UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES

Master Thesis

Identification and Transmission of Financial Shocks based on Smooth Transition Conditional Heteroskedasticity

Pablo Serrano Ascandoni

Supervised by Heino Bohn Nielsen

December 2021

Identification and Transmission of Financial Shocks based on Smooth Transition Conditional Heteroskedasticity

Pablo Serrano Ascandoni

Supervised by Heino Bohn Nielsen

December 2021

Abstract

This thesis explores three different specifications of Vector Autoregressions with the objective of choosing the most suitable form for analysing the macroeconomic effects of financial shocks in a model with inflation, production growth, monetary policy and the bond risk premium. The main specifications explored are the inclusion of conditional heteroskedasticity and non-linear transmission using a smooth transition function that discriminates between two volatility regimes.

I make use of the volatility clustering present in macro and financial variables to identify financial shocks and evaluate the theory based restrictions imposed by a Cholesky decomposition. Explicitly modelling heteroskedasticity not only allows for identification but also for a more accurate inference in the form of impulse responses and hypothesis testing. The impulse responses are compared across models to evaluate the effects of financial shocks. Testing is used to check full identification and the suitability of Cholesky restrictions. The model is augmented to include non-linear transmission of shocks and level shifts conditional on the volatility regime. The augmented model is used for comparing state dependent impulse responses to a financial shock using a Cholesky identification. Lastly, I explore the existence of an amplification mechanism in high stress regimes.

The main conclusions from this research are: (i) the regimes obtained with a smooth transition in covariances are more suitable for modelling financial shocks than those derived from a Markov Switching model, (ii) heteroskedasticity identification does not provide overall support for Cholesky restrictions but these restriction are valid for the financial shock identification, (iii) the non-linear dynamics from different volatility periods are likely to be the result of different sized shocks (heteroskedasticity) instead of an amplification mechanism, and (iv) the impulse responses to financial shocks are robust across the considered specifications and identification approaches: a risk premium shock triggers a fall in output growth and inflation followed by a monetary policy loosening.

Contents

List of Figures

List of Tables

1 Introduction

Financial crises begin with sudden adjustments in expectations that force re-evaluations of assets. The average decrease in optimism is followed by a falling demand for risky assets and the consequent credit contraction and rising credit spreads (Minsky 1992). If the correction affects widely distributed assets or systemically important financial actors, risk premiums jump both in the short and long-term as agents get rid of risky assets, pushing their prices downwards and damaging the solvency of those holding them. Institutions holding risky assets and those funding them directly face the 'survival constraint' as liquidity freezes, impeding agents to serve their debts unless they engage in fire sales (Mehrling 2010). The need for liquidity and subsequent fire sales further amplifies the downward spiral of falling asset prices. Under these conditions, Central Banks can step in to provide liquidity into the system and stop the default's cascade by rising asset valuations and counteracting the effects of fire sales. Despite these measures, high risk premiums tend to persist, signaling the tight credit conditions and making credit expensive for 'higher-risk' borrowers. This typically dampens private lending, borrowing and investment with negative consequences in production, employment, demand and prices.

The macroeconomic literature refers to these adjustments in expectations as financial shocks and they are well captured by exogenous increases in risk premiums and credit spreads. Gilchrist and Zakrajšek (2011) construct the excess bond premium, an index that contains the unpredictable component of credit spreads. This index signals the price of default risk and explains most of the variation from spreads. Using this measure, they find that financial shocks depress consumption, investment, output and prices. These conclusions are drawn from a time-invariant volatility (homoskedastic) linear VAR identified with exclusion restrictions.

A stylized fact of macroeconomic and financial time series is the existence of periods of higher and lower volatility. This invalidates the constant variance assumption of homoskedastic models, crucial for structural analysis and testing (Engle 1982). Intuitively, periods of recessions or financial stress are characterized by more volatile fluctuations: such as deflation, falling output growth or drastic reductions in interest rates. These features can be included in the model to estimate more accurately standard errors, improving statistical inference in the form of impulse response analysis and hypothesis testing. Heteroskedasticity can be further used for identification purposes without the need of theoretical justifications on restrictions.

Furthermore, it could be argued that these high volatility periods display non-linear dynamics (Hubrich, D'agostino, et al. 2013). Intuitively, higher risk premiums have a negative impact on market financing conditions. They raise the cost of debt for households and firms and reduce credit supplied (Stiglitz and Weiss 1981). This ultimately impacts long term investment, productivity, employment and prices if there are no alternative ways of obtaining funds when banks cut on lending (Chodorow-Reich 2014).

During expansions, firms can easily substitute between different forms of short/long-term funding by issuing bonds, commercial papers, shares, retaining profits or taking out bank loans. The amount of credit a bank grants to customers depends on an individual assessment of the borrower's solvency but also on the bank's capitalization and liquidity structure which is highly procyclical. The rising asset and collateral values during booms soften the credit constraint that firms and households face (Kiyotaki and Moore 1997). During recessions these dynamics revert, agents become credit constrained and cannot easily substitute between different sources of funding. Bank's profitability is damaged by scarce market liquidity and non-performing loans, forcing them to cut on lending. Investors implement readjustments to their portfolios and sell their risky assets, potentially resulting in fire sales that automatically damage the solvency of other agents (Gorton and Metrick 2012). Consumer pessimism induces precautionary saving and lowers consumption levels, aggravating the downward spiral. If overall tighter financial constraints amplify this mechanism, the inclusion of non-linear dynamics must be considered.

This paper explores various specifications of Vector Autoregressions with the objective of choosing the most suitable form for analysing the macroeconomic effects of financial shocks. The main specifications explored are the inclusion of conditional heteroskedasticity and non-linear transmission using a smooth transition function that discriminates between two volatility regimes. The time-dependent volatility present in macro and financial variables is further used to identify financial shocks and evaluate the theory based exclusion restrictions from a Cholesky decomposition. The existence of non-linear transmission of these shocks is addressed by augmenting the heteroskedastic model with regime dependent transmission and level shifts.

The rest of the paper is structured as follows. Section 2 summarizes the literature on financial shocks and Vector Autoregressions, with emphasis on Heteroskedasticity Identification and non-linear transmission modelling. Section 3 presents the baseline homoskedastic VAR, its estimation, derivation of impulse responses and the construction of confidence bands using a Wild bootstrap algorithm. It then presents the data, builds the model and imposes Cholesky restrictions for identification. The residuals are analyzed for signs of misspecification and in light of those results, alternative data based identification methods are discussed. Section 4 presents the heteroskedastic linear VAR using smooth transition in covariances and addresses Heteroskedasticity Identification, model estimation and the optimal smooth transition function for the given data. The regime's transition is compared with that of a Markov Switching VAR to show the benefits of imposing a predetermined structure for the identification of financial stress regimes. The smooth transition VAR is estimated and identified using heteroskedasticity to later conduct inference in the form of full identification testing and impulse response analysis. Lastly, an over-identified model with additional Cholesky restrictions is used to assess the validity of this identification method and compare the impulse responses once again. Section 5 presents the non-linear smooth transition VAR that includes conditional heteroskedasticity, non-linear transmission of shocks and regime-dependent level shifts.

It discusses the estimation method based on a Bayesian Monte Carlo algorithm and conducts inference in the form of regime dependent impulse responses for a financial shock identified with Cholesky restrictions. The non-linear transmission is addressed and formally tested under asymptotic normality assumptions to add further support for linear transmission of heteroskedastic shocks. Lastly, Section 6 wraps up everything and summarizes the main findings.

2 Literature Review

Section 2.1 summarizes the relevant literature on macro-financial linkages and financial frictions in relation with credit spreads and risk premiums, including evidence in favor and against non-linearities in the transmission of financial shocks. Section 2.2 briefly discusses the relevant literature on VAR modelling and structural identification with particular emphasis on Heteroskedasticity Identification and non-linear transmission modelling. It also comments on relevant papers that address inference under heteroskedasticity conditions and consider different bootstrapping techniques and Bayesian estimation methods to overcome the problems that heteroskedasticity poses for inference.

2.1 Literature on Financial Shocks

A big share of the literature on credit spreads is grounded on the theoretical and empirical findings related to financial frictions and business cycles. Empirically, credit spreads are found to negatively correlate with prices and production. This was a distinctive characteristic of Minsky's Instability of Credit hypothesis. Minsky (1992) explained that expansion driven optimism drives up prices and encourages risky lending and borrowing, reinforcing the upward trend and compressing the spreads and risk premium. This trend is reverted when investors optimism proves unjustified, increasing risk aversion, credit rationing, falling prices and rising risk premium.

A recent strand of literature dives and dissects the transmission and amplification mechanisms of financial stress. Adrian and Shin (2010) explain how households, banks and shadow banks behave differently along the cycle: while banks target a fixed leverage ratio, brokers and dealers clearly display a procyclical leverage. The procyclical behavior of shadow banks is outlined in Gregoriou, Racicot, and Théoret (2021) while Feng, Lu, and Xiao (2020) show the impact that this has on amplifying unstable price dynamics. The amplification mechanism through the balance sheet channel accounts for the role of net worth from the borrower's side (B. Bernanke and Gertler 1989, B. Bernanke and Gertler 1990, Kiyotaki and Moore 1997) and from the lender's balance sheet, liquidity needs and capital structure (B. S. Bernanke and Blinder 1988, Kashyap and Stein 2000, Heuvel 2002). Stiglitz and Weiss (1981) explain how credit rationing naturally arises from information asymmetries. In practice, house prices are highly relevant for the financial cycle, as these are the main asset that households can use as collateral for loans. Mian and Sufi (2011) explore the highly procyclical home-equity channel and how it helps to build up vulnerabilities in the boom, while Mian, Sufi, and Trebbi (2015) show how foreclosures amplify the falling prices in the downturn. All of these channels are considered in the leverage cycle model of Geanakoplos (2010). For an extensive summary of the literature on macro-financial linkages see Borio (2014) and Claessens and Kose (2018).

Credit spreads' fluctuations arise from time-varying risk premiums that signal investors' sentiments (Gilchrist and Zakrajšek 2011). The relationship between sentiment and both corporate and mortgage spreads is outlined in Becard and Gauthier (2021). They argue that pessimism forces shadow banks to deleverage and lower their demand for mortgage backed securities. This reduces bank's profits from securitization[1] activities and pushes them to cut on lending by rising collateral requirements and interest rates on their loans to both firms and households. Additionally, consumption is lowered and house prices fall, worsening the scenario. On the firm side, investment is reduced, rising unemployment and further lowering output and prices. Their 'market sentiment' shock resembles the common shock obtained by Angeletos, Collard, and Dellas (2020) which maximizes the volatility in output at business cycle frequency.

These market driven cycles tend to be further amplified by institutional features such as interconnectedness (Allen and Gale 2000, Caccioli, Barucca, and Kobayashi 2017). The recently introduced macroprudential policies are aimed to prevent and minimize the buildup of vulnerabilities in the financial sector (Arnold et al. 2012). However, once a financial crises begins, the Central Bank can prevent a cascade of defaults and counteract the deleveraging process with active management policies (Holmström and Tirole 1998, Eggertsson and Krugman 2012). Adrian, Estrella, and Shin (2019) explain how policy rate hikes can compress the ratio of profitability to funding price and induce a cut on lending. Equivalently, lowering inter-bank lending rates should affect the whole interest rate spectrum of maturities through the transmission mechanism (Kashyap and Stein 2000). If this mechanism applies, not only the price of liquidity is lowered but also that of longer-term credit. Additional to conventional interest rate policies, asset purchase programs can be used to compress spreads for any maturity or asset class without relying on the interest rate transmission mechanism. An outline of these policies can be found in Borio and Zabai (2016). Farhi and Tirole (2017) extend this intervention logic to shadow banks. In this regard, Central Banks have begun to behave as dealers of last resort (Mehrling 2010) and account for the systemic relevance of the shadow banking system.

If agents behave asymmetrically along the cycle and frictions can amplify financial shocks, there is room for non-linear modelling. A recent strand of literature explores the non-linear dynamics of spreads and macro variables. Faulwasser, Gross, and Semmler (2018) find support for non-linear dynamics based on credit spreads and asymmetric impacts of unconventional monetary policies. Bennani and Neuenkirch (2020) find that credit spread shocks have a bigger impact on growth than credit quantity shocks in a non-linear context. Their model suggests that credit standards and risk premiums are more relevant predictors of recessions than volatility indices like the VIX. Chak and Chiu (2016) find that mortgage spread shocks have deeper effects during recessions and these are amplified through asset prices. Similar conclusions are drawn for financial stress shocks in Galvão and Owyang (2018).

[1] Gorton and Metrick (2012) already explained the high correlation between credit spreads and securitization activities by shadow banks.

On the other side, Nordström (2021) finds that heteroskedasticity is the main driver of the time varying relationship between credit spreads and unemployment using a time-varying parameters model. Even though shocks are larger during recessions, the response to those shocks is very similar to that on expansions. Other papers that find support for linear transmission of financial shocks are Chan and Eisenstat (2017), Xu and Haan (2018) and Karlsson and Österholm (2020). This thesis considers this possibility in the augmented model.

2.2 Literature on Vector Autoregressions

Most of the literature discussed above make use of structural Vector Autoregressions (VAR), initially presented by Sims (1980). These are multivariate regressions that allow variables to correlate with each other across time and account for simultaneous causality once presented in structural form. Kilian and Lütkepohl (2017) provide an extensive summary on structural VARs.

These models have been extended to account for non-linearities in the transmission. These include Markov Switching VAR, Threshold VAR, Interacted VAR or Smooth Transition VAR[2]. Auerbach and Gorodnichenko (2012) build a smooth transition VAR that allows for two regimes in volatility and coefficients' transmission. These models rely on Bayesian methods for estimation. Koop and Korobilis (2009) and Miranda-Agrippino and Ricco (2019) explore Bayesian methods for VAR models.

Structural identification usually relies on restrictions motivated by theory: exclusion restrictions on the short (Sims 1980) or long run (Blanchard and Quah 1989) and sign restrictions (Uhlig 2005). Recently, structural identification techniques are using features of the data to impose restrictions. These data-driven methods range from Heteroskedasticity Identification (HI) to higher order moments identification, that rely on independence of structural shocks. A discussion over the suitability of each of these methods is provided in Olea, Plagborg-Møller, and Qian (2021). Following their argumentation and in the light of the residual's diagnostics, this paper puts emphasis on HI and does not use independence for identification.

Heteroskedasticity is a stylized fact of macro and financial time series (Engle 1982) which is neglected when the model is assumed to be homoskedastic. As shown below, this poses a problem for inference, which is crucial for structural analysis. HI was presented by Rigobon (2003) with a fixed-point volatility break, and since then it has been applied using different specifications: Normandin and Phaneuf (2004) use a GARCH model, Lanne, Lütkepohl, and Maciejowska (2010) a Markov Switching model and Lütkepohl and Netšunajev (2017b) a smooth transition model. This identification method can be combined with non-linear transmission. In this line, Bacchiocchi and Fanelli (2015) relax the assumption of constant structural parameters across regimes and derive the rank

[2] Some relevant work on smooth transition autoregressions include Timo Teräsvirta (1994), Eitrheim and Teräsvirta (1996) or Hubrich and Timo Teräsvirta (2013)

condition necessary for HI of shocks in models with non-linear transmission. A review of different specifications for HI is provided in Lütkepohl and Netšunajev (2017a). Partial Identification is addressed by Lütkepohl, Milunovich, and Yang (2020) while Lütkepohl, Meitz, et al. (2021) develop a frequentist test for partial HI using a single break point. This research considers two of the above-mentioned specifications and argues in favor of the smooth transition.

Inference in HI VAR models is addressed by Herwartz and Plödt (2016). They highlight that identification by means of volatility changes requires of large samples (> 200 observations). Then they test for overidentifying sign restrictions in a VAR with monetary policy, inflation and growth. Other applications include Guidolin, Massagli, and Pedio (2021), that use HI to disentangle the effects of unconventional monetary policies on various risk premium while avoiding the implausible Cholesky restrictions in a financial context with asset prices and strong asymmetries (Ehrmann, Fratzscher, and Rigobon 2011). This thesis adds further evidence against Cholesky restrictions but finds support for some assumptions held to identify financial shocks.

For structural models, the main statistic of analysis are impulse responses. In empirical research, bootstrap methods are applied for constructing confidence intervals. Residual based bootstrap methods relying on the $i.i.d.$ assumption are invalid under heteroskedasticity and produce very inaccurate bands (Gonçalves and Kilian 2004). The Wild bootstrap provides an alternative whenever data displays heteroskedasticity. Kreiss (1997) established the asymptotic validity of a fixed-design Wild bootstrap for stationary autoregressions under a concrete form of conditional heteroskedasticity. Gonçalves and Kilian (2004) then proved the asymptotic validity of the recursive-design scheme. However, Brüggemann, Jentsch, and Trenkler (2016) show that the Wild bootstrap has no asymptotic validity for constructing confidence bands for impulse responses because it fails to replicate the 4rth order moments of the distribution. Instead, they suggest the use of a moving-block bootstrap for these purposes. However, their simulation suggests that these method has poor sample properties for relatively small samples and is not suitable when the process contains a unit root. It is unknown if this method remains valid under heteroskedasticity identification, as they use a recursive Cholesky identification in the research. On the other hand, Wild bootstrap performs well in the simulations. For these reasons, this thesis proceeds with the Wild bootstrap scheme. An alternative approach, would be the construction of confidence bands using a SVAR-GARCH model. This is explored by Bruder (2018) and Lütkepohl and Schlaak (2019). They explore the construction of confidence bands for impulse responses identified using conditional heteroskedastcity in a SVAR-GARCH model.

Lastly, the confidence intervals for impulse responses can be constructed using various procedures. Benkwitz, Lütkepohl, and Wolters (2001) argue in favor of Hall's percentile interval (Hall 1992) for bias correcting the bootstrapped impulse responses and recentering the 2nd order moments.

3 Baseline Homoskedastic Linear VAR

Section 3.1 presents the baseline homoskedastic VAR. Specifically, Section 3.1.1 presents the reduced form VAR as well as the the Maximum Likelihood estimation methods and the limit distribution. Section 3.1.2 presents the model in structural form, derives the impulse responses and discusses the construction of confidence intervals with bootstrap methods. Section 3.1.2 briefly comments on the implications of including some unit roots. Once the theory is presented it continues with the empirical approach. Section 3.2 presents the data and lag length choice based on various information criteria. Section 3.3 uses a Cholesky decomposition for identification of structural shocks and conducts inference in the form of impulse responses. Section 3.4 presents multiple specification tests on the residual's assumptions and their implications. Lastly, Section 3.5 discusses alternative and more suitable identification schemes in accordance with the specification tests. The estimation, identification and testing of the VAR relies on the R package **vars** presented in Pfaff (2008).

3.1 Linear Homoskedastic VAR

This section presents the baseline linear homoskedastic VAR, its estimation, recovering of structural impulse responses and the construction of suitable confidence bands using bootstrap methods. Most of these derivations are then expanded or complemented in Appendix A. This analysis follows Chapter 2 in Kilian and Lütkepohl (2017), Hamilton (1994) and Nielsen (2021).

3.1.1 Reduced form VAR and Estimation

Consider the K-dimensional linear VAR(p): $Y_t \in \mathbb{R}^K$

$$Y_t = v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t \tag{1}$$

where $Y_t = [y_{1,t}, y_{2,t}, \ldots, y_{K,t}]'$ is a $(K \times 1)$ vector of observable variables, v is a $(K \times 1)$ vector of constants and the $(K \times 1)$ error term is a white noise process $u_t \sim N(0, \Omega_u)$ with no serial correlation and homoskedasticity. Ω_u is a $(K \times K)$ time-invariant positive definite matrix. The coefficient matrices $\{A_i\}_{i=1}^{p}$ are $(K \times K)$.

The system in (1) can be rewritten as:

$$Y_t = \mathbf{A} X_{t-1} + u_t \tag{2}$$

with $\mathbf{A} = [v, A_1, \ldots, A_p]$ of size $(K \times (Kp+1))$ and $X_{t-1} = [\mathbf{1}', Y_{t-1}', \ldots, Y_{t-p}']'$ of size $((Kp+1) \times 1)$, where $\mathbf{1}'$ is a vector of 1s. The Gaussian likelihood and log-likelihood density functions for a given process $Y_t \in \mathbb{R}^K$ are:

$$f(Y|\theta) = \mathcal{L}_T(\theta) = \prod_{t=1}^{T} l_t(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^K |\Omega_u|}} \exp\left(-\frac{1}{2} u_t' \Omega_u^{-1} u_t\right) \tag{3}$$

$$\log \mathcal{L}_T(\theta) = -\frac{TK}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log |\Omega_u| - \frac{1}{2} \sum_{t=1}^{T} (Y_t - \mathbf{A} X_{t-1})' \Omega_u^{-1} (Y_t - \mathbf{A} X_{t-1}) \tag{4}$$

where the given data is $Y = [Y_t, Y_{t-1}, \ldots, Y_0]$, and the parameters $\theta = (\mathbf{A}, \Omega_u) \in \Theta$. The Maximum Likelihood Estimates: $\widehat{\theta}^{MLE} = (\widehat{\mathbf{A}}^{MLE}, \widehat{\Omega}_u^{MLE})$, now denoted simply as $\widehat{\theta}$ throughout this section, are derived from the condition:

$$\widehat{\theta} := \arg\max_{\theta \in \Theta} \log \mathcal{L}_T(\theta)$$

The first order derivative is denoted as the score: $S_T(\theta) = \frac{\partial \log \mathcal{L}_T(\theta)}{\partial \theta}$. By definition, the score evaluated in the MLE must equal to zero: $S_T(\widehat{\theta}) = \mathbf{0}$, where $\mathbf{0}$ is a matrix of 0s[3]. Optimization implies:

$$S_T(\theta) = \begin{bmatrix} \frac{\partial \log \mathcal{L}_T(\mathbf{A}, \Omega_u)}{\partial \mathbf{A}} \\ \frac{\partial \log \mathcal{L}_T(\mathbf{A}, \Omega_u)}{\partial \Omega_u} \end{bmatrix} = \begin{bmatrix} -2\Omega_u^{-1} \sum_{t=1}^{T} (Y_t - \mathbf{A} X_{t-1}) X_{t-1}' \\ -\frac{T}{2} \Omega_u^{-1} + \frac{1}{2} \Omega_u^{-2} \sum_{t=1}^{T} u_t u_t' \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

For the continuous differentiable function in (4) there is a unique closed form solution[4], which we denote as the MLE. This is actually a quasi MLE, because the estimate for the variance matrix relies on the estimate of the coefficients. Using the two-step step estimation technique and assuming that Ω_u is positive definite, the MLE are:

$$\widehat{\mathbf{A}} = \left(T^{-1} \sum_{t=1}^{T} X_{t-1} X_{t-1}'\right)^{-1} \left(T^{-1} \sum_{t=1}^{T} Y_t X_{t-1}'\right) \tag{5}$$

$$\widehat{\Omega}_u = T^{-1} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t' \quad \text{where} \quad \widehat{u}_t = \widehat{Y}_t - \widehat{\mathbf{A}} X_{t-1} \tag{6}$$

The MLE are equal to the OLS estimates obtained from: $\widehat{\theta}^{OLS} := \arg\min_{\theta \in \Theta} \sum_{t=1}^{T} u_t' u_t$. The step-by-step derivation of the OLS estimates can be found in Appendix A.3.

Let Y_t be a stationary and weakly dependent process with finite fourth order moments

[3] A comment on notation: Along this paper $\mathbf{0}$ is used to refer to vectors and matrices of zeros with different dimensions. This should not be interpreted as if each $\mathbf{0}$ were identical. Their dimensions must be inferred from the counterpart. The reason for this notational inconsistency is to facilitate reading.

[4] For a given continuous differentiable function, the score also equals zero when evaluated in a local maximum (or minimum). For this reason, the MLE might be difficult or impossible to derive in practice when functions have multiple maximum points. If there is no closed form solution, we can use a grid search and further numerical optimization techniques.

(Hamilton 1994) defined by:

$$
\begin{align}
&E[Y_t] = \mu \tag{7}\\
&E[(Y_t - \mu)(Y_t - \mu)'] = \omega < \infty \tag{8}\\
&E[Y_t, Y_{t+s}] = E[Y_t, Y_{t-s}] \tag{9}\\
&\text{Corr}(Y_t, Y_{t+h}) \to 0, \quad \text{as} \quad h \to \infty \tag{10}
\end{align}
$$

where conditions (7), (8) and (9) imply covariance stationarity[5] and (10) refers to weakly dependence. A covariance stationary process has constant average and variance with stable and symmetric dynamics across time. A weakly dependent process is mean reverting as the correlation between current and future periods decays to zero for large horizons. If Y_t is stationary and weakly dependent the Law of Large Numbers applies to the score:

$$
S_T(\mathbf{A}) = \frac{\partial \log \mathcal{L}_T(\mathbf{A}, \Omega_u)}{\partial \mathbf{A}} = \sum_{t=1}^{T} \frac{\partial \log l_T(\mathbf{A}, \Omega_u)}{\partial \mathbf{A}} \xrightarrow[LLN]{P} E\left[\frac{\partial \log l_T(\mathbf{A}, \Omega_u)}{\partial \mathbf{A}}\right]
$$

then the sample average of score contributions converges towards its expected value. Specifically, the consistency condition requires the Law of Large Numbers to apply to:

$$
T^{-1} \sum_{t=1}^{T} u_t X'_{t-1} \xrightarrow[LLN]{P} E[u_t X'_{t-1}] = \mathbf{0} \tag{11}
$$

If (11) holds, then $\widehat{\mathbf{A}}$ is a consistent estimate of $\mathbf{A}$. Consistency is given by the estimator's convergence in probability to its expected value: $\widehat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$, which relies on the Law of Large Numbers. Appendix A.3 proves consistency for the OLS estimator under condition (11). If the error term was not a white noise process but displayed autocorrelation, consistency would not hold. This is the case when there are omitted significant lags of the endogenous variables.

