Jianhua Z (2001). “Concave extended linear modeling: a theoretical synthesis”. In: *Statistica Sinica*, pp. 173–197.

IMF (Oct. 2018). *Direction of Trade Statistics Introductory Notes*. International Monetary Fund.

Jackson, Matthew O and Yves Zenou (Sept. 2012). *Games on Networks*. Tech. rep. 9127. C.E.P.R. Discussion Papers.

Jackson, Matthew O. (2010). *Social and Economic Networks*. New Jersey: Princeton University Press, pp. 1–504. ISBN: 1.

Jackwerth, Jens (2004). “Option-implied risk-neutral distributions and risk aversion”. In.

Jackwerth, Jens and Marco Menner (2020). “Does the Ross Recovery Theorem Work Empirically?” In: *Journal of Financial Economics* 137.3, pp. 723–739.

Jackwerth, Jens Carsten (2000). “Recovering risk aversion from option prices and realized returns”. In: *The Review of Financial Studies* 13.2, pp. 433–451.

Kan, Raymond and Cesare Robotti (2008). “Specification tests of asset pricing models using excess returns”. In: *Journal of Empirical Finance* 15.5, pp. 816–838.

Kelly, Bryan, Seth Pruitt, and Yinan Su (2019). “Characteristics are covariances: A unified model of risk and return”. In: *Journal of Financial Economics* 134.3, pp. 501–524.

Kostakis, Alexandros, Nikolaos Panigirtzoglou, and George Skiadopoulos (2011). “Market timing with option-implied distributions: A forward-looking approach”. In: *Management Science* 57.7, pp. 1231–1249.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2018). “Interpreting factor models”. In: *The Journal of Finance* 73.3, pp. 1183–1223.

– (2020). “Shrinking the cross-section”. In: *Journal of Financial Economics* 135.2, pp. 271–292.

Lancaster, T. (2004). *Introduction to Modern Bayesian Econometrics*. Wiley and Sons. ISBN: 9781405117203.

Lanne, Markku, Helmut Lütkepohl, and Katarzyna Maciejowska (2010). “Structural vector autoregressions with Markov switching”. In: *Journal of Economic Dynamics and Control* 34.2, pp. 121–131.

Lehmann, Erich and George Casella (2006). *Theory of point estimation*. Springer Science & Business Media.

Lehmann, Erich Leo, Joseph P Romano, and George Casella (2005). *Testing statistical hypotheses*. Vol. 3. Springer.

LeSage, J P and R K Pace (2009). *Introduction to spatial econometrics*. Statistics, textbooks and monographs. CRC Press. ISBN: 9781420064247.

Lettau, Martin and Markus Pelger (2020a). “Estimating latent asset-pricing factors”. In: *Journal of Econometrics* 218.1, pp. 1–31.

– (2020b). “Factors that fit the time series and cross-section of stock returns”. In: *The Review of Financial Studies* 33.5, pp. 2274–2325.

Linn, Matthew, Sophie Shive, and Tyler Shumway (2018). “Pricing kernel monotonicity and conditional information”. In: *The Review of Financial Studies* 31.2, pp. 493–531.

Maggiori, Matteo (Oct. 2017). “Financial intermediation, international risk sharing, and reserve currencies”. In: *American Economic Review* 107.10, pp. 3038–3071. ISSN: 00028282. DOI: 10.1257/aer.20130479.

Martin, Ian (2017). “What is the Expected Return on the Market?” In: *The Quarterly Journal of Economics* 132.1, pp. 367–433.

Mrkaic, Mr. Mico, Minsuk Kim, and Rui Mano (Sept. 2020). *Do FX Interventions Lead to Higher FX Debt? Evidence from Firm-Level Data*. IMF Working Papers 2020/197. International Monetary Fund.

Mukhin, Dmitry (Feb. 2022). "An Equilibrium Model of the International Price System". In: *American Economic Review* 112.2, pp. 650–688. ISSN: 0002-8282. DOI: 10.1257/aer.20181550.

Müller, Ulrich K and Philippe-Emmanuel Petalas (2010). "Efficient estimation of the parameter path in unstable time series models". In: *The Review of Economic Studies* 77.4, pp. 1508–1539.

Newey, Whitney and Kenneth West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Onatski, Alexei (2010). "Determining the number of factors from empirical distribution of eigenvalues". In: *The Review of Economics and Statistics* 92.4, pp. 1004–1016.

Ozdagli, Ali K. and Michael Weber (2023). "Monetary Policy through Production Networks: Evidence from the Stock Market".

Pelger, Markus (2019). "Large-dimensional factor modeling based on high-frequency observations". In: *Journal of Econometrics* 208.1, pp. 23–42.

Perks, Michael et al. (Aug. 2021). *Evolution of Bilateral Swap Lines*. IMF Working Papers 2021/210. International Monetary Fund.

Rosenberg, Joshua V and Robert F Engle (2002). "Empirical pricing kernels". In: *Journal of Financial Economics* 64.3, pp. 341–372.

Ross, Stephen A (2015). "The Recovery Theorem". In: *The Journal of Finance* 70.2, pp. 615–648.

Schumaker, Larry (2007). *Spline functions: basic theory*. Cambridge University Press.

Shanken, Jay (1992). "On the estimation of beta-pricing models". In: *The Review of Financial Studies* 5.1, pp. 1–33.

Sims, Christopher A and Tao Zha (1999). "Error Bands for Impulse Responses". In: *Econometrica* 67.5, pp. 1113–1155. DOI: 10.1111/1468-0262.00071.

Stone, Charles J (1990). "Large-sample inference for log-spline models". In: *The Annals of Statistics* 18.2, pp. 717–741.

Svirydzenka, Katsiaryna (Jan. 2016). *Introducing a New Broad-based Index of Financial Development*. IMF Working Papers 2016/005. International Monetary Fund.

Ulrich, Maxim and Simon Walther (2020). "Option-implied information: What's the vol surface got to do with it?" In: *Review of Derivatives Research* 23.3, pp. 323–355.

Weierstrass, Karl (1885). "Über die analytische Darstellbarkeit sogenannter willkürlicher Funktionen einer reellen Veränderlichen". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin* 2, pp. 633–639.

Welch, Ivo and Amit Goyal (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction". In: *The Review of Financial Studies* 21.4, pp. 1455–1508.

Wooldridge, Jeffrey M (1994). "Estimation and inference for dependent processes". In: *Handbook of econometrics* 4, pp. 2639–2738.

4. Appendix of Chapter 1

4.1. Lemmas. In what follows, let $\bar{X} = X - i_T \bar{x}'$, where $\bar{x} = 1/T \sum_t X_t$, let $\|.\|$ denote the Frobenius norm, $\|.\|_{sp}$ denote the spectral norm and $\|.\|_\infty$ denote the infinity norm. Also define $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. Let $\mu = \mathbb{E}[R_t]$ and $\Sigma_T = Var(R_t) = Q\Lambda_T Q'$, where Q denotes the collection of eigenvectors and Λ_T is a diagonal matrix containing eigenvalues. Finally, for the $T \times N$ return matrix R, define the sample covariance estimator as $\hat{\Sigma}_T = 1/T\bar{R}'\bar{R}$ and the scaled sample covariance estimator as $\hat{\Sigma}_T = 1/(TN)\bar{R}'\bar{R}$. Both matrices share the same eigenvectors and their eigenvalues are connected via $\hat{\Lambda}_T = \hat{\Lambda}_{NT}$.

Lemma L.1. *Let $\hat{\Lambda}_{TN,K}$ correspond to the K largest eigenvalues of $\hat{\Sigma}_{TN}$. We have $\hat{\Lambda}_{TN,K} \xrightarrow{p} \Lambda_{TN,K}$.*

Proof. From (8) we have $\bar{R} = \bar{F}Q'_K + \bar{E}$ and therefore

$$\frac{1}{NT}\bar{R}'\bar{R} = \frac{1}{NT}(Q_K\bar{F}'\bar{F}Q'_K + Q_K\bar{F}'\bar{E} + \bar{E}'\bar{F}Q'_K + \bar{E}'\bar{E}) \tag{31}$$

The goal is to analyze the eigenvalues of the above. To do so we will analyze the Frobenius norm of the last three terms as it gives an upper bound for the spectral norm.

Note $\|Q_K\bar{F}'\bar{E}\| = \|\bar{F}'\bar{E}\|$ since $Q'_K Q_K = I_K$. Further,

$$\|\bar{F}'\bar{E}\| = \|F'E - T\bar{f}\bar{e}'\| \leq \|F'E\| + T\|\bar{f}\|\|\bar{e}\|$$

By assumption A.2 $\|F'E\| \leq O_p(\sqrt{NT})$. By assumption A.3 (i) $\|\bar{f}\| = O_p(\sqrt{N})$. By assumption A.1 (i)

$$\mathbb{E}[\|\bar{e}\|^2] = \frac{1}{T^2}\mathbb{E}[\sum_i\sum_t\sum_s E_{i,t}E_{i,s}] = \frac{N}{T^2}\sum_t\sum_s\mathbb{E}[\frac{1}{N}E'_sE_t]$$

$$\leq \frac{N}{T^2}\sum_t\sum_s|\gamma_N(s,t)| \leq M\frac{N}{T}$$

$$\mathbb{E}[\|E\|^2] = \mathbb{E}[\sum_i\sum_t E^2_{i,t}] \leq N\sum_t|\gamma_N(t,t)| \leq MNT$$

so $\|\bar{e}\| = O_p(\sqrt{\frac{N}{T}})$ and $\|E\| = O_p(\sqrt{NT})$. Therefore, $\|\bar{F}'\bar{E}\| = O_p(\sqrt{NT}) + O_p(N\sqrt{T}) = O_p(N\sqrt{T})$. This implies that

$$\|\frac{1}{NT}\bar{F}'\bar{E}\|_{sp} \leq \|\frac{1}{NT}\bar{F}'\bar{E}\| = O_p(\frac{1}{\sqrt{T}}) \xrightarrow{p} 0 \tag{32}$$

which also holds for $\bar{E}'\bar{F}Q'_K$.

Next, note that $\|\bar{E}'\bar{E}\| = \|\bar{E}\bar{E}'\|$ and $\bar{E}\bar{E}' = EE' - E\bar{e}i'_T - i_T\bar{e}'E' + i_T\bar{e}'\bar{e}i'_T$. Note $\|i_T\bar{e}'E'\| \leq \|i_t\|\|\bar{e}\|\|E\| = O_p(\sqrt{T}N)$ and $\|i_T\bar{e}'\bar{e}i'_T\| = T\|\bar{e}\|^2 = O_p(N)$ by assumption A.1 (i). Let Γ be a $T \times T$ matrix with $\gamma_N(s,t)$. By assumption A.1 (ii) we have

$$\mathbb{E}[\|EE' - N\Gamma\|^2] = \sum_t\sum_s\mathbb{E}[(\sum_j e_{j,t}e_{j,s} - \mathbb{E}[e_{j,t}e_{j,s}])^2] \leq MNT^2$$

Define $\rho_N(s,t) = \gamma_N(s,t)/(\gamma_N(s,s)\gamma_N(t,t))$. By the Cauchy-Schwartz inequality $|\rho_N(s,t)| \leq 1$. Then,

$$\begin{aligned}\|\Gamma\|^2 &= \sum_t \sum_s \gamma_N(s,t)^2 = \sum_t \sum_s \rho_N(s,t)^2 \gamma_N(s,s)\gamma_N(t,t) \\ &\leq M \sum_t \sum_s |\rho_N(s,t)| \sqrt{\gamma_N(s,s)\gamma_N(t,t)} = M \sum_t \sum_s |\gamma_N(s,t)| \leq MT\end{aligned}$$

by assumption A.1 (i) and $|\rho_N(s,t)| \leq 1$. To see this note $|\rho_N(s,t)| \geq \rho_N(s,t)^2$ by $|\rho_N(s,t)| \leq 1$. Next, $\gamma_N(s,s)\gamma_N(t,t) = \sqrt{\gamma_N(s,s)\gamma_N(t,t)}^2$ and by assumption A.1 (i) $\sqrt{\gamma_N(s,s)\gamma_N(t,t)} \leq M$. Hence, $\rho_N(s,t)^2\gamma_N(s,s)\gamma_N(t,t) \leq M|\rho_N(s,t)|\sqrt{\gamma_N(s,s)\gamma_N(t,t)}$.

Therefore, we have

$$\begin{aligned}\|\bar{E}\bar{E}' - N\Gamma\| &= \|EE' - N\Gamma\| + 2\|E\bar{e}i_T'\| + \|i_T\bar{e}'\bar{e}i_T'\| \\ &= O_p(T\sqrt{N}) + O_p(N) + O_p(\sqrt{T}N)\end{aligned}$$

and,

$$\begin{aligned}\|\bar{E}'\bar{E}\| &= \|\bar{E}\bar{E}'\| \leq \|\bar{E}\bar{E}' - N\Gamma\| + \|N\Gamma\| \\ &= O_p(T\sqrt{N}) + O_p(N) + O_p(\sqrt{T}N) = O_p(\frac{NT}{\delta_{NT}})\end{aligned} \tag{33}$$

This implies that

$$\|\frac{1}{NT}\bar{E}'\bar{E}\|_{sp} \leq \|\frac{1}{NT}\bar{E}'\bar{E}\| = O_p(\frac{1}{\delta_{NT}}) \xrightarrow{p} 0$$

Therefore, the largest eigenvalues of the three last matrices in (31) converge to zero. By the matrix pertubation theorem, the K largest eigenvalues of $\frac{1}{NT}\bar{R}'\bar{R}$ are then determined by the first matrix $\frac{1}{NT}Q_K\bar{F}'\bar{F}Q_K'$, whose eigenvalues are the same as those of $\frac{1}{NT}\bar{F}'\bar{F}$ since $Q_K'Q_K = I_K$. Since $\frac{1}{NT}\bar{F}'\bar{F}$ converges to $\Lambda_{TN,K}$ by assumption A.3 (ii) we have the desired result. Weyl's inequality can be used to show convergence of each eigenvalue[35]. □

Lemma L.2. *We have* $\|\hat{Q}_K - Q_K H\| \leq O_p(1/\sqrt{T} + 1/N)$.

Proof. I begin by establishing a preliminary convergence rate. From the definition of an eigenvector we have $1/(TN)\bar{R}'\bar{R}\hat{Q}_K = \hat{Q}_K\hat{\Lambda}_{NT,K}$ and therefore

$$\hat{Q}_K = \frac{1}{NT}(Q_K\bar{F}'\bar{F}Q_K' + Q_K\bar{F}'\bar{E} + \bar{E}'\bar{F}Q_K' + \bar{E}'\bar{E})\hat{Q}_K\hat{\Lambda}_{NT,K}^{-1}$$

Define $H = 1/(NT)\bar{F}'\bar{F}Q_K'\hat{Q}_K\hat{\Lambda}_{NT,K}^{-1}$, the rotation matrix up to which we recover Q_K based on $\hat{Q}_K$. Therefore,

$$\hat{Q}_K - Q_K H = \frac{1}{NT}(Q_K\bar{F}'\bar{E} + \bar{E}'\bar{F}Q_K' + \bar{E}'\bar{E})\hat{Q}_K\hat{\Lambda}_{NT,K}^{-1} \tag{34}$$

Note $\|Q_K\| = O_p(1)$ and $\|\hat{Q}_K\| = O_p(1)$. By lemma L.1 and the Continuous Mapping Theorem we have $\|\hat{\Lambda}_{NT,K}^{-1}\| = O_p(1)$. Using (32) and (33) we have

$$\begin{aligned}\|\hat{Q}_K - Q_K H\|^2 &\leq 2\|Q_K\|^2\|\hat{Q}_K\|^2\|\hat{\Lambda}_{NT,K}^{-1}\|^2\|\frac{\bar{F}'\bar{E}}{NT}\|^2 + \|\hat{Q}_K\|^2\|\hat{\Lambda}_{NT,K}^{-1}\|^2\|\frac{\bar{E}'\bar{E}}{NT}\|^2 \\ &= O_p(\frac{1}{T}) + O_p(\frac{1}{\delta_{NT}^2}) = O_p(\frac{1}{\delta_{NT}^2})\end{aligned}$$

[35] By Weyl's inequality we have for $k \in 1, ..., K$

$$|\lambda_k(\frac{1}{NT}\bar{R}'\bar{R}) - \lambda_k(\frac{1}{NT}Q_K\bar{F}'\bar{F}Q_K')| \leq \|\frac{1}{NT}(Q_K\bar{F}'\bar{E} + \bar{E}'\bar{F}Q_K' + \bar{E}'\bar{E})\|_{sp} \leq O_p(\frac{1}{\delta_{NT}}) \xrightarrow{p} 0$$

The above rate can be improved. Note that because $\bar{E}Q_K = 0$

$$\hat{Q}_K - Q_K H = \frac{1}{NT} Q_K \bar{F}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}^{-1}_{NT,K} + \frac{1}{NT}\bar{E}'\bar{F}Q'_K\hat{Q}_K\hat{\Lambda}^{-1}_{NT,K}$$
$$+ \frac{1}{NT}\bar{E}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}^{-1}_{NT,K}$$

Using (32), (33), the Continuous Mapping Theorem with lemma L.1 and the previously established rate

$$\begin{aligned}\|\hat{Q}_K - Q_K H\| &= \|Q_K\|\|\frac{1}{NT}\bar{F}'\bar{E}\|\|(\hat{Q}_K - Q_K H)\|\|\hat{\Lambda}^{-1}_{NT,K}\| \\ &+ \|\frac{1}{NT}\bar{E}'\bar{F}\|\|Q_K\|\|\hat{Q}_K\|\|\hat{\Lambda}^{-1}_{NT,K}\| \\ &+ \|\frac{1}{NT}\bar{E}'\bar{E}\|\|(\hat{Q}_K - Q_K H)\|\|\hat{\Lambda}^{-1}_{NT,K}\| \\ &\leq O_p(\frac{1}{\sqrt{T}\delta_{NT}}) + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\delta^2_{NT}}) = O_p(\frac{1}{\sqrt{T}} + \frac{1}{N})\end{aligned}$$

□

Lemma L.3. *We have* $\|H\| \leq O_p(1)$.

Proof. By the definition of H we have

$$\|H\| \leq \|\frac{\bar{F}'\bar{F}}{NT}\|\|Q_K\|\|\hat{Q}_K\|\|\hat{\Lambda}^{-1}_{NT,K}\| = O_p(1)$$

by assumption A.3 (ii), $\|Q_K\| = O_p(1)$, $\|\hat{Q}_K\| = O_p(1)$, lemma L.1 and the Continuous Mapping Theorem. □

Lemma L.4. *We have* $\|\hat{Q}'_K(Q_K H - \hat{Q}_K)\| \leq O_p(1/T + 1/N^2)$.

Proof. Note we can write

$$\hat{Q}'_K(Q_K H - \hat{Q}_K) = (\hat{Q}_K - Q_K H)'(Q_K H - \hat{Q}_K) + H'Q'_K(Q_K H - \hat{Q}_K)$$

Lemma L.2 establishes the rate of the first component. For the second term, we have by (34)

$$\begin{aligned}Q'_K(Q_K H - \hat{Q}_K) &= -\frac{1}{NT}(Q'_K Q_K \bar{F}'\bar{E} + Q'_K\bar{E}'\bar{F}Q'_K + Q'_K\bar{E}'\bar{E})\hat{Q}_K\hat{\Lambda}^{-1}_{NT,K} \\ &= -\frac{1}{NT}\bar{F}'\bar{E}\hat{Q}_K\hat{\Lambda}^{-1}_{NT,K} \\ &= -\frac{1}{NT}\bar{F}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}^{-1}_{NT,K}\end{aligned}$$

where I used $Q'_K\bar{E}' = 0$. Therefore,

$$\|Q'_K(Q_K H - \hat{Q}_K)\| \leq \|\frac{1}{NT}\bar{F}'\bar{E}\|\|(\hat{Q}_K - Q_K H)\|\|\hat{\Lambda}^{-1}_{NT,K}\| = O_p(\frac{1}{T} + \frac{1}{N\sqrt{T}}) \tag{35}$$

by (32), lemma L.1, L.2 and the Continuous Mapping Theorem. Therefore,

$$\|\hat{Q}'_K(Q_K H - \hat{Q}_K)\| \leq O_p\big(\max\big(\frac{1}{T}, \frac{1}{\sqrt{T}N}, \frac{1}{N^2}\big)\big) = O_p(\frac{1}{T} + \frac{1}{N^2})$$

□

Lemma L.5. *We have* $H = S + O_p(1/T + 1/N^2)$, $H^{-1} = S + O_p(1/T + 1/N^2)$ *where* S *is a diagonal matrix with* 1 *or* -1 *and* $\|H^{-1}\| \leq O_p(1)$.

Proof. Note that

$$\hat{Q}_K' Q_K = (\hat{Q}_K - Q_K H)' Q_K + H' = H' + O_p(\frac{1}{T} + \frac{1}{N\sqrt{T}})$$

by (35). Therefore, $\hat{Q}_K' Q_K H = H'H + O_p(1/(\sqrt{T}\delta_{NT}))$ by lemma L.3. Next, note that

$$\hat{Q}_K' Q_K H = \hat{Q}_K'(Q_K H - \hat{Q}_K) + I_K = I_K + O_p(\frac{1}{T} + \frac{1}{N^2})$$

by lemma L.4. Together, these imply

$$H'H = I_K + O_p(\frac{1}{T} + \frac{1}{N^2})$$

This shows that, up to $O_p(1/\delta_{NT}^2)$, H is an orthogonal matrix and therefore has eigenvalues of 1 or -1. By definition of H I have

$$H' = \hat{\Lambda}_{NT,K}^{-1} \hat{Q}_K' Q_K \bar{F}' \bar{F} \frac{1}{NT} = \hat{\Lambda}_{NT,K}^{-1} H' \bar{F}' \bar{F} \frac{1}{NT} + O_p(\frac{1}{T} + \frac{1}{N\sqrt{T}})$$

Hence,

$$\frac{\bar{F}'\bar{F}}{NT} H = H \hat{\Lambda}_{NT,K} + O_p(\frac{1}{T} + \frac{1}{N\sqrt{T}}) \tag{36}$$

As in Bai and Ng (2013) we can now conclude that H, up to a negligible term, is a matrix consisting of eigenvectors for the diagonal matrix $\bar{F}'\bar{F}/(NT)$. Thus, H consists of eigenvectors that have a single non-zero element. This implies that H is diagonal up to $O_p(1/T + 1/(N\sqrt{T}))$, or $O_p(1/T + 1/N^2)$. Since the eigenvalues are 1 or -1, we have

$$H = S + O_p(\frac{1}{T} + \frac{1}{N^2})$$

where S is a $K \times K$ diagonal matrix with 1 or -1. Note the same argument can be made for H^{-1}, by multiplying (36) by H^{-1} from the left and right to obtain $H^{-1} = S + O_p(1/T + 1/N^2)$. This further implies $\|H^{-1}\| \leq O_p(1)$. □

4.2. Proofs of Theorems.

Proof of theorem 1. The goal is to analyze the asymptotic properties of $\hat{\mu}_F$. The approach is to express $\hat{\mu}_F$ in the quantities characterized in the lemmas to establish its convergence rate. First, note that under E.1 the data generating process is given by (8) and hence

$$\hat{\mu}_F = \frac{1}{T}\sum_t \hat{F}_t = \hat{Q}_K'(Q_K\bar{f} + \bar{e})$$

Note assumption A.3 (i) implies $\frac{1}{\sqrt{N}}\bar{f} \xrightarrow{p} \mu_F$, so it will be convenient to standardize the above by $1/\sqrt{N}$. We then have

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\mu}_F &= \hat{Q}_K'(Q_K\frac{1}{\sqrt{N}}\bar{f} + \frac{1}{\sqrt{N}}\bar{e})\\ &= \hat{Q}_K'(Q_K\mu_F + Q_K(\frac{1}{\sqrt{N}}\bar{f} - \mu_F) + \frac{1}{\sqrt{N}}\bar{e})\\ &= \hat{Q}_K'(Q_K\mu_F + Q_K\bar{v} + \frac{1}{\sqrt{N}}\bar{e})\end{aligned}$$

where I defined $\bar{v} = \frac{1}{\sqrt{N}}\bar{f} - \mu_F$ in line with assumption A.4. Next, note that

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\mu}_F &= \hat{Q}_K'(Q_K H H^{-1}(\mu_F + \bar{v}) + \frac{1}{\sqrt{N}}\bar{e})\\ &= \hat{Q}_K'(\hat{Q}_K H^{-1}(\mu_F + \bar{v}) + (Q_K H - \hat{Q}_K)H^{-1}(\mu_F + \bar{v}) + \frac{1}{\sqrt{N}}\bar{e})\\ &= H^{-1}(\mu_F + \bar{v}) + \hat{Q}_K'(Q_K H - \hat{Q}_K)(H^{-1}\mu_F + H^{-1}\bar{v}) + (\hat{Q}_K - Q_K H)'\frac{1}{\sqrt{N}}\bar{e}\end{aligned}$$

where I used the fact that $Q_K'\bar{e} = 0$ and that $\hat{Q}_K'\hat{Q}_K = I_K$. Therefore,

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\mu}_F - H^{-1}\mu_F &= H^{-1}\bar{v} + \hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\mu_F\\ &\quad + \hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\bar{v} + (\hat{Q}_K - Q_K H)'\frac{1}{\sqrt{N}}\bar{e}\end{aligned} \tag{37}$$

I analyze each term on the right-hand side separately. First, we have

$$\|H^{-1}\bar{v}\| \leq O_p(1)O_p(\frac{1}{\sqrt{T}}) = O_p(\frac{1}{\sqrt{T}})$$

by assumption A.4 and lemma L.5. Next,

$$\|\hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\mu_F\| \leq O_p(\frac{1}{T} + \frac{1}{N^2})O_p(1)O_p(1) = O_p(\frac{1}{T} + \frac{1}{N^2})$$

by lemma L.4 and L.5 and since μ_F is of dimension K and finite by assumption A.3 (i). Next,

$$\|\hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\bar{v}\| \leq O_p(\frac{1}{T} + \frac{1}{N^2})O_p(1)O_p(\frac{1}{\sqrt{T}}) = O_p(\frac{1}{\sqrt{T}T} + \frac{1}{\sqrt{T}N^2})$$

by lemma L.4 and L.5 and assumption A.4. Finally,

$$\|(\hat{Q}_K - Q_K H)'\frac{1}{\sqrt{N}}\bar{e}\| \leq O_p(\frac{1}{T} + \frac{1}{\sqrt{T}N})$$

by lemma L.2 and since $\|\bar{e}\| = O_p(\sqrt{N}/\sqrt{T})$ by assumption A.1.

To conclude, we have

$$\frac{1}{\sqrt{N}}\hat{\mu}_F - H^{-1}\mu_F = H^{-1}\bar{v} + O_p(\frac{1}{T} + \frac{1}{N^2})$$

Note that by lemma L.5 we have $H^{-1} = S_K + O_p(1/T + 1/N^2)$ and therefore

$$\frac{1}{\sqrt{N}}\hat{\mu}_F - S\mu_F = S\bar{v} + O_p(\frac{1}{T} + \frac{1}{N^2})$$

which prooves the first part of the theorem.

Finally, given $\sqrt{T}/N^2 \to 0$ assumption A.4 implies that under the null

$$\sqrt{T}(\frac{1}{\sqrt{N}}\hat{\mu}_F - S\mu_F) \xrightarrow{d} N(0, \Omega_\mu)$$

where $\Omega_\mu = \Omega_v = \lim_{T\to\infty} T\mathbb{E}[\bar{v}\bar{v}']$.

□

Proof of theorem 2. The goal is to analyze the asymptotic properties of $\hat{\alpha}$. To do so assumption A.2' is imposed. This impacts several of the previous lemmas, which are given below.

Lemma L.2'. *We have* $\|\hat{Q}_K - Q_K H\| \leq O_p(1/\delta_{NT}^2)$.

Lemma L.4'. *We have* $\|\hat{Q}_K'(Q_K H - \hat{Q}_K)\| \leq O_p(1/\delta_{NT}^4)$.

The proofs follow the same steps as the original lemmas and are hence omitted for brevity.

Note that the estimate is given by

$$\hat{\alpha} = \bar{r} - \hat{Q}_K \hat{\mu}_F$$

As with $\hat{\mu}_F$, it will be convenient to scale by $1/\sqrt{N}$. Given (8) we have

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\alpha} &= \frac{1}{\sqrt{N}}\bar{r} - \hat{Q}_K \frac{1}{\sqrt{N}}\hat{\mu}_F \\ &= \frac{1}{\sqrt{N}}\bar{e} + (Q_K H - \hat{Q}_K)H^{-1}\mu_F + Q_K\bar{v} - \hat{Q}_K(\frac{1}{\sqrt{N}}\hat{\mu}_F - H^{-1}\mu_F)\end{aligned}$$

Substituting (37)

$$\begin{aligned}\frac{1}{\sqrt{N}}\hat{\alpha} = &\frac{1}{\sqrt{N}}\bar{e} + (Q_K H - \hat{Q}_K)H^{-1}\mu_F \\ &+ (Q_K H - \hat{Q}_K)H^{-1}\bar{v} - \hat{Q}_K\hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\mu_F \\ &- \hat{Q}_K\hat{Q}_K'(Q_K H - \hat{Q}_K)H^{-1}\bar{v} - \hat{Q}_K(\hat{Q}_K - Q_K H)'\frac{1}{\sqrt{N}}\bar{e}\end{aligned}$$

It will be convenient to scale the above expression by $\sqrt{NT}$ to directly analyze $\sqrt{T}\hat{\alpha}$. Doing so

$$\begin{aligned}\sqrt{T}\hat{\alpha} = &\sqrt{T}\bar{e} - \sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\mu_F \\ &- \sqrt{T}\hat{Q}_K(\hat{Q}_K - Q_K H)'\bar{e} - \sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\bar{v} \\ &+ \sqrt{NT}\hat{Q}_K\hat{Q}_K'(\hat{Q}_K - Q_K H)H^{-1}\mu_F + \sqrt{NT}\hat{Q}_K\hat{Q}_K'(\hat{Q}_K - Q_K H)H^{-1}\bar{v}\end{aligned}$$

Note that

$$\|\sqrt{T}\hat{Q}_K(\hat{Q}_K - Q_K H)'\bar{e}\| \leq O_p(1)O_p(\frac{1}{\delta_{NT}^2})O_p(\sqrt{N}) = O_p(\frac{\sqrt{N}}{\delta_{NT}^2})$$

by lemma L.2', assumption A.1 and $\|\hat{Q}_K\| = O_p(1)$. Next

$$\|\sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\bar{v}\| \leq O_p(\sqrt{NT})O_p(\frac{1}{\delta_{NT}^2})O_p(1)O_p(\frac{1}{\sqrt{T}}) = O_p(\frac{\sqrt{N}}{\delta_{NT}^2})$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.2' and assumption A.4. Next

$$\begin{aligned}\|\sqrt{NT}\hat{Q}_K\hat{Q}_K'(\hat{Q}_K - Q_K H)H^{-1}\mu_F\| &\leq O_p(\sqrt{NT})O_p(1)O_p(\frac{1}{\delta_{NT}^4})O_p(1)O_p(1) \\ &= O_p(\frac{\sqrt{NT}}{\delta_{NT}^4})\end{aligned}$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.4', assumption A.3 and $\|\hat{Q}_K\| = O_p(1)$. Finally,

$$\|\sqrt{NT}\hat{Q}_K\hat{Q}_K'(\hat{Q}_K - Q_K H)H^{-1}\bar{v}\| \leq O_p(\sqrt{NT})O_p(1)O_p(\frac{1}{\delta_{NT}^4})O_p(1)O_p(\frac{1}{\sqrt{T}})$$
$$= O_p(\frac{\sqrt{N}}{\delta_{NT}^4})$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.4', assumption A.4 and $\|\hat{Q}_K\| = O_p(1)$. Therefore,

$$\sqrt{T}\hat{\alpha} = \sqrt{T}\bar{e} - \sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\mu_F + O_p(\frac{\sqrt{N}}{\delta_{NT}^2}) + O_p(\frac{\sqrt{NT}}{\delta_{NT}^4}) \tag{38}$$

The second term can be further decomposed using (34), which yields

$$\begin{aligned}\sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\mu_F = &\frac{1}{\sqrt{NT}}Q_K\bar{F}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\\ &+ \frac{1}{\sqrt{NT}}\bar{E}'\bar{F}H\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\\ &+ \frac{1}{\sqrt{NT}}\bar{E}'\bar{F}Q_K'(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\\ &+ \frac{1}{\sqrt{NT}}\bar{E}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\end{aligned}$$

Note that

$$\|\frac{1}{\sqrt{NT}}Q_K\bar{F}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\| \leq O_p(1)O_p(1)O_p(\frac{1}{\delta_{NT}^2})O_p(1)O_p(1)O_p(1)$$
$$= O_p(\frac{1}{\delta_{NT}^2})$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.2' and L.5, assumption A.2' and A.3 and $\|Q_K\| = O_p(1)$. Similarly,

$$\|\frac{1}{\sqrt{NT}}\bar{E}'\bar{F}Q_K'(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\| \leq O_p(1)O_p(1)O_p(\frac{1}{\delta_{NT}^2})O_p(1)O_p(1)O_p(1)$$
$$= O_p(\frac{1}{\delta_{NT}^2})$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.2' and L.5, assumption A.2' and A.3 and $\|Q_K\| = O_p(1)$. Finally

$$\|\frac{1}{\sqrt{NT}}\bar{E}'\bar{E}(\hat{Q}_K - Q_K H)\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F\| \leq O_p(\frac{\sqrt{NT}}{\delta_{NT}})O_p(\frac{1}{\delta_{NT}^2})O_p(1)O_p(1)O_p(1)$$
$$= O_p(\frac{\sqrt{NT}}{\delta_{NT}^3})$$

by the Continuous Mapping Theorem with lemma L.1, lemma L.2' and L.5, assumption A.3 and equation (33). Therefore,

$$\begin{aligned}\sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\mu_F &= \frac{1}{\sqrt{NT}}\bar{E}'\bar{F}H\hat{\Lambda}_{NT,K}^{-1}H^{-1}\mu_F + O_p(\frac{\sqrt{NT}}{\delta_{NT}^3})\\ &= \frac{1}{\sqrt{NT}}E'F\hat{\Lambda}_{NT,K}^{-1}\mu_F - \sqrt{T}\bar{e}\mu_F'\hat{\Lambda}_{NT,K}^{-1}\mu_F\\ &- \sqrt{T}\bar{e}\bar{v}'\hat{\Lambda}_{NT,K}^{-1}\mu_F + O_p(\frac{\sqrt{NT}}{\delta_{NT}^3})\end{aligned}$$

where to obtain the second equality I used the fact that a diagonal matrix always commutes, so $H\hat{\Lambda}^{-1}_{NT,K}H^{-1} = \hat{\Lambda}^{-1}_{NT,K}$, the fact that $\bar{E}'\bar{F} = E'F - T\bar{e}\bar{f}'$ and the definition of $\bar{v}$.

The terms can be analyzed in infinity norm to establish tighter bounds. Specifically, all bounds obtained in Frobenius norm carry over to the infinity norm, as the above are $N \times 1$ vectors. Note

$$\begin{aligned}\|\frac{1}{\sqrt{NT}}E'F\hat{\Lambda}^{-1}_{NT,K}\mu_F\|_\infty &= \max_i |\frac{1}{\sqrt{NT}}E_i'F\hat{\Lambda}^{-1}_{NT,K}\mu_F| = \max_i \|\frac{1}{\sqrt{NT}}E_i'F\hat{\Lambda}^{-1}_{NT,K}\mu_F\| \\ &\leq \max_i \|\frac{1}{\sqrt{NT}}E_i'F\|\|\hat{\Lambda}^{-1}_{NT,K}\mu_F\| \leq O_p(\frac{1}{\sqrt{N}})\end{aligned}$$

where the first equality follows from the definition of the infinity norm, the second equality from the fact that $E_i'F\hat{\Lambda}^{-1}_{NT,K}\mu_F$ is a scalar and the first inequality by standard norm properties. The second inequality uses assumption A.2 and A.3 and the Continuous Mapping Theorem with lemma L.1. Next

$$\begin{aligned}\|\sqrt{T}\bar{e}\bar{v}'\hat{\Lambda}^{-1}_{NT,K}\mu_F\|_\infty &= \max_i |\sqrt{T}\bar{e}_i\bar{v}'\hat{\Lambda}^{-1}_{NT,K}\mu_F| = \max_i \|\sqrt{T}\bar{e}_i\bar{v}'\hat{\Lambda}^{-1}_{NT,K}\mu_F\| \\ &\leq \max_i \|\sqrt{T}\bar{e}_i\|\|\bar{v}'\hat{\Lambda}^{-1}_{NT,K}\mu_F\| \leq O_p(\frac{1}{\sqrt{T}})\end{aligned}$$

using the same arguments as before and assumption A.3, A.4 and A.5 and the Continuous Mapping Theorem with lemma L.1. Therefore,

$$\sqrt{NT}(\hat{Q}_K - Q_K H)H^{-1}\mu_F = -\sqrt{T}\bar{e}\mu_F'\hat{\Lambda}^{-1}_{NT,K}\mu_F + O_p(\frac{1}{\delta_{NT}}) + O_p(\frac{\sqrt{NT}}{\delta^3_{NT}})$$

Substituting the above in (38)

$$\begin{aligned}\sqrt{T}\hat{\alpha} &= \sqrt{T}\bar{e} + \sqrt{T}\bar{e}\mu_F'\hat{\Lambda}^{-1}_{NT,K}\mu_F + O_p(\frac{1}{\delta_{NT}}) + O_p(\frac{\sqrt{NT}}{\delta^3_{NT}}) \\ &= \sqrt{T}\bar{e}(1 + \mu_F'\hat{\Lambda}^{-1}_{NT,K}\mu_F) + O_p(\frac{1}{\delta_{NT}}) + O_p(\frac{\sqrt{NT}}{\delta^3_{NT}})\end{aligned}$$

which proves the first part of the theorem. Notice that the final term vanishes as long as $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$. For example if $T/N \to c < \infty$ these conditions are satisfied.

Finally, assumption A.5, Slutsky's theorem and lemma L.1 imply that under the null, as long as $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$

$$\sqrt{T}\hat{\alpha} \xrightarrow{d} N(0, \Omega_\alpha)$$

where $\Omega_\alpha = (1 + \mu_F'\Lambda^{-1}_{NT,K}\mu_F)^2\Omega_e$ with $\Omega_e = \lim_{T\to\infty} T\mathbb{E}[\bar{e}\bar{e}']$.

□

Proof of theorem 3. The goal is to analyze the asymptotic properties of $\hat{\alpha}'\hat{\Sigma}_T^{-1}\hat{\alpha}$. First, note that since $\hat{\alpha} = \hat{Q}_{N-K}\hat{Q}'_{N-K}\bar{r}$ the HJ-distance can equivalently be written as

$$\begin{aligned}\hat{\alpha}'\hat{\Sigma}_T^{-1}\hat{\alpha} &= \bar{r}'\hat{Q}_{N-K}\hat{\Lambda}_{T,N-K}^{-1}\hat{Q}'_{N-K}\bar{r} \\ &= \bar{r}'\hat{Q}_{N-K}\hat{Q}'_{N-K}\hat{Q}_{N-K}\hat{\Lambda}_{T,N-K}^{-1}\hat{Q}'_{N-K}\hat{Q}_{N-K}\hat{Q}'_{N-K}\bar{r} \\ &= \hat{\alpha}'\hat{\Sigma}_{\alpha,T}^{-1}\hat{\alpha}\end{aligned}$$

where the first equality follows since $\hat{\Sigma}_T = \hat{Q}\hat{\Lambda}_T\hat{Q}'$ and by the orthogonality of eigenvectors. The second equality exploits $\hat{Q}'_{N-K}\hat{Q}_{N-K} = I_{N-K}$. The final equality defines $\hat{\Sigma}_{\alpha,T}^{-1} = \hat{Q}_{N-K}\hat{\Lambda}_{T,N-K}^{-1}\hat{Q}'_{N-K}$. Importantly, $\hat{\Sigma}_{\alpha,T} \neq \hat{\Omega}_\alpha$ defined in theorem 2.

We can further write the HJ-distance as

$$\hat{\alpha}'\hat{\Sigma}_T^{-1}\hat{\alpha} = \hat{\alpha}'\hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{\Omega}_\alpha^{\frac{1}{2}}\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha^{\frac{1}{2}}\hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{\alpha}$$

Under the conditions in theorem 2

$$\sqrt{T}\hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{\alpha} \xrightarrow{d} N(0, I_N)$$

Let $\hat{\Omega}_\alpha^{\frac{1}{2}}\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha^{\frac{1}{2}} = \hat{W}\hat{D}\hat{W}'$ where $\hat{W}'\hat{W} = I_N$. Therefore,

$$\sqrt{T}\hat{W}'\hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{\alpha} \xrightarrow{d} N(0, I_N)$$

For convenience define $\hat{x} = \sqrt{T}\hat{W}'\hat{\Omega}_\alpha^{-\frac{1}{2}}\hat{\alpha}$. Combining the above, the HJ-distance can be written as

$$T\hat{\alpha}'\hat{\Sigma}_T^{-1}\hat{\alpha} = \sum_{i=1}^{N}\hat{x}_i^2\hat{d}_i$$

which is a weighted sum of variables that are asymptotically $\chi^2(1)$. The weights are given by the elements of $\hat{D}$ which are the eigenvalues of $\hat{\Omega}_\alpha^{\frac{1}{2}}\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha^{\frac{1}{2}}$, or equivalently of $\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha$.

To determine the degrees of freedom of the distribution I need to assess the number of non-zero eigenvalues of $\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha$. Since $\hat{\Sigma}_{\alpha,T}^{-1} = \hat{Q}_{N-K}\hat{\Lambda}_{T,N-K}^{-1}\hat{Q}'_{N-K}$ the first matrix has $N-K$ non-zero eigenvalues and K zero eigenvalues. Therefore, $\hat{D}$ will have only $N-K$ non-zero eigenvalues. Hence,

$$T\hat{\alpha}'\hat{\Sigma}_T^{-1}\hat{\alpha} = \sum_{i=1}^{N-K}\hat{x}_i^2\hat{d}_i$$

where the sum is taken for the $N-K$ non-zero eigenvalues.

For finite N the above would converge to a weighted χ^2 distribution[36], however N goes to infinity. As the asymptotic properties of $\hat{x}_i$ are known, so the Lindeberg CLT can be applied to the limit, which however requires an extra argument.

Denote by x_i the limit of $\hat{x}_i$. Note $\hat{x}$ converges to a vector of normally distributed independent random variables with unit variance. Hence x_i is independent of x_j and x_i^2 is $\chi^2(1)$ distributed. Therefore, $\mathbb{E}[x_i^2] = 1$ and $Var(x_i^2) = 2$. Similar to Bai (2003), the almost sure representation theorem can be applied. Because, $\hat{x}_i^2 \xrightarrow{d} x_i^2$ the almost sure representation theorem asserts that there exists a random variable $\hat{x}_i^{*2}$ and x_i^{*2} with the

[36] Specifically, a weighted χ^2 distribution with weights equal to the non-zero eigenvalues of $\Sigma_{\alpha,T}^{-1}\Omega_\alpha$.

same distributions as $\hat{x}_i^2$ and x_i^2 respectively such that $\hat{x}_i^{*2} \xrightarrow{d} x_i^{*2}$ almost surely. Therefore

$$\sum_i^{N-K} \hat{x}_i^{*2}\hat{d}_i = \sum_i^{N-K} x_i^{*2}\hat{d}_i + \sum_i^{N-K} (x_i^{*2} - \hat{x}_i^{*2})\hat{d}_i$$

and the last term is $o_p(1)$ because of almost sure convergence.