If additionally the expected value of $\widehat{\mathbf{A}}$ was equal to the true underlying parameter $\mathbf{A}$ from the DGP, then the estimator would be unbiased: $E[\widehat{\mathbf{A}}] = \mathbf{A}$. This relies on the strict exogeneity condition: $E[u_i X'_j] = 0, \ \forall i \neq j$. However this cannot be fulfilled in times series models, meaning that MLE is biased by construction. See Appendix A.3 for a more detailed explanation. In practice, the formula for the covariance in (6) can be adapted to account for the degrees of freedom used in the two step estimation[6]. Nevertheless, the bias converges towards zero as $T \to \infty$ and becomes less of a problem in large samples.

[5] This should not be confused with strict stationarity. A process $Y_t : t = 0, 1, \ldots$ is strictly stationary if the joint distribution $(Y_t, \ldots, Y_{t+h}), \forall t, h \geq 0$ does not depend on t.

[6] The MLE estimate for the covariance is: $\widehat{\Omega}_u = T^{-1} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}'_t$. However, this estimate is biased. In practice, many applications use: $\widehat{\Omega}_u = (T - Kp - 1)^{-1} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}'_t$. Specifically, this correction accounts for the degrees of freedom used to estimate $\widehat{\mathbf{A}}$ in the 1st stage of quasi MLE. It can be shown that correcting for the degrees of freedom reduces the bias in small samples, where the MLE underestimates the true variance.

Now consider the second order condition:

$$H_T(\theta) = \frac{\partial^2 \log \mathcal{L}_T(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\mathbf{A})\partial \text{vec}(\mathbf{A})'} & \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\mathbf{A})\partial \text{vec}(\Omega_u)'} \\ \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\Omega_u)\partial \text{vec}(\mathbf{A})'} & \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\Omega_u)\partial \text{vec}(\Omega_u)'} \end{bmatrix}$$

The block diagonal Information matrix is then obtained under the assumptions $E[u_t] = 0$ and $E[u_t u_t'] = \Omega_u$, which implies that residuals must have zero mean and constant variance, ruling out heteroskedasticity for the estimation of $I(\theta)$:

$$I(\theta) = -E[H_T(\theta)] = \begin{bmatrix} E[X_{t-1}X_{t-1}'] \otimes \Omega_u^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\Omega_u^{-2} \end{bmatrix}$$

Under general regularity conditions it can be shown that the MLE have the following asymptotic distribution:

$$\sqrt{T} \begin{bmatrix} \text{vec}(\widehat{\mathbf{A}} - \mathbf{A}) \\ \text{vec}(\widehat{\Omega}_u - \Omega_u) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0},\ \begin{bmatrix} E[X_{t-1}X_{t-1}'] \otimes \Omega_u^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\Omega_u^{-2} \end{bmatrix}\right)$$

The regularity conditions needed for the derivation of this result can be found in Appendix A.2. It relies on the assumption of homoskedasticity for the asymptotic variance of the MLE estimates. If the error term was heteroskedastic, MLE estimates would still be consistent but inefficient, as their variance would be imprecisely estimated.

The asymptotic Normality of MLE allows hypothesis testing on the estimated parameters. Wald and LR tests have χ^2 asymptotic distributions while t-ratios for any parameter in A_i have a standard normal asymptotic distribution: $(\widehat{a}_{mn,i} - a_{mn,i})/\widehat{\Sigma}_{a_{mn,i}} \sim N(0,1)$ where $\Sigma_{a_{mn,i}}$ is the corresponding standard error. If these t-ratios had fatter tails we should use a t-distribution instead. But a t-distribution with $T - Kp - 1$ degrees of freedom is very close to the standard normal whenever T is large enough. Once again, a large sample helps in the application of these tools. See Appendix A.1 for a more detailed explanation of MLE and hypothesis testing.

3.1.2 Structural form VAR and Impulse Responses

The model in (1) can be represented in structural form, allowing variables to correlate with each other contemporaneously as follows:

$$\begin{aligned} Y_t &= v + \sum_{i=1}^{p} A_i Y_{t-i} + B\epsilon_t \\ B^{-1}Y_t &= B^{-1}v + \sum_{i=1}^{p} B^{-1}A_i Y_{t-i} + \epsilon_t \end{aligned} \tag{12}$$

where B contains the contemporaneous relationships between variables such that $u_t = B\epsilon_t$. Now ϵ_t is a vector of homoskedastic and serially uncorrelated structural shocks with variance normalized to one: $\epsilon_t \sim N(\mathbf{0}, I_K)$. Estimation of (5) and (6) is not directly informative on the underlying process if $B \neq I_K$, as there is simultaneous causality. Intuitively, the observable outcomes of the underlying process are a combination of various shocks to $\epsilon_{i,t}$ and their propagation across variables and over time. We cannot directly estimate these effects by OLS, as first we need to recover B.

Structural analysis requires recovering B but this can be any matrix satisfying:

$$\Omega_u := E[u_t u_t'] = E[B\epsilon_t \epsilon_t' B'] = BI_K B' = BB'$$

The matrix B is not uniquely determined by the relationship $\Omega_u = BB'$, because the symmetric covariance Ω_u has $K(K+1)/2$ different elements and then the matrix B has K^2 different elements. In order to uniquely determine the transformation matrix B we need to impose $K(K-1)/2$ further restrictions on B. In practice, there are multiple identification strategies for imposing these restrictions. Once the unique matrix B is recovered, we can obtain the structural shocks ϵ_t. After the shocks are identified, $u_t = B\epsilon_t$ can be combined with (1) such that:

$$A(L)Y_t = \left(I_k - \sum_{i=1}^{p} A_i L^i \right) Y_t = v + B\epsilon_t$$

where L is the lag operator such that: $L^j x_t = x_{t-j}$. The system is stable if the characteristic polynomial has no roots inside or on the unit circle of complex numbers:

$$|A(z)| = \left| I_K - \sum_{i=1}^{p} A_i z^i \right| \neq 0 \quad \forall z \in \mathbb{C}, \quad |z| \leq 1 \tag{13}$$

where the roots of the polynomial are every z_j that solves $|A(z)| = 0$. Condition (13) means that all the characteristic roots are outside of the complex unit circle $|z_j| > 1$, or equivalently, all the eigenvalues are inside $|\lambda_j| = |1/z_j| < 1$. Then $A(L)$ has full rank and all variables are stationary. Now $A(L)$ can be inverted to obtain the Moving Average Representation of the process (1):

$$Y_t = \left(I_K - \sum_{i=1}^{p} A_i L^i \right)^{-1} (v + B\epsilon_t) = \widetilde{v} + \sum_{h=0}^{t} \zeta_h \epsilon_{t-h} \tag{14}$$

where $\zeta_h = \partial Y_t / \partial \epsilon_{t-h}'$ is the $(K \times K)$ matrix of responses from the vector of variables Y_t to an exogenous impulse on the vector of structural shocks ϵ_t. In practice, the Impulse Response Functions (IRF) are obtained for a finite horizon $h = 0, 1, \ldots, \aleph$. The h period's ahead response of the mth variable to an nth shock is given by: $\zeta_{mn,h} = \partial y_{m,t} / \partial \epsilon_{n,t-h}$. IRFs converge to zero as $h \to \infty$ under the stability condition (13) because the effects of the shocks die out. Kilian and Lütkepohl (2017) enumerate three restrictions that linear VAR models impose on the linear IRF: (i) proportional increases with the magnitude of the shock, (ii) time and state invariant and (iii) symmetry to positive and negative shocks.

In order to see how IRFs (ζ_h) are constructed in practice it is useful to rewrite the process in (1) in companion form:

$$Z_t = \nu + \Pi Z_{t-1} + U_t$$

such that Z_t is a VAR(1) process with Z_t and Z_{t-1} of dimensions $(pK \times K)$, ν and U_t of dimensions $(pK \times 1)$ and Π of dimensions $(pK \times pK)$:

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} v \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_K & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & I_K & \mathbf{0} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ \vdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

The Moving Average solution can be derived by recursive substitution:

$$\begin{aligned} Z_t &= \nu + \Pi Z_{t-1} + U_t \\ &= \nu + \Pi(v + \Pi Z_{t-2} + U_{t-1}) + U_t \\ &= (I_K + \Pi)\nu + \Pi^0 U_t + \Pi U_{t-1} + \Pi^2 Z_{t-2} \\ &\vdots \\ Z_t &= (I_K + \Pi + \ldots + \Pi^{t-1})\nu + \sum_{i=0}^{t-1} \Pi^i U_{t-i} + \Pi^t Z_0 \end{aligned}$$

such that $\Pi^0 = I_{pK},\ \Pi^2 = \Pi\Pi,\ \Pi^3 = \Pi\Pi\Pi$. The Moving Average representation consists on a deterministic part, a moving average of the error terms and the initial values. If the matrices $\Pi^t \xrightarrow{P} \mathbf{0}$ as $T \to \infty$, then the effect of the effect of the initial values vanish: $\Pi^t Z_0 \xrightarrow{P} \mathbf{0}$. This is the case when the model is stable. Consider the spectral decomposition of Π:

$$\Pi = V \Lambda V^{-1}$$

where the eigenvectors are given by $V = [v_1, \ldots, v_{pK}]$ while the matrix of eigenvalues is $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{pK})$. These can be obtained as the solution to the problem: $|\Pi - \lambda_i I_p| = 0$. It holds that: $\Pi^n = V \Lambda^n V^{-1}$, where $\Lambda^n = \text{diag}(\lambda_1^n, \ldots, \lambda_{pK}^n)$. If the eigenvalues are all inside the unit circle and the stability condition is satisfied, then $\Pi^n \to \mathbf{0}$ exponentially fast. For that reason the effect of the initial values vanish form the Moving Average representation of Z_t.

Let Π^i_{mn} be the mth row and nth column element from Π^i. If we are interest in recovering Y_t we only need Π^i_{11}, the first element of each matrix Π^i. The process becomes:

$$Y_t = (\Pi^0_{11} + \Pi_{11} + \Pi^2_{11} + \Pi^3_{11} + \ldots)v + \Pi^0_{11} u_t + \Pi_{11} u_t + \Pi^2_{11} u_{t-2} + \Pi^3_{11} u_{t-3} + \ldots$$

where $\Pi_{11}^0 = I_K,\ \Pi_{11}^1 = \Pi_{11}^0 A_1,\ \Pi_{11}^2 = \Pi_{11}^1 A_1 + \Pi_{11}^0 A_2,\ \Pi_{11}^3 = \Pi_{11}^2 A_1 + \Pi_{11}^1 A_2 + \Pi_{11}^0 A_3$ and so on. See Appendix A.5 for a detailed explanation on how to derive these values. For an horizon of $\aleph$ the moving average process becomes:

$$\begin{aligned} Y_t &= \sum_{h=0}^{\aleph} \Pi_{11}^h \upsilon + \sum_{h=0}^{\aleph} \Pi_{11}^h u_{t-h} \\ &= \sum_{h=0}^{\aleph} \Pi_{11}^h \upsilon + \sum_{h=0}^{\aleph} \Pi_{11}^h B\epsilon_{t-h} \\ &= \widetilde{\upsilon} + \sum_{h=0}^{\aleph} \zeta_h \epsilon_{t-h} \end{aligned}$$

which is equivalent to the Impulse Responses in (14). $\zeta_h = \Pi_{11}^h B = \partial Y_t / \partial \epsilon'_{t-h}$ is the $(K \times K)$ response of the vector of variables Y_t to an exogenous impulse to the vector of structural shocks ϵ_t. Above it can be seen that the IRFs are continuous functions of the parameters:

$$\zeta_h = \zeta_h(\upsilon, A_1, \ldots, A_p, B)$$

Applying the continuous mapping theorem, the IRFs are consistently estimated if the parameters are consistent (Kilian and Lütkepohl 2017).

Under general conditions the IRF have asymptotically normal distributions that may be used to construct confidence intervals. However in practice bootstrap methods are used, as they work better in small samples. These methods also rely on asymptotic theory[7] but allow for the construction of bands without deriving the analytical solutions. The Wild bootstrap is suggested by Kreiss (1997) as a suitable candidate whenever data displays heteroskedasticity. Gonçalves and Kilian (2004) find that this method also performs well when data is homoskedastic, however methods that assume *i.i.d.* errors perform badly when data is heteroskedastic. Brüggemann, Jentsch, and Trenkler (2016) show that the Wild bootstrap does not properly capture the higher order moments of the underlying distribution and therefore is not suitable for constructing confidence bands for IRFs under conditional heteroskedasticity. Alternatively they suggest a Moving Block bootstrap that is asymptotically valid. However, it performs badly in small samples and only works if residuals are uncorrelated and there are no I(1) variables included in the VAR whereas the Wild bootstrap performs well in simulation studies. Because of these shortcomings, this paper follows Lütkepohl and Netšunajev (2017b) and chooses a Wild bootstrap with fixed design. Appendix A.6 explains the Wild bootstrap algorithm and how it is used to construct confidence bands for the IRFs.

The set of bootstrapped IFRs: $\{\widehat{\zeta}_{mn,h}^{(m)}\}_{m=1}^{M}$ is composed out of $h = 0, \ldots, \aleph$ pointwise distributions for each $\widehat{\zeta}_{mn}$: $f(\widehat{\zeta}_{mn,h})$. The 90% confidence interval for IRFs can be created using Hall's percentile interval from Hall (1992):

$$CI^{90\%} = [\widehat{\zeta}_{mn,h} - (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{5\%}),\ \widehat{\zeta}_{mn,h} + (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{95\%})] \tag{15}$$

[7] See Cavaliere, Nielsen, and Rahbek (2020) for an overview of bootstrap methods and theory.

where $f(\widehat{\zeta}_{mn,h})^{Q\%}$ is the Qth quantile of the bootstrap distribution of $\widehat{\zeta}_{mn}$ at point h. This method is suggested by Benkwitz, Lütkepohl, and Wolters (2001) to center the confidence bands around the IRFs and act as a bias corrector. If IRFs were biased, the bootstrap algorithm amplifies this bias. This method provides a simple bias adjustment mechanism.

3.1.3 Unit Roots

If there are integrated variables included in the VAR such that $Y_t \sim I(1)$[8], the process is not stable as defined by (13). This can happen if one variable contains a unit root, making the model unstable: $A(z) = 0$ for $z = 1$, and therefore non-stationary. If variables can be linearly combined such that there exists a stable long run equilibrium relationship between them, the $I(1)$ the process is said to be cointegrated (Johansen 1988). This calls for an alternative specification that accounts for this long-term stable relationship and allows the model to be stationary in the short-term. This possibility is not explored on this paper and therefore I refer to the literature on cointegration.

The case of some unit roots is of more relevance to this paper. When the process is $I(1)$ the LLN and CLT do not apply in their standard versions. In this case, MLE and OLS can still be used and are asymptotically normal under general conditions. Sims, Stock, and Watson (1990) show that coefficient estimates of VAR parameters with unit roots are consistently estimated under MLE and OLS assuming that the error term is a martingale difference sequence and provided that the model includes an intercept and enough lags. However, the estimated coefficients converge in probability at different rates: the estimators of the unit roots converge at speed T to non-normal distributions, while the estimators of stationary variables converge at the usual $\sqrt{T}$ rate to Normal distributions. Under these conditions, individual Wald tests have the usual $\chi^2_{(q)}$ asymptotic distributions.

However, Kilian and Lütkepohl (2017) point out that joint tests on $\mathbf{A}$ are not asymptotically valid under the presence of a unit root or near-unit root. The reason is that the covariance matrix of the asymptotic distribution $\widehat{\Sigma}_{\mathbf{A}}$ is singular because some estimators converge at rate T instead of $\sqrt{T}$ (Toda and P. C. Phillips 1993).

P. C. B. Phillips (1998) then shows that if coefficient estimates are consistent, then IRFs are also point-wise consistent as they are functions of the coefficients. However IRFs are inconsistent in long horizons h and the limit. Inoue and Kilian (2020) further prove the uniform validity of impulse responses without the need of lag-augmentation provided that the model includes $p > 1$. However this uniform validity cannot be extended to inference in the form of Wald tests, which require the use of lag-augmented VAR models.

[8] A process $W_t \sim I(d)$ is said to be integrated of order d. By differentiating d times, it can be converted into a stationary process without a unit root: $\Delta^d W_t \sim I(0)$, therefore integrated of order 0.

3.2 Data and Model Selection

In order to correctly interpret the effects of financial shocks, it is necessary to find the best VAR specification and ensure a robust estimation across models. For this purpose I use various models that add complexity to the baseline homoskedastic linear VAR explored in this section. Every model is a simplification of the economy that reduces the main macroeconomic variables of interest to four: inflation, output growth, monetary policy rate and the bond risk premium. There are many measures for these variables but the Heteroskedasticity Identification explored in Section 4 is not robust to all of them. The reason is that this method requires the relative variance in high volatility periods to be different for each structural shock, as further explained in Sections 3.5 and 4.1.1. The variables have been selected according to whether they fully identify the system under Heteroskedasticity Identification[9]. These are: year-on-year Producer Price Index growth, year-on-year industrial production growth, Shadow Federal Funds Rate (Wu and Xia 2016) and Excess Bond Premium (Gilchrist and Zakrajšek 2011). The latter can be interpreted as the price of default risk and contains the unpredictable components that explains most of the variation from credit spreads. On the other hand, the shadow rate combines the nominal Federal Funds rate with unconventional monetary policies, meaning that this index comprises multiple tools therefore having a ambiguous interpretation. The reason for using this measure instead of the nominal rate is to account for the effects of unconventional policies which significantly lower risk premiums and are found to affect other macroeconomic variables in similar ways (Borio and Zabai 2016)[10]. A detailed description of each variable can be found in Appendix B. The series are displayed on Figure 1 and cover the period from 1980-01 to 2020-01 with monthly frequency and length $T = 481$. Because these variables are not seasonally adjusted, a way to get rid of seasonal autocorrelation while de-trending industrial production and prices is to include them in year-on-year growth rates with: $x_t = \log(X_t) - \log(X_{t-12})$.

The following analysis requires estimating the model presented in Section 3.1 (OLS) with the current set of variables. For the linear homoskedastic VAR model, the lag length $p = 3$ is chosen on the basis of 3 information Criteria displayed in Table 1. These IC

[9]This means that other variables were taken into account but were discarded because their relative volatility across the regimes of interest was not different enough. Among the considered measures of inflation, the Consumer Price Index and Core Consumer Price Index failed to achieve full identification. The ten year bond credit spread was also considered instead of the bond premium, but also failed to fully identify the system. Fortunately, there are many proxies for these variables in macroeconomics, allowing to exchange them without significantly altering the conclusions. Nevertheless, it should be kept in mind that these choices have some implications for the conclusions.

[10]The use of one or the other rate for monetary policy has implications in the interpretation of results. Each monetary policy tool is different and therefore differently affects the economy, but ignoring unconventional policies seems less appropriate that including them, considering their presence in the sample and their direct effects on risk premiums. There is a trade off between accurately differentiating these tools and their different effects or ignoring the relevance of some of these measures.

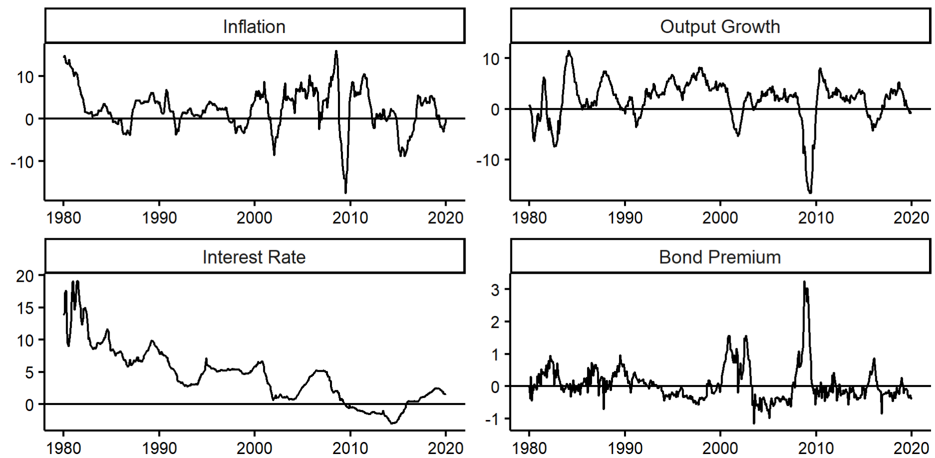


Figure 1: Endogenous variables over time at monthly frequency. Sample: 1980-2020.

are computed as:

$$
\begin{aligned}
AIC(p) &= \log(|\widehat{\Omega}_u(p)|) + 2T^{-1}(pK^2) \\
HQC(p) &= \log(|\widehat{\Omega}_u(p)|) + 2T^{-1}\log(\log(T))(pK^2) \\
SC(p) &= \log(|\widehat{\Omega}_u(p)|) + T^{-1}\log(T)(pK^2)
\end{aligned}
$$

This lag length will be used for the following models in Sections 4 and 5 to facilitate the comparability of results across different model specifications.

p	1	2	3	4	5	6	7	8	9	10
$AIC(p)$	-4.44	-4.95	-5.04	-5.06	-5.09	-5.08	-5.11	-5.11	-5.13	**-5.17**
$HQC(p)$	-4.37	-4.83	**-4.86**	-4.83	-4.79	-4.73	-4.71	-4.65	-4.62	-4.60
$SC(p)$	-4.26	**-4.64**	-4.58	-4.46	-4.34	-4.20	-4.09	-3.94	-3.82	-3.72

Table 1: Information Criteria of linear VAR(p)

3.3 Cholesky Identification and Inference

Consider the 4-dimensional VAR($p = 3$) with variables: inflation (π_t), output growth (x_t), interest rate (i_t) and bond premium (bp_t). Identification can be achieved by imposing theory-based restrictions, in this case a Cholesky decomposition: $u_t = B\epsilon_t$ with the following variable ordering:

$$
\begin{bmatrix} u_t^{\pi} \\ u_t^{x} \\ u_t^{i} \\ u_t^{bp} \end{bmatrix} = \begin{bmatrix} B_{11} & 0 & 0 & 0 \\ B_{21} & B_{22} & 0 & 0 \\ B_{31} & B_{32} & B_{33} & 0 \\ B_{41} & B_{42} & B_{43} & B_{44} \end{bmatrix} \begin{bmatrix} \epsilon_t^{\pi} \\ \epsilon_t^{x} \\ \epsilon_t^{i} \\ \epsilon_t^{bp} \end{bmatrix} = \begin{bmatrix} B_{11}\epsilon_t^{\pi} \\ B_{21}\epsilon_t^{\pi} + B_{22}\epsilon_t^{x} \\ B_{31}\epsilon_t^{\pi} + B_{32}\epsilon_t^{x} + B_{33}\epsilon_t^{i} \\ B_{41}\epsilon_t^{\pi} + B_{42}\epsilon_t^{x} + B_{43}\epsilon_t^{i} + B_{44}\epsilon_t^{bp} \end{bmatrix}
$$

Here, the bond premium reacts contemporaneously to shocks on every variable while the interest rate reacts on impact to all but bond premium shocks. Output growth responds to contemporaneous shocks on inflation but only responds to financial variables' shocks one period after. This way, inflation is the least reactive variable and only responds to shocks on other variables one month later. The reasoning behind this ordering is that financial markets incorporate faster the information than agents in the good's markets. The interest rate is determined in the interbank lending market, which responds to liquidity needs of banks and is affected by the Central Bank policy. Financial information and events occurring in other segments of the financial sector only affect the short-term rate when other financial actors change their behavior and this later affects banks or regulators. Examples would include both hidden information inferred from changing prices (Evans and Lyons 2002) or events affecting shadow banks that change their behavior, affecting the balance sheet of banks and therefore their liquidity needs or risk perception (Becard and Gauthier 2021). This would mean that the interest rate is less reactive than the bond premium. Lastly, if prices are sticky, inflation reacts to mismatches in aggregate demand and supply with a lag.

Using the above-mentioned Cholesky ordering, the recovered $\widehat{B}$ from the estimated reduced model is shown in Table 2. Standard errors are not provided as $B(\mathbf{A}, \Omega_u)$ is a function of the variance and parameter estimates and therefore not directly estimated. The signs of the effects are in line with the previous intuition: an inflationary (π_t) shock has a positive effect on inflation, negative effect on growth, positive response of the policy rate and increase in risk premium. A demand shock (x_t) affects positively output and interest rates but negatively the risk premium. An interest rate (i_t) negatively affects positively affects itself and the risk premium. The financial shock (bp) is restricted to zero on impact for each variable.

	π_t	x_t	i_t	bp_t
π_t	1.11	0	0	0
x_t	-0.05	0.84	0	0
i_t	0.00	0.12	0.45	0
bp_t	-0.02	-0.01	0.01	0.23

Table 2: $\widehat{B}$ under Cholesky Identified VAR(3)

IRF are derived as explained in Section 3.1.2. Following Lütkepohl and Netšunajev (2017b), estimation of 2nd order moments is done using a fixed design Wild bootstrap with 2000 iterations. See Appendix A.6 for a detailed explanation of the algorithm. Gonçalves and Kilian (2004) suggest this method whenever data displays heteroskedasticity as the fixed design preserves the pattern of the original sample, allowing a suitable comparison between the actual IRFs and the bootstrapped ones. While recursive design is more accurate than fixed design, they find that differences vanish in large samples. The reasoning behind this bootstrap choice against alternative methods like the moving-block bootstrap[11] is explained in Sections 2.2 and 3.1.2. Using the bootstrapped IRFs, the 90% confidence bands are built using Hall's percentile interval as explained in (15). This interval accounts for the bias present in the estimated IRFs and acts as bias corrector for the confidence bands (Benkwitz, Lütkepohl, and Wolters 2001).

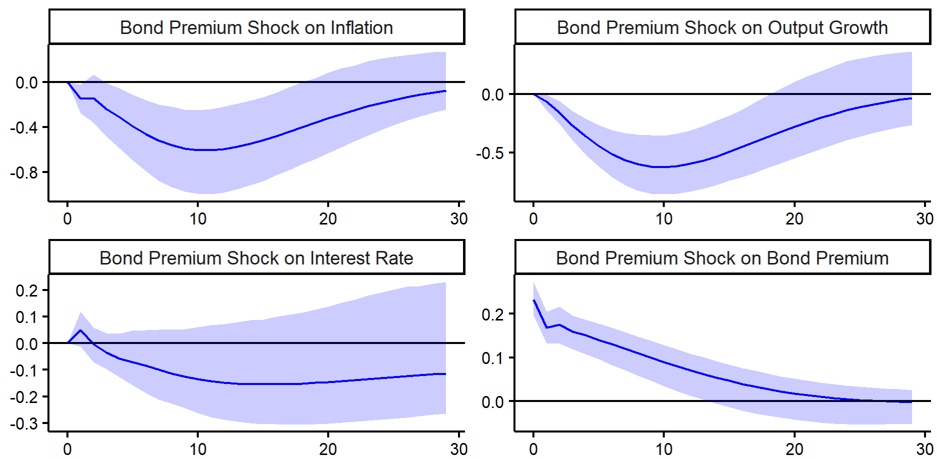


Figure 2: Impulse responses to a bond premium shock identified with Cholesky restrictions. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws. The full set of impulse responses is displayed in Figure 12.

Figure 2 displays the IRFs for financial shocks. A bond premium shock triggers a fall in output growth, increases deflation and consequently lowers the interest rate. Exogenous shocks to the bond premium can be interpreted as unexpected increases in the price of risk or sudden decreases in the expected future solvency of risky firms relative to the government. Consequently lenders constrain credit supply and rise the interest rate. The overall tighter financial conditions make borrowing expensive, forcing firms to reduce investment and employment with a negative impact on aggregate demand and therefore downward pressure on prices. Being US a big open economy, this reduction in demand also affects commodity and energy prices. Consistent with the Federal Reserve mandate on price stability and maximum sustainable employment, the interest rate is reduced until inflation and output recover and the bond premium falls. This reduction in interest rates has wide confidence bands because the shadow rate combines the nominal Federal Funds rate with forward guidance and unconventional monetary policies, implying heterogeneous reactions during the considered period. These results are in line with Gilchrist and Zakrajšek (2011).

[11]Brüggemann, Jentsch, and Trenkler (2016) argue against the asymptotic validity of Wild bootstrap as it fails to replicate the 4th order moments. Instead they argue for a moving-block bootstrap but it does not work whenever data is I(1) and the interest rate series displays high persistence. Furthermore, the Wild bootstrap performs well in their simulations.

3.4 Diagnostics

This section evaluates some of the most relevant assumptions held by the model presented in 3.1. Stability is checked and tests for no autocorrelation, normality and homoskedasticity are considered. These tests are displayed in Table 3. The purpose of this section is not to provide a detailed explanation for each test but to evaluate some of the assumptions held along Section 3.1. I refer to external papers for a detailed analysis. Both Hamilton (1994) and Kilian and Lütkepohl (2017) expand on these tools.

Covariance stationarity was given by conditions (7), (8) and (9). These are not satisfied if the system is not stable or the parameters change over time, that is, if there are structural breaks. If the eigenvalues are inside of the complex unit circle, the system is stable in line with (13). Eigenvalues are allowed to be close to unity, as far as they are not equal or higher than one in modulus. This means that shocks are allowed to have persistence as far as its impact decays over time and the process reverts to its expected value. Table 4 shows that the modulus of all eigenvalues $|\lambda_j|$ is inside of the unit circle with one very close to unity, signaling a potential unit root.

test	JB Normality	Skewness	Kurtosis	$\text{Portman}_{(132)}$	$\text{ARCH}_{(20)}$
statistic	27639	492	2073	2100	3182
p-value	0.00	0.00	0.00	0.29	0.00
degrees of freedom	8	4	4	2064	2000

Table 3: Specification tests for unrestricted VAR(3)

Real	Imaginary	Modulus
0.933	0.000	0.933
0.871	0.096	0.876
0.871	-0.096	0.876
0.807	0.000	0.807
0.299	-0.488	0.573
0.299	0.488	0.573
0.358	0.000	0.358
0.341	0.000	0.341
0.140	-0.115	0.182
0.140	0.115	0.182
-0.075	0.078	0.108
-0.075	-0.078	0.108

Table 4: Eigenvalues of companion matrix from linear VAR(p)

The near root comes from the high persistency of the interest rate[12]. It is out of the scope of this paper to solve this problem and therefore I chose to follow the American

[12]One problem arises from the 'short' sample that begins in the 1980, when interest rates were

literature (Sims 1980, Uhlig 2005) and proceed with the interest rate in levels. If this variable had a unit root, parameter estimates would still be consistently estimated (Sims, Stock, and Watson 1990) and so would impulse responses (P. C. B. Phillips 1998) as explained in Section 3.1.3.

On the other hand, if parameters changed over time conditional on different states of the economy or due to structural breaks, then non-linear modelling should be considered. In the context of a VAR that aims to model financial cycles' dynamics, this could arise if the transmission of shocks was amplified in periods of financial stress (Galvão and Owyang 2018). If not only the transmission but also the expected value changed, this would signal the existence of level shifts. These two possibilities will be explored in Section 5.

Consider the case of an autocorrelated error term:

$$u_t = \sum_{i=1}^{s} \rho_i u_{t-i} + e_t \quad \text{with} \quad e_t \sim i.i.d.N(0, \Sigma_e)$$

Consider the simpler case of $s = 1$, then the consistency condition (11) is not satisfied:

$$E[u_t X'_{t-1}] = E[(\rho_1 u_{t-1} + e_t) X'_{t-1}] \neq \mathbf{0}$$

because $X_{t-1} = [\mathbf{1}', Y'_{t-1}, \ldots, Y'_{t-p}]'$ contains: $Y_{t-1} = v + \sum_{i=2}^{p+1} A_i Y_{t-i} + u_{t-1}$, and $E[u_t u'_{t-1}] \neq \mathbf{0}$. Residual's autocorrelation is tested with a Portmanteu test (Box and Pierce 1970) which evaluates the null hypothesis of no autocorrelation $H_0 : E[u_t u'_{t-i}] = \mathbf{0},\ i = 1, 2, \ldots, s$. The statistic is:

$$Q_s^{Portman} = T \sum_{i=1}^{s} \mathrm{tr}(\widehat{C}_i' \widehat{C}_0^{-1} \widehat{C}_i \widehat{C}_0^{-1}) \sim \ \chi^2_{(K^2(s-p))}$$

where $\widehat{C}_i = \frac{1}{T} \sum_{t=i+1}^{T} \widehat{u}_t \widehat{u}'_{t-i}$ and $\widehat{u}_t$ are the OLS residuals. Under the null and for large values of s, the test statistic follows approximately a χ^2. For this reason the test must be evaluated in large horizons (Box and Pierce 1970, Kilian and Lütkepohl 2017). Here $s = 132$ is chosen to ensure convergence[13]. Table 3 shows that the null cannot be rejected, favoring the hypothesis of no autocorrelation: $E[u_t X'_{t-1}] = \mathbf{0}$. This is necessary to consistently estimate $\mathbf{A}$.

increased repeatedly to account for high inflation. However, for longer horizons the interest rates seems more stationary. The American literature tend to include this variable in levels arguing that consistency and inference remain valid for a wide range of tools (Sims, Stock, and Watson 1990, P. C. B. Phillips 1998). Another strand of literature argue that interest rates have a long-run equilibrium relationship with other macroeconomic variables and can be better modelled following Johansen (1988). Christensen and Nielsen (2005) find a stable cointegration relationship between the Federal Funds, employment and the bond rate which can be interpreted as a Taylor rule.

[13]The chosen length $s = 132$ covers 11 times a year consisting of 12 observations. The test requires high values of s for convergence, but this also undermines the estimation accuracy. This paper favors using a longer horizon due to the high frequency (monthly) and relatively big sample size employed $T = 481$.

The analysis in Section 3.1.1 relied on a normally distributed error term. The normality assumption can be tested using the univariate tests from Jarque and Bera (1987) based on the 3rd order moment (skewness) and the 4rth order moment (kurtosis) of the residuals, which are contrasted with those of a Normal distribution[14]. Kilian and Lütkepohl (2017) explain that the residual error of each variable needs to be orthogonal. This can be done by using the structural errors $\epsilon_t = B^{-1}u_t$ obtained with a Cholesky decomposition of the covariance. The skewness (ξ^S), kurtosis (ξ^K) test statistics are given by:

$$\xi^S = \frac{T}{6}\left(T^{-1}\sum_{t=1}^{T}\left((\widehat{\epsilon}_t - \bar{\epsilon}_t)'\widehat{\Omega}_\epsilon(\widehat{\epsilon}_t - \bar{\epsilon}_t)\right)^3\right)^2 \sim \chi^2_{(K)}$$

$$\xi^K = \frac{T}{24}\left(T^{-1}\sum_{t=1}^{T}\left((\widehat{\epsilon}_t - \bar{\epsilon}_t)'\widehat{\Omega}_\epsilon(\widehat{\epsilon}_t - \bar{\epsilon}_t)\right)^4 - 3\right)^2 \sim \chi^2_{(K)}$$

Each test is independent asymptotically and therefore can be combined into the joint test (ξ^{JB}) from Jarque and Bera (1987):

$$\xi^{JB} = \xi^S + \xi^K \sim \chi^2_{(2K)}$$

These are implemented with Pfaff (2008) package. Table 3 show that the tests reject the null hypothesis of normality, skewness and kurtosis. Normality is not necessary for consistency but if residuals are not normally distributed and the likelihood was assumed to be Gaussian, efficiency is undermined. Lanne and Saikkonen (2007) show that processes where u_t is conditionally non-Gaussian can be consistently estimated by Gaussian MLE which turns to be asymptotically Gaussian. However, this estimator may be inefficient relative to a full Gaussian MLE. Figure 3 shows that the financial variables have excess kurtosis while skewness seems to be a minor problem. Non-normality can arise from conditional heteroskedastcity.

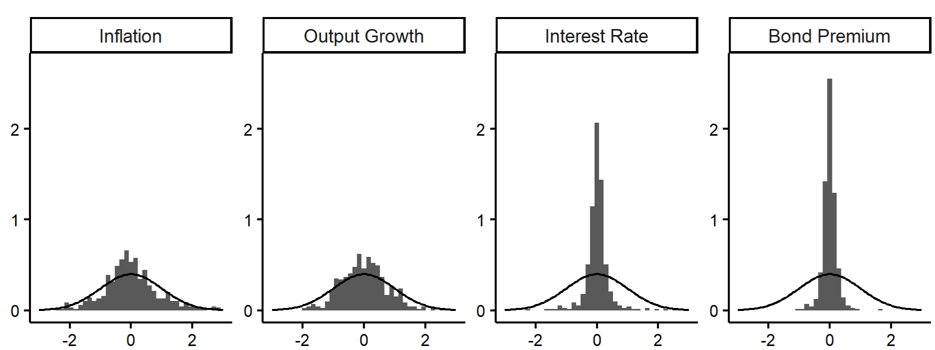


Figure 3: Histogram with residuals of unrestricted VAR(3) against Normal distribution. The histogram is re-scaled to cover an area that sums up to one.

Lastly, conditional heteroskedasticity can be tested using an the ARCH-LM test from Engle (1982). Dufour, Khalaf, and Beaulieu (2010) emphasize that univariate statistics

[14] A Normal distribution has zero skewness and kurtosis of three.

are not independent if the errors are contemporaneously correlated in a multivariate setup. A suitable multivariate test is used in Doornik and Hendry (1997). Consider the model given by:

$$\text{vech}(u_t u_t') = b_0 + \sum_{i=1}^{s} b_i \text{vech}(u_{t-i}u_{t-i}') + e_t$$

where b_0 is $(\frac{1}{2}K(K+1) \times 1)$, b_i are $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ and vech is an operator that stacks the columns from the main diagonal on downwards. This means that the 2nd order moments of the u_t are serially correlated and therefore the error term is conditionally heteroskedastic. The null hypothesis of no heteroskedasticity is given by $H_0 : b_i = 0, \ \forall i = 1, \ldots, s$. The corresponding LM statistic from Doornik and Hendry is:

$$Q_s^{ARCH} = \frac{1}{2}TK(K+1) - T\text{tr}(\widehat{\Omega}_e\widehat{\Omega}_u^{-1}) \sim \chi^2_{(sK^2(K+1)^2/4)}$$

with $\widehat{\Omega}_e$ the estimated residual covariance of e_t, and $\widehat{\Omega}_u$ is the equivalent covariance in the restricted model which coincides with the residual covariance of the VAR model. This means that residuals are not $i.i.d.$ due to the existence of conditional heteroskedasticity. This can be visually appreciated in Figure 4, where residuals are shown to have volatility clustering. This feature is specially noticeable in the bond premium but also on the other series. Despite the interest rate has an extremely high volatility during the Volker's disinflation era in the early 80s, it also displays high and low periods of volatility. Heteroskedasticity undermines the efficiency of MLE/OLS and result in incorrect inference.

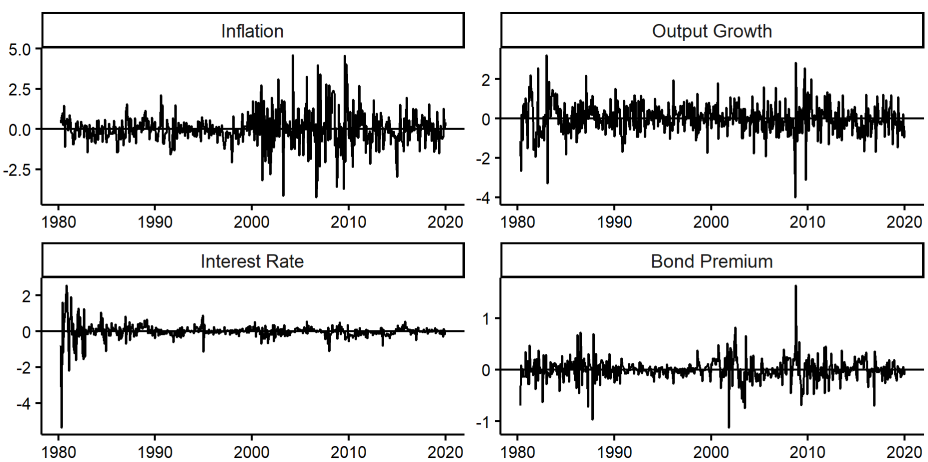


Figure 4: Residuals of unrestricted VAR(3)

These results open the door to alternative specification methods that account for non-normality or conditional heteroskedasticity in the error term and use these features for structural identification of shocks.

3.5 Theory vs Data-driven Identification

Identification requires imposing restrictions in order to recover the structural shocks from the reduced form residuals. Restrictions can be motivated by economic theory that imposes a specific structure, such as short-run (Sims 1980) or long-run exclusion restrictions (Blanchard and Quah 1989) and sign restrictions (Uhlig 2005). Instrumental variables are also used for identification as they are assumed to correlate with the shock of interest while being orthogonal to the other shocks (Gertler and Karadi 2015). Alternatively, restrictions can be motivated by statistical features of the data and make use of the 2nd order moments (heteroskedasticity) or higher order moments (independence and non-gaussianity).

In the light of the results from Section 3.4, higher order moments could be exploited for imposing additional restrictions that allow structural identification without the need of theory-based restrictions (Lanne and Lütkepohl 2010). The key assumption for higher order moments' identification is that shocks $\epsilon_{i,t}$ are mutually independent and non-Gaussian. If shocks are non-Gaussian, 3rd and 4th order moments can be used to impose further restrictions. Olea, Plagborg-Møller, and Qian (2021) argue that this implies not being able to identify non purely orthogonal shocks, ruling out the possibility of identifying a process whose structural error term is affected by a common disturbance. Specifically, they show that identification by independence would fail for the following white noise process:

$$\epsilon_{i,t} = \varkappa_t \eta_{i,t}, \ \forall i = 1, \ldots, K \tag{16}$$

where $\eta_{i,t}$ are the *i.i.d.* orthogonal structural shocks, while $\varkappa_t$ is a shared scalar volatility process that is independent but affects each of the orthogonal structural shocks $\eta_{i,t}$. This way, the VAR process Y_t is not only affected by the shocks $\eta_{i,t}$ but also by $\varkappa_t$ shocks. Traditional methods for obtaining IRFs such as those described in Section 3.1.2 would fail to estimate impulse responses to $\eta_{i,t}$ shocks but they can consistently estimate responses to shocks from the white noise processes defined in (16). ϵ_t is not *i.i.d.* but it is still interesting. Many macroeconomic processes are highly correlated and very likely jointly generated by not independent process, making independence an implausible assumption. One could question whether a macroeconomic shock can be independent of other shocks at all. It is out of the scope of this paper to address the philosophical arguments behind independence and their justification, but strengthening this assumption requires of a good reasoning. The reliance on higher order moments require the DGP to be sufficiently non-Gaussian, otherwise suffering from weak identification (Sims 2020, Olea, Plagborg-Møller, and Qian 2021). Figure 3 shows that inflation and growth residuals are relatively Gaussian and Table 3 confirms the presence of heteroskedasticity, which is at odds with independence identification. All these factors, make higher moments identification relatively hard to justify.

On the other hand, identification through heteroskedasticity does not impose strict but weak independence. It assumes that the error term of the DGP is serially uncorrelated with conditional or unconditional heteroskedasticity:

$$E(u_t) = 0 \text{ and } E(u_t u_t') = \Omega_t$$

This seems reasonable to describe the process displayed in Figure 4. Furthermore, the tests results from Table 3 support the assumptions of (conditional) heteroskedastcity and no autocorrelation. Heteroskedasticity Identification relies on weaker assumptions than Independent Components as it does not require Gaussian nor mutually independent shocks for SVAR estimation. Shocks are assumed to be conditionally orthogonal:

$$E(\epsilon_t|F_{t-1}) = 0 \text{ and } Cov(\epsilon_{i,t}, \epsilon_{j,t}|F_{t-1}) = 0, \; \forall i \neq j$$

where F_{t-1} is the information set available in $t-1$. It relies on the assumption that impulse responses are time-invariant while the volatility of shocks is not, providing with further identifying assumptions. This way if shocks have different relative changes in volatility they can be uniquely identified. This method is explained in detail in Section 4.1.1.

Following the reasoning of Olea, Plagborg-Møller, and Qian (2021) and in light of the residuals' diagnostics from Section, identification from heteroskedasticity seems to be a suitable strategy for the current analysis. Section 4 expands on this method, exploring new specification for the model.

4 Conditionally Heteroskedastic Linear VAR

The previous section presented the baseline model and discussed its limitations to later argue for explicitly modelling heteroskedasticity as a way to correct for residuals' misspecification and to impose identifying restrictions based on the data instead of the theory. This section elaborates on the conditional heteroskedastic VAR where the variance is modelled as a smooth transition between two volatility regimes. Section 4.1 presents the ST-VAR model. Specifically, section 4.1.1 explores the model in reduced form, the reasons for modelling conditional heteroskedasticity and the logic behind Heteroskedasticity Identification (HI). Then, section 4.1.2 presents the estimation algorithm: an iterative Feasible Weighted Least Squares method. For a given data set, section 4.2 finds the optimal transition function. Section 4.3 explores an alternative specification for conditional heteroskedasticity: the MS-VAR, and discusses the benefits of the ST-VAR over the MS-VAR model for the research. Section 4.4 explores the estimated covariances under HI, tests for full identification and shows the impulse responses. Lastly, section 4.5 estimates the model under both Cholesky and HI, tests for the suitability of Cholesky restrictions and shows the impulse responses with over-identifying restrictions. The estimation and identification scheme presented in this Section relies on the R package **svars** presented in A. Lange et al. (2021).

4.1 Linear VAR with Smooth Transition in Covariances (ST-VAR)

4.1.1 ST-VAR and Heteroskedasticity Identification

Consider the linear model in (1) where the error term is now conditionally heteroskedastic and modelled with a smooth transition in covariances (ST-VAR) like the model from Lütkepohl and Netšunajev (2017b):

$$Y_t = v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t = \mathbf{A} X_{t-1} + u_t \tag{17}$$

$$u_t \sim N(0, \Omega_t) \tag{18}$$

$$\Omega_t = (1 - F(s_t, \gamma, c))\Omega^L + F(s_t, \gamma, c)\Omega^H \tag{19}$$

$$F(s_t, \gamma, c) = \frac{1}{1 + \exp(-\exp(\gamma)(s_t - c))} \tag{20}$$

where both the intercept v and the coefficient matrices A_i are linear and independent of the volatility state while shocks can be of different sizes conditional on the expected volatility regime $j = \{L, H\}$. The expected time-dependent covariance fluctuates between two volatility regimes weighted by the bounded logistic transition function $0 < F(s_t, \gamma, c) < 1$. This function depends on two parameters: the speed of adjustment γ and the intercept of transition c. For a large enough speed of adjustment, the transition is abrupt as most of the observations would be in either of the two regimes[15]. The transition variable s_t has been normalized such that: $s_t \in [0,1]$. The probability of being in regime H is given by: $P(j = H) = F(s_t, \gamma, c)$, and the probability of being in regime L: $P(j = L) = (1 - F(s_t, \gamma, c))$.

Including time-dependence in the covariance of the model has multiple advantages over assuming homoskedastic residuals:

- If there is heteroskedasticity, OLS estimation of the reduced form parameters is inefficient as the standard errors are biased. This is problematic for statistical inference as the construction of confidence intervals in hypothesis testing, impulse responses and other statistics is based on incorrect assumptions, potentially leading to wrong conclusions. Data from high volatility periods is weighted equally than data from low volatility periods and therefore has a larger impact on the minimization of the unweighted criterion function. If OLS is inefficient then then there are more suitable estimators such as the Generalized Least Squares or the Weighted Least Squares (Appendix A.4), which minimizes a weighted criterion function that differentiates between different volatility periods. Using OLS residuals largely underestimates the standard errors, creating the potential for incorrect inference. This paper is mostly interested in hypothesis testing and impulse responses, so it becomes crucial to account for heteroskedasticity.

- Periods of different variance can be explicitly modelled and used to impose restrictions that allow to identify the structural shocks without the need of theory driven restrictions. As the cloud of observed data contains the simultaneous realization of multiple shocks OLS is biased in systems of simultaneous equations (Hamilton 1994) and restrictions are needed to identify the model and recover the underlying structural parameters. Rigobon (2003) explains that heteroskedasticity can be used similar to the instrumental variables identification approach, which makes use of an 'exogenous' variable that measures a shift in one of the underlying system's curves without affecting the other curve. This allows to measure the slope of the curve that was unaffected by the shock. Equivalently, we can split the cloud of realizations into two sub samples where one of them contains more volatile shocks to one variable. This implies enlarging the cloud of observations around the other curve. A requirement for this would be that the relative change in variances is different to have different cloud ellipses. This way, heteroskedasticity can be used for structural identification. Accurately estimating the 2nd order moment structure is crucial when it is being used for identification of structural parameters (Bruder 2018).

Assuming time-invariant coefficients, Heteroskedasticity Identification (HI) is possible

[15] For large enough γ, the ST-VAR behaves closer to a Threshold VAR. This means that there are two differentiated regimes, however it does not benefit from the larger sample that the ST-VAR uses. The latter uses every observation weighted by the transition function for the estimation of the covariances. This implies a larger data set used for estimation than that of a Threshold VAR.

if the error covariances can split into two volatility regimes:

$$\Omega^L = BB' \quad \text{and} \quad \Omega^H = B\Lambda B' \tag{21}$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ for which λ_k is the variance of the structural shock ϵ_k on the high volatility regime ($j = H$) relative to the variance of that shock on the low volatility regime ($j = L$), where the variance has been normalized to one. This means that the structural shocks have variance given by:

$$E[\epsilon_t \epsilon_t'] = \begin{cases} I_K & \text{with probability} \quad P(j = L) = 1 - F(s_t, \gamma.c) \\ \Lambda & \text{with probability} \quad P(j = H) = F(s_t, \gamma.c) \end{cases}$$

As both I_K and Λ are diagonal matrices, structural shocks are instantaneously uncorrelated by construction. Lanne, Lütkepohl, and Maciejowska (2010) show that B is unique apart from changes in signs and ordering of the shocks if $\lambda_i \neq \lambda_j, \forall i \neq j$. This condition implies that the structural model is fully identified by means of HI. Then, by $\epsilon_t = B^{-1}u_t$ the structural shocks can be recovered from the residuals assuming that the same matrix of contemporaneous relationships B applies for the whole sample periods, such that the impact of the shocks is time-invariant and only their volatility fluctuates over time.

An example on how HI works is shown by Rigobon (2003) and Lütkepohl (2012)) with a two-dimensional VAR model. Consider the following two-dimensional error term to illustrate the example:

$$u_t = B\epsilon_t = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

where the two regimes' covariances are:

$$\Omega^L = BB' = \begin{bmatrix} B_{11}^2 + B_{12}^2 & B_{11}B_{21} + B_{12}B_{22} \\ B_{11}B_{21} + B_{12}B_{22} & B_{21}^2 + B_{22}^2 \end{bmatrix}$$

$$\Omega^H = B\Lambda B' = \begin{bmatrix} \lambda_1 B_{11}^2 + \lambda_2 B_{12}^2 & \lambda_1 B_{11}B_{21} + \lambda_2 B_{12}B_{22} \\ \lambda_1 B_{11}B_{21} + \lambda_2 B_{12}B_{22} & \lambda_1 B_{21}^2 + \lambda_2 B_{22}^2 \end{bmatrix}$$

There are six equations with six structural parameters to estimate $\{B_{11}, B_{12}, B_{21}, B_{22}, \lambda_1, \lambda_2\}$. The solution is unique if λ_1 and λ_2 are distinct and ordered some way. For this reason, shocks might need re-ordering or changes in sign (under the assumption that positive and negative shocks are symmetric) after being estimated. In the actual system of equations that this paper uses, the matrices (Ω^L, Ω^H) are symmetric with $K(K+1)/2$ distinct elements while B has K^2 elements and Λ has K elements. Using the data set from Section 3.2, $K = 4$. There is no need to impose a lower triangular structure in B. The assumption of common contemporaneous relationships across regimes already provides with the necessary restrictions for a unique solution. Similar to a Cholesky decomposition, these restrictions do not necessarily result in economically meaningful shocks and only result in purely orthogonal shocks by construction. Further theoretical arguments are needed to interpret them in an economically meaningful way.