Note that the first term is a weighted sum of independent random variables with unit mean and finite variance. So as $N \to \infty$ the Lindeberg CLT applies. Assume $\hat{d}_i$ converges to some d_i, so Slutsky's theorem applies. Therefore, as $N, T \to \infty$

$$\frac{\sum_i^{N-K} x_i^{*2}\hat{d}_i - \sum_i^{N-K} \hat{d}_i}{\sqrt{2\sum_i^{N-K} \hat{d}_i^2}} \xrightarrow{d} N(0,1)$$

which implies

$$\frac{\sum_i^{N-K} \hat{x}_i^{*2}\hat{d}_i - \sum_i^{N-K} \hat{d}_i}{\sqrt{2\sum_i^{N-K} \hat{d}_i^2}} \xrightarrow{d} N(0,1)$$

Finally, since $\hat{x}_i^{*2}$ and $\hat{x}_i^2$ have the same distribution I obtain

$$\frac{\sum_i^{N-K} \hat{x}_i^{2}\hat{d}_i - \sum_i^{N-K} \hat{d}_i}{\sqrt{2\sum_i^{N-K} \hat{d}_i^2}} \xrightarrow{d} N(0,1)$$

To summarize, the above establishes that under the conditions in theorem 2, for consistent estimates of Ω_α and $\Sigma_{\alpha,T}$

$$\frac{T\hat{\alpha}\hat{\Sigma}_T^{-1}\hat{\alpha} - \sum\limits_i^{N-K} \hat{d}_i}{\sqrt{2\sum\limits_i^{N-K} \hat{d}_i}} \xrightarrow{d} N(0,1)$$

where $\hat{d}_i$ are the nonzero eigenvalues of $\hat{\Sigma}_{\alpha,T}^{-1}\hat{\Omega}_\alpha$.

□

4.3. Selection Rules for RP-PCA. I here demonstrate that selection of RP-principal component factors based on risk premia is approximately optimal for squared pricing errors and selection based on Sharpe ratios is optimal for the HJ-distance and the Sharpe ratio of the associated tangency portfolio. The key difference to section 1.3 is that for RP-PCA eigenvectors are no longer equal to betas. However, results carry over.

Let $\tilde{\Sigma} = \Sigma + (1+\gamma)\mu\mu' = \tilde{Q}\tilde{\Lambda}\tilde{Q}'$ denote the modified covariance matrix in Lettau and Pelger (2020b) where γ is a hyperparameter controlling penalization of pricing errors, $\tilde{Q}$ denotes the collection of eigenvectors and $\tilde{\Lambda}$ is a diagonal matrix containing the eigenvalues. For a selection of K eigenvectors, factors are constructed as $\tilde{F}_K = R\tilde{Q}_K$ and let $\tilde{\mu}_F = \tilde{Q}_K'\mu$ denote their means.

Pricing errors are given by

$$\alpha = \mu - \tilde{\beta}\tilde{\mu}_F$$

where $\tilde{\beta} = Cov(R_t, \tilde{F}_t)Var(\tilde{F}_t)^{-1}$. Using the definition of $\tilde{F}_K$ and since $\tilde{Q}_K$ are fixed

$$\tilde{\beta} = \Sigma\tilde{Q}_K(\tilde{Q}_K'\Sigma\tilde{Q}_K)^{-1}$$

Note by definition of $\tilde{\Sigma}$

$$\begin{aligned}\tilde{Q}_K'\Sigma\tilde{Q}_K &= \tilde{Q}_K'\tilde{\Sigma}\tilde{Q}_K - (1+\gamma)\tilde{Q}_K'\mu\mu'\tilde{Q}_K \\ &= \tilde{\Lambda}_K - (1+\gamma)\tilde{\mu}_F\tilde{\mu}_F'\end{aligned}$$

where the second equality follows from the eigendecomposition of $\tilde{\Sigma}$ and the definition of $\tilde{\mu}_F$. By similar arguments,

$$\begin{aligned}\Sigma\tilde{Q}_K &= \tilde{\Sigma}\tilde{Q}_K - (1+\gamma)\mu\mu'\tilde{Q}_K \\ &= \tilde{Q}_K\tilde{\Lambda}_K - (1+\gamma)\mu\tilde{\mu}'_F\end{aligned}$$

Therefore, $\tilde{\beta}$ can be expressed as

$$\begin{aligned}\tilde{\beta} &= (\tilde{Q}_K\tilde{\Lambda}_K - (1+\gamma)\mu\tilde{\mu}'_F)(\tilde{\Lambda}_K - (1+\gamma)\tilde{\mu}_F\tilde{\mu}'_F)^{-1} \\ &= (\tilde{Q}_K\tilde{\Lambda}_K - (1+\gamma)\mu\tilde{\mu}'_F)(\tilde{\Lambda}_K^{-1} + (1+\gamma)\tilde{\Lambda}_K^{-1}\tilde{\mu}_F\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}(1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F)^{-1}) \\ &= \tilde{Q}_K + \tilde{Q}_K(1+\gamma)\tilde{\mu}_F\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}(1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F)^{-1} \\ &\quad - (1+\gamma)\mu\tilde{\mu}'_F\tilde{\Lambda}_K^{-1} - (1+\gamma)^2\mu\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}(1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F)^{-1}\end{aligned}$$

where the second equality follows by the Sherman-Morrison equality. It is straightforward to show using the definition of $\tilde{\mu}_F$ that

$$\tilde{\beta}\tilde{\mu}_F = \tilde{Q}_K\tilde{Q}'_K\mu - \tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu\frac{(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F}{1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F}$$

where $\tilde{Q}_{N-K}$ denotes the collection of eigenvectors omitted from the factor model.

Since $\tilde{Q}$ forms an orthonormal basis $\mu = \tilde{Q}_K\tilde{Q}'_K\mu + \tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu$ and hence

$$\alpha = \tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu\frac{1}{1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F} \tag{39}$$

and since $\tilde{Q}'_{N-K}\tilde{Q}_{N-K} = I_{N-K}$ it follows

$$\alpha'\alpha = \mu'\tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu\frac{1}{(1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F)^2} \tag{40}$$

Note that $\tilde{Q}'_{N-K}\mu$ are the risk premia of the RP-principal component factors of the omitted factors. The second term depends on $\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F$ which is approximately equal to the squared Sharpe ratio of the included factors. Depending on $\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F$ and γ the second term can be above or below 1. Therefore, the second term introduces a non-trivial non-linearity in optimal factor selection. To minimize the first term factors with low risk premia should be excluded. Depending on γ and $\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F$ this is however not guaranteed to minimize $\alpha'\alpha$ due to the second term.

Nevertheless, (40) shows that selection based on risk premia seems beneficial. To demonstrate that selection based on risk premia is at least approximately optimal, note that in the data typically $\tilde{\beta} \approx \tilde{Q}_K$ and hence $\alpha \approx \tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu$ so $\alpha'\alpha \approx \mu'\tilde{Q}_{N-K}\tilde{Q}'_{N-K}\mu$. This suggests that in practice the second term is less important and hence selection of RP-principal component factors based on squared risk premia is approximately optimal for minimizing the sum of squared pricing errors $\alpha'\alpha$.

Turning to the HJ-distance, note that by the Sherman-Morrison equality and the definition of $\tilde{\Sigma}$

$$\Sigma^{-1} = \tilde{\Sigma}^{-1} + (1+\gamma)\tilde{\Sigma}^{-1}\mu\mu'\tilde{\Sigma}^{-1}(1-(1+\gamma)\mu'\tilde{\Sigma}^{-1}\mu)^{-1}$$

Using (39) it is then straightforward to show that

$$\alpha'\Sigma^{-1}\alpha = \mu'\tilde{Q}_{N-K}\tilde{\Lambda}_{N-K}^{-1}\tilde{Q}'_{N-K}\mu\frac{1}{1-(1+\gamma)\tilde{\mu}'_F\tilde{\Lambda}_K^{-1}\tilde{\mu}_F}\frac{1}{1-(1+\gamma)\mu'\tilde{\Sigma}^{-1}\tilde{\mu}}$$

note that the first term resembles the expression in section 1.3.2 and can be thought of as equal to the sum of the squared Sharpe ratios of the omitted factors. The above is decreasing in the sum of the squared Sharpe ratios of the RP-principal component factors included in the factor model. Hence, including RP-principal component factors with the largest squared Sharpe ratios, $(\mu'\tilde{q}_k)^2/\tilde{\lambda}_k$, leads to a factor model that minimizes the HJ-distance.

The return on the associated tangency portfolio is given by $R^*_{t+1} = \tilde{\mu}_F \tilde{\Sigma}_F^{-1} \tilde{F}_{K,t+1}$. Hence, the squared Sharpe ratio is

$$\begin{aligned} SR^2 &= \tilde{\mu}_F' \tilde{\Sigma}_F^{-1} \tilde{\mu}_F \\ &= \mu' \tilde{Q}_K (\tilde{Q}_K' \Sigma \tilde{Q}_K)^{-1} \tilde{Q}_K' \mu \end{aligned}$$

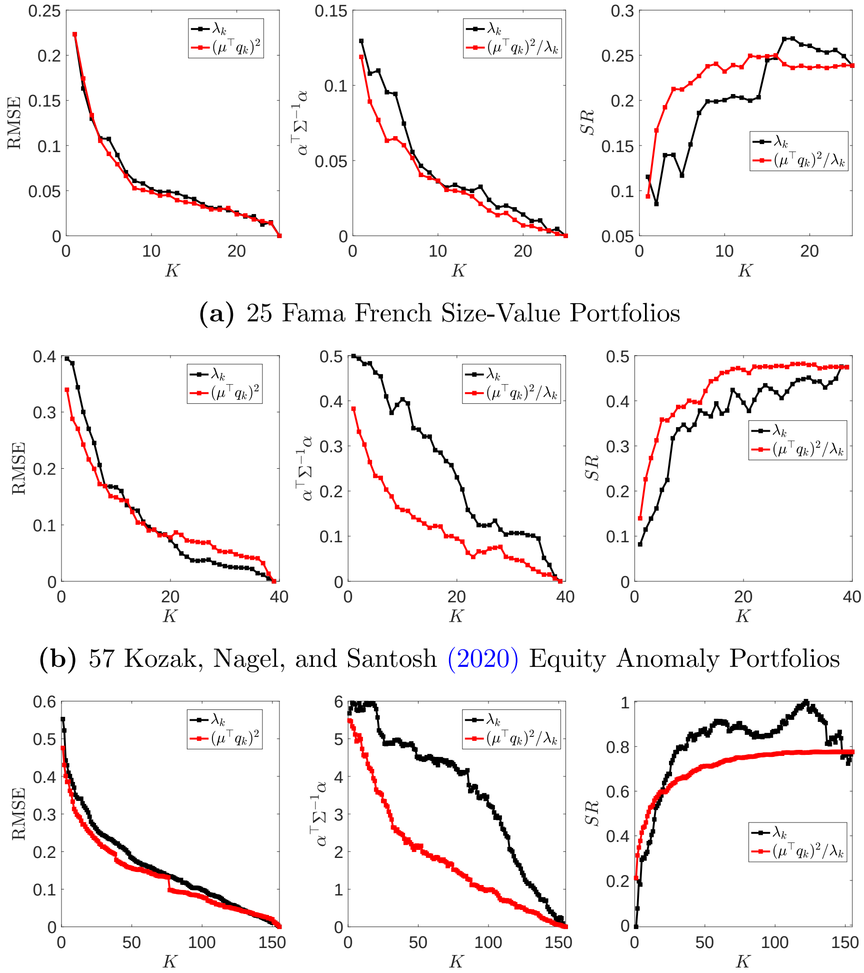
by definition of $\tilde{F}_K$. By the Sherman-Morrison equality and the definition of $\tilde{\Sigma}$

$$SR^2 = \tilde{\mu}_F' \tilde{\Lambda}_K^{-1} \tilde{\mu}_F \frac{1}{1 - (1+\gamma)\tilde{\mu}_F' \tilde{\Lambda}_K^{-1} \tilde{\mu}_F}$$

which is increasing in the sum of squared Sharpe ratios of the included factors, $\tilde{\mu}_F' \tilde{\Lambda}_K^{-1} \tilde{\mu}_F$. Hence, including RP-principal component factors with the largest squared Sharpe ratios, $(\mu'\tilde{q}_k)^2/\tilde{\lambda}_k$, leads to a factor model that maximizes the Sharpe ratio of the associated tangency portfolio.

4.4. Additional Figures.

Figure 14. Full OOS Performance – 24 months



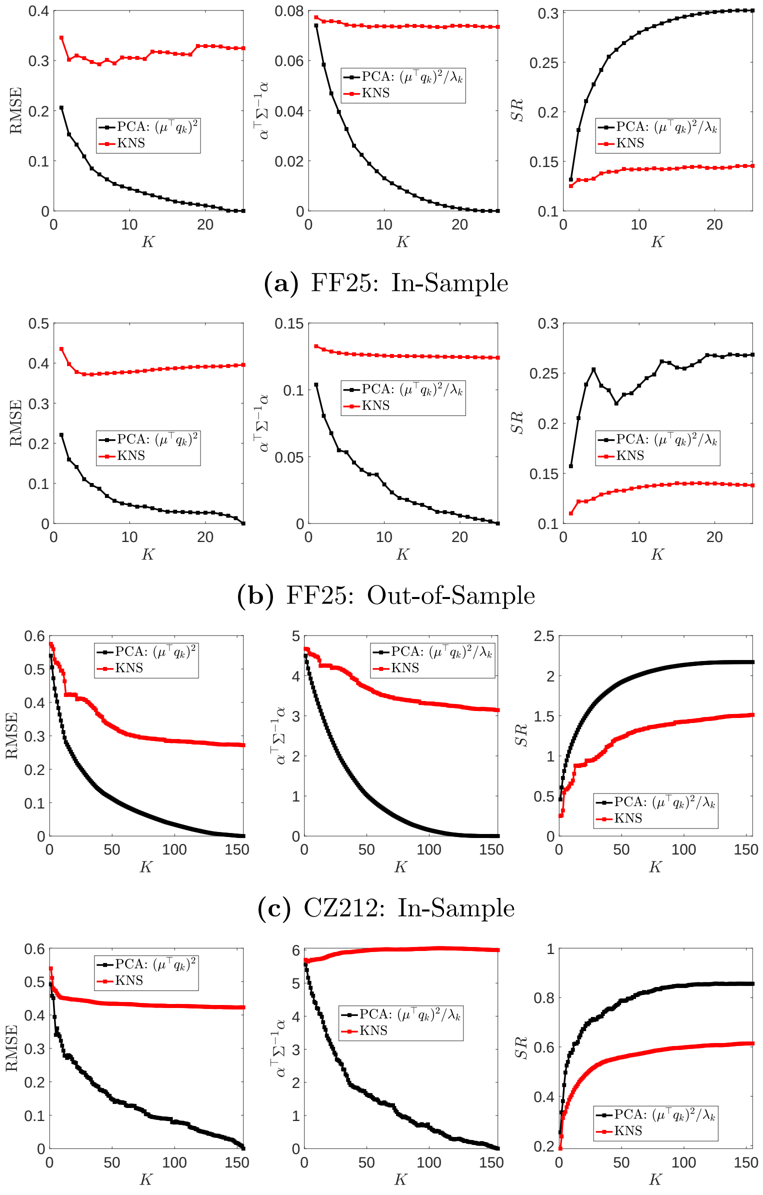
(a) 25 Fama French Size-Value Portfolios

(b) 57 Kozak, Nagel, and Santosh (2020) Equity Anomaly Portfolios

(c) 212 A. Y. Chen and Zimmermann (2022) Equity Anomaly Portfolios

Note: The figure depicts the out-of-sample RMSE, the HJ-distance and Sharpe ratio of R^*_{t+1} for factor models subsequently increasing K, the number of principal components selected. Out-of-sample statistics are constructed in a rolling fashion using 20 years of data to estimate parameters to then construct the pricing errors or return on R^*_{t+1} over the next 24 months before re-estimation. For each test asset set the principal components estimated on the training data are extracted and sorted based on λ_k, $(\mu' q_k)^2$ or $(\mu' q_k)^2/\lambda_k$ to construct the factor models and the tangency portfolio weights.

Figure 15. Performance Comparison with Robust SDF Estimator – Other Datasets



(a) FF25: In-Sample

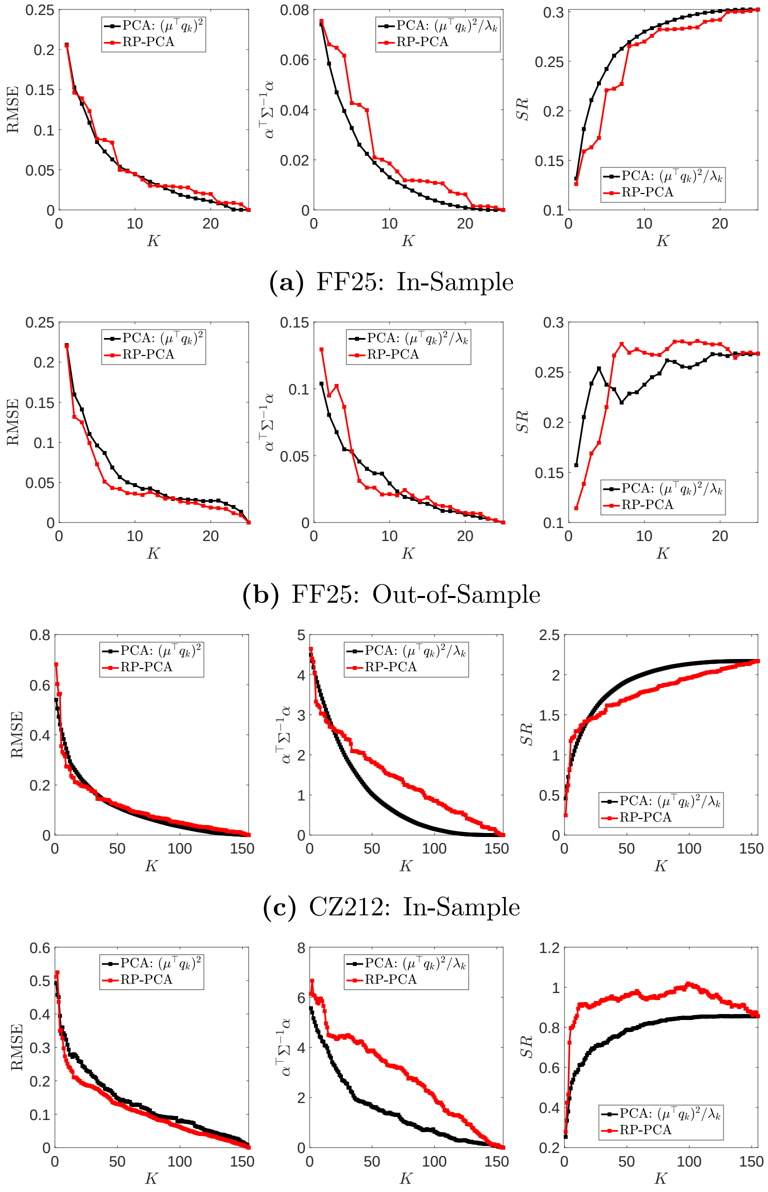
(b) FF25: Out-of-Sample

(c) CZ212: In-Sample

(d) CZ212: Out-of-Sample

Note: The figure replicates figure 5 for the FF25 and CZ212 dataset.

Figure 16. Performance Comparison with Baseline RP-PCA – Other Datasets



(a) FF25: In-Sample

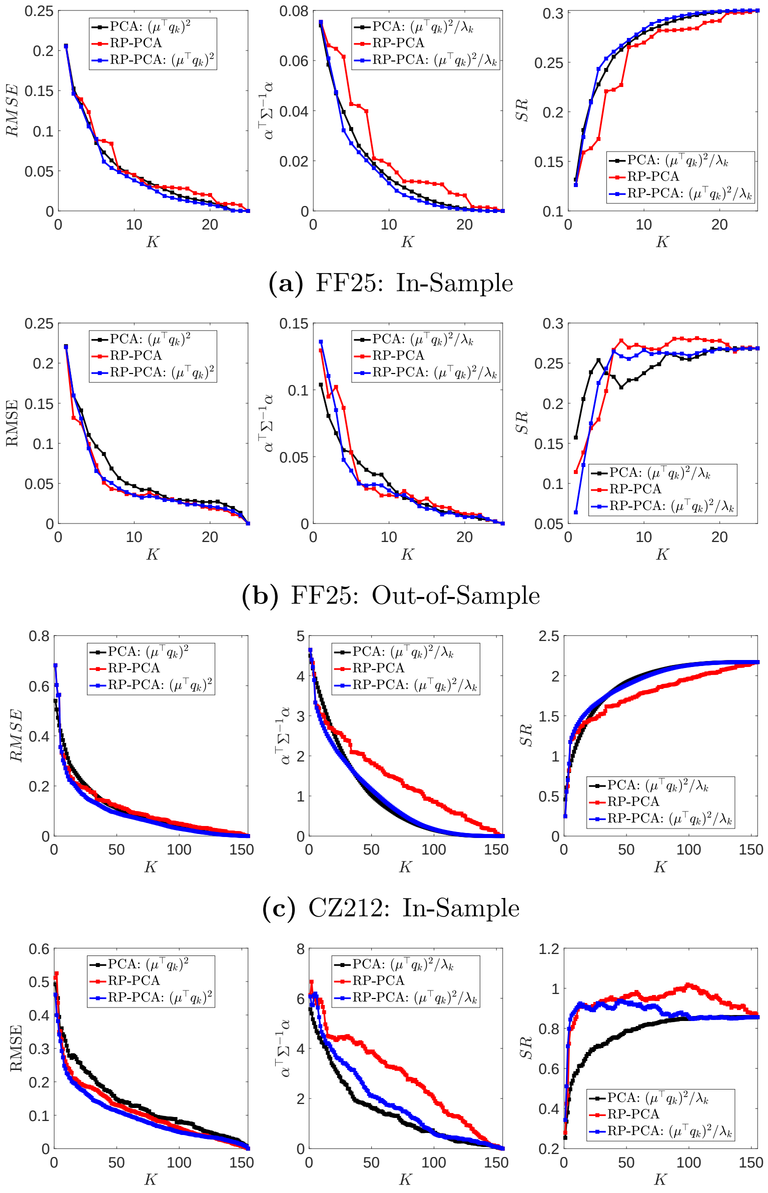
(b) FF25: Out-of-Sample

(c) CZ212: In-Sample

(d) CZ212: Out-of-Sample

Note: The figure replicates figure 5 for the FF25 and CZ212 dataset.