4.1.2 ST-VAR Estimation

The probability density function of the model in (18), (19) and (20) is given by the Likelihood:

$$f(Y|\theta) = \mathcal{L}_T(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^K |\Omega_t|}} \exp\left(-\frac{1}{2} u_t' \Omega_t^{-1} u_t\right) \tag{22}$$

where Y is the available dataset $Y = [Y_t, Y_{t-1}, \ldots, Y_0]$ and $\theta = [\mathbf{A}, \Omega_t] \in \Theta$ the vector of unknown parameters. The model is nonlinear and has no closed form solution so estimation is done with an algorithm that obtains $\widehat{\Omega}_t$ with Maximum Likelihood (MLE) and $\widehat{\mathbf{A}}$ with Feasible Weighted Least Squares (WLS), following Lütkepohl and Netšunajev (2017b) and A. Lange et al. (2021). The reason for this estimation method is that OLS fitting treats outliers equally regardless of whether they appear in regions of high or low variance. Under heteroskedasticity, WLS takes into account the volatility of data in each region in order to estimate the coefficients. An outlier would change the slope of the parameters significantly only if it is in a low volatility region but not that much when it is in a high volatility region.

The estimation algorithm is as follows:

- **Step 1:** Generate starting values using the homoskedastic linear VAR estimates: $\widehat{B} = \text{chol}(\widehat{\Omega}_u^{OLS})$ and $\widehat{\Lambda} = I_K$ using the OLS estimates of the residuals: $\widehat{u}_t = \widehat{u}_t^{OLS} = Y_t - \widehat{\mathbf{A}}^{OLS} X_{t-1}$.
- **Step 2:** Obtain the MLE: $(\widehat{B}^{MLE}, \widehat{\Lambda}^{MLE})$ using a Newton-type algorithm for optimization that minimizes the negative log-Likelihood function $(-\log \mathcal{L}_T^*)$ derived from (22). This is equivalent to maximizing: $\log \mathcal{L}_T^*$[16] :

$$\arg\max_{B,\Lambda} \log \mathcal{L}_T^* = \frac{TK}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widehat{u}_t'\Omega_t^{-1}\widehat{u}_t$$
$$\text{where } \Omega_t = (1 - F(s_t, \gamma, c))BB' + F(s_t, \gamma, c)B\Lambda B'$$

The optimization is done for a given $F(s_t, \gamma, c)$ and restriction matrix if applicable (not the case now). The starting values are the estimates: $B^{(0)} = \widehat{B}$ and $\Lambda^{(0)} = \widehat{\Lambda}$ from the previous step.

[16]Note that the actual log-Likelihood would be:

$$\log \mathcal{L}_T = -\frac{TK}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\Omega_u| - \frac{1}{2}\sum_{t=1}^{T}\widehat{u}_t'\Omega_t^{-1}\widehat{u}_t$$

such that the sign of the constant is changed. This paper uses a proportional likelihood with a positive constant. This is the criterion used in Lütkepohl and Netšunajev (2017b) and A. Lange et al. (2021) and I stick to their choice to allow comparability.

- **Step 3:** Construct the Feasible WLS estimates using:

$$\widehat{\Omega}_t = (1 - F(s_t, \gamma, c))\widehat{B}\widehat{B}' + F(s_t, \gamma, c)\widehat{B}\widehat{\Lambda}\widehat{B}'$$

$$\text{vec}\left(\widehat{\mathbf{A}}^{WLS}\right) = \sum_{t=1}^{T}\left(X_{t-1}X_{t-1}' \otimes \widehat{\Omega}_t^{-1}\right)^{-1}\left(\left(X_{t-1} \otimes \widehat{\Omega}_t^{-1}\right)Y_t\right)$$

- **Step 4:** Generate new (homoskedastic) residuals: $\widehat{u}_t = Y_t - \widehat{\mathbf{A}}^{WLS}X_{t-1}$ and iterate until convergence[17] over steps 2-5 using the last MLE estimates as starting values for the optimization in **Step 2**: $B^{(0)} = \widehat{B}^{MLE}$ and $\Lambda^{(0)} = \widehat{\Lambda}^{MLE}$. Take the last estimates of $(\widehat{\mathbf{A}}^{WLS}, \widehat{B}^{MLE}, \widehat{\Lambda}^{MLE})$ as final values.

4.2 Optimal Smooth Transition Function

For a given exogenous s_t, the optimal $F(s_t, \gamma, c)$ can be found by applying a grid search over (γ, c) to the algorithm described above. These parameters are searched within the intervals: $\gamma \in [0, 10]$ and $c \in [0, 1]$.[18] For each pair of values the model is re-estimated and the final parameters are chosen to be those that maximize the log-Likelihood. s_t is chosen such that the system is fully identified. A feature of the bond premium is that whenever it displays high values it also has higher volatility. This feature supports the logic of the smooth transition, so that as long as the risk premium is high, so is the probability of being in a high volatility regime. The high correlation of bond premium with systemic risk implies higher volatility of other macro variables when $P(j = H)$ is high (Gilchrist and Zakrajšek 2011, Galvão and Owyang 2018). s_t is set to be a two month lag of the normalized excess bond premium where the lag length is chosen such that the log-Likelihood is maximized. Table 5 shows the log-likelihood values for each lag length[19] and its corresponding parameters. Figure 5 shows the transition variable s_t and the corresponding transition function $F(s_t, \gamma = 3.5, c = 0.3)$ over time.

Other papers use systemic risk to characterize regimes of financial stress (Galvão and Owyang 2018) or various indicators of recessions and expansions like the GDP gap. Sometimes a moving average transformation is used to smooth the otherwise volatile

[17] The algorithm iterates over these steps until the log-Likelihood changes by less than 0.001 or the maximum number of iterations is reached. I use a maximum of 50 iterations because the model is computationally demanding. This becomes more important when applying the bootstrap algorithm. Nevertheless, this maximum is enough to ensure convergence, as results are unaffected when increasing this ceiling.

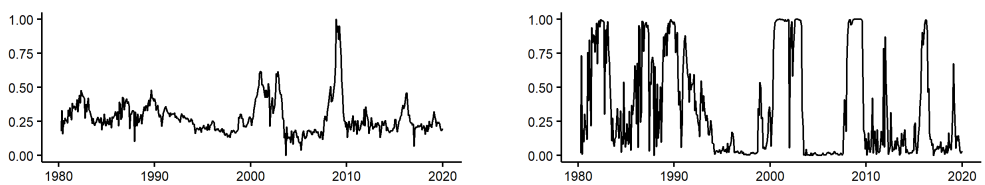
[18] Negative values of the slope $\gamma < 0$ would simply switch the interpretation of the regimes. The interval considered for the slope is wider than the interval considered by Lütkepohl and Netšunajev (2017a). Moreover, the interval for c covers every possible break point, allowing for the existence of only one regime if $c = 0$ or $c = 1$. The grid steps used for γ are of size 0.5 while for c the steps are of size 0.1.

[19] A lag of zero is not considered to avoid contemporaneous endogeneity of the regime. This is done in Auerbach and Gorodnichenko (2012) to avoid that policy changes affect contemporaneously the state in which the economy is.

transition between regimes (Chak and Chiu 2016). All of these approaches offer interesting alternatives that allow clear interpretations of results. However, for these transition variables to work in this context they must fully identify the system. Only the case case of a systemic stress index has been tried for this research, but it failed to fully identify the current system of equations. For this reason, I continue with the bond premium.

variable	γ	c	$\log \mathcal{L}_T^*$
s_{t-1}	3.0	0.5	-1498
$\mathbf{s_{t-2}}$	**3.5**	**0.3**	**-1438**
s_{t-3}	2.5	0.5	-1493

Table 5: Choice of lag length in s_t and (γ, c) based on the maximum log-Likelihood. Each combination of (γ, c) maximizes the log-Likelihood for a given transition variable.



(a) Transition variable (s_t): normalized version of the two period lag of the Excess Bond Premium

(b) Optimal smooth transition function: $F(s_t, \gamma = 3.5, c = 0.3)$ as given by (20). Probability of being in regime $j = H$.

Figure 5: ST-VAR(3) regime transition

4.3 Comparison with Markov Switching VAR

Another candidate to model conditionally heteroskedastic regimes that imposes less predetermined structure is a Markov Switching process. Let's now consider VAR with Markov Switching between two covariance regimes (MS(2)-VAR):

$$Y_t = v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t = \mathbf{A} X_{t-1} + u_t \tag{23}$$

$$u_t | \vartheta_t \sim N(0, \Omega(\vartheta_t)) \tag{24}$$

$$\Omega(\vartheta_t) = \begin{cases} \Omega^L & \text{with probability} \quad \mathrm{P}(\vartheta_t = L) \\ \Omega^H & \text{with probability} \quad \mathrm{P}(\vartheta_t = H) \end{cases} \tag{25}$$

where the distribution of the error term depends on a discrete unobserved random process ϑ_t that follows a Markov Chain with two regimes $\{L, H\}$. This way, the error is normal conditional on the state of the Markov process. The probability of moving from regime i to regime j is given by $p_{j|i}$ which only depends on the state of the previous period:

$$p_{j|i} = \mathrm{P}(\vartheta_t = j | \vartheta_{t-1} = i, \vartheta_{t-2}, \ldots, \vartheta_0) = \mathrm{P}(\vartheta_t = j | \vartheta_{t-1} = i), \; i, j \in \{L, H\} \tag{26}$$

This way, there is no gain from knowing the full history of states. The system must be in either of the two states, such that:

$$p_{L|L} + p_{H|L} = 1 \quad \text{and} \quad p_{L|H} + p_{H|H} = 1$$

The matrix of conditional probabilities is then given by:

$$\mathbf{P} = \begin{bmatrix} \mathrm{P}(\vartheta_t = L|\vartheta_{t-1} = L) & \mathrm{P}(\vartheta_t = L|\vartheta_{t-1} = H) \\ \mathrm{P}(\vartheta_t = H|\vartheta_{t-1} = L) & \mathrm{P}(\vartheta_t = H|\vartheta_{t-1} = H) \end{bmatrix} = \begin{bmatrix} p_{L|L} & p_{L|H} \\ p_{H|L} & p_{H|H} \end{bmatrix} = \begin{bmatrix} p_{L|L} & 1 - p_{H|H} \\ 1 - p_{L|L} & p_{H|H} \end{bmatrix}$$

The Markov Chain is weakly mixing if $\mathbf{P}$ satisfies:

$$\text{(i) } p_{L|L},\ p_{H|H} < 1 \quad \text{and} \quad \text{(ii) } p_{L|L} + p_{H|H} > 0$$

Condition (i) implies that the system is irreducible to a one regime system. This would happen if the probability of staying in one regime is one or equivalently, if the probability of switching from one regime is zero. Condition (ii) implies that the irreducible system is aperiodic and therefore the probability of switching is smaller than one. Under these conditions, the Markov Chain switches will occur but are not predictable.

For some given starting values, the unconditional density is given by:

$$\begin{aligned} f(Y, \vartheta_t, \theta) = &\prod_{t=1}^{T} f(Y_t|\vartheta_t = L, \theta) f(\vartheta_t = L|\vartheta_{t-1}, Y_{t-1}, \ldots, \vartheta_0, Y_0) \\ &\times f(Y_t|\vartheta_t = H, \theta) f(\vartheta_t = H|\vartheta_{t-1}, Y_{t-1}, \ldots, \vartheta_0, Y_0) \\ = &f(Y_0|\vartheta_0 = L)\mathrm{P}(\vartheta_0 = L) \times f(Y_0|\vartheta_0 = H)\mathrm{P}(\vartheta_0 = H) \\ &\times \prod_{t=1}^{T} f(Y_t|\vartheta_t = L, \theta)\mathrm{P}(\vartheta_t = L) \times f(Y_t|\vartheta_t = H, \theta)\mathrm{P}(\vartheta_t = H) \end{aligned}$$

where the unconditional probabilities[20] are:

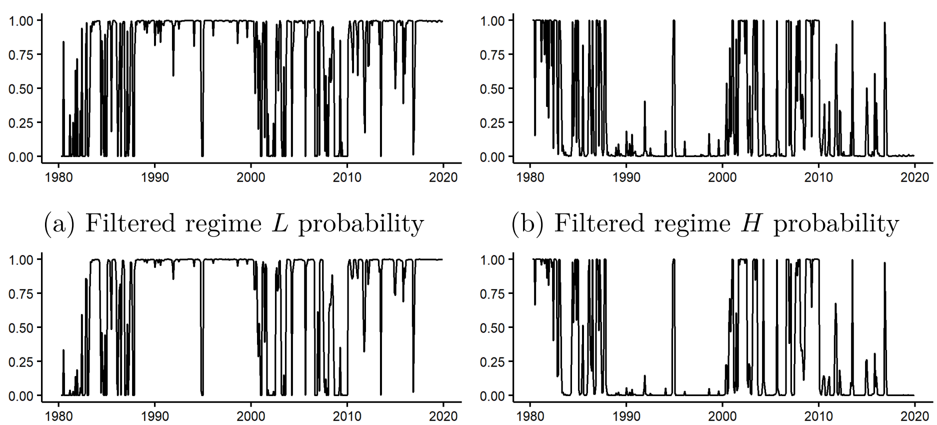
$$\begin{aligned} \begin{bmatrix} \mathrm{P}(\vartheta_t = L) \\ \mathrm{P}(\vartheta_t = H) \end{bmatrix} &= \begin{bmatrix} \mathrm{P}(\vartheta_t = L|\vartheta_{t-1} = L)\mathrm{P}(\vartheta_{t-1} = L) + \mathrm{P}(\vartheta_t = L|\vartheta_{t-1} = H)\mathrm{P}(\vartheta_{t-1} = H) \\ \mathrm{P}(\vartheta_t = H|\vartheta_{t-1} = L)\mathrm{P}(\vartheta_{t-1} = L) + \mathrm{P}(\vartheta_t = H|\vartheta_{t-1} = H)\mathrm{P}(\vartheta_{t-1} = H) \end{bmatrix} \\ &= \mathbf{P} \begin{bmatrix} \mathrm{P}(\vartheta_{t-1} = L) \\ \mathrm{P}(\vartheta_{t-1} = H) \end{bmatrix} \\ &= \begin{bmatrix} (2 - p_{L|L} - p_{H|H})^{-1}(1 - p_{H|H}) \\ (2 - p_{L|L} - p_{H|H})^{-1}(1 - p_{L|L}) \end{bmatrix} \end{aligned}$$

The Likelihood contribution is then given by:

$$f(Y_t|\vartheta_t, \theta) = \frac{1}{\sqrt{(2\pi)^K |\Omega(\vartheta_t)|}} \exp\left(-\frac{1}{2} u_t' \Omega(\vartheta_t)^{-1} u_t\right)$$

[20] See Hamilton (1994) for the derivation of this result.

where the vector of parameters to estimate is: $\theta = \{\mathbf{A}, \Omega^H, \Omega^L, p_{H|H}, p_{L|L}\}$. Because the switching process ϑ_t is unobserved, it is difficult to evaluate the likelihood for the observed values. For that reason estimation must be carried out by an Expectation-Maximization algorithm (EM) where the MLE is obtained by iteratively maximizing the expectation of the log-Likelihood (see Hamilton 1994, T. Lange and Rahbek 2009, Doornik 2013). The optimization must include the constraints for irreducibility and aperiodicity. This paper only uses this model for comparing the historical transition of both the filtered and smoothed transition probabilities to the transition function of the ST-VAR model presented before. The filtered probabilities are estimates of the probabilities at period t using all available data up to t, whereas the smoothed probabilities are estimated for each period t using all the data in the sample.



(a) Filtered regime L probability (b) Filtered regime H probability

(c) Smoothed regime L probability (d) Smoothed regime H probability

Figure 6: MS(2)-VAR(3) transition probability for low (L) and high (H) volatility regimes.

Figure 6 displays the regimes' transitions as given by the MS(2)-VAR(3) filtered and smoothed probabilities while Figure 5 showed the optimal transition function from the ST-VAR(3) in the previous section. The MS-VAR transition is too abrupt and the model remains for very little time in each regime. Moreover, the reason for the switches is solely based on a constant probability evaluated at each point. These probability estimates are also influenced for every variable's volatility change, and not only the bond spread, meaning that inflation during the 80s is affecting the regimes significantly. The regimes displayed in Figure 6 seem to have unconnected to the periods of financial stress that I aim to capture. The ST-VAR seems to better capture periods of high volatility in bond premium, with differentiated regimes that persist over time. This research focuses on financial shocks, so the predetermined structure of the ST-VAR better captures the desired regimes while the MS(2)-VAR fails given the chosen variables and sample. For this reason, the paper continues with the ST-VAR specification.

Despite the model choice it should be highlighted that misspecification in the form of the regimes does not affect consistency if data displays heteroskedasticity, given that the rank condition is satisfied (Rigobon 2003, Sims 2020).

4.4 Heteroskedasticity Identification Testing and Inference

Consider a 4-dimensional ST-VAR(p=3) with the same endogenous variables as before (see Section 3.2). The existence of different covariance matrices allows to recover the structural shocks that are more likely to take place in high volatility regimes using Heteroskedasticity Identification (HI). The identification condition was defined in equation (21). This method works if the error term is either conditionally or unconditionally heteroskedastic. Using the estimation algorithm described in section 4.1.2, I estimate the matrices B (Table 6c) and Λ (Table 6d) that define the covariance matrices (Tables 6a and 6b). The covariances Ω^H, Ω^L are not just differently scaled matrices, as the HI scheme does not impose a lower triangular structure on B. The estimated Λ is clearly that of a high volatility regime as every $\lambda_i > 1$. This can also be seen in Ω^H.

	π_t	x_t	i_t	bp_t
π_t	0.98	-0.10	0.02	-0.01
x_t	-0.10	0.45	0.00	-0.01
i_t	0.02	0.00	0.02	0.00
bp_t	-0.01	-0.01	0.00	0.02

(a) $\widehat{\Omega}^L = \widehat{B}\widehat{B}'$

	π_t	x_t	i_t	bp_t
π_t	1.56	0.01	-0.08	-0.05
x_t	0.01	1.14	0.36	0.01
i_t	-0.08	0.36	0.79	0.02
bp_t	-0.05	0.01	0.02	0.10

(b) $\widehat{\Omega}^H = \widehat{B}\widehat{\Lambda}\widehat{B}'$

	π_t	x_t	i_t	bp_t
π_t	0.91***	0.38**	-0.02	-0.05
	(0.10)	(0.18)	(0.02)	(0.09)
x_t	-0.35**	0.57***	0.07***	0.03
	(0.14)	(0.07)	(0.02)	(0.10)
i_t	0.02	0.01	0.15***	-0.03
	(0.02)	(0.02)	(0.01)	(0.02)
bp_t	0.01	-0.02	0.01	0.15***
	(0.02)	(0.03)	(0.01)	(0.01)

(c) $\widehat{B}$

	λ_i
π_t	1.42***
	(0.22)
x_t	2.41***
	(0.39)
i_t	36.19***
	(9.07)
bp_t	4.40***
	(0.69)

(d) diag($\widehat{\Lambda}$)

Significance assessed with a two-sided t-test: $t = (\widehat{\beta}_i - 0)/\widehat{\Sigma}_{\beta_i} \sim N(0,1)$. *Critical values (confidence level): 1.65 (90%*), 1.96 (95%**), 2.57 (99%***).*

Table 6: ST-VAR(3) estimates under Heteroskedasticity Identification after reordering of shocks

HI does not necessarily obtain economically meaningful shocks. After estimation, the shocks have to be re-ordered by visual inspection[21] in order to label them as each vari-

[21] In this case the order of the 3rd and 4th shocks was swapped and so was the sign of the responses to a financial shock. Recall that impulse responses are assumed to be symmetric for positive and negative shocks. These changes affect B, Λ and their respective standard errors.

able's specific shock. Labelling shocks obtained by means of HI is only possible when combined with a priori exogenous information. In this case, each variable's shock is assumed to display the highest variability on impact from that variable's responses. This choice results in comparable contemporaneous effects (B) and visually similar impulse responses to those recovered from a Cholesky decomposition. Table 6c shows that the diagonal elements display the highest values of each column, such that the effects of each shock (columns) has the largest value for the corresponding variable (rows) of the labelled shock[22]. Among the significant parameters, the signs of the contemporaneous correlations are in line with intuition for every variable but the interest rate. An inflationary (π_t) shock has a positive contemporaneous effect on inflation and negative effect on growth. A demand shock (x_t) affects positively inflation and output on impact. Surprisingly, the interest rate (i_t) shock is contemporaneously positively correlated with output. Lastly, the bond premium (bp_t) shock only affects the bond premium on impact, implying that every other variable reacts with a lag, in line with the Cholesky restrictions imposed for structural identification of financial shocks. All the other shocks have different contemporaneous relationships than those assumed on the Cholesky ordering. Nevertheless, the suitability of Cholesky restrictions will be formally tested afterwards.[23].

	Test statistic	p-value
$\lambda_1 = \lambda_2$	4.98**	0.03
$\lambda_1 = \lambda_3$	14.69***	0.00
$\lambda_1 = \lambda_4$	16.88***	0.00
$\lambda_2 = \lambda_3$	13.85***	0.00
$\lambda_2 = \lambda_4$	6.36**	0.01
$\lambda_3 = \lambda_4$	12.21***	0.00

Table 7: Heteroskedasticity Identification Wald test after re-ordering of shocks

The system is fully identified if $\lambda_i \neq \lambda_j, \forall i \neq j$. The null hypothesis of proportional variance shifts $H_0 : \lambda_i = \lambda_j, \forall i \neq j$ can be tested against a χ^2 distribution with 1 degree of freedom using the following Wald test statistic:

$$W^\lambda = \frac{(\lambda_i - \lambda_j)^2}{Var(\lambda_i) + Var(\lambda_j) - 2Cov(\lambda_i, \lambda_j)} \sim \chi^2_{(1)}$$

[22]The contemporaneous effect of a shock labelled as bond premium (f_t) shock is the highest value that the (f_t) displays on impact, and the same applies for the other variables. This is expected to be the case considering variables have cross-correlations smaller than their standard deviation.

[23]The non statistically significant contemporaneous correlations are mostly in line with narrative that introduced this paper. If these effects were significant: the inflationary shock would trigger a positive response of the policy rate and the risk premium, the demand shock (x_t) would increase interest rates and decrease the risk premium, the interest rate would reduce inflation but increases the risk premium, and lastly the financial shock would decrease inflation and the interest rate but increase output on impact.

Table 7 shows the results of the Wald tests. These support full identification of the system. Alternatively, other tests have been developed, as Lütkepohl and Netšunajev (2017b) argue that the system is unidentified under the null hypothesis, and therefore the statistic has an unknown limiting distribution. Despite this issue, the Wald test has continued to be used in research (Herwartz and Plödt 2016).

Once the underlying structural shocks are identified, the Impulse Responses can be obtained as in the linear case. The confidence bands are simulated using a Wild bootstrap with fixed design and 2000 draws. These are then re-centered using Hall's percentile interval for a 90% confidence interval, as in the linear case. The bootstrap algorithm used for the ST-VAR is explained in Appendix A.7. Figure 7 shows the IRFs for HI financial shocks.

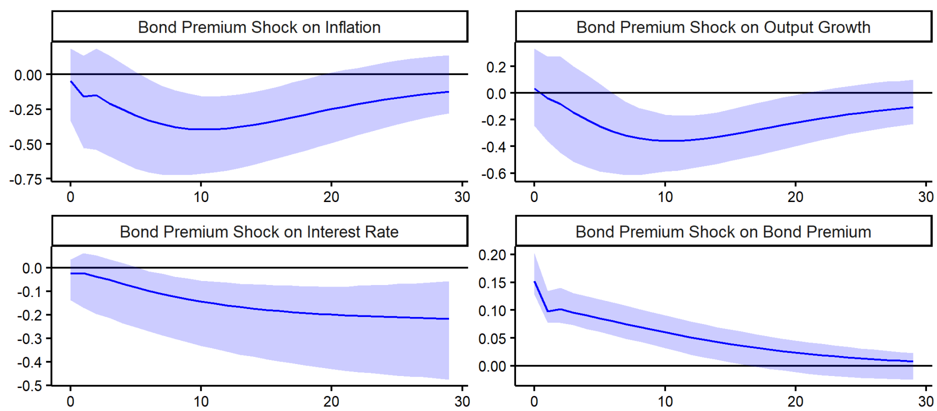


Figure 7: Impulse responses to a bond premium shock identified with heteroskedasticity. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws. The full set of impulse responses is displayed in Figure 13.

An orthogonal shock to the bond premium identified with heteroskedasticity triggers a fall in inflation and growth that converge back to zero after 20-30 months due to the falling interest rate in line with the Federal Reserve mandate. These responses are very similar to those obtained for the financial shock identified with exclusion Cholesky restrictions from Figure 2. However, now the interest rate is found to have a negative and significant response that persists over long horizons, despite the recovery in production and the price level. The responses to the other shocks can be found in Appendix C.

4.5 Testing Over-identifying Restrictions

The system is now identified by combining Cholesky and HI. Tables 8a and 8b show the new covariances. These are more alike than those under the just HI scheme (Tables 6a and 6b). The reason is that now the matrix B is forced to be lower triangular, reducing the difference mostly to the scale.

Tables 8c and 8d show the corresponding estimates for $\widehat{B}$ and $\widehat{\Lambda}$. While $\widehat{\Lambda}$ matrix is similar to the one estimated using only HI, $\widehat{B}$ differs significantly. Some previously significant estimates are now restricted to be zero while other insignificant estimates are now significant. Specifically, the difference between $\widehat{B}$ in the model with Cholesky restrictions (Table 8c) and without (Table 6c) is given by the restrictions of the significant contemporaneous effect from output shocks on inflation and from interest rates on output. If these effects are significant, then monetary policy effects on production are immediate, and prices are flexible within a one month period for demand shocks. The latter could be explained by predictability of demand shocks and adjustments in inflation expectations, which are not explicitly modelled. However, it can also be argued that these estimates are not reflecting the interpreted shock. In the case of inflation, a contemporaneous reaction to demand shocks might capture the above-mentioned effect of inflationary expectations. If Cholesky restrictions were justified, parameter estimates should be more accurately estimated now.

	π_t	x_t	i_t	bp_t
π_t	0.97	-0.06	0.01	-0.02
x_t	-0.06	0.45	0.03	-0.01
i_t	0.01	0.03	0.03	0.00
bp_t	-0.02	-0.01	0.00	0.02

(a) $\widehat{\Omega}^L = \widehat{B}\widehat{B}'$

	π_t	x_t	i_t	bp_t
π_t	1.57	-0.10	0.02	-0.03
x_t	-0.10	1.12	0.08	-0.01
i_t	0.02	0.08	0.74	0.01
bp_t	-0.03	-0.01	0.01	0.10

(b) $\widehat{\Omega}^H = \widehat{B}\widehat{\Lambda}\widehat{B}'$

	π_t	x_t	i_t	bp_t
π_t	0.99***	0	0	0
	(0.04)			
x_t	-0.06***	0.67***	0	0
	(0.03)	(0.03)		
i_t	0.01	0.05***	0.16***	0
	(0.01)	(0.01)	(0.01)	
bp_t	-0.02**	-0.01	0.00	0.15***
	(0.01)	(0.01)	(0.00)	(0.01)

(c) $\widehat{B}$

	λ_i
π_t	1.62***
	(0.24)
x_t	2.48***
	(0.42)
i_t	29.47***
	(7.07)
bp_t	4.41***
	(0.70)

(d) diag($\widehat{\Lambda}$)

Significance assessed with a two-sided t-test: $t = (\widehat{\beta}_i - 0)/\widehat{\Sigma}_{\beta_i} \sim N(0,1)$. *Critical values (confidence level): 1.65 (90%*), 1.96 (95%**), 2.57 (99%***).*

Table 8: ST-VAR(3) estimates under Heteroskedasticity + Cholesky Identification

Table 9 shows the Wald test for the new estimates of Λ, pointing towards full identification by heteroskedasticity. In the case when HI just-identifies the system any theory-based restriction imposed can be tested against the data.

Let B_r denote the restricted matrix of contemporaneous effects that combine Cholesky

	Test statistic	p-value
$\lambda_1 = \lambda_2$	3.15*	0.08
$\lambda_1 = \lambda_3$	15.49***	0.00
$\lambda_1 = \lambda_4$	14.50***	0.00
$\lambda_2 = \lambda_3$	14.52***	0.00
$\lambda_2 = \lambda_4$	5.64**	0.02
$\lambda_3 = \lambda_4$	12.43***	0.00

Table 9: Heteroskedasticity Identification Wald of Cholesky restricted ST-VAR(3)

and HI (matrix 8c). Using a Likelihood Ratio (LR) we can test the null hypothesis of the likelihood of the restricted model being equal to that of the unrestricted model $H_0 : \mathcal{L}(\widehat{B}_r) = \mathcal{L}(\widehat{B})$. This can be tested with the following LR statistic:

$$LR = 2\left(\log \mathcal{L}^*(\text{vec}(\widehat{B})) - \log \mathcal{L}^*(\text{vec}(\widehat{B}_r))\right) \sim \chi^2_{(q)} \tag{27}$$

where $q = 6$ is the number of restrictions. Table 10 shows that the null is rejected. Therefore, Cholesky restrictions are not supported by the data. However, from the significance of the estimated parameters in the covariances from Section 4.4 I argued in favor of Cholesky restrictions for the financial shock identification. Consider now a B matrix from HI with over-identifying restrictions on the financial shock only:

$$B_r^F := \begin{bmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & * \end{bmatrix} \tag{28}$$

where $*$ refer to unrestricted entries while 0 entries are exclusion restrictions. By repeating the same LR test from (27) using the new restriction matrix B_r^F, we can test the suitability of Cholesky over-identifying restrictions for the financial shock. The second test in Table 10 shows that the null hypothesis of equal likelihood cannot be rejected, providing support for these restrictions in the identification of financial shocks.

H_0	Test statistic	p-value
$\mathcal{L}(B) = \mathcal{L}(B_r)$	18.154**	0.006
$\mathcal{L}(B) = \mathcal{L}(B_r^F)$	3.386	0.496

Table 10: Likelihood Ratio tests on over-identifying restrictions in HI ST-VAR(3). Additional to the heteroskedasticity restrictions, B_r contains Cholesky restrictions while B_r^F contains exclusion restrictions for the contemporaneous effects of a bond premium shock to each variable but the bond premium.

The Impulse Responses of the over-identified system are shown in Figure 8, which displays the case of over-identifying Cholesky restrictions which are supported only for the financial shock. The confidence bands are built using the Hall's percentile interval over the set of bootstrapped impulse responses in a Wild bootstrap algorithm with fixed design and 2000 iterations. The responses to a bond premium shock are still very similar across models, but this specification displays the lowest standard errors among the considered specifications and identification methods. A bond premium shock has a significant impact on every variable and triggers a decrease in production, inflation and interest rates. The decrease in interest rates occurs with a lag but persists over time. It takes approximately 20 months for the price level, production and the bond premium to converge back to their initial levels.

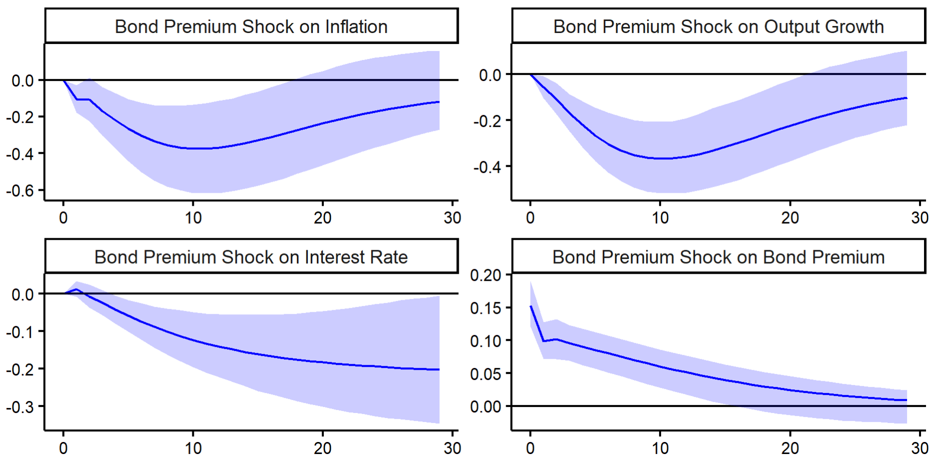


Figure 8: Impulse responses to a bond premium shock identified with heteroskedasticity and Cholesky over-identifying restrictions. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws. The full set of impulse responses is displayed in Figure 14.

5 Conditionally Heteroskedastic Non-linear VAR

The previous section explored the VAR model with smooth transition between volatility regimes. Using over-identifying restrictions derived from the conditional heteroskedasticity, the Cholesky identification scheme was found to be overall inconsistent with the data. However, Cholesky restrictions were indeed supported for the identification of the financial shock. Using this identification scheme, this section explores the existence of amplification mechanisms in the transmission of financial shocks. A natural extension of the conditionally heteroskedastic model is to allow for structural changes and model non-linear dynamics in the transmission of financial shocks. Section 5.1 presents the non-linear smooth transition VAR (NL-ST-VAR) model. Specifically, section 5.1.1 presents the reduced form while Section 5.1.2 presents the structural form. Section 5.2 addresses estimation using Bayesian methods, where Section 5.2.1 explains the prior choices and Section 5.2.2 presents the Metropolis-Hastings algorithm (see Appendix A.8 for a detailed description). Section 5.3 presents the state dependent impulse responses to a financial shock obtained from a Cholesky decomposition, and compares the covariance matrices with those from the ST-VAR. Lastly, section 5.4 formally tests the non-linear transmission of shocks. Using a Wald test, there is no evidence for non-linear transmission. The estimation of the non-linear model in this section made use of a modified version of the publicly available code from Auerbach and Gorodnichenko (2012).

5.1 Non-linear Smooth Transition VAR (NL-ST-VAR)

5.1.1 Reduced form NL-ST-VAR

Consider now a VAR(p) process $Y_t \in \mathbb{R}^K$ with conditional heteroskedasticity and conditional structural changes in parameters and regime dependent level shits. This way, not only we allow for different volatility periods but also different transmission of shocks during those periods and different conditional expectations. Let both heteroskedasticity and non-linear dynamics be predetermined by a smooth transition function that fluctuates between two regimes:

$$Y_t = (1 - F(s_t, \gamma, c)) \left(v^L + \sum_{i=1}^{p} A_i^L Y_{t-i} \right) + F(s_t, \gamma, c) \left(v^H + \sum_{i=1}^{p} A_i^H Y_{t-i} \right) + u_t, \quad (29)$$

$$u_t \sim N(0, \Omega_t) \quad (30)$$

$$\Omega_t = (1 - F(s_t, \gamma, c))\Omega^L + F(s_t, \gamma, c)\Omega^H \quad (31)$$

$$F(s_t, \gamma, c) = \frac{1}{1 + \exp(-\exp(\gamma)(s_t - c))} \quad (32)$$

In this model, both the intercept v^j, the coefficient matrices A_i^j and the covariance matrices Ω^j smoothly fluctuate between two regimes: $j = \{L, H\}$ allowing them to be in transition. Now the model in (29) represents a probability weighted sum of two linear VAR processes whereas (31) is the weighted sum of two covariance matrices. This way the error term in (30) is normally distributed conditional on the regime. This model extends the linear and conditionally heteroskedastic ST-VAR from Section 4.1 to allow for non-linear transmission of shocks and different sample averages depending on the regime. This means that for high values of s_t such that $F(s_t, \gamma, c) = 1$, the process Y_t is expected to display variance given by Ω^H, an average given by v^H and transmission given by A_i^H.

The model includes the same vector of variables as before and for a lag length $p = 3$. The transition between regimes is given by the function $F(s_t, \gamma = 3.5, c = 0.3)$ that was chosen in Section 4.2. This enables clearer comparisons across models. When the risk premium is high, the economy is modelled to be in a high financial stress regime with probability given by $P(j = H) = F(s_t, \gamma, c)$. This implies a higher volatility of variables but also different transmission of shocks. Section 2.1 presents multiple papers that research on the non-linear dynamics of financial shocks. This section intends to estimate these effects, conduct inference in the form of conditional impulse responses and test for the suitability of non-linear transmission modelling against the simpler model presented on Section 4.1.

5.1.2 Structural form NL-ST-VAR

In the light of the results from Section 4.5 it seems valid to apply a Cholesky decomposition for the identification of financial shocks. The conditions for Heteroskedasticity Identification in non-linear models have been established in Bacchiocchi and Fanelli (2015) making this scheme applicable for the current non-linear VAR. The model above is the most flexible among those considered in this paper but also the most complex in terms interpretation and computation. This calls for using simpler identification schemes whenever these are supported by the data. This is the case in the exclusion restrictions for the shock of interest.

Conditional on the regime j, the model is linear and can be set in structural form as a function of the state dependent IRFs:

$$\begin{aligned}
Y_t|j &= \begin{cases} v^L + \sum_{i=1}^p A_i^L Y_{t-i} + B^L \epsilon_t^L & \text{with probability} \quad P(j=L) \\ v^H + \sum_{i=1}^p A_i^H Y_{t-i} + B^H \epsilon_t^H & \text{with probability} \quad P(j=H) \end{cases} \\
&= \begin{cases} \left(I_K - \sum_{i=1}^p A_i^L L^i\right)^{-1} (v^L + B^L \epsilon_t^L) & \text{with probability} \quad P(j=L) \\ \left(I_K - \sum_{i=1}^p A_i^H L^i\right)^{-1} (v^H + B^H \epsilon_t^H) & \text{with probability} \quad P(j=H) \end{cases} \\
&= \begin{cases} \tilde{v}^L + \sum_{h=0}^t \zeta_h^L \epsilon_{t-h}^L & \text{with probability} \quad P(j=L) \\ \tilde{v}^H + \sum_{h=0}^t \zeta_h^H \epsilon_{t-h}^H & \text{with probability} \quad P(j=H) \end{cases}
\end{aligned}$$

These IRFs (ζ_h^j) are state dependent, meaning that they ignore feedback effects that affect the state of the economy on which they rely. The state dependent IRFs allow to estimate different effects of shocks in periods of high and low financial stress as defined by higher values of the bond premium. This means that the economy stays in each regime for the horizon $\aleph$ in which impulse responses are evaluated, allowing to differentiate between regimes more clearly at the expense of imposing an unrealistic restriction. For a modelling strategy that allows the shocks in a low stress regime to switch the state of the economy to a high stress regime, see Galvão and Owyang (2018). Because the model is linear conditional on the regime, the IRFs are not functions of history (Koop, Pesaran, and Potter 1996) and can be computed as in the linear case (see Appendix A.5) using each regimes' parameters $v^j, \mathbf{A}^j$ and B^j for $j = \{L, H\}$. Now there are two matrices of contemporaneous impacts (B^L, B^H) but both follow the Cholesky decomposition presented in Section 3.3.

5.2 Estimation of NL-ST-VAR

The Likelihood and log-Likelihood functions of the model in (18), (19), (31) and (20) are given by:

$$f(Y|\theta) = \mathcal{L}_T(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^K |\Omega_t|}} \exp\left(-\frac{1}{2} u_t' \Omega_t^{-1} u_t\right) \tag{33}$$

$$\log \mathcal{L}_T(\theta) \propto \log \mathcal{L}_T^*(\theta) = \frac{TK}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log |\Omega_t| - \frac{1}{2} \sum_{t=1}^{T} u_t' \Omega_t^{-1} u_t \tag{34}$$

$$u_t = Y_t - \Big[(1 - F(s_t, \gamma, c))\mathbf{A}^L X_{t-1} + F(s_t, \gamma, c)\mathbf{A}^H X_{t-1}\Big] \tag{35}$$

where $Y = [Y_t, Y_{t-1}, \ldots, Y_0]$, $X_{t-1} = [\mathbf{1}', Y_{t-1}', \ldots, Y_{t-p}']'$ and for each regime $j \in \{L, H\}$ we have $\mathbf{A}^j = [v^j, A_1^j, \ldots, A_p^j]$. For a given function $F(s_t, \gamma, c)$, the vector of parameters to be estimated is: $\theta = \{\Omega^L, \Omega^H, \mathbf{A}^L, \mathbf{A}^H\} \in \Theta$. The model is highly non-linear, and estimation requires making use of Bayesian methods. Specifically, the model is estimated with the code from Auerbach and Gorodnichenko (2012).

Bayesian estimation techniques simulate the posterior distribution of the parameters of interest θ conditional on a given data set Y. This means that parameters are now treated as a random variable and the data as given, contrary to the frequentist approach employed in the previous sections. Deriving the posterior requires of a conditional density of the data given θ and some prior beliefs on the data set Y that shrink the parameter space for the posterior. Let $f(\theta|Y)$ be the posterior probability density function, $f(\theta)$ the unconditional density with prior beliefs and $f(Y|\theta)$ the conditional density given by the likelihood function. Under Bayes Theorem it holds that:

$$f(\theta|Y) = f(Y, \theta) f(Y) = \frac{f(Y|\theta) f(\theta)}{f(Y)}$$

where the unconditional sample probability function given by $f(Y)$. Using the Law of Total Probability this unconditional density integrates the parameter's space such that $f(Y) = \int_\Theta f(Y|\theta) f(\theta) \partial\theta$ becomes an integral in a continuous event space, acting as a scaling constant. For this reason, the denominator can be disregarded for the optimization method and the quasi-posterior density can be used instead:

$$f(\theta|Y) = \frac{f(Y|\theta)f(\theta)}{\int_\Theta f(Y|\theta)f(\theta)\partial\theta} \propto f(Y|\theta)f(\theta)$$

5.2.1 Prior Selection

In practice, the choice of the priors $f(\theta)$ has been discussed extensively for multiple macroeconomic applications. See Miranda-Agrippino and Ricco (2019) for a summary on prior selection within macroeconomics. For this analysis I impose little structure for some of the parameters of estimate by using uninformative priors that barely shrink the parameter space.

The prior on the covariance matrices is chosen to be an uninformative flat prior with density: $f(\Omega^L) = f(\Omega^H) \sim N(\underline{\Omega}, \underline{\Sigma}_\Omega)$. The prior is flat, as it is centered around the OLS variance estimate from the homoskedastic VAR $\underline{\Omega} = \widehat{\Omega}_u$ and is uninformative because the variance $\underline{\Sigma}_\Omega = 10^6$ is chosen to be large enough, implying low penalization for deviating from prior beliefs. This allows the data to speak for itself when estimating the variances in each regime, allowing for comparisons with the linear models from the previous sections.

The VAR parameters follow a Minnesota prior (Litterman 1986) that approximates the coefficients to those of a multivariate random walk with drift. This prior is not motivated by theory but it is suitable for highly persistent processes, specially the interest variable that might contain a unit root. Additionally, this prior is computationally convenient. Consider the prior beliefs of the parameters to be given by: $f(\mathbf{A}^L) = f(\mathbf{A}^H) \sim N(\underline{\mathbf{A}}, \underline{\Sigma}_\mathbf{A})$. For a lag length $i = 1, \ldots, p$, the Minnesota prior assumes that the coefficients in $A_1, \ldots, A_p$ are normally distributed with:

$$\underline{A}_{mn,i} = \begin{cases} 1 & \text{if } \ m = n, \ i = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\underline{\Sigma}_{A_{mn,i}} = \begin{cases} (\varphi_1/i)^2 & \text{if } \ m = n, \forall i \\ (\varphi_1\varphi_2\widehat{\Sigma}_{A_{mm,i}})^2/(\widehat{\Sigma}_{A_{nn,i}}i)^2 & \text{if } \ m \neq n, \forall i \end{cases}$$

where $A_{mn,i}$ is the mth and nth column element form the matrix A_i, and equivalently $\Sigma_{A_{mn,i}}$ is the mth and nth column element form the matrix Σ_{A_i} which contains the estimated variance of coefficients in A_i for the baseline linear VAR. The hyperparameter φ_1 indicates the tightness of the parameter space and therefore the strength of the prior beliefs. A value of $\varphi_1 = 0$ implies that prior beliefs dominate and the model is set to behave as a vector of random walk variables, while setting $\varphi_1 \to \infty$ implies a high variance and therefore less weight of the prior. On the other hand, $0 < \varphi_2 < 1$ sets the weights assigned to information from other variables, and therefore the random walk behavior. Following Auerbach and Gorodnichenko (2012), the hyperparameters are chosen to be $\{\varphi_1 = 1, \varphi_2 = 0.5\}$.

Under the prior beliefs[24], the VAR(3) is expected to behave as a random walk with drift:

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \\ Y_{4,t} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\underline{A}_1} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \\ Y_{4,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{t,3} \\ u_{4,t} \end{bmatrix}$$

such that $\underline{A}_2$ and $\underline{A}_3$ are matrices of zeros.

In practice, the vector of parameters to estimate (θ) includes γ with the following prior density: $f(\gamma) \sim N(\underline{\gamma}, \underline{\Sigma}_\gamma)$. The variance is chosen small enough to ensure an estimate equal to the prior expected value: $\widehat{\gamma} = \underline{\gamma} = 3.5$. For simplicity, the following exposition ignores this fact, as doing so has no impact in the understanding of the estimation method.

This model is equivalent to that in Auerbach and Gorodnichenko (2012). In these models it is not possible to analytically derive a closed form solution for the posterior and therefore numerical methods are needed for simulating the posterior.

5.2.2 Metropolis-Hastings Algorithm

The parameter estimates in a Bayesian framework are defined as the expectation of the quasi-posterior. Applying the Bayes Theorem and substituting the Likelihood and log-Likelihood functions, we get:

$$\begin{aligned} \widehat{\theta} &= \int_\Theta \theta f(\theta|Y) \partial\theta \\ &= \int_\Theta \theta \left(\frac{f(Y|\theta) f(\theta)}{\int_\Theta f(Y|\theta) f(\theta) \partial\theta} \right) \partial\theta \\ &= \int_\Theta \theta \left(\frac{\mathcal{L}_T(\theta) f(\theta)}{\int_\Theta \mathcal{L}_T(\theta) f(\theta) \partial\theta} \right) \partial\theta \\ &= \int_\Theta \theta \left(\frac{e^{\log \mathcal{L}_T(\theta)} f(\theta)}{\int_\Theta e^{\log \mathcal{L}_T(\theta)} f(\theta) \partial\theta} \right) \partial\theta \end{aligned}$$

which for the given likelihood and prior densities, has no closed form solution. Chernozhukov and Hong (2003) suggest a Markov Chain Monte Carlo algorithm to recover

[24] The parameter's variance is then believed to follow:

$$\underline{\Sigma}_{A_1} = \begin{bmatrix} 1 & 0.44 & 1.37 & 5.57 \\ 0.14 & 1 & 0.78 & 3.16 \\ 0.04 & 0.08 & 1 & 1.01 \\ 0.01 & 0.02 & 0.06 & 1 \end{bmatrix}, \quad \underline{\Sigma}_{A_2} = \begin{bmatrix} 0.25 & 0.11 & 0.34 & 1.39 \\ 0.03 & 0.25 & 0.19 & 0.79 \\ 0.01 & 0.02 & 0.25 & 0.25 \\ 0.00 & 0.00 & 0.01 & 0.25 \end{bmatrix} \quad \text{and} \quad \underline{\Sigma}_{A_3} = \begin{bmatrix} 0.11 & 0.05 & 0.15 & 0.62 \\ 0.01 & 0.11 & 0.08` & 0.35 \\ 0.01 & 0.01 & 0.11 & 0.11 \\ 0.00 & 0.00 & 0.01 & 0.11 \end{bmatrix}$$

the first and second order moment point estimates. For some starting values $\theta^{(0)}$, the Metropolis-Hastings algorithm generates an ergodic Markov Chain $S = (\theta^{(1)}, \ldots, \theta^{(M)})$ by iteratively optimizing an instrumental density penalty function $(Q(\theta))$. By minimizing $Q(\theta)$, the algorithm minimizes the negative log-Likelihood and penalizes for deviations from prior expectations weighted by prior variances. The process iteratively draws $\theta^{(m)}$ for a large number M until the Markov Chain converges towards the posterior distribution of the parameters $f(\theta|Y)$. Equivalently, for a given continuous differentiable function $g(\theta) : \Theta \to \mathbb{R}$:

$$\frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) \to_p \int_{\Theta} g(\theta) f(\theta|Y)\partial\theta \quad \text{as} \quad M \to \infty \tag{36}$$

This implies, statistics that are continuous and differentiable functions of the parameters converge towards their expected values (Chernozhukov and Hong 2003). Then, the estimate $\widehat{\theta}$ can be obtained as the quasi-posterior mean $\widehat{\theta} = \bar{\theta}$ with variance given by $\widehat{\Sigma}_\theta$:

$$\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M} \theta^{(m)} \tag{37}$$

$$\widehat{\Sigma}_\theta = \frac{1}{M}\sum_{m=1}^{M} (\theta^{(m)} - \bar{\theta}^{(m)})(\theta^{(m)} - \bar{\theta}^{(m)})' \tag{38}$$

If (36) holds, then these are consistent estimates for θ and Σ_θ. Chernozhukov and Hong (2003) show that these estimates are consistent under fairly general regularity conditions assumed for MLE (see Appendix A.1).

The starting values $\theta^{(0)}$ for the algorithm are computed in a two-step WLS method by maximizing the likelihood in (34) and (35) with respect to $\{\Omega^{(L)}, \Omega^{(H)}\}$ and then deriving the WLS estimates for the coefficients $\{\mathbf{A^{(L)}}, \mathbf{A^{(H)}}\}$. This is possible because the model is linear conditional on the covariances. Using these starting values, the algorithm iteratively draws candidate values for the covariances and compute the WLS coefficients to evaluate the criterion function with the new draw. If the candidate draw results in lower values for the criterion than the draw from the previous period, then there is a higher probability of accepting the new candidate. This process is repeated for a large enough number of draws to ensure the convergence of the Markov Chain. The number of iterations is set to be $M = 100000$ and in order to avoid excessive dependence on the priors, the algorithm discards the first 20% of the draws as burn-in.

The Metropolis-Hastings algorithm proceeds as follows:

- **Step 1:** Set $m = 0$ and some initial values $\theta^{(0)} = \{\Omega^{L(0)}, \Omega^{H(0)}, \mathbf{A}^{L(0)}, \mathbf{A}^{H(0)}\}$. The initial values of the regime covariances can be constructed with the MLE estimates of B^j for $i = \{L, H\}$:

$$\Omega^{L(0)} = \widehat{B}^{L,MLE}(\widehat{B}^{L,MLE})' \quad \text{and} \quad \Omega^{H(0)} = \widehat{B}^{H,MLE}(\widehat{B}^{H,MLE})'$$

where the MLE are obtained by maximizing the log-Likelihood with respect to (B^L, B^H) using a Newton-type algorithm. Conditional on the covariances, the model is linear. The initial values for the coefficients $\mathbf{A}^L, \mathbf{A}^H$ are obtained from the WLS estimator. More technical details can be found in Appendix A.8.