Figure 17. Performance Comparison with Baseline and Optimally Selected RP-PCA – Other Datasets



(a) FF25: In-Sample

(b) FF25: Out-of-Sample

(c) CZ212: In-Sample

(d) CZ212: Out-of-Sample

Note: The figure replicates figure 6 for the FF25 and CZ212 dataset.

4.5. Additional Tables.

Table 11. IS Performance

		RMSE			HJ			SR		
	Factors:	1	3	5	1	3	5	1	3	5
FF25	λ_k	0.206	0.147	0.100	0.076	0.067	0.052	0.125	0.155	0.198
	$(\mu' q_k)^2$	**0.206**	**0.132**	**0.085**	0.076	0.057	0.036	0.125	0.186	0.235
	$(\mu' q_k)^2/\lambda_k$	0.900	0.179	0.107	**0.074**	**0.047**	**0.033**	**0.132**	**0.211**	**0.242**
	KNS	0.346	0.310	0.297	0.077	0.076	0.074	0.125	0.131	0.138
	PL	0.205	0.139	0.089	0.075	0.065	0.043	0.126	0.163	0.221
KNS57	λ_k	0.445	0.420	0.368	0.637	0.628	0.588	0.092	0.135	0.241
	$(\mu' q_k)^2$	**0.432**	**0.316**	**0.243**	0.545	0.511	0.441	0.317	0.368	0.453
	$(\mu' q_k)^2/\lambda_k$	0.432	0.404	0.396	**0.545**	**0.423**	**0.327**	**0.317**	**0.472**	**0.565**
	KNS	0.503	0.416	0.371	0.633	0.578	0.539	0.317	0.331	0.411
	PL	0.446	0.367	0.183	0.630	0.495	0.384	0.129	0.389	0.512
CZ212	λ_k	0.610	0.490	0.456	4.704	4.626	4.562	0.054	0.284	0.380
	$(\mu' q_k)^2$	**0.540**	**0.472**	**0.421**	4.644	4.470	4.214	0.250	0.486	0.702
	$(\mu' q_k)^2/\lambda_k$	0.599	0.570	0.559	**4.498**	**4.182**	**3.928**	**0.457**	**0.724**	**0.883**
	KNS	0.575	0.559	0.519	4.670	4.649	4.532	0.250	0.319	0.581
	PL	0.681	0.561	0.355	4.646	4.320	3.330	0.246	0.621	1.173

Note: The table depicts the RMSE, HJ-distance and Sharpe ratio of R^*_{t+1} for factor models with different K, the number of principal components selected, matching the size of frequently encountered factor models in the literature. Bold letters indicates the model with the best performance for a specific metric across the different selection methods for standard PCA by dataset and approximation degree.

Table 12. OOS Performance

		RMSE			HJ			SR		
	Factors:	1	3	5	1	3	5	1	3	5
FF25	λ_k	0.221	**0.129**	0.103	0.130	0.108	0.088	0.112	0.159	0.157
	$(\mu' q_k)^2$	**0.221**	0.141	**0.096**	0.130	0.104	0.053	0.112	0.160	0.237
	$(\mu' q_k)^2/\lambda_k$	0.607	0.526	0.439	**0.104**	**0.068**	**0.053**	**0.157**	**0.239**	**0.237**
	KNS	0.435	0.378	0.372	0.133	0.129	0.127	0.110	0.122	0.129
	PL	0.220	0.125	0.073	0.130	0.102	0.053	0.114	0.169	0.215
KNS57	λ_k	0.388	0.343	**0.241**	0.489	0.458	0.441	0.070	0.118	0.218
	$(\mu' q_k)^2$	**0.347**	**0.268**	0.250	0.388	0.381	0.349	**0.303**	0.295	0.329
	$(\mu' q_k)^2/\lambda_k$	0.411	0.364	0.356	**0.385**	**0.274**	**0.242**	0.208	**0.347**	**0.366**
	KNS	0.428	0.384	0.339	0.457	0.431	0.427	0.223	0.326	0.326
	PL	0.404	0.222	0.148	0.514	0.373	0.278	0.144	0.358	0.444
CZ212	λ_k	0.560	**0.446**	0.414	5.683	5.936	5.813	-0.041	0.214	0.316
	$(\mu' q_k)^2$	**0.492**	0.451	**0.340**	5.828	5.460	5.422	0.176	0.349	**0.524**
	$(\mu' q_k)^2/\lambda_k$	0.568	0.566	0.564	**5.564**	**5.162**	**4.858**	**0.254**	**0.380**	0.496
	KNS	0.539	0.479	0.472	5.704	5.652	5.679	0.188	0.312	0.337
	PL	0.511	0.436	0.352	6.136	6.090	5.951	0.279	0.465	0.796

Note: The table depicts the out-of-sample RMSE, HJ-distance and Sharpe ratio of R^*_{t+1} for factor models with different K, the number of principal components selected, matching the size of frequently encountered factor models in the literature. Bold letters indicates the model with the best performance for a specific metric across the different selection methods for standard PCA by dataset and approximation degree.

Table 13. Prices of Risk under Different Selection Rules

	FF25			KNS57			CZ212		
PC	λ_k	$(\mu' q_k)^2$	$(\mu' q_k)^2/\lambda_k$	λ_k	$(\mu' q_k)^2$	$(\mu' q_k)^2/\lambda_k$	λ_k	$(\mu' q_k)^2$	$(\mu' q_k)^2/\lambda_k$
PC1	0.356^{***}	0.356^{***}	5.081^{***}	0.536^{**}	5.841^{***}	5.841^{***}	0.189	1.628^{***}	9.988^{***}
	(0.000)	(0.000)	(0.000)	(0.028)	(0.000)	(0.000)	(0.238)	(0.000)	(0.000)
PC2	1.070^{***}	1.070^{***}	0.356^{***}	0.537^{*}	0.536^{**}	24.297^{***}	0.640^{***}	0.640^{***}	7.120^{***}
	(0.001)	(0.001)	(0.000)	(0.088)	(0.028)	(0.000)	(0.005)	(0.005)	(0.000)
PC3	0.477	3.010^{***}	3.010^{***}	0.916^{*}	2.848^{***}	6.050^{***}	1.628^{***}	7.120^{***}	34.554^{***}
	(0.251)	(0.000)	(0.000)	(0.055)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PC4	0.960^{**}	0.960^{**}	1.070^{***}	1.565^{***}	6.050^{***}	11.199^{***}	1.088^{***}	9.988^{***}	10.077^{***}
	(0.050)	(0.050)	(0.001)	(0.006)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)
PC5	3.010^{***}	5.081^{***}	3.717^{***}	2.848^{***}	1.565^{***}	33.412^{***}	2.855^{***}	2.855^{***}	24.397^{***}
	(0.000)	(0.000)	(0.003)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)
PC6	0.932	0.477	4.087^{***}	5.841^{***}	0.916^{*}	12.594^{***}	1.414^{**}	3.595^{***}	39.477^{***}
	(0.289)	(0.251)	(0.001)	(0.000)	(0.055)	(0.000)	(0.025)	(0.000)	(0.000)
PC7	1.210	3.717^{***}	0.960^{**}	0.655	3.957^{***}	42.616^{***}	3.039^{***}	0.189	62.409^{***}
	(0.184)	(0.003)	(0.050)	(0.516)	(0.000)	(0.000)	(0.000)	(0.238)	(0.000)
PC8	5.081^{***}	4.087^{***}	4.619^{**}	3.957^{***}	0.537^{*}	3.957^{***}	3.595^{***}	1.088^{***}	18.429^{***}
	(0.000)	(0.001)	(0.044)	(0.000)	(0.088)	(0.000)	(0.000)	(0.004)	(0.000)
PC9	1.654	1.210	4.188^{*}	6.050^{***}	11.199^{***}	2.848^{***}	0.299	3.039^{***}	21.477^{***}
	(0.178)	(0.184)	(0.076)	(0.000)	(0.000)	(0.000)	(0.641)	(0.000)	(0.000)
PC10	1.786	0.932	3.472^{*}	2.853^{**}	2.853^{**}	13.149^{***}	-0.037	5.664^{***}	7.693^{***}
	(0.116)	(0.289)	(0.090)	(0.015)	(0.015)	(0.000)	(0.965)	(0.000)	(0.000)

Note: The table depicts the estimated prices of risk, $\hat{b}_k = \hat{\mu}'\hat{q}_k/\hat{\lambda}_k$ for the first ten principal component factors under different selection rules. In brackets below the price of risk estimates the p-value associated with $H_0 : b_k = 0$ is reported based on the asymptotic distribution derived in section 1.4. The Newey and West (1987) covariance estimator is employed with $T^{1/4}$ lags. Estimates significant at the 10%, 5% and 1% are indicated by *, ** and *** respectively. Factors have been normalized to have positive mean.

5. Appendix of Chapter 2

5.1. Data.

5.1.1. *Excessive Currency Invoicing.* We define excessive currency invoicing as in (21), that is, as the aggregate exports (imports) per country and currency in excess of exports (imports) to (from) countries that have such currency as their base currency, in USD equivalent amounts. To construct excessive currency invoicing we will rely on data on the value of merchandise imports and exports disaggregated according to a country's trading partners and on data on the shares of aggregate exports (imports) invoiced in USD and EUR by countries.

The data on exports (imports) trades between countries over time is obtained from the Direction of Trade Statistics database of the International Monetary Fund at monthly frequency in USD equivalent amounts. Import trades are quoted as cost, insurance and freight and translated to free-on-board by dividing reported values by 1.1 before January 2000 and by 1.06 thereafter, following IMF (2018). Export trades are quoted as free-on-board. Where possible, missing import (export) values of a country from (to) its counterparty are filled with observed export (import) values of the counterparty to (from) the country. This leaves us with monthly unbalanced import (export) trade data for 216 different countries from January 1960 to January 2022.

The data on the shares of aggregate imports (exports) invoiced by USD and EUR are taken from Boz et al. (2022) and are at annual frequency. We keep only payment share data on the USD and EUR due to data coverage issues. This leaves us with data on payment shares for the USD, EUR, and "*Other excluding EUR and USD*" for imports and exports. We found that in seven cases the last category for exports reported a negative value. In these cases we set the value to zero and renormalize the other shares.

Data for several key countries, such as China, Mexico, and Canada, are missing or poorly covered by the Boz et al. (2022) dataset. For this reason, we augment the dataset with proprietary data obtained from SWIFT on payment settlements across borders. These data are available to us for several countries at monthly frequency, broken down by counterparty country, currency and message type. We focus on message types MT400, which is an advice of payment, and MT700, which is an issue of a documentary credit. Both message types are directly related to trade activity. From this we calculate for each country the shares of aggregate payments sent (imports) or received (exports) in USD and EUR at monthly frequency. To combine the SWIFT payment share data with the annual payment share data by Boz et al. (2022), we average the SWIFT data for each year. This allows us to augment the payment share dataset with data on eleven countries[37] from 2011 to 2022.

To combine the annual payment share data with the monthly import (export) data, we assume that payment shares are constant throughout the reporting year. This leaves us with annual unbalanced payment share data for 119 different countries from 1989 to 2019.

[37] The countries are Mexico, Singapore, the United Arab Emirates, Vietnam, China, Hong Kong, Canada, Taiwan, Libya, Cuba, and Sri Lanka. Data on more countries are available, however, were disregarded due to data quality concerns. Specifically, to add a country, we first calculate the total imports or exports of a country implied by the SWIFT dataset. We then calculate the correlation of changes in SWIFT-based imports or exports with changes in imports or exports reported in the Direction of Trade Statistics database. We keep a country only if the correlation is above 0.2 for both imports and exports. We have made two exceptions to this rule: Canada exhibited a correlation of 0.195 for imports and Mexico exhibited a correlation of 0.12 for exports. Due to the size and role within international trade we kept these in the final dataset.

Following our definition in (21), we first aggregate import (export) trades of a country across counterparties. We then multiply the resulting value by the aggregate import (export) invoicing share for the USD or EUR. Finally, we obtain USD excessive currency invoicing by deducting the import (export) trade with the United States. To obtain EUR excessive currency invoicing, we deduct the import (export) trades with countries that were members of the Euro Area at the respective point in time. This leaves us with monthly unbalanced excessive currency invoicing data for 119 different countries from January 1989 to December 2022.

5.1.2. *Consumer Price Index Data.* We construct inflation and inflation volatility of the consumer price index (CPI) of a country.

The data on the CPI of each country is obtained from the International Financial Statistics and Consumer Price Index (CPI) database of the International Monetary Fund at monthly frequency. For New Zealand and Australia, the CPI was not available at monthly frequency. Hence, we used CPI data at quarterly frequency. For Argentina no CPI data were available. Hence, consumer price implied inflation was taken from the Instituto Nacional de Estadistica y Censos Republica Argentina at monthly frequency.

We first linearly interpolate the raw data to monthly frequency where necessary. We then calculate the implied inflation rate of a country as the percentage change in the inflation index. To this we add the Argentinian inflation data. This leaves us with monthly unbalanced inflation data for 185 different countries from January 1989 to December 2022.

To obtain inflation volatility, a GARCH(1,1) model is fit to the inflation data. The GARCH(1,1) model is fit in an expanding fashion for each country separately and requires at least 24 initial inflation datapoints. If inflation data are missing within the estimation window, due to the data being unbalanced, we linearly interpolate the inflation data. From the fitted model we then infer the monthly volatility at the end of the estimation window and multiply it by $\sqrt{12}$. This leaves us with monthly unbalanced inflation volatility data for 185 different countries from December 1990 to December 2022.

5.1.3. *Exchange Rate Data.* We construct the change and realised volatility of the exchange rate of a country with the USD and the EUR.

The exchange rate data are obtained from Reuters and the Bank of International Settlement Statistics Warehouse at daily frequency. We use the latter database for Euro Area countries as it in our experience accurately accounts for changes in the home currency of countries, i.e. the adoption of the EUR by some countries. We measure exchange rates as the amount of a country's currency per USD (EUR).

Based on the raw data, we calculate the average exchange rate prevailing within each month. We then calculate the change in monthly average exchange rates with the USD (EUR) for each country. Given how we measure exchange rates, a positive change corresponds to a country's currency depreciating, whilst a negative change corresponds to a countries currency appreciating with respect to the USD (EUR). This leaves us with monthly change in exchange rate data for 149 different countries from February 2000 to August 2022.

To obtain realised exchange rate volatility with the USD (EUR), we calculate the daily log return for each exchange rate. We then calculate realised exchange rate volatility within each month. Finally, we multiply the volatility measure with $\sqrt{252}$. To calculate the volatility measure, we require at least 15 observations within each month. This leaves us with monthly realised exchange rate volatility data for 149 different countries from January 2000 to August 2022.

5.1.4. *Trade Share Data.* We construct the percentage trade share of a country out of aggregate import (export) trades.

Data on import (export) trades between countries over time is obtained from the Direction of Trade Statistics database of the International Monetary Fund at monthly frequency in USD equivalent amounts. We clean the raw data following the same steps as in section 5.1.1 and are left with monthly trade data.

To obtain the import (export) trade share of a country out of total trades, we first calculate the total aggregate imports (exports), aggregated across counterparties and countries, at each point in time. We then aggregate the imports (exports) of a country across counterparties. Finally, we divide a country's aggregate imports (exports) by the total aggregate imports (exports). This leaves us with monthly trade shares for 216 different countries from January 1960 to January 2022.

5.1.5. *Foreign Debt by Firms.* We use data on the aggregate amount of firm-level debt denominated in USD and the EUR and its change, measured in USD equivalents.

Data on firm-level debt-instruments aggregated by industry are obtained from Mrkaic, Kim, and Mano (2020) at annual frequency. We first aggregate the raw data by industry and debt-instrument type, leaving us with annual data for USD and EUR denominated debt by country. We then linearly interpolate the raw data to monthly frequency. This leaves us with monthly unbalanced aggregate firm-level debt data for 141 different countries from December 2005 to December 2020.

Based on the aggregate firm-level debt data we also calculate its change. This leaves us with monthly unbalanced change in aggregate firm-level debt data for 139 different countries from January 2006 to December 2020.

5.1.6. *Swap Line Data.* We construct a dummy variable indicating whether a swap line existed between a country and the United States Federal Reserve (FED) or the European Central Bank (ECB).

We obtain data on swap line agreements from Perks et al. (2021) and from The Yale Program on Financial Stability swap line database. To construct our dummy variable, we checked in both databases whether throughout a month a country had a swap line in place with the FED or ECB. If a swap line was in place, the dummy variable takes value one. We construct the swap line dummy variable at monthly frequency for the 119 countries in our excessive currency invoicing sample from April 1994 to December 2019.

5.1.7. *Financial Market Index Data.* We construct the change of the financial market index as developed by Svirydzenka (2016). The index aims to summarise the development of financial institutions and financial markets in terms of their depth, access, and efficiency.

Data on the index by country is obtained from the Financial Development Index database of the International Monetary Fund at annual frequency. We first linearly interpolate the index to monthly frequency and then calculate its change. This leaves us with monthly unbalanced change in financial market index data for 187 different countries from January 1981 to December 2019.

5.1.8. *Foreign Direct Investment Data.* We use foreign direct investment equity flows into or out of the reporting economy, measured in USD equivalents.

Data on foreign direct investment flows by country are obtained from the Balance of Payments database of the World Bank at annual frequency. We linearly interpolate the data to monthly frequency. This leaves us with monthly unbalanced data on foreign direct investment inflows (outflows) for 193 (187) different countries from December 1970 to December 2021.

5.1.9. *Gross Domestic Product Data.* We use nominal gross domestic product data, measured in USD equivalents.

Data on nominal gross domestic product by country is obtained from the "*International Financial Statistics*" database by the International Monetary Fund at annual and quarterly frequencies in local currency. First, we combine the annual and quarterly databases to obtain better country coverage. To do so, we divide the annual figures by four and repeat them throughout the quarters within a year. We then linearly interpolate the data to monthly frequency. Lastly, using our average monthly exchange rate data (see section 5.1.3), we translate the gross domestic product to USD equivalents. This leaves us with monthly unbalanced data on nominal gross domestic product for 104 different countries from January 2000 to June 2022.

5.2. Additional Estimation Details. We estimate four different models (Panel, SEM, SLM, and SDM) using Bayesian methods in our empirical analysis. Throughout we assume a flat prior on $\beta := [\delta^\top, \rho^\top]^\top$, θ, and σ^2 and a uniform prior for ϕ over $[-1, 1]$. Below we describe the posterior sampling algorithms for the three spatial model specifications (SDM, SLM and SEM), while we omit for brevity the one of the simple panel model since it is readily available in the literature (see, e.g., Lancaster (2004)).

It will be convenient to define some notation. As we allow for unbalanced samples in our estimation approach, the number of cross-sectional observation available per period, N_t, changes over time. Let $N = \sum_{t=1}^{T} N_t$ denote the total number of observations in our sample. At each point in time, let $y_t = [y_{1,t}, ..., y_{N_t,t}]^\top$ be the $N_t \times 1$ vector containing our dependent variable observations, let $X_t = [x_{1,t}, ..., x_{N_t,t}]^\top$ be the $N_t \times k$ matrix containing our independent variable observations, and let $\mathbf{G}_t$ be the $N_t \times N_t$ matrix containing the row standardized network weights (hence, $\mathbf{G}_t$ is always a right stochastic matrix). Define the $N \times 1$ vector $y = [y_1, ..., y_T]^\top$, the $N \times k$ matrix $X = [X_1^\top, ..., X_T^\top]^\top$, and the block-diagonal $N \times N$ matrix $\mathbf{G}$ containing $\mathbf{G}_t$ $\forall$ t as its diagonal elements. Furthermore, let I be the $N \times N$ identity matrix, $\tilde{y} = (I - \phi\mathbf{G})y$, and $\tilde{X} = (I - \phi\mathbf{G})X$. We will always be conditioning on the matrix of independent variables X and the network matrix $\mathbf{G}$. Hence, for brevity, we will leave this conditioning implicit in the notation.