- **Step 2:** Consider $\Psi^{(m)} = \{B^{L(m)}, B^{H(m)}\}$ such that each covariance matrix can be constructed by: $\Omega^{j(m)} = B^{j(m)}(B^{j(m)})', j = \{L, H\}$. Draw a candidate vector $\Upsilon^{(m)}$ where: $\Upsilon^{(m)} = \Psi^{(m)} + \psi^{(m)}$. We are interested in $\Psi^{(m)}$ but the draw is constructed to include the noise from $\psi \sim i.i.d.N(0, \Sigma_\Psi)$. The variance of ψ is initially set to $\Sigma_\Psi = 0.0025$ and adjusted on the fly based on the acceptance rate of the algorithm.[25] Discard the draw $\Upsilon^{(m)}$ if any of the eigenvalues of $\Omega^{L(m)}$ and $\Omega^{H(m)}$ are equal to zero and repeat **Step 2**. If the eigenvalues are greater than zero, continue to **Step 3**. The reason for this is to ensure that the Cholesky decomposition is feasible.

- **Step 3:** Conditional on the covariances in $\Upsilon^{(m)}$, compute $\{\mathbf{A}^{L(0)}, \mathbf{A}^{H(0)}\}$ using Feasible WLS (see Appendix A.8). Let now $Q(\theta)$ be the criterion function to minimize which includes the negative log-Likelihood and penalty functions for each parameter within θ to penalize for deviations from prior beliefs:

$$\begin{aligned} Q(\theta) = &- \log \mathcal{L}_T^*(\theta) + \mathbf{1}'(\mathbf{A}^L - \underline{\mathbf{A}})\underline{\Sigma}_{\mathbf{A}}(\mathbf{A}^L - \underline{\mathbf{A}})'\mathbf{1} + \mathbf{1}'(\mathbf{A}^H - \underline{\mathbf{A}})\underline{\Sigma}_{\mathbf{A}}(\mathbf{A}^H - \underline{\mathbf{A}})'\mathbf{1} \\ &+ \mathbf{1}'(\Omega^L - \underline{\Omega})\underline{\Sigma}_{\Omega}(\Omega^L - \underline{\Omega})'\mathbf{1} + \mathbf{1}'(\Omega^H - \underline{\Omega})\underline{\Sigma}_{\Omega}(\Omega^H - \underline{\Omega})'\mathbf{1} \end{aligned}$$

using the prior parameters given in Section 5.2.1 and the $(K \times 1)$ vector of 1s $\mathbf{1} = [1, \ldots, 1]'$ that convert each penalty functions into a single number[26]. Now define the $(m+1)$th vector of covariances $\Psi^{(m+1)}$ to be equal to $\Upsilon^{(m)}$ with a higher probability[27] if the criterion $Q(\theta)$ evaluates the candidate vector as more suitable than $\Psi^{(m)}$. Set the $m+1$ state of the chain as:

$$\Psi^{(m+1)} = \begin{cases} \Upsilon^{(m)} & \text{with probability} \quad P(\Psi, \psi) = \min\{1, \ \exp[Q(\Upsilon^{(m)}) - Q(\Psi^{(m)})]\} \\ \Psi^{(m)} & \text{with probability} \quad 1 - P(\Psi, \psi) \end{cases}$$

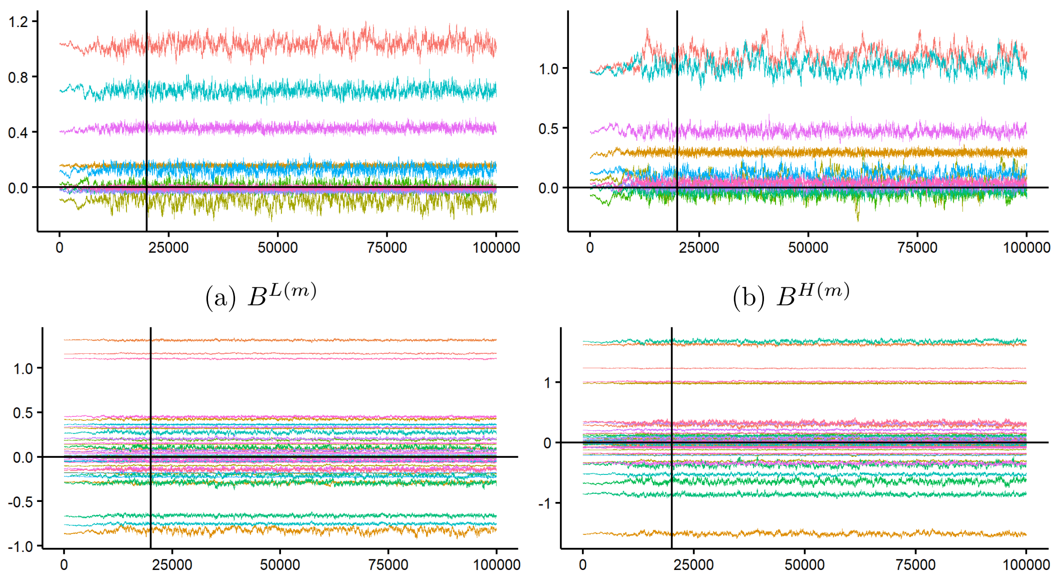
where $Q(\Upsilon^{(m)}) = Q(\Upsilon^{(m)}, \mathbf{A}^{L(m)}, \mathbf{A}^{H(m)})$ evaluates the criterion for the candidate vector of parameters while $Q(\Psi^{(m)}) = Q(\Psi^{(m)}, \mathbf{A}^{L(m)}, \mathbf{A}^{H(m)})$ it at the current state of the Markov Chain. This means that the candidate could be accepted even if it yields a higher value of the criterion to minimize. This facilitates the coverage of the entire distribution by considering candidates that are further away from the optimum.

[25] When the acceptance rate is too low and falls below 0.25, the variance is tightened to search for candidates closer to the priors. When the acceptance rate is too high and rises above 0.35 then the variance is increased to search further away from the priors. This way the algorithm ensures a high coverage of the distribution of considered candidates. These are the values chosen by Auerbach & Gorodnichenko (2012).

[26] This makes use of: $\mathbf{1}'X\mathbf{1}$ being equal to the sum of the elements within the $(K \times K)$ matrix X.

[27] Computationally, this is done by drawing z from the process $z \sim i.i.d.N(0, 1)$ and accepting the candidate vector $\Upsilon^{(m)}$ if $P(\Psi, \psi) > z$.

- **Step 4:** Set $m = m + 1$ and return to **Step 2**. Repeat the loop for $m = 1, \ldots, M$ where $M = 100000$ to ensure convergence towards the quasi-posterior distribution.



(a) $B^{L(m)}$ (b) $B^{H(m)}$

(c) $\mathbf{A}^{L(m)}$ (d) $\mathbf{A}^{H(m)}$

Figure 9: Convergence of the parameters within the Markov Chain towards their posterior distribution as defined by the uncertainty generated within the chain. On the vertical axis, the values of each accepted parameter draw. On the horizontal axis, the iterations $m = 1, \ldots, M$ with $M = 100,000$. The horizontal line separates the burn-in period with 20% of the observations.

This algorithm is explained in more detail in Appendix A.8. Figure 9 displays the convergence of the Markov Chain $S = (\theta^{*(0)}, \ldots, \theta^{*(M)})$ towards the expected values of $\theta^* = (B^L, B^H, \mathbf{A}^L, \mathbf{A}^H)$[28]. By construction, the Chain fluctuates around the optimum with a given volatility that does not decay for larger horizons of the Chain. This results in a distribution over S that can be thought as an approximation to the quasi-posterior density $f(S|Y)$. The burn-in period effectively discards those values that heavily rely on initial values $\theta^{(0)}$ until convergence is achieved.

Figure 10 shows the acceptance ratio of the draws $\Upsilon^{(m)}$ for the B^j matrices and their respective covariances and WLS parameters. Initially, there is a high acceptance ratio because the algorithm takes small steps towards the optimum. The decay in the acceptance ratio illustrates how the algorithm finds a local minimum for the criterion function to minimize. The new draws in the surroundings of this minimum are rarely accepted,

[28] Recall that the vector θ had previously been defined as $\theta = (\Omega^L, \Omega^H, \mathbf{A}^L, \mathbf{A}^H)$, but the Monte Carlo algorithm integrates the Cholesky identification scheme within the estimation process and instead optimizes for $\theta^* = (B^L, B^H, \mathbf{A}^L, \mathbf{A}^H)$. The vector $\theta = (B^L, B^H, \mathbf{A}^L, \mathbf{A}^H)$ can easily be constructed by computing the covariances as: $\Omega^j = B^j(B^j)', j = \{L, H\}$.

and the variance Σ_ψ is increased to allow the algorithm to draw from further away of the optimum.

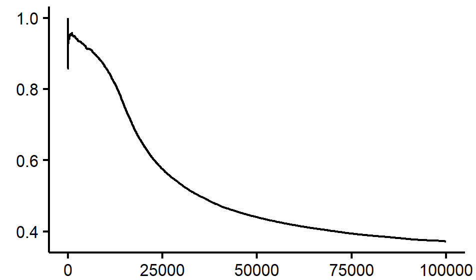


Figure 10: Acceptance Ratio of the Markov Chain Monte Carlo draws $\Upsilon^{(m)}$ for each iteration $m = 1, \dots, M$ where $M = 100000$.

IRFs can be created using the each of the accepted draws in the vector of parameters after the burn-in of 20%. This means generating a series of $0.8M$ point-wise IRFs $(\zeta_h^{(0.2M)}, \dots, \zeta_h^{(M)})$ equivalent to deriving the density of IRFs. As ζ_h are continuously differentiable functions of θ, convergence arises from (36). This way, the point-wise mean is used as the 'true' impulse response coefficient while confidence bands can be generated using the 5% and 95% of the distribution. These intervals use the uncertainty generated within the chain.

5.3 Inference in NL-ST-VAR

Table 11 shows the NL-ST-VAR average estimates of the covariances and B^j matrices. Now the model is estimated with Cholesky restrictions in each regime and therefore two B^j matrices are obtained. The covariances in Table 11a and 11b are comparable[29] to those derived under HI in the ST-VAR (see Tables 6a and 6b). The two covariances are different and Ω^H is clearly that of the high volatility regime, as can be deduced from the larger diagonal elements.

Figure 11 displays the state dependent impulse responses to a financial shock as given by the point-wise average of accepted values from the Markov Chain and the subsequent derivation of $\zeta_h^{(m)}$. The confidence bands are derived from the 5^{th} and the 95^{th} percentile of the simulated point-wise distributions. These bands are very narrow by construction and should not be interpreted as the uncertainty over the responses in each regime (Auerbach and Gorodnichenko 2012). Instead, they solely signal the small uncertainty generated within the chain after convergence. This can be the result of a very accurate estimation of the parameters, of the stating values or by construction, due to the low uncertainty generated because of a small variance Σ_ψ in the draws from the chain. Each impulse responses is constructed under the assumption that the economy is in

[29]All the covariances' elements have the same signs. The magnitudes are different but equivalent.

one of the regimes and not in transition (as most of the observations). For that reason we would expect that in reality the response would be somewhere in between the two extremes. In this regard, the conditional impulse responses can be interpreted as bounds for the actual response of the variable. Because they are built under the assumption that the economy remains on the same regime throughout time, these impulse responses become inconsistent in longer horizons[30], where they display explosive dynamics. This is a problem in state dependent impulse responses that do not allow feedback effects to endogenously switch the regime as the shock effects vanish[31]. Accounting for this problem is out of the scope of this paper.

	π_t	x_t	i_t	bp_t
π_t	1.08	-0.10	0.01	-0.01
x_t	-0.10	0.50	0.09	-0.01
i_t	0.01	0.09	0.20	-0.00
bp_t	-0.01	-0.01	0.00	0.02

(a) $\widehat{\Omega}^L = \widehat{B}^L(\widehat{B}^L)'$

	π_t	x_t	i_t	bp_t
π_t	1.23	0.05	-0.04	-0.03
x_t	0.05	1.04	0.11	0.01
i_t	-0.04	0.11	0.23	0.02
bp_t	-0.03	0.01	0.02	0.08

(b) $\widehat{\Omega}^H = \widehat{B}^H(\widehat{B}^H)'$

	π_t	x_t	i_t	bp_t
π_t	1.03	0	0	0
x_t	-0.09	0.70	0	0
i_t	0.01	0.13	0.43	0
bp_t	-0.01	0.02	-0.00	0.15

(c) $\widehat{B}^L$

	π_t	x_t	i_t	bp_t
π_t	1.11	0	0	0
x_t	0.04	1.02	0	0
i_t	-0.04	0.11	0.47	0
bp_t	-0.03	0.01	0.04	0.29

(d) $\widehat{B}^H$

Table 11: NL-ST-VAR(3) estimates under Cholesky Identification

AS can be seen in Figure 11, financial shocks (as measured by orthogonal bond premium shocks) are bigger in size during periods of high stress by definition of the volatile regime. Recall that the bond premium values are determining the probability of being in a high volatility regime. In periods of high financial stress, a bond premium shock causes a sudden fall in output growth, inflation and the interest rate. Intuitively, the higher risks and funding costs lower investment and increase defaults lowering output. The reduction in investment can also arise from tighter lending constraints and credit rationing. The reduction in aggregate demand is accompanied by falling prices as agents adjust their inflationary expectations. The interest rate falls in line the policy mandate of the Federal Reserve signaling short term interest rate cuts and/or other unconventional monetary policies. Following the policy response, output growth is the first aggregate to recover, strengthening aggregate demand and therefore reducing deflationary pressures.

[30]Note that IRFs also tend to be inconsistently estimated for long horizons when using non-stationary VAR (P. C. B. Phillips 1998). If the interest rate is recognised as a unit root, this poses a problem for impulse response analysis in long horizons.

[31]Galvão and Owyang (2018) solve this problem by allowing shocks to switch regimes if the responses through feedback effects.

As inflation and growth are recovered, the risk premium is reduced and the policy rate returns to the pre-crisis state but remains low until inflation and growth are positive and the risk premium has fallen enough. On the other hand, financial shocks in periods of low stress are smaller in size and cause a small fall in output and prices and no clear response from the Central Bank. Indeed, it seems like the policy rate is slightly increased. This raises doubts on whether the model is actually capturing financial shocks during low stress regimes.

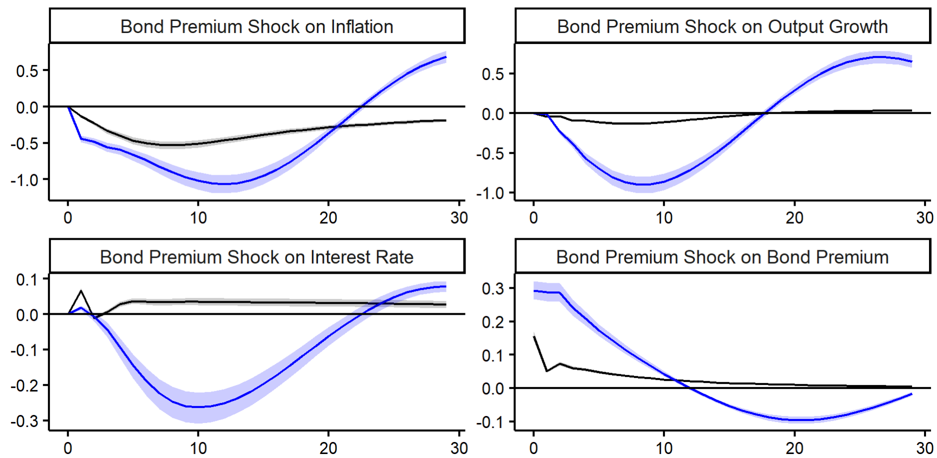


Figure 11: Conditional Impulse Response Functions with Cholesky restrictions in NL-ST-VAR(3) for a bond premium shock. In blue, responses to a financial shock in high volatility regime. In black, responses in low volatility regime. 90% CI with uncertainty generated by the MCMC algorithm.

These results are in line with most of the literature (Chak and Chiu 2016, Galvão and Owyang 2018, Bennani and Neuenkirch 2020) and seem to signal a clear difference in the transmission of financial shocks between periods of high and low financial stress. However, it could be argued that the different dynamics arise from different sized shocks, making heteroskedasticity the main driver of non-linearities.

5.4 Testing for Non-linear Transmission

The non-linear transmission of shocks implies the rejection of the joint null hypothesis:

$$H_0 : \mathbf{A}_*^L = \mathbf{A}_*^H \quad \rightarrow \quad H_0 : \mathrm{vec}(\mathbf{A}_*^L) - \mathrm{vec}(\mathbf{A}_*^H) = 0 \tag{39}$$

with $\mathbf{A}_*^j = [A_1^j, \ldots, A_p^j]$ of size $(K \times pK)$ and $\mathrm{vec}(\mathbf{A}_*^j)$ of size $(pK^2 \times 1)$. This can be formally tested if the limiting distribution of the parameters of interest is:

$$\widehat{\mathbf{A}}_*^j \sim N(\mathbf{A}_*^j, \Sigma_{\mathbf{A}_*^j}), \; j = \{L, H\}$$

The hypothesis on (39) can be rewritten as:

$$H_0 : R\boldsymbol{\alpha} = 0 \tag{40}$$

where:

$$\boldsymbol{\alpha} = \begin{bmatrix} \text{vec}(A_1^L, \ldots, A_p^L) \\ \text{vec}(A_1^H, \ldots, A_p^H) \end{bmatrix} \tag{41}$$

$$R = [I_{pK^2}, (-1)I_{pK^2}] = \begin{bmatrix} 1 & 0 & \ldots & 0 & -1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & -1 \end{bmatrix} \tag{42}$$

Here $\boldsymbol{\alpha}$ is $(2pK^2 \times 1)$ sized and R has $(4pK^2 \times 2pK^2)$ dimensions. Let first assume that the distribution of the estimator $\widehat{\boldsymbol{\alpha}}$ is given by a Normal Gaussian in the limit:

$$\widehat{\boldsymbol{\alpha}} \sim N(\boldsymbol{\alpha}, \Sigma_{\boldsymbol{\alpha}}) \tag{43}$$

where the $(2pK^2 \times 2pK^2)$ matrix of standard errors is given by:

$$\widehat{\Sigma}_{\boldsymbol{\alpha}} = \frac{1}{M} \sum_{m=1}^{M} (\boldsymbol{\alpha}^{(m)} - \bar{\boldsymbol{\alpha}}^{(m)})(\boldsymbol{\alpha}^{(m)} - \bar{\boldsymbol{\alpha}}^{(m)})'$$

Under the asymptotic normality assumption (43), we can test the assumption of time-invariant effects held in Sections 3 and 4. Under the null hypothesis, shocks are linearly transmitted and there is no amplification mechanism. In this case, the heteroskedastic linear model would be more suitable than the current augmented non-linear model. This would be in line with Chan and Eisenstat (2017), Xu and Haan (2018), Karlsson and Österholm (2020) and Nordström (2021).

H_0	Test statistic	p-value
$\mathbf{A}_*^L = \mathbf{A}_*^H$	0.99	1

Table 12: Wald test for non-linear transmission in the NL-ST-VAR(3) model. Under the null H_0, there is no linear transmission of shocks. The test relies on the assumption of asymptotic normally distributed vector of parameters.

Under the assumption of asymptotic normality (43), the Wald test statistic is given by:

$$W^{\boldsymbol{\alpha}} = (R\hat{\boldsymbol{\alpha}})'(R\widehat{\Sigma}_{\boldsymbol{\alpha}}R')^{-1}(R\hat{\boldsymbol{\alpha}}) \sim \chi^2_{(pK^2)}$$

Table 12 shows the results of this test. There is no support for non-linear modeling in the form of different transmission of shocks. However, we should be cautious in interpreting this result. The Minnesota prior shrunk the coefficients towards a random walk and the interest rate variable could contain a unit root, meaning that the asymptotic distribution of the Wald joint hypothesis test would not be normal Gaussian (Toda and P. C. Phillips 1993, Kilian and Lütkepohl 2017).

Despite having evidence against non-linear modelling, the evidence might be based on wrong assumptions over the limit distribution of the joint Wald test. A deeper analysis is needed to test the existence of non-linear transmission. However, this is beyond of the scope of this paper.

6 Conclusion

In this thesis I have explored multiple specifications for modelling financial shocks with the objective of choosing the most suitable one. The main features explored are heteroskedasticity and non-linearities, including different heteroskedasticity specifications and using the preferred form for structural identification of shocks. This allows testing traditional theory based restrictions. The system is augmented to include level shifts and different transmission of shocks across regimes.

First, a linear homoskedastic VAR is estimated. Using a Cholesky identification method, I analyze the impulse responses. Bond premium shocks are found to trigger a fall in output growth and inflation and the subsequent reduction in monetary policy rate, where this rate captures both unconventional policies and the nominal Federal Funds rate. After analyzing the residuals of the unrestricted model, I conclude that the previous model is wrongly specified. Residuals are not normally distributed with excess kurtosis and display conditional heteroskedasticity. This implies that inference in the form of impulse responses is inaccurate. Furthermore, conditional heteroskedasticity can be used to impose restrictions without the need of theoretically justified assumptions on the contemporaneous relationships between variables.

Second, a linear but heteroskedastic VAR is estimated using smooth transition in covariances. By differentiating between two volatility regimes, identification of shocks is possible without the need of theory based restrictions. Regimes of financial stress are also determined by high values of the bond premium, acknowledging that whenever the bond premium displays higher values there is also higher volatility. The regimes' probabilities from the smooth transition function are compared with the filtered and smoothed probabilities of a Markov Switching VAR. Unlike the Markov Switching model, the regimes from the smooth transition function are connected to periods of financial stress. Using the smooth transition in covariances, I estimate the VAR model and conduct inference. Full identification is confirmed using Wald tests. When estimating the model subject to over-identifying restrictions, a Likelihood Ratio test does not provide support for Cholesky restrictions. The reason is that inflation contemporaneously reacts to demand shocks and output growth reacts on impact to monetary policy shocks. These were assumptions from the Cholesky decomposition which are not supported by the data. However, the same test supports the exclusion restrictions for the identification of financial shocks. For this reason, the impulse responses are very similar across models.

Third, the heteroskedastic model is augmented to include different intercepts and coefficients across volatility regimes. Using Bayesian methods that impose an uninformative prior on the covariance and a Minnesota prior in the coefficients, I implicitly account for unit roots. Using a Cholesky decomposition for the identification of financial shocks I find that the responses are larger for every variable when the shocks occur in periods of high stress. However, it is unclear whether this is the result of different sized shocks or an amplification mechanism derived from imperfect substitutability of sources of funding and other financial frictions. A Wald joint hypothesis test cannot reject the null hypothesis of linear transmission of heteroskedastic shocks, however this test might not have the desired limit distribution if the system contains unit roots. More information is needed to draw conclusions on the existence of a non-linear transmission.

Overall, the main conclusions from this research are: (i) the regimes obtained with a smooth transition in covariances are more suitable for modelling financial shocks than those derived from a Markov Switching model, (ii) heteroskedasticity identification does not provide overall support for Cholesky restrictions but these restriction are valid for the financial shock identification, (iii) the non-linear dynamics from different volatility periods are likely to be the result of different sized shocks (heteroskedasticity) instead of an amplification mechanism, but more evidence is needed to conclude anything, and (iv) the impulse responses to financial shocks are robust across the considered specifications and identification approaches: a risk premium shock triggers a fall in output growth and inflation followed by a monetary policy loosening.

References

Adrian, Tobias, Arturo Estrella, and Hyun Song Shin (Sept. 2019). “Risk-taking channel of monetary policy”. In: *Financial Management* 48 (3), pp. 725–738.

Adrian, Tobias and Hyun Song Shin (2010). “Liquidity and Leverage”. In: *Journal of financial intermediation* 19 (3), pp. 418–437.

Allen, Franklin and Douglas Gale (2000). “Financial Contagion”. In: *The Journal of Political Economy* 108 (1), pp. 1–33.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2020). “Business-cycle anatomy”. In: *American Economic Review* 110 (10), pp. 3030–70.

Arnold, Bruce et al. (Dec. 2012). “Systemic risk, macroprudential policy frameworks, monitoring financial systems and the evolution of capital adequacy”. In: *Journal of Banking and Finance* 36 (12), pp. 3125–3132.

Auerbach, Alan J. and Yuriy Gorodnichenko (May 2012). “Measuring the output responses to fiscal policy”. In: *American Economic Journal: Economic Policy* 4 (2), pp. 1–27.

Bacchiocchi, Emanuele and Luca Fanelli (Dec. 2015). “Identification in Structural Vector Autoregressive Models with Structural Changes, with an Application to US Monetary Policy”. In: *Oxford Bulletin of Economics and Statistics* 77 (6), pp. 761–779.

Becard, Yvan and David Gauthier (2021). *Banks, shadow banks, and business cycles.* Tech. rep. Bank of England.

Benkwitz, Alexander, Helmut Lütkepohl, and Jürgen Wolters (2001). “Comparison of bootstrap confidence intervals for impulse responses of German monetary systems”. In: *Macroeconomic Dynamics* 5 (1), pp. 81–100.

Bennani, Hamza and Matthias Neuenkirch (2020). *The Financial Accelerator in the Euro Area: New Evidence Using a Mixture VAR Model.* Tech. rep. University of Trier Research Papers in Economics, (11/20).

Bernanke, Ben and Mark Gertler (1989). “Agency Costs, Net Worth, and Business Fluctuations”. In: *The American Economic Review* 79 (1), pp. 14–31.

– (1990). “Financial Fragility and Economic Performance”. In: *Source: The Quarterly Journal of Economics* 105 (1), pp. 87–114.

Bernanke, Ben S and Alan S Blinder (1988). “Credit, money, and aggregate demand”. In: *The American Economic Review* 78 (2), pp. 435–439.

Blanchard, Oliver Jean and Danny Quah (1989). “The Dynamic Effects of Aggregate Demand and Supply Disturbances.” In: *The American Economic Review* 79 (4), pp. 655–673.

Borio, Claudio (2014). “The financial cycle and macroeconomics: What have we learnt?” In: *Journal of Banking and Finance* 45 (1), pp. 182–198.

Borio, Claudio and Anna Zabai (2016). *Unconventional monetary policies: a re-appraisal.* Tech. rep. Bank for International Settlements.

Box, G E P and David A Pierce (1970). “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models”. In: *Source: Journal of the American Statistical Association* 65 (332), pp. 1509–1526.

Bruder, Stefan (2018). *Inference for structural impulse responses in SVAR-GARCH models*. Tech. rep. Department of Economics, University of Zurich.
Brüggemann, Ralf, Carsten Jentsch, and Carsten Trenkler (Mar. 2016). “Inference in VARs with conditional heteroskedasticity of unknown form”. In: *Journal of Econometrics* 191 (1), pp. 69–85.
Caccioli, Fabio, Paolo Barucca, and Teruyoshi Kobayashi (Oct. 2017). “Network models of financial systemic risk: A review”. In: *Journal of Computational Social Science* 1 (1), pp. 81–114.
Cavaliere, Giuseppe, Heino Bohn Nielsen, and Anders Rahbek (2020). *An Introduction to Bootstrap Theory in Time Series Econometrics*. Tech. rep. Department of Economics. University of Copenhagen.
Chak, Hung Jack Cheng and Ching-Wai Chiu (2016). *Nonlinearities of mortgage spreads over the business cycles*. Tech. rep. Bank of England.
Chan, Joshua C C and Eric Eisenstat (2017). “Bayesian Model Comparison for Time-Varying Parameter VARs with Stochastic Volatility”. In: *Journal of Applied Econometrics* 33 (4), pp. 509–532.
Chernozhukov, Victor and Han Hong (Aug. 2003). “An MCMC approach to classical estimation”. In: *Journal of Econometrics* 115 (2), pp. 293–346.
Chodorow-Reich, Gabriel (Feb. 2014). “The employment effects of credit market disruptions: Firm-level evidence from the 2008-9 financial crisis”. In: *Quarterly Journal of Economics* 129 (1), pp. 1–59.
Christensen, Anders Møller and Heino Bohn Nielsen (2005). *US Monetary Police 1988-2004 An Empirical Analysis*. Tech. rep. Department of Economics, University of Copenhagen.
Claessens, Stijn. and M. Ayhan. Kose (2018). *Frontiers of macrofinancial linkages*. Tech. rep. Bank for International Settlements. Monetary and Economic Department, p. 191.
Doornik, Jurgen A (2013). “Econometric Analysis with Markov-Switching Models.” In: *Timberlake Consultants Ltd, London*.
Doornik, Jurgen A and David F Hendry (1997). “Modelling Dynamic Systems Using PcFiml 9.0 for Windows”. In: *International Thomson Business Press*.
Dufour, Jean Marie, Lynda Khalaf, and Marie Claude Beaulieu (Mar. 2010). “Multivariate residual-based finite-sample tests for serial dependence and arch effects with applications to asset pricing models”. In: *Journal of Applied Econometrics* 25 (2), pp. 263–285.
Eggertsson, Gauti B. and Paul Krugman (Aug. 2012). “Debt, deleveraging, and the liquidity trap: A fisher-minsky-koo approach”. In: *Quarterly Journal of Economics* 127 (3), pp. 1469–1513.
Ehrmann, Michael, Marcel Fratzscher, and Roberto Rigobon (2011). “Stocks, Bonds, Money Markets and Exchange Rates: Measuring International Financial Transmission”. In: *Journal of Applied Econometrics* 26 (6), pp. 948–974.
Eitrheim, Ø and T Teräsvirta (1996). “Testing the adequacy of smooth transition autoregressive models”. In: *Journal of Econometrics* 1 (74), pp. 59–75.

Engle, Robert F (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”. In: *Econometrica: Journal of the econometric society* 50 (4), pp. 987–1007.

Evans, Martin D D and Richard K Lyons (2002). “Order Flow and Exchange Rate Dynamics”. In: *Journal of political economy* 110 (1), pp. 170–180.

Farhi, Emmanuel and Jean Tirole (2017). “Shadow Banking and the Four Pillars of Traditional Financial Intermediation”. In: *The Review of Economic Studies* 88 (6), pp. 2622–2653.

Faulwasser, Timm, Marco Gross, and Willi Semmler (2018). “Credit Cycles, Regime Switching and Monetary Policy”. In: *The New School - Department of Economics; Universitaet Bielefeld*.

Feng, Xu, Lei Lu, and Yajun Xiao (Feb. 2020). “Shadow banks, leverage risks, and asset prices”. In: *Journal of Economic Dynamics and Control* 111.

Galvão, Ana Beatriz and Michael T. Owyang (Oct. 2018). “Financial Stress Regimes and the Macroeconomy”. In: *Journal of Money, Credit and Banking* 50 (7), pp. 1479–1505.

Geanakoplos, John (2010). “The Leverage Cycle”. In: *National Bureau of Economic Research* 24.

Gertler, Mark and Peter Karadi (2015). “Monetary policy surprises, credit costs, and economic activity”. In: *American Economic Journal Macroeconomics* 7 (1), pp. 44–76.

Gilchrist, Simon and Egon Zakrajšek (2011). “Credit Spreads and Business Cycle Fluctuations”. In: *American economic review* 102 (4), pp. 1692–1720.

Gonçalves, Sílvia and Lutz Kilian (Nov. 2004). “Bootstrapping autoregressions with conditional heteroskedasticity of unknown form”. In: *Journal of Econometrics* 123 (1), pp. 89–120.

Gorton, Gary and Andrew Metrick (June 2012). “Securitized banking and the run on repo”. In: *Journal of Financial Economics* 104 (3), pp. 425–451.

Gregoriou, Greg N., François Éric Racicot, and Raymond Théoret (Jan. 2021). “The response of hedge fund tail risk to macroeconomic shocks: A nonlinear VAR approach”. In: *Economic Modelling* 94, pp. 843–872.

Guidolin, Massimo, Valentina Massagli, and Manuela Pedio (2021). “Does the Cost of Private Debt Respond to Monetary Policy? Heteroskedasticity-Based Identification in a Model with Regimes”. In: *The European Journal of Finance*, pp. 1–30.

Hall, Peter (1992). “Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density”. In: *Source: The Annals of Statistics* 20 (2), pp. 675–694.

Hamilton, James D (1994). *Time Series Analysis*. Princeton University Press.

Herwartz, Helmut and Martin Plödt (Feb. 2016). “Simulation Evidence on Theory-based and Statistical Identification under Volatility Breaks”. In: *Oxford Bulletin of Economics and Statistics* 78 (1), pp. 94–112.

Heuvel, Skander J Van Den (2002). *The Bank Capital Channel of Monetary Policy*. Tech. rep. Wharton School, University of Pennsylvania.

Holmström, Bengt and Jean Tirole (1998). “Private and public supply of liquidity”. In: *Journal of Political Economy* 106 (1), pp. 1–40.
Hubrich, Kirstin, Antonello D’agostino, et al. (2013). *Financial Shocks and the Macroeconomy: Heterogeneity and Non-Linearities*. Tech. rep. European Central Bank.
Hubrich, Kirstin and Timo Teräsvirta (2013). “Thresholds and Smooth Transitions in Vector Autoregressive Models”. In: *CREATES Research Papers. Department of Economics and Business Economics, Aarhus University.*
Inoue, Atsushi and Lutz Kilian (Sept. 2020). “The Uniform Validity of Impulse Response Inference in Autoregressions”. In: *Journal of Econometrics* 215 (2), pp. 450–472.
Jarque, Carlos M and Anil K Bera (1987). “A Test for Normality of Observations and Regression Residuals”. In: *International Statistical Review* 55 (2), pp. 163–172.
Johansen, Soren (1988). “Statistical analysis of cointegration vectors”. In: *Journal of Economic Dynamics and Control* 12 (2-3), pp. 231–254.
Karlsson, Sune and Pär Österholm (Jan. 2020). “The relation between the corporate bond-yield spread and the real economy: Stable or time-varying?” In: *Economics Letters* 186.
Kashyap, Anil K and Jeremy C Stein (2000). “What do a million observations on banks say about the transmission of monetary policy?” In: *American Economic Review* 90 (3), pp. 407–428.
Kilian, Lutz and Helmut Lütkepohl (Nov. 2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.
Kiyotaki, Nobuhiro and John Moore (1997). “Credit cycles”. In: *Journal of Political Economy* 105 (2), pp. 211–248.
Koop, Gary and Dimitris Korobilis (2009). *Bayesian Multivariate Time Series Methods for Empirical Macroeconomics*. Tech. rep. MPRA Paper 20125.
Koop, Gary, M Hashem Pesaran, and Simon M Potter (1996). “Impulse response analysis in nonlinear multivariate models”. In: *Journal of Econometrics* 74, pp. 119–147.
Kreiss, Jens-Peter (1997). *Asymptotical Properties of Residual Bootstrap for Autoregressions*. Tech. rep. Institute für Mathematik, Technische Universitat.
Lange, Alexander et al. (Mar. 2021). “svars: An r package for data-driven identification in multivariate time series analysis”. In: *Journal of Statistical Software* 97 (1), pp. 1–34.
Lange, Theis and Anders Rahbek (2009). “An Introduction to Regime Switching Time Series Models”. In: *Handbook of Financial Time Series*. Ed. by Thomas Mikosch et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 871–887.
Lanne, Markku and Helmut Lütkepohl (Jan. 2010). “Structural vector autoregressions with nonnormal residuals”. In: *Journal of Business and Economic Statistics* 28 (1), pp. 159–168.
Lanne, Markku, Helmut Lütkepohl, and Katarzyna Maciejowska (Feb. 2010). “Structural vector autoregressions with Markov switching”. In: *Journal of Economic Dynamics and Control* 34 (2), pp. 121–131.

Lanne, Markku and Pentti Saikkonen (Jan. 2007). “A multivariate generalized orthogonal factor GARCH model”. In: *Journal of Business and Economic Statistics* 25 (1), pp. 61–75.

Litterman, Robert B (1986). “Forecasting with Bayesian Vector Autoregressions: Five Years of Experience”. In: *Source: Journal of Business & Economic Statistics* 4 (1), pp. 25–38.

Lütkepohl, Helmut (2012). *Identifying Structural Vector Autoregressions via Changes in Volatility*. Tech. rep. German Institute for Economic Research.

Lütkepohl, Helmut, Mika Meitz, et al. (Mar. 2021). “Testing identification via heteroskedasticity in structural vector autoregressive models”. In: *The Econometrics Journal* 24 (1), pp. 1–22.

Lütkepohl, Helmut, George Milunovich, and Minxian Yang (Oct. 2020). “Inference in partially identified heteroskedastic simultaneous equations models”. In: *Journal of Econometrics* 218 (2), pp. 317–345.

Lütkepohl, Helmut and Aleksei Netšunajev (Jan. 2017a). “Structural vector autoregressions with heteroskedasticity: A review of different volatility models”. In: *Econometrics and Statistics* 1, pp. 2–18.

– (Nov. 2017b). “Structural vector autoregressions with smooth transition in variances”. In: *Journal of Economic Dynamics and Control* 84, pp. 43–57.

Lütkepohl, Helmut and Thore Schlaak (Apr. 2019). “Bootstrapping impulse responses of structural vector autoregressive models identified through GARCH”. In: *Journal of Economic Dynamics and Control* 101, pp. 41–61.

Mehrling, Perry (2010). *The New Lombard Street: How the Fed Became the Dealer of Last Resort*. Princeton University Press.

Mian, Atif and Amir Sufi (Aug. 2011). “House prices, home equity-based borrowing, and the US household leverage crisis”. In: *American Economic Review* 101 (5), pp. 2132–2156.

Mian, Atif, Amir Sufi, and Francesco Trebbi (2015). “Foreclosures, House Prices, and the Real Economy”. In: *The Journal of Finance* 70 (6), pp. 2587–2633.

Minsky, Hyman P (1992). *The Financial Instability Hypothesis*. Tech. rep. Levy Economics Institute of Bard College.

Miranda-Agrippino, Silvia and Giovanni Ricco (2019). “Bayesian Vector Autoregressions: Estimation”. In: *Oxford Research Encyclopedia of Economics and Finance.*

Nielsen, Heino Bohn (2021). *A Course in Time Series Econometrics*. Tech. rep.

Nordström, Martin (2021). “Credit spread and employment growth–a time-varying relationship?” In: *Applied Economics Letters* 28 (1), pp. 23–31.

Normandin, Michel and Louis Phaneuf (Sept. 2004). “Monetary policy shocks: Testing identification conditions under time-varying conditional volatility”. In: *Journal of Monetary Economics* 51 (6), pp. 1217–1243.

Olea, José Luis Montiel, Mikkel Plagborg-Møller, and Eric Qian (2021). *SVAR Identification From Higher Moments: Has the Simultaneous Causality Problem Been Solved?* Tech. rep. Papers and proceedings.

Pfaff, Bernhard (2008). “VAR, SVAR and SVEC Models: Implementation Within R Package vars”. In: *Journal of Statistical Softwaredoi10.18637jss.v027.i04.* 27 (4), pp. 1–32.
Phillips, Peter C B (1998). “Impulse response and forecast error variance asymptotics in nonstationary VARs”. In: *Journal of Econometrics* 83, pp. 21–56.
Rigobon, Roberto (2003). “Identification through heteroskedasticity”. In: *Review of Economics and Statistics* 85 (4), pp. 777–792.
Sims, Christopher A (1980). “Macroeconomics and Reality”. In: *Econometrica* 48 (1), pp. 1–48.
– (2020). “SVAR Identification Through Heteroskedasticity With Misspecified Regimes”. In.
Sims, Christopher A, James H Stock, and Mark W Watson (1990). “Inference in Linear Time Series Models with some Unit Roots”. In: *Econometrica* 58 (1), pp. 113–144.
Stiglitz, Joseph and Andrew Weiss (1981). “Credit Rationing in Markets With Imperfect Information”. In: *American Economic Review* 71, pp. 393–410.
Teräsvirta, Timo (1994). “Specification, estimation, and evaluation of smooth transition autoregressive models.” In: *Source: Journal of the American Statistical Association* 89 (425), pp. 208–218.
Toda, Hiro Y and Peter CB Phillips (1993). “The spurious effect of unit roots on vector autoregressions: an analytical study”. In: *Journal of Econometrics* 59.3, pp. 229–255.
Uhlig, Harald (Mar. 2005). “What are the effects of monetary policy on output? Results from an agnostic identification procedure”. In: *Journal of Monetary Economics* 52 (2), pp. 381–419.
Wu, Jing Cynthia and Fan Dora Xia (2016). “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound”. In: *Journal of Money, Credit and Banking* 48 (2-3), pp. 253–291.
Xu, Yimin and Jakob de Haan (Sept. 2018). “The time-varying relationship between credit spreads and employment growth”. In: *Applied Economics* 50 (41), pp. 4387–4401.

Appendices

A Theory, Methods and Algorithms

A.1 Maximum Likelihood

The derivations from this section summarize and follow the Lecture notes on Maximum Likelihood from Nielsen (2021).

Consider the probability density function of the sample data $f(Y) := f(Y_1, \dots, Y_T)$ and the $(K \times 1)$ dimensional vector of parameters $\theta \in \Theta$. The probability of observing the data conditional on θ is given by the likelihood:

$$\mathcal{L}_T(\theta) := f(Y|\theta) = \prod_{t=1}^{T} f(Y_t|Y_1, \dots, Y_{t-1}; \theta) = \prod_{t=1}^{T} l_t(\theta)$$

The Maximum Likelihood Estimator of θ is then obtained as follows:

$$\widehat{\theta}^{ML} = \arg\ \max_{\theta \in \Theta} \log \mathcal{L}(\theta|Y)$$

The MLE $(\widehat{\theta})$ implies by definition a first order derivative of zero in the maximum, if the likelihood is a continuous and differentiable function. The first order derivative is denoted as the score. For a multivariate model with K variables, the score S_T is a $(K \times 1)$ vector of score contributions. The score evaluated at the maximum likelihood is zero, $S_T(\widehat{\theta}) = 0$ where:

$$S_T(\theta) = \frac{\partial \log \mathcal{L}_T(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \log l_t(\theta)}{\partial \theta} = 0$$

but this is also the case in local maximum/minimum of the function. For this reason, the MLE might be difficult or impossible to derive in practice when functions have multiple maximum points. If there is no closed form solution, we can use a grid search and further numerical optimization techniques.

The second derivative of the likelihood is the Hessian matrix:

$$H_T(\theta) = \frac{\partial^2 \log \mathcal{L}_T(\theta)}{\partial \theta \partial \theta'} = \sum_{t=1}^{T} \frac{\partial^2 \log l_t(\theta)}{\partial \theta \partial \theta'}$$

which is a $(K \times K)$ matrix. If the Hessian is negative at the MLE, then $\widehat{\theta}$ is a local maximum of the log-likelihood. Let first define the Fisher Information matrix as the inverse of the Hessian:

$$I(\theta) = -E[H_t(\theta)]$$

This will later be used for deriving the second order moments of the MLE.

Using a Taylor polynomial expansion on $S_T(\widehat{\theta})$ around the true value θ_0:

$$S_T(\widehat{\theta}) = S_T(\widehat{\theta}_0) + H_T(\theta_0)(\widehat{\theta} - \theta_0) + O_T$$

where O_T include higher order moments. Assuming $O_T \xrightarrow{p} 0$, then:

$$(\widehat{\theta} - \theta_0) = (-H_T(\theta_0))^{-1} S_T(\theta_0) = \left(-\sum_{t=1}^{T} \frac{\partial^2 \log l_t(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(\sum_{t=1}^{T} \frac{\partial \log l_t(\theta)}{\partial\theta} \right)$$

Considering the following assumptions:

$$\text{(i) } \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \log l_t(\theta)}{\partial\theta}\bigg|_{\theta_0} \xrightarrow[LLN]{p} E\left[\frac{\partial \log l_t(\theta)}{\partial\theta}\right] = 0 \tag{44}$$

$$\text{(ii) } \frac{\sqrt{T}}{T}\sum_{t=1}^{T} \frac{\partial \log l_t(\theta)}{\partial\theta}\bigg|_{\theta_0} \xrightarrow[CLT]{d} N(0, J(\theta_0)) \tag{45}$$

$$\text{(iii) } \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log l_t(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta_0} \xrightarrow[LLN]{p} E\left[\frac{\partial^2 \log l_t(\theta)}{\partial\theta\partial\theta'}\right] = -I(\theta_0) \tag{46}$$

$$\text{(iv) } \max_{i,j,k} \left| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^3 \log l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} \right| \le c_T \xrightarrow[LLN]{p} c < \infty \tag{47}$$

This way, using the Law of Large Numbers (LLN) the estimator of the score evaluated at its true value converges towards its expected value which is zero. The score follows the Central Limit Theorem (CLT) and converges in distribution to a Normal Gaussian distribution. The Hessian converges to its expectation by the LLN, which is the negative Fisher Information matrix. Lastly, it is assumed that the third order derivative is bounded by a constant in a neightborhood value of θ_0.

Scaling by $\sqrt{T}$ it can be shown that the MLE is asymptotically normal:

$$\sqrt{T}(\widehat{\theta} - \theta_0) = \left(-\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log l_t(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T} \frac{\partial \log l_t(\theta)}{\partial\theta} \right) \xrightarrow{d} N\left(0, I(\theta_0)^{-1}\right)$$

where $J(\theta_0) = E\left[\frac{\partial \log l_t(\theta)}{\partial\theta}\left(\frac{\partial \log l_t(\theta)}{\partial\theta}\right)'\right]$. Consistency is then given by: $\widehat{\theta} \xrightarrow{p} \theta_0$. This way, the MLE is efficient as there is no other estimator with lower variance.

However in practice, the distribution is never fully Gaussian. If the distribution has been approximated by a Gaussian we are obtaining the quasi MLE. Using the same assumptions as before we now have that the quasi MLE:

$$\sqrt{T}(\widehat{\theta} - \theta_0) \xrightarrow{d} N\left(0, I(\theta_0)^{-1} J(\theta_0) I(\theta_0)^{-1}\right)$$

where the variance uses the robust errors. Estimates are still consistent but inference will now be affected as the second order moments are different.

The asymptotic distribution of the MLE is:

$$\widehat{\theta} \sim N(\theta_0, \Sigma_\theta) \tag{48}$$

Now we can conduct inference in the form of hypothesis testing. Consider the null H_0 and alternative H_A hypothesis:

$$H_0 : R\theta_0 = a, \quad \text{and} \quad H_0 : R\theta_0 \neq a \tag{49}$$

where R is a $(q \times K)$ matrix that imposes q linear restrictions.

The Likelihood Ratio test statistic is defined as:

$$LR = -2[\log \mathcal{L}(R\widehat{\theta}) - \log \mathcal{L}(\widehat{\theta})] \sim \chi^2_q$$

A Wald test on the MLE estimates given by (48) is asymptotically distributed as:

$$R\widehat{\theta} \sim N(R\theta_0, R\Sigma_\theta R') \tag{50}$$

Under the null, Wald test W statistic is:

$$W = (R\widehat{\theta} - a)'(R\Sigma_\theta R')^{-1}(R\widehat{\theta} - a) \sim \chi^2_q \tag{51}$$

In practice this can be calculated with some estimator of the variance: $\widehat{\Sigma}_\theta$. The Wald test is two-sided because is based on the squared distance between the restricted estimates and the null, and normalized by the variance.

The one-sided t-test statistic for a single restriction $H_0 : \theta_{0,i} = b$, is then computed with:

$$t = \Sigma^{-1/2}_{\theta_{0,i}}(\widehat{\theta}_{0,i} - b) \sim N(0, 1) \tag{52}$$

for a given estimator of $\widehat{\Sigma}_{\theta_{0,i}}$. This t-ratio equals a Wald test with one restriction.

A.2 Maximum Likelihood in VAR

Consider the VAR(p) model $Y_t \in \mathbb{R}$:

$$\begin{aligned} Y_t &= v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t = \mathbf{A}X_{t-1} + u_t, \\ u_t &\sim N(0, \Omega_u) \end{aligned}$$

with $\mathbf{A} = [v, A_1, \ldots, A_p]$ and $X_{t-1} = [\mathbf{1}', Y'_{t-1}, \ldots, Y'_{t-p}]'$ and where $\mathbf{1}'$ is a vector of 1s. The Gaussian likelihood and log-likelihood density functions are given by:

$$f(Y|\theta) = \mathcal{L}_T(\theta) = \prod_{t=1}^{T} l_t(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^K |\Omega_u|}} \exp\left(-\frac{1}{2} u'_t \Omega_u^{-1} u_t\right)$$

$$\log \mathcal{L}_T(\theta) = -\frac{TK}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \log|\Omega_u| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - \mathbf{A}X_{t-1})'\Omega_u^{-1}(Y_t - \mathbf{A}X_{t-1})$$

where the given data is $Y = [Y_t, Y_{t-1}, \ldots, Y_0]$, and the parameters $\theta = (\mathbf{A}, \Omega_u) \in \Theta$. The Maximum Likelihood Estimates $\widehat{\theta} = (\widehat{\mathbf{A}}, \widehat{\Omega}_u)$ are derived from the condition:

$$\widehat{\theta} := \text{arg max}_{\theta \in \Theta} \log \mathcal{L}_T(\theta)$$

Optimization implies:

$$\begin{aligned} S_T(\theta) = \begin{bmatrix} S_T(\mathbf{A}) \\ S_T(\Omega_u) \end{bmatrix} &= \begin{bmatrix} \frac{\partial \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \mathbf{A}} \\ \frac{\partial \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \Omega_u} \end{bmatrix} \\ &= \begin{bmatrix} -2\Omega_u^{-1} \sum_{t=1}^T (Y_t - \mathbf{A}X_{t-1})X'_{t-1} \\ -\frac{T}{2\Omega_u} + \frac{1}{2}\Omega_u^{-2} \sum_{t=1}^T u_t u'_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \end{aligned}$$

and assuming Ω_u is positive definite, the first order condition is satisfied. The MLE is actually a quasi MLE, because the estimate for the variance matrix relies on the estimate of the coefficients. These MLE are:

$$\widehat{\mathbf{A}} = \left(\sum_{t=1}^T X_{t-1}X'_{t-1} \right)^{-1} \left(\sum_{t=1}^T Y_t X'_{t-1} \right) \quad \text{and} \quad \widehat{\Omega}_u = \frac{1}{T} \sum_{t=1} \widehat{u}_t \widehat{u}'_t$$

After correcting for degrees of freedom in the variance estimate, the MLE is equal to the OLS estimates obtained from: $\widehat{\theta}^{OLS} = \text{arg min} \sum_{t=1}^T u'_t u_t$. The OLS estimates are derived in Section A.3.

The first order condition for the coefficient's can be rewritten as:

$$T^{-1} \sum_{t=1}^T \widehat{u}_t X'_{t-1} = \mathbf{0}$$

which is derived from the regularity condition (44).