5.2.1. *Spatial Error Model.* The model takes the form $(I - \phi\mathbf{G})y = (I - \phi\mathbf{G})X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$. Conditional on ϕ, a flat prior on β and σ^2 yields the normal-inverse-gamma posterior distribution for β and σ^2 of a linear regression model of $\tilde{y}$ on $\tilde{X}$. That is,

$$p(\beta|y, \sigma^2, \phi) \sim N(\hat{\beta}, (\tilde{X}^\top \tilde{X})^{-1}\sigma^2)$$

$$p(\sigma^2|y, \phi) \sim \text{Inv-}\Gamma((N - k)/2 - 1, N\hat{\sigma}^2/2)$$

where $\hat{\beta}$ is the OLS coefficient of a regression of $\tilde{y}$ on $\tilde{X}$ and $\hat{\sigma}^2$ is the corresponding OLS estimate of the residual variance.

The posterior of ϕ conditional on β and σ^2 is non-standard, but can be readily obtained by writing the likelihood[38] for y and dropping the terms that do not affect the posterior shape. This gives us

$$p(\phi|y,\beta,\sigma^2) \propto |I-\phi\mathbf{G}|exp\Big(-\frac{1}{2\sigma^2}[\tilde{y}-\tilde{X}\beta]^\top[\tilde{y}-\tilde{X}\beta]\Big)$$

The above is a non-standard distribution, but we can take draws from it using a Metropolis-Hastings (MH) approach. To do so, we use a Gaussian proposal distribution. To ensure that $|\phi|<1$ (see proposition 1), we always discard draws outside of the support $[-1,1]$ by modifying the acceptance rate.

The Gibbs sampling algorithm, with a nested MH component, to draw from the posterior distribution is then as follows:

(1) Initialization:
- Set $b=1$ and set a starting value ϕ_0

(2) OLS step:
- Compute $\hat{\beta}=(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top\tilde{y}$ and $\hat{\sigma}^2=(\hat{\epsilon}^\top\hat{\epsilon})/(N-k)$, where $\hat{\epsilon}=\tilde{y}-\tilde{X}\hat{\beta}$

(3) Draw β and σ^2:
- Draw σ_b^2 from Inv-$\Gamma((N-k)/2-1, N\hat{\sigma}^2/2)$
- Draw β_b from $N(\hat{\beta},(\tilde{X}^\top\tilde{X})^{-1}\sigma_b^2)$

(4) Draw ϕ using MH:
- Draw ϕ_c from $N(\phi_{b-1},c^2)$
- Calculate the acceptance rate $r=min(1,\frac{p(\phi_c|y,\beta_b,\sigma_b^2)q(\phi_{b-1}|\phi_c)}{p(\phi_{b-1}|y,\beta_b,\sigma_b^2)q(\phi_c|\phi_{b-1})},\mathbb{1}(|\phi_c|\leq 1))$, where $q(\phi_{b-1}|\phi_c)$ is $N(\phi_c,c^2)$ evaluated at ϕ_{b-1} and $p(\phi_c|y,\beta_b,\sigma_b^2)$ is the derived posterior of ϕ evaluated at ϕ_c, β_b and σ_b^2. $q(\phi_c|\phi_{b-1})$ and $p(\phi_{b-1}|y,\beta_b,\sigma_b^2)$ are defined similarly
- Set $\phi_b=\phi_c$ with probability r, else set $\phi_b=\phi_{b-1}$. If $\phi_b=\phi_c$, set $ac_b=1$
- Calculate $acr=\sum_{j=1}^b ac_j/b$. If $acr<0.4$, set $c=c/1.1$. If $acr>0.6$, set $c=1.1c$

(5) Increase b by one and repeat from point 2 above.

Repeating the above B times, after discarding an initial set of draws, leaves us with a set of parameter draws from the posterior. We always set $B=50000$, discard the first 5000 draws, set $\phi_0=0.5$, and initialize $c=0.2$.

5.2.2. *Spatial Lag Model.* The model takes the form $(I-\phi G)y=X\beta+\epsilon$, where $\epsilon\sim N(0,\sigma^2 I)$. Conditional on ϕ, a flat prior on β and σ^2 yields the normal-inverse-gamma posterior distribution for β and σ^2 of a linear regression model of $\tilde{y}$ on X. That is,

$$\begin{aligned} p(\beta|y,\sigma^2,\phi) &\sim N(\hat{\beta},(X^\top X)^{-1}\sigma^2) \\ p(\sigma^2|y,\phi) &\sim \text{Inv-}\Gamma((N-k)/2-1,N\hat{\sigma}^2/2) \end{aligned}$$

where $\hat{\beta}$ is the OLS coefficient of $\tilde{y}$ on X and $\hat{\sigma}^2$ is the corresponding OLS estimate of the residual variance.

The posterior of ϕ conditional on β and σ^2 is non-standard, but can be readily obtained by writing the likelihood[39] for y and dropping the terms that do not affect the posterior

[38] This is simply the likelihood of $y=X\beta+\eta$, where $\eta\sim N(0,\sigma^2(I-\phi G)^{-1}[(I-\phi G)^{-1}]^\top)$.

[39] This is simply the likelihood of $y=(I-\phi G)^{-1}X\beta+\eta$, where $\eta\sim N(0,\sigma^2(I-\phi G)^{-1}[(I-\phi G)^{-1}]^\top)$.

shape. This gives us

$$p(\phi|y,\beta,\sigma^2) \propto |I-\phi\mathbf{G}|exp\bigg(-\frac{1}{2\sigma^2}[\tilde{y}-X\beta]^\top[\tilde{y}-X\beta]\bigg)$$

The above is a non-standard distribution, but we can take draws from it using a Metropolis-Hastings (MH) approach. To do so, we use a Gaussian proposal distribution. To ensure that $|\phi| < 1$ (see proposition 1), we always discard draws outside of the support $[-1,1]$ by modifying the acceptance rate.

The Gibbs sampling algorithm, with a nested MH component, to draw from the posterior distribution is then as follows:

(1) Initialization:
- Set $b=1$ and set a starting value ϕ_0

(2) OLS step:
- Compute $\hat{\beta}=(X^\top X)^{-1}X^\top\tilde{y}$ and $\hat{\sigma}^2=(\hat{\epsilon}^\top\hat{\epsilon})/(N-k)$, where $\hat{\epsilon}=\tilde{y}-X\hat{\beta}$

(3) Draw β and σ^2:
- Draw σ_b^2 from Inv-$\Gamma((N-k)/2-1, N\hat{\sigma}^2/2)$
- Draw β_b from $N(\hat{\beta},(X^\top X)^{-1}\sigma_b^2)$

(4) Draw ϕ using MH:
- Draw ϕ_c from $N(\phi_{b-1},c^2)$
- Calculate the acceptance rate $r=min(1,\frac{p(\phi_c|y,\beta_b,\sigma_b^2)q(\phi_{b-1}|\phi_c)}{p(\phi_{b-1}|y,\beta_b,\sigma_b^2)q(\phi_c|\phi_{b-1})},\mathbb{1}(|\phi_c|\leq 1))$, where $q(\phi_{b-1}|\phi_c)$ is $N(\phi_c,c^2)$ evaluated at ϕ_{b-1} and $p(\phi_c|y,\beta_b,\sigma_b^2)$ is the derived posterior of ϕ evaluated at ϕ_c, β_b and σ_b^2. $q(\phi_c|\phi_{b-1})$ and $p(\phi_{b-1}|y,\beta_b,\sigma_b^2)$ are defined similarly
- Set $\phi_b=\phi_c$ with probability r, else set $\phi_b=\phi_{b-1}$. If $\phi_b=\phi_c$, set $ac_b=1$
- Calculate $acr=\sum_{j=1}^{b}ac_j/b$. If $acr<0.4$, set $c=c/1.1$. If $acr>0.6$, set $c=1.1c$

(5) Increase b by one and repeat from point 2 above.

Repeating the above B times, after discarding an initial set of draws, leaves us with a set of parameter draws from the posterior. We always set $B=50000$, discard the first 5000 draws, set $\phi_0=0.5$ and initialize $c=0.2$.

5.2.3. *Spatial Durbin Model.* The model takes the form $(I-\phi\mathbf{G})y=X\beta+\mathbf{G}X_s\theta+\epsilon$, where $\epsilon\sim N(0,\sigma^2 I)$. The matrix X_s is of dimension $N\times s$ with $s\leq k$ and contains a subset of the matrix X. This allows for spatial lags of some independent variables to enter the model. It will be convenient to define $Z=[X,\mathbf{G}X_s]$ and $\gamma=[\beta^\top,\theta^\top]^\top$. Conditional on ϕ, a flat prior on γ and σ^2 yields the normal-inverse-gamma posterior distribution for γ and σ^2 of a linear regression model of $\tilde{y}$ on X. That is,

$$p(\gamma|y,\sigma^2,\phi)\sim N(\hat{\gamma},(Z^\top Z)^{-1}\sigma^2)$$
$$p(\sigma^2|y,\phi)\sim \text{Inv-}\Gamma((N-k-s)/2-1,N\hat{\sigma}^2/2)$$

where $\hat{\gamma}$ is the OLS coefficient of $\tilde{y}$ on Z and $\hat{\sigma}^2$ is the corresponding OLS estimate of the residual variance.

The posterior of ϕ conditional on γ and σ^2 is non-standard, but can be readily obtained by writing the likelihood[40] for y and dropping the terms that do not affect the posterior

[40] This is simply the likelihood of $y=(I-\phi G)^{-1}Z\gamma+\eta$, where $\eta\sim N(0,\sigma^2(I-\phi G)^{-1}[(I-\phi G)^{-1}]^\top)$.

shape. This gives us

$$p(\phi|y,\gamma,\sigma^2) \propto |I-\phi G| exp\bigg(-\frac{1}{2\sigma^2}[\tilde{y}-Z\gamma]^\top[\tilde{y}-Z\gamma]\bigg)$$

The above is a non-standard distribution, but we can take draws from it using a Metropolis-Hastings (MH) approach. To do so, we use a Gaussian proposal distribution. To ensure that $|\phi| < 1$ (see proposition 1), we always discard draws outside of the support $[-1, 1]$ by modifying the acceptance rate.

The Gibbs sampling algorithm, with a nested MH component, to draw from the posterior distribution is then as follows:

1. Initialization:
   - Set $b = 1$ and set a starting value ϕ_0
2. OLS step:
   - Compute $\hat{\gamma} = (Z^\top Z)^{-1} Z^\top \tilde{y}$ and $\hat{\sigma}^2 = (\hat{\epsilon}^\top \hat{\epsilon})/(N-k-s)$, where $\hat{\epsilon} = \tilde{y} - Z\hat{\gamma}$
3. Draw γ and σ^2:
   - Draw σ^2_b from Inv-$\Gamma((N-k-s)/2-1, N\hat{\sigma}^2/2)$
   - Draw γ_b from $N(\hat{\gamma}, (Z^\top Z)^{-1}\sigma^2_b)$
4. Draw ϕ using MH:
   - Draw ϕ_c from $N(\phi_{b-1}, c^2)$
   - Calculate the acceptance rate $r = min(1, \frac{p(\phi_c|y,\gamma_b,\sigma^2_b)q(\phi_{b-1}|\phi_c)}{p(\phi_{b-1}|y,\gamma_b,\sigma^2_b)q(\phi_c|\phi_{b-1})}, \mathbb{1}(|\phi_c| \leq 1))$, where $q(\phi_{b-1}|\phi_c)$ is $N(\phi_c, c^2)$ evaluated at ϕ_{b-1} and $p(\phi_c|y,\gamma_b,\sigma^2_b)$ is the derived posterior of ϕ evaluated at ϕ_c, γ_b and σ^2_b. $q(\phi_c|\phi_{b-1})$ and $p(\phi_{b-1}|y,\gamma_b,\sigma^2_b)$ are defined similarly
   - Set $\phi_b = \phi_c$ with probability r, else set $\phi_b = \phi_{b-1}$. If $\phi_b = \phi_c$, set $ac_b = 1$
   - Calculate $acr = \sum_{j=1}^{b} ac_j/b$. If $acr < 0.4$, set $c = c/1.1$. If $acr > 0.6$, set $c = 1.1c$
5. Increase b by one and repeat from point 2 above.

Repeating the above B times, after discarding an initial set of draws, leaves us with a set of parameter draws from the posterior. We always set $B = 50000$, discard the first 5000 draws, set $\phi_0 = 0.5$, and initialize $c = 0.2$.

5.2.4. Γ *Estimation via Heterogeneous SVAR* . For exposition we consider the SDM case. Let x denote the dataset, i.e. $x \in \{EUR-Im, USD-Im, EUR-Ex, USD-Ex\}$. Define $Z^x_t = [X_t, G^x_t X_t]$, $\gamma^x = [(\beta^x)^\top, (\theta^x)^\top]^\top$, and $\epsilon^x_t = (I - \phi^x G^x_t) y^x_t - Z^x_t \gamma^x$. Denote by $\boldsymbol{\epsilon_{i,t}} = [\epsilon^{Im}_{USD,i,t}, \epsilon^{Im}_{EUR,i,t}, \epsilon^{Ex}_{USD,i,t}, \epsilon^{Ex}_{EUR,i,t}]^\top$ the 4×1 vector of residuals for a specific country. Let $\boldsymbol{\epsilon_i}$ be the corresponding $4 \times T_i$ matrix of residuals for country i. Define $\boldsymbol{\tilde{y}}_i$ as the $T_i \times 4$, matrix where the element in row t and column x is $\tilde{y}^x_{i,t} = \{(I - \phi^x G^x_t)\}_{i.} y^x_t$. Define $\boldsymbol{Z_i} = [Z^{Im}_{USD,i}, Z^{Im}_{EUR,i}, Z^{Ex}_{USD,i}, Z^{Ex}_{EUR,i}]$ as the $T_i \times 4k$ matrix of independent variables, and define $\boldsymbol{\Psi}$ as a blockdiagonal matrix of dimension $4k \times 4$ with the respective γ^x on its diagonal. Note that we have $\boldsymbol{\epsilon_i}^\top = \boldsymbol{\tilde{y}}_i - \boldsymbol{Z_i \Psi}$. For convenience, we denote the collection of parameters across the different datasets by $\boldsymbol{\alpha} = [\alpha^{Im}_{USD}, \alpha^{Im}_{EUR}, \alpha^{Ex}_{USD}, \alpha^{Ex}_{EUR}]$.

We assume that $\epsilon_{i,t} \sim N(0, \Sigma_i)$, where $\Sigma_i = \Gamma^{-1}\Lambda_i(\Gamma^{-1})^\top$ for all i. Note that the likelihood is proportional to

(41)

$$q(\{\Sigma_i\}, \boldsymbol{\Psi}, \boldsymbol{\phi}) \propto \prod_{i=1}^{N} |\Sigma_i|^{-T_i/2} \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T_i} \epsilon_{i,t}^\top \Sigma_i^{-1}\epsilon_{i,t}\Bigg)$$
$$= \prod_{i=1}^{N} |\Sigma_i|^{-T_i/2} \exp\Bigg(-\frac{1}{2} trace(\Sigma_i^{-1}[\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}{}^\top + (\boldsymbol{\Psi} - \hat{\boldsymbol{\Psi}})^\top \boldsymbol{Z}_{\boldsymbol{i}}^\top \boldsymbol{Z}_{\boldsymbol{i}}(\boldsymbol{\Psi} - \hat{\boldsymbol{\Psi}})])\Bigg)$$

where $\hat{\boldsymbol{\Psi}}$ are the OLS estimates of the regression parameters and $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}$ the corresponding OLS residuals.

Similar to Lanne, Lütkepohl, and Maciejowska (2010) and Brunnermeier et al. (2021), an identification issue remains. We can multiply Γ and Λ by scale factors without changing the likelihood. Following Brunnermeier et al. (2021), we impose the restriction

$$\frac{1}{N}\sum_{i=1}^{N} \lambda_{x,i} = 1 \quad \forall \quad x \in 1, ..., 4$$

where in slight abuse of notation, $\lambda_{x,i}$ is the x^{th} diagonal element of Λ_i. The interpretation of this normalization is that we make the cross-country average structural variance one in each equation. Given the normalization and the technical condition that each pair of equations differs in variance in at least one country, we can uniquely identify Γ, up to the sign of a row. See Lanne, Lütkepohl, and Maciejowska (2010) for details and Brunnermeier et al. (2021) for a similar application in the context of time series heteroskedasticity.

Following Brunnermeier et al. (2021), we use a Dirichlet prior for $\lambda_{x.}/N$. This restricts each $\lambda_{x,i}$ to lie in $[0, N]$ and enforces our normalization constraint that for each structural shock the $\lambda_{x,i}$ average to one across countries. We further introduce a prior $p(\Gamma) = |\Gamma|^{4k}$ and integrate out $\boldsymbol{\Psi}$ such that the posterior becomes

$$p(\{\Lambda_i\}, \Gamma|\boldsymbol{\phi}) \propto \prod_{i=1}^{N} |\Gamma|^{T_i}|\Lambda_i|^{-(T_i-4k)/2} \exp\Bigg(-\frac{1}{2} trace(\Gamma^\top\Lambda_i^{-1}\Gamma\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}{}^\top)\Bigg) \prod_{j=1}^{4}\frac{\mathit{\Gamma}(\alpha N)}{\mathit{\Gamma}(\alpha)^N}\prod_{i=1}^{N}\frac{\lambda_{j,i}}{N}^{\alpha-1}$$

where, in slight abuse of notation, $\mathit{\Gamma}(.)$ refers to the Gamma function. The above is a non-standard distribution, but we can take draws from it using a Metropolis-Hastings (MH) approach. To do so we will employ random walks for all parameters[41].

Note that the distribution of $\{\Lambda_i\}$ and Γ is conditional on the $\boldsymbol{\phi}$. Specifically, the $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}$ are OLS residuals computed conditional on the draw of $\boldsymbol{\phi}$. Hence, given draws for $\boldsymbol{\phi}$, we can sample $\{\Lambda_i\}$ and Γ. The draws of $\{\Lambda_i\}$ and Γ do not affect the draws of the other parameters. Hence, to make drawing efficient, we first carry out the estimation of the models on the different ECI datasets and store the corresponding parameter draws, or, equivalently, OLS residuals $\hat{\epsilon}_t^x$. Put differently, we can estimate $\{\Lambda_i\}$ and Γ and the other parameters in two separate stages.

The sampling algorithm to draw $\{\Lambda_i\}$ and Γ from the posterior distribution is as follows:

(1) Pre-estimation:
- Estimate the model on the different ECI datasets and store $\hat{\epsilon}_t^x$ for different draws

(2) Initialization:

[41] For Γ this is straightforward. Note that Λ_i needs to be sampled subject to the normalization constraint. We enforce the constraint, by sampling $\lambda_{x,1:N-1}$, using the random walk and set $\lambda_{x,N}$ such that the constraint is satisfied.

- Set $b = 1$ and set a starting value Γ_0 and $\Lambda_{i,0}$ for all i

(3) Draw $\{\Lambda_i\}$ and Γ using MH:

- Denote the collection of parameters as $\theta = [vec(\Gamma)^\top, diag(\Lambda_1)^\top, ..., diag(\Lambda_{N-1})^\top]^\top$
- Draw $vec(\theta_c)$ from $N(vec(\theta_{b-1}, cV_\theta)$
- Determine $\Lambda_{N,c}$ from the constraint
- Calculate the acceptance rate $r = min(1, \frac{p(\{\Lambda_{i,c}\},\Gamma_c|\boldsymbol{\phi}_{b-1})}{p(\{\Lambda_{i,b-1}\},\Gamma_{b-1}|\boldsymbol{\phi}_{b-1})})$, where $p(\{\Lambda_i\}, \Gamma|\boldsymbol{\phi})$ was defined previously
- Set $\{\Lambda_{i,b}\} = \{\Lambda_{i,c}\}$ and $\Gamma_b = \Gamma_c$ with probability r, else set $\{\Lambda_{i,b}\} = \{\Lambda_{i,b-1}\}$ and $\Gamma_b = \Gamma_{b-1}$. If accepted, set $ac_b = 1$
- Calculate $acr = \sum_{j=1}^{b} ac_j/b$. If $acr < 0.4$, set $c = max(c/1.1, 10^{-6})$. If $acr > 0.6$, set $c = min(1.1c, 5)$

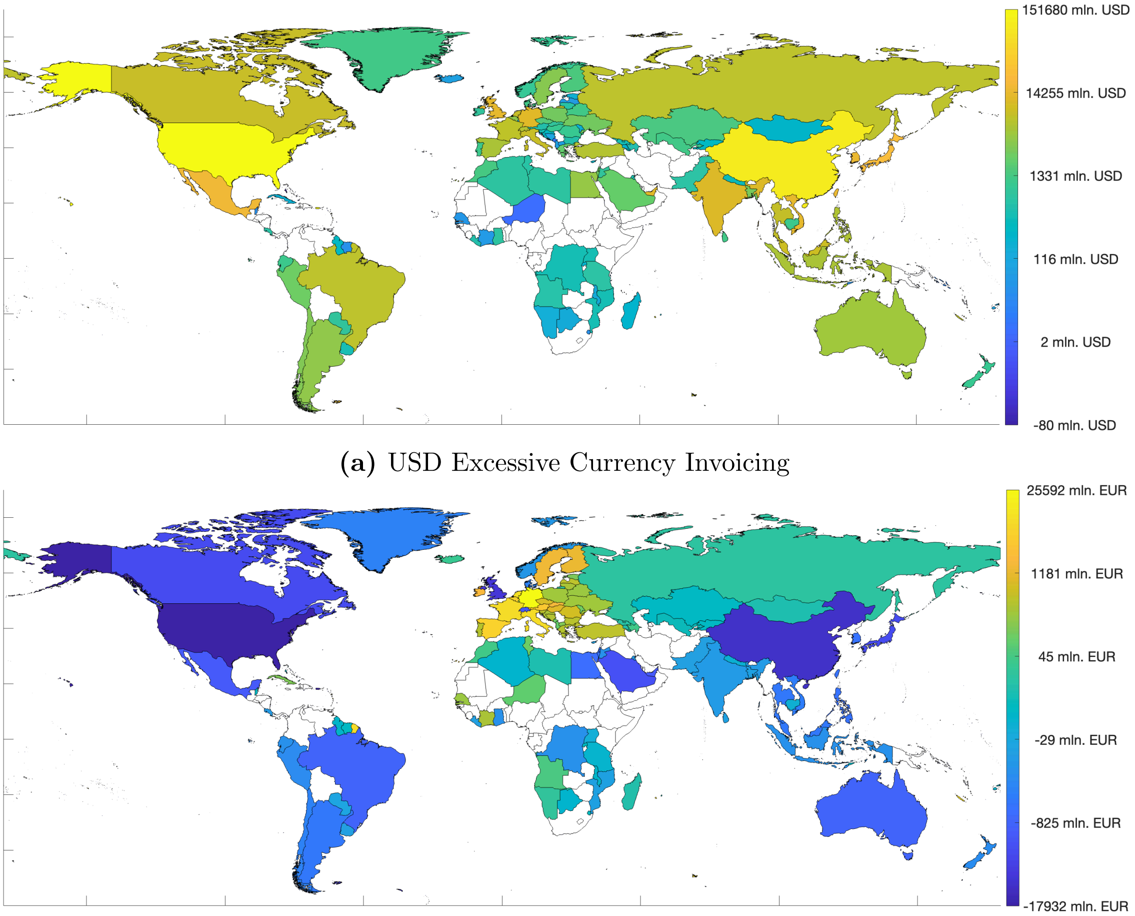
(4) Increase b by one and repeat from point 3 above.

Repeating the above B times, after discarding an initial set of draws, leaves us with a set of parameter draws from the posterior. We always set $B = 50000$ and discard the first 5000 draws. To initialize the algorithm we first compute the posterior mode[42]. We then set Γ_0 and $\Lambda_{i,0}$ to the respective estimates. We start with $c = 0.1$, set V_θ to the inverse Hessian obtained from the estimation of the posterior mode, and set the prior for $\{\Lambda_i\}$ with $\alpha = 2$. We have found that since we use an estimate of the posterior mode satisfying the normalizations, no ex-post normalization (sign-flipping or row permutation) was necessary. To obtain a reliable sample from the posterior distribution, we ran the above algorithm 100 times and pooled the final draws across all MCMC chains.

[42] The posterior mode is obtained from a constrained optimization for $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{i}}$ evaluated at the posterior mean of $\boldsymbol{\phi}$. The constraints ensure that the diagonal elements of Γ are positive. Finally the rows of the resulting estimates for Γ and Λ_i are permuted to ensure that the large elements are on the diagonal of Γ.

5.3. Additional Figures.

Figure 18. Import-Based Excessive Currency Invoicing across Countries

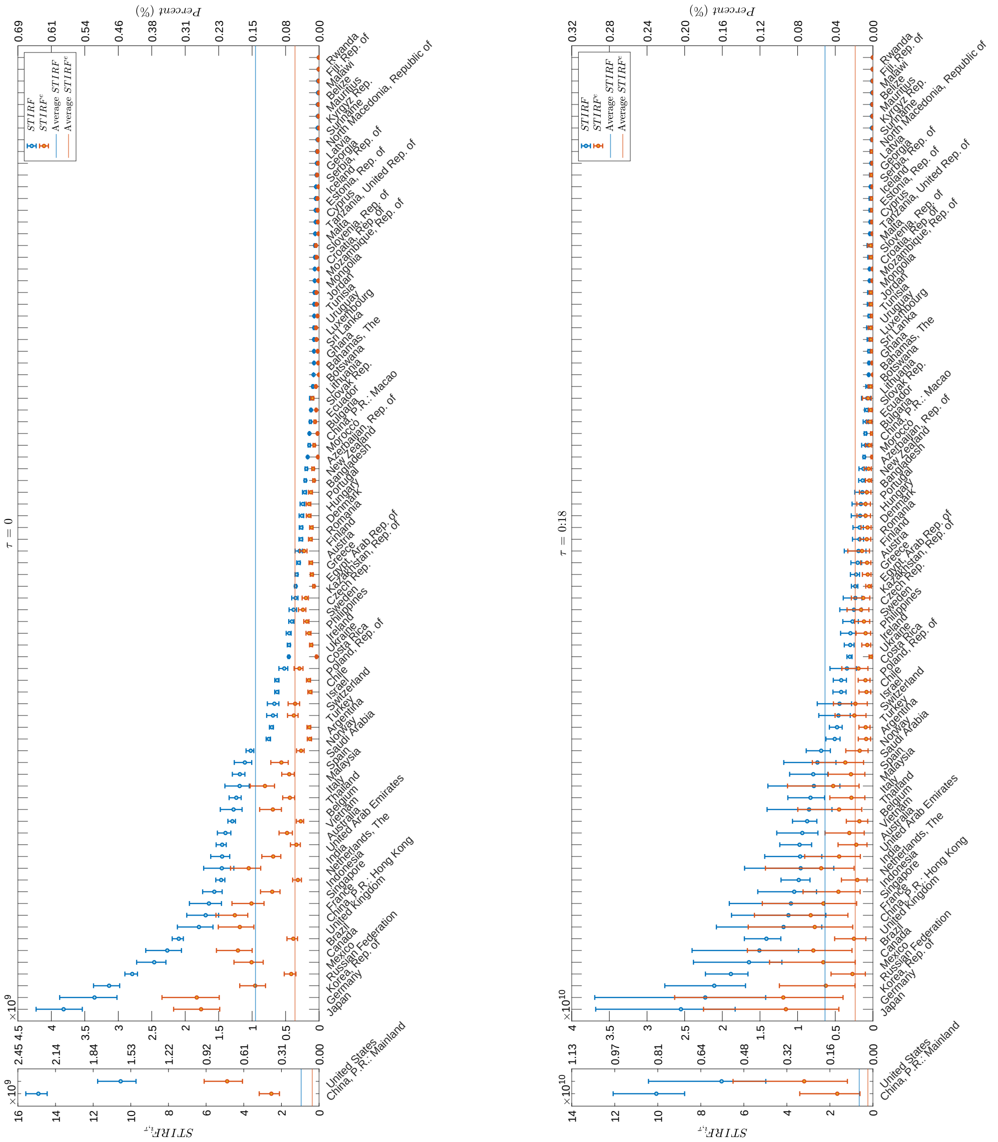


(a) USD Excessive Currency Invoicing

(b) EUR Excessive Currency Invoicing

The figure depicts the average monthly excessive currency invoicing across countries over our sample. All amounts are in USD equivalents. The countries marked in white are not included in our sample due to missing observations. The top ten countries by import-based excessive USD invoicing positions in our sample are: the United States, China, Hong Kong, Japan, Singapore, Mexico, Taiwan, South Korea, Vietnam, and the United Kingdom. The top ten countries by import-based excessive EUR invoicing positions in our sample are: Germany, the Netherlands, Italy, France, Belgium, Spain, Austria, Ireland, the Slovak Republic, and Sweden. Panel (a): USD excessive currency invoicing. Panel (b): EUR excessive currency invoicing.

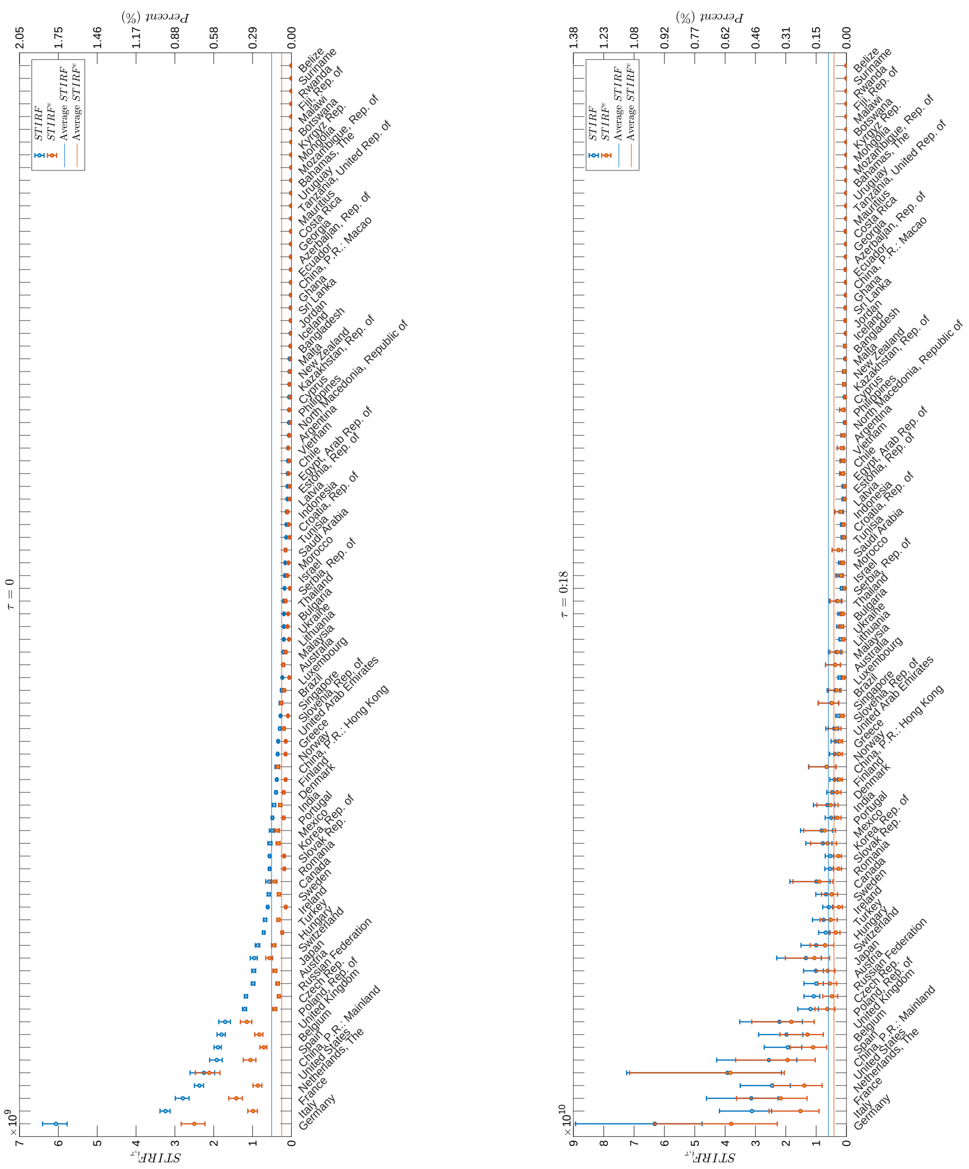
Figure 19. Impulse-Response Functions of USD Aggregate Currency Invoicing



(a) Contemporaneous Effect **(b)** Cumulated Effect over 18 Months

Spatiotemporal impulse-response functions to a domestic one standard deviation shock. Left axis = USD. Right axis = percentage of monthly total aggregate currency invoicing in USD over same horizon. Panel (a): contemporaneous effect. Panel (b): cumulative effect after 18 months. Box-plots report posterior means and centered 95% posterior coverage.

Figure 20. Impulse-Response Functions of EUR Aggregate Currency Invoicing

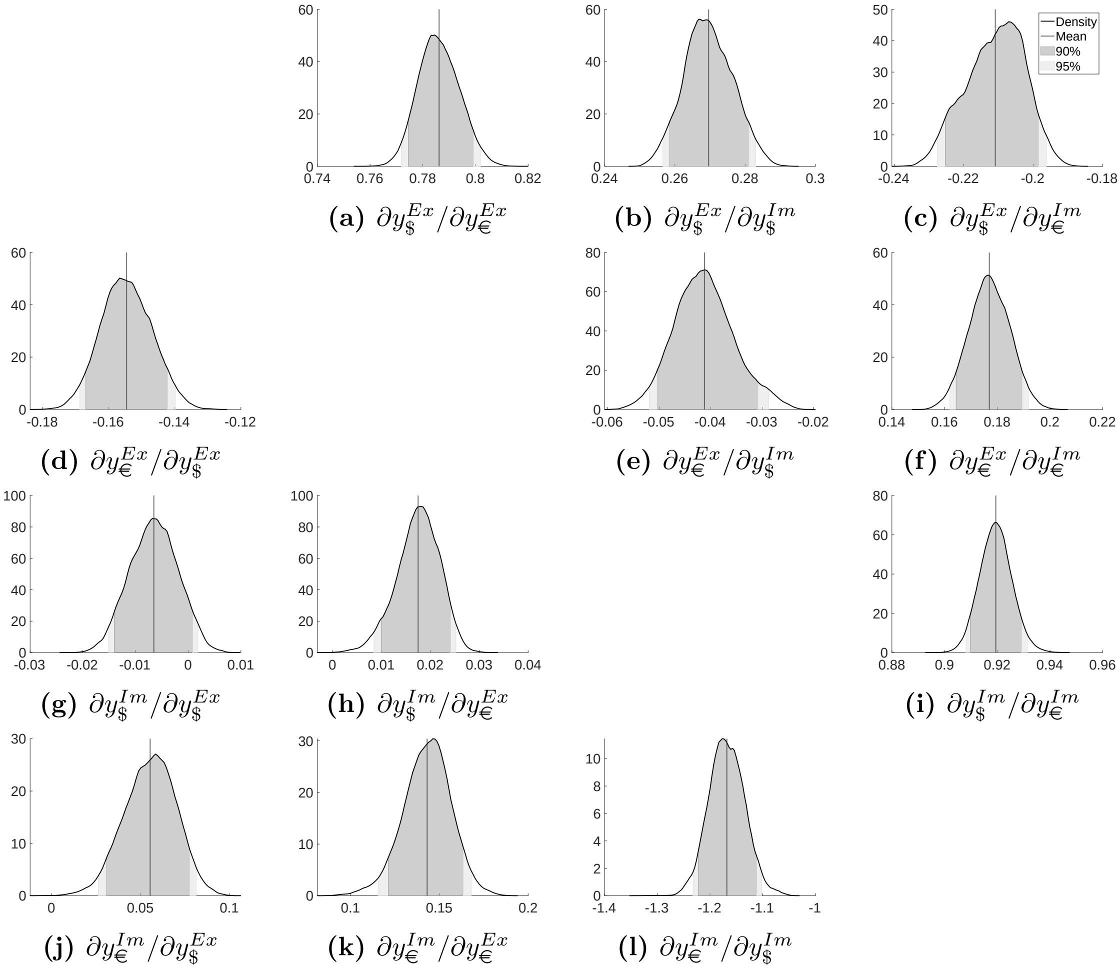


(a) Contemporaneous Effect

(b) Cumulated Effect over 18 Months

Spatiotemporal impulse-response functions to a domestic one standard deviation shock. Left axis = EUR. Right axis = percentage of monthly total aggregate currency invoicing in EUR over same horizon. Panel (a): contemporaneous effect. Panel (b): cumulative effect after 18 months. Box-plots report posterior means and centered 95% posterior coverage.

Figure 21. Cross-Currency and Export-Import Spillovers – Aggregate Currency Invoicing



(a) $\partial y_{\$}^{Ex}/\partial y_{\text{€}}^{Ex}$ **(b)** $\partial y_{\$}^{Ex}/\partial y_{\$}^{Im}$ **(c)** $\partial y_{\$}^{Ex}/\partial y_{\text{€}}^{Im}$

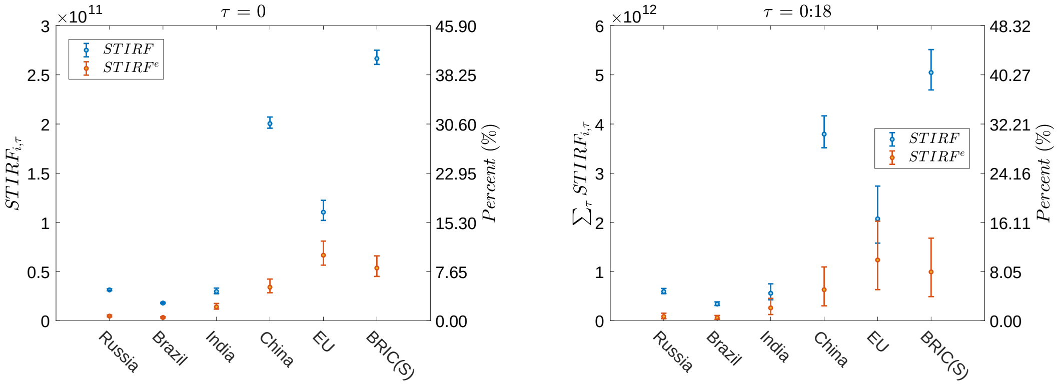
(d) $\partial y_{\text{€}}^{Ex}/\partial y_{\$}^{Ex}$ **(e)** $\partial y_{\text{€}}^{Ex}/\partial y_{\$}^{Im}$ **(f)** $\partial y_{\text{€}}^{Ex}/\partial y_{\text{€}}^{Im}$

(g) $\partial y_{\$}^{Im}/\partial y_{\$}^{Ex}$ **(h)** $\partial y_{\$}^{Im}/\partial y_{\text{€}}^{Ex}$ **(i)** $\partial y_{\$}^{Im}/\partial y_{\text{€}}^{Im}$

(j) $\partial y_{\text{€}}^{Im}/\partial y_{\$}^{Ex}$ **(k)** $\partial y_{\text{€}}^{Im}/\partial y_{\text{€}}^{Ex}$ **(l)** $\partial y_{\text{€}}^{Im}/\partial y_{\$}^{Im}$

The figure depicts the posterior distribution of the elements of Γ, identified via cross-sectional heteroskedasticity. For interpretation, we scaled the draws of Γ_b such that the diagonal only contains ones and then multiplied each row by negative one. Additionally, the figure depicts the posterior mean, as well as 90% and 95% confidence intervals.

Figure 22. Counterfactual: Abandonment of USD as Vehicle Currency – Aggregate Currency Invoicing



(a) Contemporaneous Effect **(b)** Cumulated Effect over 18 Months

Spatiotemporal impulse-response functions to a shock sequence that sets the aggregate currency invoicing of the specified countries to zero permanently. EU contains all 19 EUR-Area countries while BRIC(S) contain the BRICS countries excluding South Africa due to missing observations. Left axis = USD. Right axis = percentage of monthly total excess currency invoicing in USD. Panel (a): contemporaneous effect. Panel (b): cumulative effect after 18 month. Box-plots report posterior means and centered 95% posterior coverage.

5.4. Additional Tables.

Table 14. The Posterior Likelihood of Trade-Network Spillovers

Specification:		ACI_{USD}^{Ex}	ACI_{EUR}^{Ex}	ACI_{USD}^{Im}	ACI_{EUR}^{Im}
Panel	$\ln p_m$	1602.892	2260.553	1298.417	163.474
	$prob_m$	0.000	0.000	0.000	0.000
SEM	$\ln p_m$	1612.020	2429.686	1335.039	-27.280
	$prob_m$	0.000	0.000	0.000	0.000
SLM	$\ln p_m$	1626.628	2348.159	1308.943	273.662
	$prob_m$	0.000	0.000	0.000	0.000
SDM	$\ln p_m$	1769.339	2592.921	1543.097	562.501
	$prob_m$	1.000	1.000	1.000	1.000

The table reports the logarithm of the marginal likelihood ($\ln p_m$) of the data, given the model and the posterior model probabilities ($prob_m$). Note that the marginal likelihoods are adjusted by subtracting the logarithm of the number of observations. The models are separately estimated on each dataset using our baseline specification. Depending on the dataset, the baseline specification uses respectively USD or EUR export or import-based aggregate currency invoicing as the dependent variable. As independent variables, we include lags of inward foreign direct investments, a USD SWAP line dummy, exchange rate changes with the USD and EUR, realized exchange rate volatility with the USD and EUR, the share of aggregate exports, CPI-based inflation and CPI-based inflation volatility, USD export-, USD import-, EUR export-, and EUR import-based aggregate currency invoicing, and country- and time-fixed effects.

Table 15. The Baseline Spatial Durbin Model

		Independent Variables												
		ECI^{Im}_{USD}	ECI^{Im}_{EUR}	ECI^{Ex}_{USD}	ECI^{Ex}_{EUR}	FDI^{In}	$FXChng_{USD}$	$FXChng_{EUR}$	$FXVol_{USD}$	$FXVol_{EUR}$	$SWAP_{USD}$	TS^{Ex}	CPI	$CPIVol$
Panel A. Dependent Variable: Export Excess Currency Invoicing in USD (ECI^{Ex}_{USD})														
Short-term	Direct Effect	0.015**	0.027***	0.791***	-0.035***	0.015**	-0.015	-0.013	-0.004	-0.023**	-0.055**	-0.101***	-0.009	0.019***
		(0.059)	(0.000)	(0.000)	(0.000)	(0.023)	(0.243)	(0.264)	(0.658)	(0.020)	(0.074)	(0.000)	(0.188)	(0.004)
	Total Effect	-0.013	0.061***	0.846***	-0.058***	0.053**	-0.075	0.078	-0.111***	-0.103***	0.188**	-0.145***	-0.075***	0.005
		(0.630)	(0.004)	(0.000)	(0.005)	(0.015)	(0.148)	(0.153)	(0.000)	(0.001)	(0.044)	(0.000)	(0.004)	(0.633)
Long-term	Direct Effect	0.065**	0.122***		-0.155***	0.068**	-0.066	-0.054	-0.022	-0.104**	-0.236**	-0.451***	-0.040	0.084***
		(0.060)	(0.000)		(0.000)	(0.022)	(0.228)	(0.288)	(0.603)	(0.018)	(0.084)	(0.000)	(0.165)	(0.005)
	Total Effect	-0.081	0.337***		-0.321***	0.294**	-0.411	0.426	-0.610***	-0.568***	1.031**	-0.798***	-0.411***	0.030
		(0.608)	(0.008)		(0.010)	(0.019)	(0.153)	(0.158)	(0.001)	(0.002)	(0.047)	(0.000)	(0.005)	(0.629)
	ϕ							0.241***						
								(0.000)						
	R^2							0.949						
	NObs							10871						
	log marginal							284.0414						
Panel B. Dependent Variable: Export Excess Currency Invoicing in EUR (ECI^{Ex}_{EUR})														
Short-term	Direct Effect	0.122***	-0.056***	-0.079***	0.668***	-0.010	0.029**	-0.002	-0.001	0.022**	-0.203***	0.038***	-0.007	-0.005
		(0.000)	(0.000)	(0.000)	(0.000)	(0.234)	(0.053)	(0.895)	(0.930)	(0.062)	(0.000)	(0.000)	(0.353)	(0.554)
	Total Effect	0.310***	-0.105***	-0.034	0.738***	-0.076***	-0.026	-0.030	-0.021	0.066**	0.043	0.077***	0.074***	-0.053***
		(0.000)	(0.000)	(0.225)	(0.000)	(0.002)	(0.652)	(0.618)	(0.520)	(0.050)	(0.677)	(0.000)	(0.009)	(0.000)
Long-term	Direct Effect	0.369***	-0.169***	-0.233***		-0.030	0.086**	-0.006	-0.004	0.068**	-0.599***	0.113***	-0.020	-0.015
		(0.000)	(0.000)	(0.000)		(0.212)	(0.056)	(0.881)	(0.920)	(0.058)	(0.000)	(0.000)	(0.395)	(0.523)
	Total Effect	1.147***	-0.395***	-0.126		-0.282***	-0.095	-0.110	-0.076	0.246**	0.158	0.287***	0.273**	-0.197***
		(0.000)	(0.000)	(0.231)		(0.002)	(0.653)	(0.618)	(0.523)	(0.053)	(0.678)	(0.000)	(0.010)	(0.000)
	ϕ							0.159***						
								(0.000)						
	R^2							0.963						
	NObs							10872						
	$\ln p_m$							-1854.050						

The table reports the posterior means of the estimated effects and their respective p-values. Coefficient estimates significant at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively. Estimation is carried out separately for the two different datasets. On top of the depicted independent variables, we always include country- and time-fixed effects.

Table 16. The Baseline Spatial Durbin Model

		Independent Variables												
		ACI^{Im}_{USD}	ACI^{Im}_{EUR}	ACI^{Ex}_{USD}	ACI^{Ex}_{EUR}	FDI^{In}	$FXChng_{USD}$	$FXChng_{EUR}$	$FXVol_{USD}$	$FXVol_{EUR}$	$SWAP_{USD}$	TS^{Ex}	CPI	$CPIVol$
		Panel A. Dependent Variable: Export Aggregate Currency Invoicing in USD (ACI^{Ex}_{USD})												
Short-term	Direct Effect	0.028^{***}	-0.003	0.861^{***}	-0.055^{***}	0.020^{***}	-0.019^{**}	-0.014	0.005	-0.029^{***}	-0.076^{***}	-0.121^{***}	0.003	0.013^{**}
		(0.000)	(0.711)	(0.000)	(0.000)	(0.001)	(0.081)	(0.167)	(0.553)	(0.001)	(0.004)	(0.000)	(0.617)	(0.029)
	Total Effect	-0.053	-0.020	0.858^{***}	0.009	0.070^{***}	-0.090	0.090	-0.126^{***}	-0.132^{***}	0.063	-0.178^{***}	-0.078^{***}	-0.015
		(0.111)	(0.472)	(0.000)	(0.686)	(0.004)	(0.104)	(0.117)	(0.000)	(0.000)	(0.564)	(0.000)	(0.004)	(0.182)
Long-term	Direct Effect	0.160^{***}	-0.015		-0.323^{***}	0.114^{***}	-0.110^{**}	-0.080	0.029	-0.170^{***}	-0.441^{***}	-0.701^{***}	0.017	0.074^{**}
		(0.000)	(0.715)		(0.000)	(0.001)	(0.082)	(0.167)	(0.552)	(0.001)	(0.005)	(0.000)	(0.615)	(0.030)
	Total Effect	-0.318	-0.115		0.052	0.408^{***}	-0.519	0.522	-0.733^{***}	-0.770^{***}	0.357	-1.030^{***}	-0.453^{***}	-0.088
		(0.142)	(0.481)		(0.700)	(0.008)	(0.110)	(0.123)	(0.000)	(0.001)	(0.571)	(0.000)	(0.007)	(0.189)
	ϕ	0.382^{***}												
		(0.000)												
	R^2	0.964												
	NObs	10871												
	$\ln p_m$	1778.633												
		Panel B. Dependent Variable: Export Aggregate Currency Invoicing in EUR (ACI^{Ex}_{EUR})												
Short-term	Direct Effect	0.027^{***}	-0.005	-0.056^{***}	0.870^{***}	0.014^{**}	-0.025^{**}	-0.006	-0.004	-0.017^{**}	-0.094^{***}	-0.102^{***}	0.011^{**}	0.010^{**}
		(0.000)	(0.478)	(0.000)	(0.000)	(0.012)	(0.012)	(0.531)	(0.650)	(0.035)	(0.000)	(0.000)	(0.030)	(0.075)
	Total Effect	-0.030	-0.092^{***}	-0.061^{**}	0.942^{***}	0.003	-0.104^{**}	0.036	-0.179^{***}	-0.120^{***}	0.207^{**}	-0.184^{***}	-0.004	-0.023^{**}
		(0.410)	(0.004)	(0.050)	(0.000)	(0.912)	(0.093)	(0.574)	(0.000)	(0.001)	(0.090)	(0.000)	(0.886)	(0.067)
Long-term	Direct Effect	0.159^{***}	-0.033	-0.338^{***}		0.082^{**}	-0.156^{**}	-0.033	-0.031	-0.106^{**}	-0.548^{***}	-0.620^{***}	0.068^{**}	0.056^{**}
		(0.000)	(0.407)	(0.000)		(0.014)	(0.010)	(0.561)	(0.511)	(0.026)	(0.000)	(0.000)	(0.034)	(0.088)
	Total Effect	-0.254	-0.791^{**}	-0.516^{**}		0.029	-0.873	0.303	-1.497^{***}	-1.005^{***}	1.732^{**}	-1.546^{***}	-0.037	-0.199^{**}
		(0.416)	(0.019)	(0.065)		(0.902)	(0.102)	(0.579)	(0.000)	(0.002)	(0.098)	(0.000)	(0.887)	(0.083)
	ϕ	0.492^{***}												
		(0.000)												
	R^2	0.965												
	NObs	10872												
	$\ln p_m$	2602.215												

The table reports the posterior means of the estimated effects and their respective p-values. Coefficient estimates significant at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively. Estimation is carried out separately for the two different datasets. On top of the depicted independent variables, we always include country- and time-fixed effects.

6. Appendix of Chapter 3

6.1. Assumptions, Lemmas and Proofs. I here present the accompanying assumptions, lemmas and proofs to section 3.3.1.

6.1.1. *Assumptions and Lemmas.* In what follows, for $h \in \mathcal{M}$ or $h \in \mathcal{M}_K$, let $l_t(h) = log(\pi_t(R_{t+1}; h)$, $L(h) = \mathbb{E}[l_t(h)]$ and $L_T(h) = 1/T \sum_{t=1}^{T} l_t(h)$. Further, define $u_t(h) = \frac{\partial}{\partial h} l_t(h)$, $i_t(h) = \frac{\partial}{\partial h} u_t(h)$, $U_T(h) = \frac{\partial}{\partial h} L_T(h)$ and $I_T(h) = \frac{\partial}{\partial h} U_T(h)$. Subscripts t indicate that variables are defined with respect to filtration $\mathcal{F}_t$ which is defined as usually.

Assumption A.6. *Let $\mathcal{M}$ be compact under the sup-norm.*

Assumption A.7. *Assume $l_t(h)$ satisfies the standard measurability and continuity requirements on $\mathcal{R} \times \mathcal{M}$.*

Assumptions A.6 and A.7 are fairly standard. Assumption A.6 may be violated if $\mathcal{M}$ is too complex, however, can be easily relaxed following X. Chen (2007). For a definition of the standard measurability and continuity requirements see Definition A.2 in the appendix of Wooldridge (1994).

Assumption A.8. *Assume that i) there is a $m \in \mathcal{M}$ such that $\pi_t^o(R) = \pi_t(R; m)\ \forall\ t$, ii) $L(m) > -\infty$ and iii) for $h_1, h_2 \in \mathcal{M}$ that $Pr(R_{t+1} \in \mathcal{R} : l_t(h_1) \neq l_t(h_2)) > 0$.*

Assumption A.8 i) formalizes that $M_t(R)$ is proportional to a time-invariant function and the implied probability measure is correct. Assumption A.8 iii) requires the log-likelihood ratios to be non-constant random variables.

Assumption A.9. *For all $K \geq 1$ and $h \in \mathcal{M}_K$ let $\underset{T\to\infty}{\text{plim}} \sup_{h \in \mathcal{M}_K} |L_T(h) - L(h)| = 0$.*

Assumption A.9 requires $l_t(h)$ to satisfy a uniform weak law of large numbers.

Assumption A.10. *Assume the interchange of derivatives and integrals holds for all $h \in \mathcal{M}$.*

Assumption A.11. *Assume $\mathbb{E}[|\frac{\partial}{\partial h} l_t(h)|^r] < \infty$ for $r > 2$ for all t.*

Assumption A.11 implies that $u_t(h)$ and $i_t(h)$ are uniformly integrable and also ensures that the relevant expectations exist.

Lemma L.6. *Under assumptions A.10 and A.11 (i) $\mathbb{E}_{t-1}[u_t(m)] = 0$, (ii) $\mathbb{E}_{t-1}[u_t(m)u_\tau(m)] = 0$ for $\tau < t$ and (iii) $\mathbb{E}_{t-1}[-i_t(m)] = Var_{t-1}(u_t(m))$. By the law of iterated expectations, (iv) $\mathbb{E}[u_t(m)] = 0$ for all t and (v) $\mathbb{E}[u_t(m)u_\tau(m)] = 0$ for $\tau \neq t$. By the law of total variance, (vi) $Var(u_t(m)) = \mathbb{E}[-i_t(m)] = \mathbb{E}[u_t(m)^2]$.*

Assumption A.12. *Assume $\frac{1}{T}\sum_{t=1}^{T} Var(u_t(m)) \xrightarrow{p} b_m$ and $\frac{1}{T}\sum_{t=1}^{T} u_t(m)^2 \xrightarrow{p} b_m$.*

Lemma L.7. *Under assumptions A.10 and A.11 it follows by lemma L.6 that $u_t(m)$ is a martingale difference sequence and a L^1-mixingale with respect to $\mathcal{F}_t$. Furthermore, $\frac{1}{T}\sum_{t=1}^{T} u_t(m) \xrightarrow{p} 0$. Additionally, under A.12 it follows that $T^{-1/2}\sum_{t=1}^{T} u_t(m) \xrightarrow{d} N(0, b_m)$.*

The proofs for the lemmas are given in section 6.1.2. Lemma L.7 establishes that the law of large numbers for L^1-mixingales and the central limit theorem for a martingale difference sequence hold for $u_t(m)$ (see Hamilton (2020) proposition 7.6 and 7.8 for definitions).

Define the optimal sieve approximation on $\mathcal{M}_K$ in population by $m^* = \pi_K m$. The difference between m^* and m is due to approximation error. The difference between m^* and $\hat{m}$ is due to estimation error. Let $\bar{m}^* \in [m^*, \hat{m}]$.

Assumption A.13. *Assume (i) $I_T(m) \xrightarrow{p} i_m$ and (ii) $||I_T(\bar{m}^*) - I_T(m)||_\infty = o_p(1)$.*

Assumption A.14. *Assume (i) $||m - m^*||_\infty = o(T^{-1/2})$ and (ii) $||U_T(m) - U_T(m^*)||_\infty = o_p(T^{-1/2})$.*

Assumption A.13 (i) makes standard assumptions regarding the convergence of the Hessian. Due to theorem 4 assumption A.13 (ii) can be relaxed. Assumption A.14 (i) is often referred to as an under-smoothing condition, which ensures that the approximation error is asymptotically negligible (see X. Chen (2007) and Christensen (2017)). Assumption A.14 (ii) is similar to condition 4.4 (i) in section 4.2.1 of X. Chen (2007).

6.1.2. *Proofs of Lemmas.*

Proof of lemma L.6. I will show results (i), (ii), and (iii) of lemma L.6. (iv), (v) and (vi) are self-contained. Let $\frac{\partial}{\partial h} f(R; h)\big|_{h=m} = f'(m)$. Assumption A.11 is required throughout for the existence of the relevant integrals. To show (i) note that $u_t(m) = \pi'_{t-1}(m)/\pi_{t-1}(m)$. Therefore,

$$\mathbb{E}_{t-1}[u_t(m)] = \int_{\mathcal{R}} \frac{\pi'_{t-1}(m)}{\pi_{t-1}(m)} \pi_{t-1}(m) dR = \frac{\partial}{\partial h} \int_{\mathcal{R}} \pi_{t-1}(m) dR = 0$$

where the second equality follows from assumption A.10. To show (ii) notice that the information set $\mathcal{F}_{t-1}$ contains $u_\tau(m)$ for $\tau < t$. Hence, (ii) is an immediate implication of (i). To show (iii), notice $Var_{t-1}(u_t(m)) = \mathbb{E}_{t-1}[u_t(m)^2]$ by (i). Further, note that

$$\begin{aligned}
\mathbb{E}_{t-1}[i_t(m)] &= \int_{\mathcal{R}} \left(\frac{\pi''_{t-1}(m)}{\pi_{t-1}(m)} - \left(\frac{\pi'_{t-1}(m)}{\pi_{t-1}(m)} \right)^2 \right) \pi_{t-1}(m) dR \\
&= \frac{\partial^2}{\partial h^2} \int_{\mathcal{R}} \pi_{t-1}(m) dR - \int_{\mathcal{R}} u_t(m)^2 \pi_{t-1}(m) dR \\
&= -\mathbb{E}_{t-1}[u_t(m)^2]
\end{aligned}$$

where the interchange of partial derivative and integrals follows from assumption A.10 and the definition of $u_t(m)$ was used. This establishes (iii). □

Proof of lemma L.7. By lemma L.6 (i) and (iv) it follows that $u_t(m)$ is a martingale difference sequence with respect to $\mathcal{F}_t$ (see Hamilton (2020) p. 189). By assumption A.11 it follows that $\mathbb{E}[|u_t(m)|^r] < M' \ \forall \ t$ for some $r > 2$ and $M' < \infty$, so $u_t(m)$ is uniformly integrable (see Hamilton (2020) proposition 7.7). This implies $\mathbb{E}[|u_t(m)|] < M$ for some $M < \infty$.

To show that $u_t(m)$ satisfies the law of large numbers for L^1-mixingales, note that since $\mathbb{E}[u_t(m)] = 0 \ \forall \ t$ and $\mathbb{E}_{t-1}[u_t(m)] = 0 \ \forall \ t$, it follows by choosing $d_t = M$ and $\zeta_0 = 1$ and $\zeta_n = 0 \ \forall \ n \geq 1$ that $u_t(m)$ satisfies the definition of an L^1-mixingale with respect to $\mathcal{F}_t$ (see Hamilton (2020) example 7.9). Therefore, by proposition 7.6 in Hamilton (2020) it follows that $\frac{1}{T} \sum_{t=1}^T u_t(m) \xrightarrow{p} 0$.

Since $u_t(m)$ is a martingale difference sequence, $\mathbb{E}[u_t(m)^2] = Var(u_t(m)) > 0$, and by assumptions A.11 and A.12 the conditions of proposition 7.8 Hamilton (2020) are satisfied, i.e. $u_t(m)$ satisfies the central limit theorem for a martingale difference sequence.

Therefore,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t(m) \xrightarrow{d} N(0, b_m)$$

□

6.1.3. *Proofs of Theorems.*

Proof of theorem 4. I here demonstrate that under assumptions A.6-A.9 the consistency of the estimator $\hat{m}$ defined in equation (26) follows as $K, T \to \infty$ and $K/T \to 0$. The strategy is to show that conditions 3.1", 3.2, 3.4 and 3.5(i) of X. Chen (2007) hold and therefore theorem 3.1 of X. Chen (2007) applies. This establishes consistency.

First, assumption A.6-A.8 imply condition 3.1" in X. Chen (2007). To see this, consider $\pi_t(R; m)$, the probability measure induced by m given the observed $a_t(R)$ as defined by the fundamental pricing rule representation theorem. The Kullback-Leibler information inequality implies $L(m) > L(h)\ \forall\ h \in \mathcal{M}$ where the strict inequality is due to assumption A.8. Hence, m is the unique solution.

Second, the linear sieve spaces $\mathcal{M}_K$ defined for the estimator in equation (26) in combination with a version of the Weierstrass approximation theorem imply condition 3.2 in X. Chen (2007). Note that $\mathcal{M}_K$ as defined for the estimator satisfies $\mathcal{M}_K \subset \mathcal{M}_{K+1} \subset \mathcal{M}$ for all $K \geq 1$. Denote by $\pi_K m$ the approximation of m on $\mathcal{M}_K$. The Weierstrass approximation theorem establishes that m can be approximated arbitrarily well through polynomials. By Jackson's theorem (theorem 13.3 Schumaker (2007)[43]) it follows that $||\pi_K m - m||_\infty \to 0$ as $K \to \infty$.

Third, condition 3.4 in X. Chen (2007) is not required since $L(h)$ is concave for $h \in \mathcal{M}_K$. Stone (1990) has shown that the estimator as defined in equation (26) is concave for $h \in \mathcal{M}_K$; see also Huang (2001). This is because, as stated in (26), the associated $\pi_t(R; h)$ for $h \in \mathcal{M}_K$ belongs to the exponential family. Therefore, condition 3.4 is not required (see section 2.2.3 of X. Chen (2007)).

Fourth, condition 3.5(i) of X. Chen (2007) follows directly from assumption A.9.

Therefore, the conditions of theorem 3.1 in X. Chen (2007) p. 5591 are satisfied and it follows that $||\hat{m} - m||_\infty = o_p(1)$. □

Proof of theorem 5. Define $m^* = \pi_K m$ as the optimal sieve approximation on $\mathcal{M}_K$ in population. A mean value expansion of $U_T(\hat{m})$ around m^* gives rise to

$$U_T(\hat{m}) = 0 = U_T(m^*) + I_T(\bar{m}^*)(\hat{m} - m^*)$$

where $\bar{m}^* \in [m^*, \hat{m}]$. It follows that

$$\sqrt{T}(\hat{m} - m) = \underbrace{-I_T(\bar{m}^*)^{-1}\sqrt{T}U_T(m)}_{(i)} - \underbrace{I_T(\bar{m}^*)^{-1}\sqrt{T}(U_T(m) - U_T(m^*))}_{(ii)} - \underbrace{\sqrt{T}(m - m^*)}_{(iii)}$$

By assumption A.14 it follows that $||\sqrt{T}(m - m^*)||_\infty = o(1)$ and that $||\sqrt{T}(U_T(m) - U_T(m^*))||_\infty = o_p(1)$. By assumption A.13 $I_T(\bar{m}^*) \xrightarrow{p} i_m$. Therefore term (ii) and (iii) are negligible.

By lemma L.7, the continuous mapping theorem and Slutsky's theorem it follows that

$$-I_T(\bar{m}^*)^{-1}\sqrt{T}U_T(m) \xrightarrow{d} N\left(0, \frac{b_m}{i_m^2}\right)$$

[43] Jackson's theorem gives rise to theorem 13.3 in Schumaker (2007): for every continuous, real-valued function f on the interval $[a, b]$ there exists a sequence of polynomials $\pi_K f \in \mathcal{M}_K$ with $||\pi_K f - f||_\infty \to 0$ as $K \to \infty$ where K is the order of the polynomial.

Therefore, $\sqrt{T}(\hat{m} - m) \overset{d}{\rightarrow} N(0, b_m/i_m^2) + o_p(1)$.

$\square$

6.2. Data Processing. I here detail the data-cleaning steps and inter-/ extrapolation methods implemented to arrive at the full set of Arrow-Debreu security prices.

6.2.1. *Cleaning Steps.* Several pre-processing steps of the raw data are undertaken before the inter-/ extrapolation methods are applied. These steps have the goal to mitigate data errors and to extract a representative dataset that best approximates the markets' aggregation of individual risk-neutral subjective probability beliefs or Arrow-Debreu security prices in the investor population. An incomprehensive list of studies working with similar datasets is Ait-Sahalia and Lo (1998), J. C. Jackwerth (2000), Rosenberg and Engle (2002), Bliss and Panigirtzoglou (2004), Figlewski (2008), Martin (2017), Ulrich and Walther (2020) and J. Jackwerth and Menner (2020). A survey discussing the practical details of extracting Arrow-Debreu securities is given in Figlewski (2018).

I focus on options on the S&P 500, which are actively traded and cash-settled with European exercise style, obtained from Optionmetrics. All observations with a bid price of zero and all observations with a bid price larger than their ask price are dropped.

I then begin to drop observations that are known to suffer significantly from poor liquidity. First, all observations with a residual maturity less than seven days are dropped. At each date, the implied volatility using Black and Scholes (1973) based on the mid, bid, and ask price using the dividend yield and zero coupon rate reported by Optionmetrics are calculated. Where necessary, the zero coupon rates are matched to the residual maturity of the observations by shape-preserving piecewise cubic interpolation[44]. Second, all observations are dropped for which the Black-Scholes implied volatility computation did not converge. Third, I drop all observations where the difference between the bid and ask based implied volatility is larger than 50%. This tends to eliminate options significantly in the money or sporadic outliers. Fourth, I drop all observations with a bid-ask spread above 100$, which eliminates outliers[45].

In some cases, I am left with multiple option prices per strike. For all call and put options separately, on each date, by maturity, I only keep the observations with the lowest bid-ask spread, mid price and implied volatility.

I proceed to check observations for violations of no-arbitrage conditions. For all call and put options separately, on each date, by maturity, I drop all observations that violate vertical spread arbitrage and the monotonicity condition of option price curves (see Rosenberg and Engle (2002)).

Finally, since inter-/ extrapolation will be performed, I require on each date for each maturity at least seven observations, else the observations are dropped.

The inter-/ extrapolation of the raw data is done by translating out-of-the-money option mid prices to implied volatilities using Black and Scholes (1973) and then inter-/ extrapolating the volatility surface. For a brief discussion of the advantages of inter-/ extrapolation of the volatility surface instead of the call or put price surface, see Figlewski (2018).

[44] I have found this to yield a better fit than the commonly used linear interpolation on multiple cases.

[45] These filters are similar to filters in existing studies that drop observations with "too large" bid-ask spreads or implied volatilities. For example Ait-Sahalia and Lo (1998) delete observations with implied volatilities larger than 70% and J. C. Jackwerth (2000) deletes observations with bid-ask spreads larger than 2$. I have found such existing filters to be too restrictive, however, dropping them all together leaves several outliers. The described thresholds have been chosen based on several case by case inspections to guarantee outliers are deleted efficiently.

Working with out-of-the-money implied volatilities based of call and put options has several challenges. First, to determine the moneyness of an option forward prices for a given maturity at each date are required. Second, switching from out-of-the-money put to out-of-the-money call implied volatilities leads to a discontinuity in the implied volatility curve at a specific maturity, also documented in Figlewski (2008), Figlewski (2018) and Ulrich and Walther (2020). It tends to be the case that implied volatilities for put options are higher than for call options. Explanations appeal to limits of arbitrage due to short sale constraints (see Atmaz and Basak (2019)) or lending and borrowing constraints (see Bergman (1995)). Third, to compute the correct implied volatilities, consistent with investors aggregated beliefs, the interest rate and dividend yield used by investors are required.

To address the first concern, forward prices are determined by the intersection of the mid price call and put curve. Based on put-call-parity, the two intersect at the strike that is equal to the forward. Where necessary, the curves are extrapolated over the strike price grid by maximally 25$ in either direction.

To address the second and third concern, I compute the implied dividend yield from the forward price using Optionmetrics reported zero coupon rates. Atmaz and Basak (2019) argue that short sale constraints can be thought of as "artificially" increasing the dividend yield[46]. Assuming that the LIBOR rate reported by Optionmetrics reasonably approximates the interest rate available to arbitrageurs, the market implied dividend yield is defined as

$$\delta_\tau^* = r_\tau - \frac{log(\frac{F_{t,\tau}}{S_t})}{\tau}$$

where δ_τ^* is the dividend yield, r_τ the zero coupon rate for the appropriate horizon τ, $F_{t,\tau}$ the forward price and S_t the spot price of the underlying.

Since the forward price is based on the intersection of put and call price curves, using δ_τ^* the implied volatilities based on Black and Scholes (1973) practically equates the mid price based implied volatilities of put and call options at the money, i.e. eliminates the discontinuity[47]. As a consequence, by using δ_τ^* in the computation of implied volatilities based on out-of-the-money put and call options, I am left with a smooth implied volatility curve for each maturity[48].

This leaves me with a large daily dataset of implied volatilities for irregularly spaced strikes and maturities that can be inter-/ extrapolated.

[46] They show that an increase in short-sale costs of the underlying increases put prices and decreases call prices ceteris paribus. During periods of high short-selling costs one would therefore observe a large discontinuity between implied volatilities computed from call and put prices using standard Black and Scholes (1973). They further show that short sale costs manifest themselves as an adjustment to the dividend yield and once the adjustment is taken into account implied volatilities coincide.

[47] I here only allow for an adjustment of the dividend yield and do not adjust interest rates. Suppose observed put price implied volatilities are higher than call price implied volatilities. Increasing the dividend yield, leads to a decrease of the put based implied volatility and an increase of the call based implied volatility. On the other hand, assuming lending rates are below the reported rate, decreasing the interest rate leads to the same effect. Adjustments of the dividend yield or interest rate have roughly the same impact on implied volatilities.

[48] An alternative procedure suggested by Figlewski (2008) is to average the implied volatilities of put and call prices calculated using the reported dividend yield and interest rate in an area around the forward to create a smooth transition in implied volatilities. The area in which the average is taken is ad hoc. Using δ_τ^* on the other hand leads to an implied volatility curve that lies between the put and call based implied volatilities, i.e. is averaged as well, however, can be motivated economically.

6.2.2. *Inter-/ Extrapolation.* Multiple approaches for inter/- extrapolation of the volatility surface exist. An incomprehensive list of studies is Ait-Sahalia and Lo (1998), J. C. Jackwerth (2000), J. Jackwerth (2004), Figlewski (2008) and Gatheral and Jacquier (2014). I implement four different methods to conduct robustness checks of the main results. All methods allow me to obtain implied volatilities on an arbitrarily fine grid of strikes, however, only some allow for inter-/ extrapolation across maturities.

The first – and preferred – method is based on Gatheral and Jacquier (2014). The approach fits at each date a stochastic volatility inspired parameterization of the volatility surface to the raw data that is free of calendar spread and butterfly arbitrage. The approach therefore creates arbitrage free implied volatility estimates, taking into account all available information across strikes and maturities on that date and allows for inter-/ extrapolation across strikes and maturities. Specifically, the power-law parameterization is implemented (see Gatheral and Jacquier (2014)).

The second method is based on Ulrich and Walther (2020), which is a modified version of the method in Ait-Sahalia and Lo (1998). On each date, an implied volatility curve is fit for each maturity separately, through a kernel regression over implied volatilities and moneyness. Extrapolation of the tails is done via linear interpolation, which is performed before the kernel regression (see Ulrich and Walther (2020) for details). The approach therefore creates implied volatility estimates, taking into account only information across strikes for a given maturity and allows for inter-/ extrapolation across strikes. Inter-/ extrapolation across maturities is performed through the approach described in section 5.3 of Gatheral and Jacquier (2014).

The third approach is inspired by J. Jackwerth and Menner (2020). On each date, I fit a volatility surface to the data using thin-plated smoothing splines. The approach therefore creates implied volatility estimates, taking into account information across strikes and maturities on that date and allows for inter/- extrapolation across strikes and maturities.

The fourth approach is based on Figlewski (2008). The approach fits on each date a 4^{th} order spline with a single knot placed at the money for the observed range of implied volatilities, for each maturity separately. Next the implied risk-neutral probability density is computed based on Breeden and Litzenberger (1978) from the fitted and interpolated implied volatilities. Then, motivated by the Fisher-Tippett theorem, a generalized extreme value distribution is fit to the left and right tail separately, to extrapolate the tails (see Figlewski (2008) and Figlewski (2018) for details). The approach therefore directly creates Arrow-Debreu security price curves, taking into account only information across strikes for a given maturity and allows for inter-/ extrapolation across strikes. Inter-/ extrapolation across maturities is performed through the approach described in section 5.3 of Gatheral and Jacquier (2014).

After the inter-/ extrapolation, where necessary, the call price curves are computed using Black and Scholes (1973) and the Arrow-Debreu security price curves are computed following Breeden and Litzenberger (1978) by numerically approximating the second derivative of the call price curve with respect to strikes. Each approach, except the Gatheral and Jacquier (2014) approach, is implemented such that I am left with Arrow-Debreu security price curves corresponding to one month of maturity over a discrete return grid ranging from 59% to 141% with 1% increments. For Gatheral and Jacquier (2014) Arrow-Debreu security price curves are extracted corresponding to one, two, three, six, and twelve months of maturity and over a grid ranging from 29% to 171% – this is necessary, as for higher maturities the historically observed return range is wider.

Finally, when importing the Arrow-Debreu security price curves for estimation the Arrow-Debreu security price curve is resampled onto a strike grid corresponding to gross returns that range from the minimum to the maximum of observed returns of the S&P 500 over the sample period for each maturity. They are then rescaled to ensure that the price curve sums to the reciprocal of the risk-free rate for the respective maturity. Before the estimation is conducted, Arrow-Debreu security price curves containing negative observations are dropped – this step is irrelevant for the Gatheral and Jacquier (2014) based curves. Furthermore, for computational convenience, all Arrow-Debreu security prices equal to 0 are set to $1e^{-15}$.

6.3. Goyal Welch Procedure. To assess the predictive performance of $\mathbb{E}_t[R_{t+1}]$ implied by the estimation, I follow Welch and Goyal (2008) and compute the out-of-sample R^2.

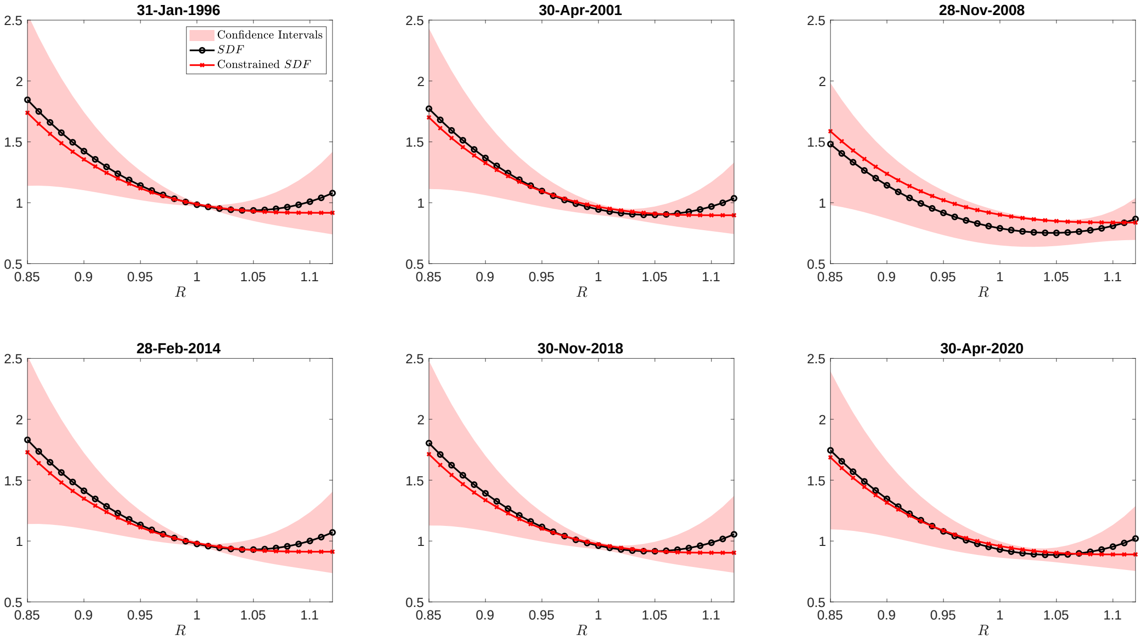
The out-of-sample R^2, R^2_{OOS}, is constructed in an expanding window fashion. I start the out-of-sample test in January 2011, i.e. 15 years after my dataset begins. To do so $\widehat{\mathbb{E}}_t[R_{t+1}]$ and the sample mean $\hat{\mu}_t$ are re-estimated each period as the training sample expands. Forecasting errors are computed as $\epsilon_{t+1} = R_{t+1} - \widehat{\mathbb{E}}_t[R_{t+1}]$ and $v_{t+1} = R_{t+1} - \hat{\mu}_t$. This is repeated until I reach T. R^2_{OOS} is then defined as

$$R^2_{OOS} = 1 - \frac{\sum_t \epsilon^2_{t+1}}{\sum_t v^2_{t+1}}$$

R^2_{OOS}, therefore benchmarks the predictor $\widehat{\mathbb{E}}_t[R_{t+1}]$ against the forecast based on the expanding window sample mean $\hat{\mu}_t$. A positive value indicates that the predictor outperforms the expanding window sample mean based forecast. A negative value indicates that the expanding window sample mean based forecast outperforms the predictor.

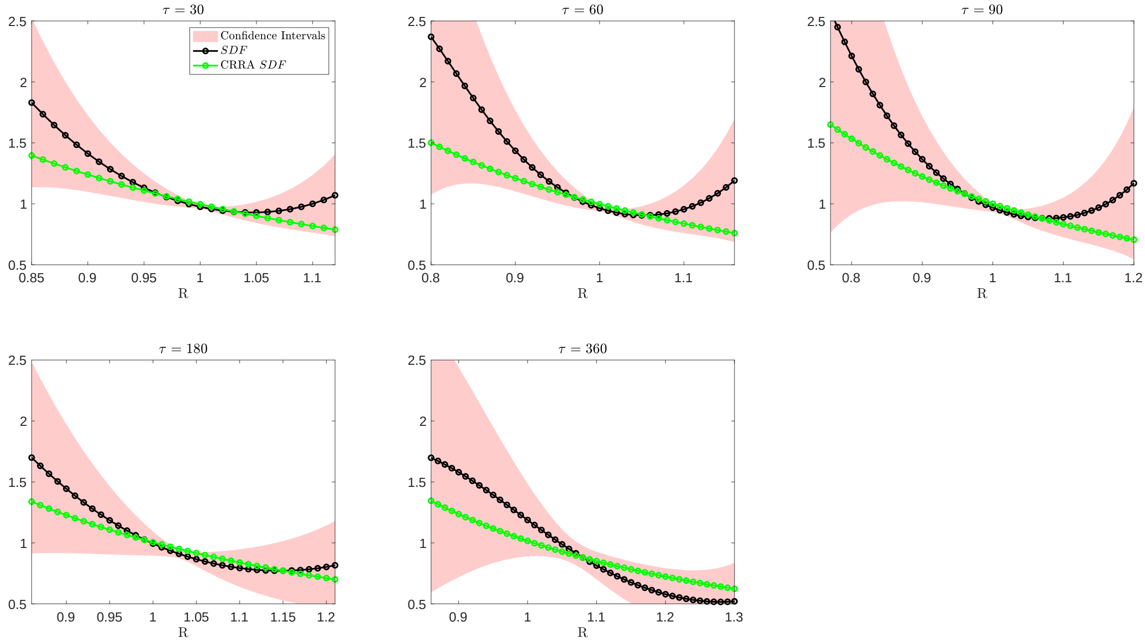
6.4. Additional Figures.

Figure 23. Unconstrained and Constrained Estimated Risk Preferences



Note: The figure depicts the unconstrained estimated stochastic discount factor function given in equation (27) in black on several snapshot dates using regular polynomial basis functions with $K = 4$. Further, 95% confidence intervals obtained via the delta method are reported. I employ a Newey and West (1987) type estimator with lag length 21 for the covariance matrix of the scores. Additionally, the figure depicts the constrained estimated stochastic discount factor function in red, corresponding to the solution of (26) subject to the constraint in (29).

Figure 24. Estimated Risk Preferences and CRRA Preferences across Maturity Horizons



Note: The figure depicts the unconstrained estimated stochastic discount factor function given in equation (27) in black for different maturity horizons using regular polynomial basis functions with $K = 4$. The estimated risk preferences correspond to 04-Jan-1996. Further, 95% confidence intervals obtained via the delta method are reported. I employ a Newey and West (1987) type estimator with lag length corresponding to the respective maturity translated to trading days for the covariance matrix of the scores. Additionally, the figure depicts the stochastic discount factor corresponding to CRRA preferences estimated as in (25) in green.

6.5. Additional Tables.

Table 17. Information Criteria

	K	1	2	3	4	5	6
Regular	AIC	35490.6078	35443.6361	35439.0596	35440.8709	35436.7712	**35430.437**
	BIC	35497.3481	**35457.1167**	35459.2806	35467.8322	35470.4729	35470.8791
Legendre	AIC	35490.6078	35443.6361	35439.0596	35440.8709	35436.7712	**35430.437**
	BIC	35497.3481	**35457.1167**	35459.2806	35467.8322	35470.4729	35470.8791
Laguerre	AIC	35490.6078	35443.6361	35439.0596	35440.8709	**35436.7712**	35438.589
	BIC	35497.3481	**35457.1167**	35459.2806	35467.8322	35470.4729	35479.0311
Hermite	AIC	35490.6078	35443.6361	35439.0596	35440.8709	35436.7712	**35430.437**
	BIC	35497.3481	**35457.1167**	35459.2806	35467.8322	35470.4729	35470.8791

Note: The table reports the Bayesian and Akaike information criterion corresponding to the estimator in (26) for various sieve dimensions and basis functions. The lowest values are highlighted in bold.

Table 18. Test of Parameter Stability – $c = 5$

K	2	3	4	5	6
qLL	-6.391	-6.487	-6.452	-10.086	-10.068
p-value	(0.091)	(0.346)	(0.712)	(0.435)	(0.721)

Note: The table repots the estimated qLL test statistic following Müller and Petalas (2010) for the estimator in (27) using non-overlapping observations and regular polynomial basis functions. The different columns correspond to the number of parameters, or the approximation degree, K. The qLL test is calculated following Müller and Petalas (2010) using $c = 5$ for *robustified* scores. The reported p-values are calculated by simulating the distribution of the qLL test statistic with $c = 5$ from the random variable in Elliott and Müller (2006) lemma 2.

Table 19. Test of Parameter Stability – $c = 25$

K	2	3	4	5	6
qLL	-38.791	-41.045	-41.305	-66.002	-65.864
p-value	(0.006)	(0.188)	(0.841)	(0.206)	(0.768)

Note: The table repots the estimated qLL test statistic following Müller and Petalas (2010) for the estimator in (27) using non-overlapping observations and regular polynomial basis functions. The different columns correspond to the number of parameters, or the approximation degree, K. The qLL test is calculated following Müller and Petalas (2010) using $c = 25$ for *robustified* scores. The reported p-values are calculated by simulating the distribution of the qLL test statistic with $c = 25$ from the random variable in Elliott and Müller (2006) lemma 2.

Table 20. Pricing Kernel Puzzle Test for Different Volatility Surface Inter- / Extrapolation Methods

	K	2	3	4	5	6
Jacquier Gatheral	LM	3.400	0.883	3.611	3.298	3.482
	p-value	(0.183)	(0.830)	(0.461)	(0.654)	(0.746)
Jackwerth Menner	LM	6.387	3.538	5.047	6.264	7.168
	p-value	(0.041)	(0.316)	(0.282)	(0.281)	(0.306)
Ulrich Walther	LM	6.072	2.921	4.375	4.204	6.422
	p-value	(0.048)	(0.404)	(0.358)	(0.520)	(0.378)
Figlewski	LM	4.661	2.130	2.985	4.230	4.404
	p-value	(0.097)	(0.546)	(0.560)	(0.517)	(0.622)

Note: The table presents the Lagrange multiplier test statistic for the estimator in (27) subject to the constraint in (29) using regular polynomial basis functions for various K as indicated in the table columns. The rows correspond to different Arrow-Debreu security extraction methods following either Gatheral and Jacquier (2014), J. Jackwerth and Menner (2020), Ulrich and Walther (2020) and Figlewski (2008) as detailed in appendix 6.2.2. The p-values are based on a χ^2 distribution with K degrees of freedom.

Table 21. Pricing Kernel Puzzle Test for Arrow-Debreu Securities Extracted from Different Option Quotes

	K	2	3	4	5	6
MID	LM	3.400	0.883	3.611	3.298	3.482
	p-value	(0.183)	(0.830)	(0.461)	(0.654)	(0.746)
BID	LM	1.592	0.421	0.513	0.555	0.583
	p-value	(0.451)	(0.936)	(0.972)	(0.990)	(0.997)
ASK	LM	6.239	1.841	2.808	2.187	2.542
	p-value	(0.044)	(0.606)	(0.591)	(0.823)	(0.864)

Note: The table presents the Lagrange multiplier test statistic for the estimator in (27) subject to the constraint in (29) using regular polynomial basis functions for various K as indicated in the table columns. The rows correspond to Arrow-Debreu securities extracted following Gatheral and Jacquier (2014) using either mid, bid, or ask based option data. The p-values are based on a χ^2 distribution with K degrees of freedom.

Table 22. Pricing Kernel Puzzle Test with Different Basis Functions

	K	2	3	4	5	6
Regular Polynomial	LM	3.400	0.883	3.611	3.298	3.482
	p-value	(0.183)	(0.830)	(0.461)	(0.654)	(0.746)
Legendre Polynomial	LM	3.400	4.125	1.564	1.294	1.839
	p-value	(0.183)	(0.248)	(0.815)	(0.935)	(0.934)
Laguerre Polynomial	LM	3.400	0.597	0.868	0.785	0.334
	p-value	(0.183)	(0.897)	(0.929)	(0.978)	(0.999)
Hermite Polynomial	LM	3.400	0.862	1.365	0.647	0.727
	p-value	(0.183)	(0.835)	(0.850)	(0.986)	(0.994)

Note: The table presents the Lagrange multiplier test statistic for the estimator in (27) subject to the constraint in (29) for various K as indicated in the table columns. The rows correspond to different basis functions employed. The p-values are based on a χ^2 distribution with K degrees of freedom.

Table 23. Pricing Kernel Puzzle Test across Maturities

τ	30	60	90	180	360
LM	3.611	3.300	2.108	1.052	0.449
p-value	(0.461)	(0.509)	(0.716)	(0.902)	(0.978)

Note: The table presents the Lagrange multiplier test statistic for the estimator in (27) subject to the constraint in (29) using regular polynomial basis functions for various $K = 4$. The columns correspond to different maturity horizons for the Arrow-Debreu securities and returns. The p-values are based on a χ^2 distribution with K degrees of freedom.