Let the Hessian be the 2nd order derivative with respect to the vectorized coefficient and covariance matrices:

$$H_T(\theta) = \begin{bmatrix} \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\mathbf{A}) \partial \text{vec}(\mathbf{A})'} & \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\mathbf{A}) \partial \text{vec}(\Omega_u)'} \\ \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\Omega_u) \partial \text{vec}(\mathbf{A})'} & \frac{\partial^2 \log \mathcal{L}_T(\mathbf{A},\Omega_u)}{\partial \text{vec}(\Omega_u) \partial \text{vec}(\Omega_u)'} \end{bmatrix}$$

The block diagonal Information matrix is then obtained under the assumptions $E[u_t] = 0$ and $E[u_t u'_t] = \Omega_u$, which implies that residuals must have zero mean and constant variance, ruling out heteroskedasticity for the estimation of $I(\theta)$:

$$I(\theta) = -E[H_T(\theta)] = \begin{bmatrix} E[X_{t-1}X'_{t-1}] \otimes \Omega_u^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\Omega_u^{-2} \end{bmatrix}$$

Now the Fisher Information matrix is block diagonal.

Consider that the VAR(p) is stable such that Y_t is stationary and weakly dependent with homoscedastic error term $u_t \sim N(0, \Omega_u)$. Consider the general regularity conditions given by (44), (45) and (46):

$$\text{(i)}\ \frac{1}{T}\begin{bmatrix} -\Omega_u^{-1}\sum_{t=1}^{T} u_t X_{t-1}' \\ -\frac{T}{2\Omega_u} + \frac{1}{2}\Omega_u^{-2}\sum_{t=1}^{T} u_t u_t' \end{bmatrix} \xrightarrow[LLN]{P} \begin{bmatrix} -\Omega_u^{-1} E[u_t X_{t-1}'] \\ -\Omega_u^{-1} + \Omega_u^{-2} E[u_t u_t'] \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$\text{(ii)}\ \frac{\sqrt{T}}{T}\begin{bmatrix} \sum_{t=1}^{T} u_t X_{t-1}' \\ -\frac{T}{2\Omega_u} + \frac{1}{2}\Omega_u^{-2}\sum_{t=1}^{T} u_t u_t' \end{bmatrix} \xrightarrow[CLT]{d} N(\mathbf{0}, J(\theta))$$

$$\text{(iii)}\ \frac{1}{T} H_T(\theta) \xrightarrow[LLN]{p} E[H_T(\theta)] = -I(\theta)$$

which require the LLN and CLT to apply. The third derivative from condition (47) equals zero for every element in $\mathbf{A}$. Given these conditions, MLE have asymptotic distribution given by:

$$\sqrt{T}\begin{bmatrix} \text{vec}(\widehat{\mathbf{A}} - \mathbf{A}) \\ \text{vec}(\widehat{\Omega}_u - \Omega_u) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0},\ \begin{bmatrix} E[X_{t-1}X_{t-1}'] \otimes \Omega_u^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\Omega_u^{-2} \end{bmatrix}\right)$$

This allows hypothesis testing on the estimated parameters. Wald and LR tests have χ^2 asymptotic distributions while t-ratios for any parameter in A_i have a standard normal asymptotic distribution. $(\widehat{a}_{mn,i} - a_{mn,i})/\omega_{a_{mn,i}} \sim N(0,1)$ where $\omega_{a_{mn,i}}$ is the corresponding standard error. If these t-ratios had fatter tails we should use a t-distribution instead. But a t-distribution with $T - Kp - 1$ degrees of freedom is very close to the standard normal whenever T is large enough.

A.3 Ordinary Least Squares in VAR

The derivations from this section summarize and follow the Lecture notes on Ordinary Least Squares from Nielsen (2021).

Consider the VAR(p) model:

$$Y_t = v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t = \mathbf{A} X_{t-1} + u_t,$$
$$u_t \sim N(0, \Omega_u)$$

with $\mathbf{A} = [v, A_1, \ldots, A_p]$ and $X_{t-1} = [\mathbf{1}', Y_{t-1}', \ldots, Y_{t-p}']'$ and where $\mathbf{1}'$ is a vector of 1s.

The OLS estimation minimizes:

$$
\begin{aligned}
\widehat{\mathbf{A}}^{OLS} &:= \arg\min_{\mathbf{A}} \sum_{t=1}^{T} u_t' u_t \\
&= \arg\min_{\mathbf{A}} \sum_{t=1}^{T} (Y_t - \mathbf{A}X_{t-1})'(Y_t - \mathbf{A}X_{t-1}) \\
&= \arg\min_{\mathbf{A}} \sum_{t=1}^{T} Y_t'Y_t - Y_t'\mathbf{A}X_{t-1} - X_{t-1}'\mathbf{A}'Y_t - X_{t-1}'\mathbf{A}'\mathbf{A}X_{t-1} \\
&= \arg\min_{\mathbf{A}} \sum_{t=1}^{T} Y_t'Y_t - 2X_{t-1}'\mathbf{A}'Y_t - X_{t-1}'\mathbf{A}'\mathbf{A}X_{t-1}
\end{aligned}
$$

minimization requires making the first order derivative equal to zero:

$$
\frac{\partial \sum_{t=1}^{T} u_t' u_t}{\partial \mathbf{A}} = \sum_{t=1}^{T} -2Y_t X_{t-1}' - 2\mathbf{A}X_{t-1}X_{t-1}' = 0
$$

and we can obtain the OLS estimates:

$$
\widehat{\mathbf{A}}^{OLS} = \sum_{t=1}^{T} \left(X_{t-1}X_{t-1}'\right)^{-1} \left(Y_t X_{t-1}'\right)
$$

if $X_{t-1}X_{t-1}'$ is non-singular and invertible. The OLS estimate of the parameters equals the MLE.

The standard errors can be derived with the 2nd derivative:

$$
\frac{\partial^2 \sum_{t=1}^{T} u_t' u_t}{\partial \mathbf{A} \partial \mathbf{A}'} = \sum_{t=1}^{T} 2X_{t-1}X_{t-1}'
$$

This is the $(K \times K)$ Hessian matrix, which is positive definite by construction. This means that the OLS estimates were a minimum and the expression above has been minimized instead of maximized.

Then OLS estimates are consistent if the LLN applies to the sample average estimates to converge towards their respective expectations:

$$
\begin{aligned}
T^{-1} \sum_{t=1}^{T} Y_t X_{t-1}' &\xrightarrow[LLN]{P} E[Y_t X_{t-1}'] \\
T^{-1} \sum_{t=1}^{T} X_{t-1} X_{t-1}' &\xrightarrow[LLN]{P} E[X_{t-1} X_{t-1}']
\end{aligned}
$$

The OLS estimator's consistency is derived from the two conditions above. Similarly, if we assume that plim $T^{-1} \sum_{t=1}^{T} u_t X_{t-1}' = E[u_t X_{t-1}'] = \mathbf{0}$, it can be show that the probability limit of the estimator is its true value:

$$
\begin{aligned}
\text{plim } \widehat{\mathbf{A}} &= \left(\text{plim } T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(\text{plim } T^{-1}\sum_{t=1}^{T} Y_t X'_{t-1}\right) \\
&= \left(\text{plim } T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(\text{plim } T^{-1}\sum_{t=1}^{T} (\mathbf{A}X_{t-1}+u_t) X'_{t-1}\right) \\
&= \mathbf{A} + \left(\text{plim } T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(\text{plim } T^{-1}\sum_{t=1}^{T} u_t X'_{t-1}\right) \\
&= \mathbf{A}
\end{aligned}
$$

For OLS to be unbiased, we need the unconditional expectation of the estimator to be equal to the true value: $E[\widehat{\mathbf{A}}] = \mathbf{A}$. A brief proof follows:

$$
\begin{aligned}
E[\widehat{\mathbf{A}}] &= E\left[\left(T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} Y_t X'_{t-1}\right)\right] \\
&= \left[\left(T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} (\mathbf{A}X_{t-1}+u_t) X'_{t-1}\right)\right] \\
&= \ \mathbf{A} + E\left[\left(T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} u_t X'_{t-1}\right)\right]
\end{aligned}
$$

For the second term to be equal to zero, strict exogeneity is needed: $E[u_i X'_j | X_0, \ldots, X_T] = \mathbf{0}, \forall i, j$. Then the error term has zero unconditional expectation: $E[u_t | X_0, \ldots, X_T) = 0$, which implies no correlation between the error term with past, current and future values of the regressors X_t. Then:

$$
\begin{aligned}
E[\widehat{\mathbf{A}}] &= \ \mathbf{A} + E\left[\left(T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} u_t X'_{t-1}\right)\right] \\
&= \ \mathbf{A}\left(T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} E[u_t | X_0, \ldots, X_T] X'_{t-1}\right) \\
&= \ \mathbf{A}
\end{aligned}
$$

However this cannot be fulfilled in times series models where regressors are lagged values of the endogenous variable:

$$
E[u_t Y'_t] = E\left[u_t \left(\mathbf{A}X_{t-1} + u_t\right)'\right] \neq 0 \ \text{ because } \ E[u_t u'_t] = \Omega_u > 0
$$

This means that OLS is biased by construction in time series.

A.4 Weighted Least Squares in VAR

Consider the VAR(p) model:

$$\begin{aligned}
Y_t &= v + \sum_{i=1}^{p} A_i Y_{t-i} + \Omega_t^{1/2} v_t = \mathbf{A} X_{t-1} + \Omega_t^{1/2} v_t, \\
u_t &\sim N(\mathbf{0}, I_K)
\end{aligned}$$

where $\Omega_t = \text{diag}(\omega_1, \ldots, \omega_K)$ such that residuals are serially uncorrelated ($\omega_{ij} = 0$) and heteroskedastic ($\omega_i \neq \omega_j$). For $v_t = \Omega_t^{-1/2}(Y_t - \mathbf{A}X_{t-1})$, the WLS estimate minimizes:

$$\begin{aligned}
\widehat{\mathbf{A}}^{WLS} &= \arg\min \sum_{t=1}^{T} v_t' v_t \\
&= \arg\min \sum_{t=1}^{T} (Y_t - \mathbf{A}X_{t-1})' \Omega_t^{-1} (Y_t - \mathbf{A}X_{t-1})
\end{aligned}$$

and the WLS estimate is such that: $\frac{\partial \sum_{t=1}^{T} v_t' v_t}{\partial \mathbf{A}} = \mathbf{0}$, hence:

$$\text{vec}\left(\widehat{\mathbf{A}}^{WLS}\right) = \sum_{t=1}^{T} \left(X_{t-1} X_{t-1}' \otimes \widehat{\Omega}_t^{-1}\right)^{-1} \left(\left(X_{t-1} \otimes \widehat{\Omega}_t^{-1}\right) Y_t\right)$$

In practice Ω_t is unknown and we rely on an estimate. The Feasible WLS (FWLS) can be obtained in two steps:

- Obtain $\widehat{\Omega}_t$ with OLS or ML.
- Obtain: $\text{vec}\left(\widehat{\mathbf{A}}^{WLS}\right) = \sum_{t=1}^{T} \left(X_{t-1} X_{t-1}' \otimes \widehat{\Omega}_t^{-1}\right)^{-1} \left(\left(X_{t-1} \otimes \widehat{\Omega}_t^{-1}\right) Y_t\right)$

OLS treats outliers equally in the minimization regardless of whether they appear in regions of high or low variance. Under hetersokedastcity, WLS takes into account the volatility of data in each region in order to estimate the coefficients. An outlier would change the slope of the parameters significantly only if it is in a low volatility region but not that much when it is in a low volatility region.

A.5 Impulse Responses

This section expands on the derivation of Impulse Responses by extending the univariate derivation from Nielsen (2021) lecture notes to the VAR case.

Consider the K-dimensional linear VAR(p): $Y_t \in \mathbb{R}^K$

$$Y_t = v + \sum_{i=1}^{p} A_i Y_{t-i} + u_t$$

where $Y_t = [y_{1,t}, y_{2,t}, \ldots, y_{k,t}]'$ is a $(K \times 1)$ vector of observable variables with length T, v is a $(K \times 1)$ vector of constants and the $(K \times 1)$ error term is a white noise process $u_t \sim N(0, \Omega_u)$.

That process can be rewritten in companion form:

$$Z_t = \nu + \Pi Z_{t-1} + U_t$$

such that Z_t is a VAR(1) process with Z_t and Z_{t-1} of dimensions $(pK \times K)$, ν and U_t of dimensions $(pK \times 1)$ and Π of dimensions $(pK \times pK)$:

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} v \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_K & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & I_K & \mathbf{0} \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ \vdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

The MA solution can be derived by recursive substitution:

$$\begin{aligned} Z_t &= \nu + \Pi Z_{t-1} + U_t \\ &= \nu + \Pi(v + \Pi Z_{t-2} + U_{t-1}) + U_t \\ &= (I_K + \Pi)\nu + \Pi^0 U_t + \Pi U_{t-1} + \Pi^2 Z_{t-2} \\ &\vdots \\ Z_t &= (I_K + \Pi + \ldots + \Pi^{t-1})\nu + \sum_{i=0}^{t-1} \Pi^i U_{t-i} + \Pi^t Z_0 \end{aligned}$$

such that $\Pi^0 = I_{pK}$, $\Pi^2 = \Pi\Pi$, $\Pi^3 = \Pi\Pi\Pi$. The MA representation consists on a deterministic part, a moving average of the error terms and the initial values. For some given starting values and $E[U_t|Z_0] = \mathbf{0}$, the conditional expectation is given by:

$$\begin{aligned} E[Z_t|Z_0] &= (I_K + \Pi + \ldots + \Pi^{t-1})\nu + \sum_{i=0}^{t-1} \Pi^i E[U_{t-i}|Z_0] + \Pi^t Z_0 \\ &= (I_K + \Pi + \ldots + \Pi^{t-1})v + \Pi^t Z_0 \end{aligned}$$

If the matrices $\Pi^t \xrightarrow{P} \mathbf{0}$ as $T \to \infty$, then the effect of the effect of the initial values vanish: $\Pi^t Z_0 \xrightarrow{P} \mathbf{0}$. This is the case when the model is stable. Consider the spectral decomposition of Π:

$$\Pi = V \Lambda V^{-1}$$

where the eigenvectors are given by $V = [v_1, \ldots, v_{pK}]$ while the matrix of eigenvalues is $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{pK})$. These can be obtained as the solution to the problem: $|\Pi - \lambda_i I_p| = 0$. It holds that: $\Pi^n = V \Lambda^n V^{-1}$, where $\Lambda^n = \mathrm{diag}(\lambda_1^n, \ldots, \lambda_{pK}^n)$. If the eigenvalues are all inside the unit circle and the stability condition is satisfied, then the $\Pi^n \to \mathbf{0}$

exponentially fast. For that reason the effect of the initial values vanish form the MA representation of Z_t.

For a length of $p = 3$ we have that:

$$\Pi^2 = \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_K & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_K & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$
$$= \begin{bmatrix} A_1A_1 + A_2 & A_1A_2 + A_3 & A_1A_3 + A_4 & \dots & A_1A_p + A_p \\ A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$

$$\Pi^3 = \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_K & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_K & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_K & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_K & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$
$$= \begin{bmatrix} \Pi^3_{11} & \Pi^3_{12} & \Pi^3_{13} & \dots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$

where:

$$\begin{aligned} \Pi^3_{11} &= A_1A_1A_1 + A_1A_2 + A_1A_2 + A_3 \\ \Pi^3_{12} &= A_1A_1A_2 + A_2A_2 + A_1A_3 + A_4 \\ \Pi^3_{13} &= A_1A_1A_3 + A_2A_3 + A_1A_4 + A_5 \\ &\vdots \end{aligned}$$

Let Π^i_{mn} be the mth row and nth column element from Π^i. If we are interest in recovering Y_t we only need Π^i_{11}, the first element of each matrix Π^i. The process becomes:

$$\begin{aligned} Y_t &= (\Pi^0_{11} + \Pi_{11} + \Pi^2_{11} + \Pi^3_{11} + \dots)v + \Pi^0_{11}u_t + \Pi_{11}u_t + \Pi^2_{11}u_{t-2} + \Pi^3_{11}u_{t-3} + \dots \\ &= (I_K + A_1 + (A_1A_1 + A_2) + (A_1A_1A_1 + A_1A_2 + A_1A_2 + A_3) + \dots)v \\ &+ I_Ku_t + A_1u_t + (A_1A_1 + A_2)u_{t-2} + (A_1A_1A_1 + A_1A_2 + A_1A_2 + A_3)u_{t-3} + \dots \end{aligned}$$

where $\Pi_{11}^0 = I_K,\ \Pi_{11}^1 = \Pi_{11}^0 A_1,\ \Pi_{11}^2 = \Pi_{11}^1 A_1 + \Pi_{11}^0 A_2,\ \Pi_{11}^3 = \Pi_{11}^2 A_1 + \Pi_{11}^1 A_2 + \Pi_{11}^0 A_3$ and so on. For an horizon of $\aleph$ the moving average process becomes:

$$
\begin{aligned}
Y_t &= \sum_{h=0}^{\aleph} \Pi_{11}^h v + \sum_{h=0}^{\aleph} \Pi_{11}^h u_{t-h} \\
&= \sum_{h=0}^{\aleph} \Pi_{11}^h v + \sum_{h=0}^{\aleph} \Pi_{11}^h B \epsilon_{t-h} \\
&= \widetilde{v} + \sum_{h=0}^{\aleph} \zeta_h \epsilon_{t-h}
\end{aligned}
$$

which is equivalent to the Impulse Responses in (14). $\zeta_h = \Pi_{11}^h B = \partial Y_t / \partial \epsilon_{t-h}'$ is the $(K \times K)$ response of the vector of variables Y_t to an exogenous impulse to the vector of structural shocks ϵ_t. Above it can be seen that the IRFs are functions of the parameters:

$$\zeta_h = \zeta_h(v, A_1, \dots, A_p, B)$$

A.6 Bootstrapping IRFs in linear VAR

Residual based bootstrap methods fail. The Wild bootstrap provides a suitable alternative without imposing structure on the heteroskedasticity form.

The fixed design Wild bootstrap algorithm for simulation of Confidence Intervals for structural IRFs is as follows:

- **Step 1:** Obtain OLS estimates of $\widehat{\mathbf{A}}$ and $\widehat{u}_t$ using the full sample. Then, draw $m = 1, \dots, M$ observations from the scalar random process given by:

  $$\eta_t \sim i.i.d.N(0, 1)$$

  Construct a sequence of M bootstrap residuals $(u_t^{(1)}, \dots, u_t^{(M)})$ by applying the draws of η_t obtained before to the following model:

  $$u_t^{(m)} = \widehat{u}_t \eta_t^{(m)}$$

- **Step 2:** Generate a sequence of synthetic data $(Y_t^{(1)}, \dots, Y_t^{(M)})$ by fitting the following model:

  $$Y_t^{(m)} = \widehat{\mathbf{A}} X_{t-1} + u_t^{(m)}$$

  using the original estimates $\widehat{\mathbf{A}}$ and the bootstrap residuals from **Step 1** and the original sample regressors X_{t-1}.

- **Step 3:** Estimate bootstrap parameters $(\widehat{\mathbf{A}}^{(1)}, \ldots, \widehat{\mathbf{A}}^{(M)})$, residuals $(\widehat{u}_t^{(1)}, \ldots, \widehat{u}_t^{(M)})$ and covariances $(\widehat{\Omega}_u^{(1)}, \ldots, \widehat{\Omega}_u^{(M)})$ using the new synthetic data:

$$
\begin{aligned}
\widehat{\mathbf{A}}^{(m)} &= \left(\sum_{t=1}^{T} X_{t-1}^{(m)} (X_{t-1}^{(m)})'\right)^{-1} \left(\sum_{t=1}^{T} Y_t^{(m)} (X_{t-1}^{(m)})'\right) \\
\widehat{u}_t^{(m)} &= Y_t^{(m)} - \widehat{\mathbf{A}}^{(m)} X_{t-1}^{(m)} \\
\widehat{\Omega}_u^{(m)} &= \frac{1}{T - Kp - 1} \sum_{t=1} \widehat{u}_t^{(m)} (\widehat{u}_t^{(m)})'
\end{aligned}
$$

  where sequence of synthetic regressors $(X_{t-1}^{(1)}, \ldots, X_{t-1}^{(M)})$ is constructed from the synthetic data generated in **Step 2** as follows: $X_{t-1}^{(m)} = [\mathbf{1}', (Y_{t-1}^{(m)})', \ldots, (Y_{t-p}^{(m)})']'$. Then recover the matrix of contemporaneous effects: $(\widehat{B}^{(1)}, \ldots, \widehat{B}^{(M)})$ from a Cholesky decomposition of each of the covariances in $(\widehat{\Omega}_u^{(1)}, \ldots, \widehat{\Omega}_u^{(M)})$.

- **Step 4:** Construct M bootstrap estimates of the IRFs: $(\zeta_h^{(1)}, \ldots, \zeta_h^{(M)})_{h=0}^{\aleph}$, as these are functions of the bootstrap parameter estimates $\zeta_h^{(m)} = \zeta_h(\widehat{\mathbf{A}}^{(m)}, \widehat{B}^{(m)})$. The set of bootstrapped IFRs: $\{\widehat{\zeta}_{mn,h}^{(m)}\}_{m=1}^{M}$ is composed out of $h = 0, \ldots, \aleph$ point-wise distributions for each $\widehat{\zeta}_{mn}$: $f(\widehat{\zeta}_{mn,h})$.

- **Step 5:** The 90% confidence interval for IRFs can be created using Hall (1992) percentile interval:

$$
CI^{90\%} = [\widehat{\zeta}_{mn,h} - (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{5\%}),\ \widehat{\zeta}_{mn,h} + (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{95\%})] \tag{53}
$$

  where $f(\widehat{\zeta}_{mn,h})^{Q\%}$ is the Qth quantile of the bootstrap distribution of $\widehat{\zeta}_{mn}$ at horizon h.

A.7 Bootstrapping IRFs in ST-VAR

The fixed design Wild bootstrap algorithm for simulation of Confidence Intervals for structural IRFs in the ST-VAR is as follows:

- **Step 1:** For a given transition function $F(s_t, \gamma, c)$, obtain feasible WLS estimates of $\widehat{\mathbf{A}}$ and $\widehat{u}_t$ using the full sample and the multiple-steps algorithm for the estimation of ST-VAR described in Section 4.1.2. Recall that the algorithm combines MLE and WLS for obtaining these estimates after using OLS as initial values. Then, draw $m = 1, \ldots, M$ observations from the scalar random process given by:

$$
\eta_t \sim i.i.d.N(0,1)
$$

  Construct a sequence of M bootstrap residuals $(u_t^{(1)}, \ldots, u_t^{(M)})$ by applying the draws of η_t obtained before to the following model:

$$
u_t^{(m)} = \widehat{u}_t \eta_t^{(m)}
$$

- **Step 2:** Generate a sequence of synthetic data $(Y_t^{(1)}, \dots, Y_t^{(M)})$ by fitting the following model:
$$Y_t^{(m)} = \widehat{\mathbf{A}} X_{t-1} + u_t^{(m)}$$
using the original estimates $\widehat{\mathbf{A}}$ and the bootstrap residuals from **Step 1** and the original sample regressors X_{t-1}.

- **Step 3:** Construct the sequence of synthetic regressors $(X_{t-1}^{(1)}, \dots, X_{t-1}^{(M)})$ using the synthetic data generated in **Step 2** as follows: $X_{t-1}^{(m)} = [\mathbf{1}', (Y_{t-1}^{(m)})', \dots, (Y_{t-p}^{(m)})']'$. Estimate The bootstrap matrices of contemporaneous effects $(\widehat{B}^{(1)}, \dots, \widehat{B}^{(M)})$, eigenvalues $(\widehat{\Lambda}^{(1)}, \dots, \widehat{\Lambda}^{(M)})$, bootstrap parameters $(\widehat{\mathbf{A}}^{(1)}, \dots, \widehat{\mathbf{A}}^{(M)})$, residuals $(\widehat{u}_t^{(1)}, \dots, \widehat{u}_t^{(M)})$ and covariances $(\widehat{\Omega}_u^{(1)}, \dots, \widehat{\Omega}_u^{(M)})$ by applying the ST-VAR algorithm (Section 4.1.2) to the new synthetic data sets.

- **Step 4:** Construct M bootstrap estimates of the IRFs: $(\zeta_h^{(1)}, \dots, \zeta_h^{(M)})_{h=0}^{\aleph}$, as these are functions of the bootstrap parameter estimates $\zeta_h^{(m)} = \zeta_h(\widehat{\mathbf{A}}^{(m)}, \widehat{B}^{(m)})$. The set of bootstrapped IFRs: $\{\widehat{\zeta}_{mn,h}^{(m)}\}_{m=1}^{M}$ is composed out of $h = 0, \dots, \aleph$ point-wise distributions for each $\widehat{\zeta}_{mn}$: $f(\widehat{\zeta}_{mn,h})$.

- **Step 5:** The 90% confidence interval for IRFs can be created using Hall (1992) percentile interval:
$$CI^{90\%} = [\widehat{\zeta}_{mn,h} - (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{5\%}),\ \widehat{\zeta}_{mn,h} + (\widehat{\zeta}_{mn,h} - f(\widehat{\zeta}_{mn,h})^{95\%})] \tag{54}$$
where $f(\widehat{\zeta}_{mn,h})^{Q\%}$ is the Qth quantile of the bootstrap distribution of $\widehat{\zeta}_{mn}$ at horizon h.

A.8 Metropolis-Hastings Algorithm

This section explains the Metropolis-Hastings algorithm taken from Auerbach and Gorodnichenko (2012) and Chernozhukov and Hong (2003) used for the estimation of the NL-ST-VAR model. The algorithm constructs a Markov Chain with $M = 100000$ iterations and then discards the first 20% a burn-in. Consider a posterior distribution $f(\theta|Y)$ known up to a constant scaling factor $f(\theta|Y) \propto f(Y|\theta)f(\theta) = e^{\log \mathcal{L}(\theta)} f(\theta)$ for the vector of parameters θ, where the likelihood is given by $f(Y|\theta) = \mathcal{L}_T(\theta)$ and the prior density $f(\theta)$ is chosen as explained in Section 5.2.1.

The Metropolis-Hastings algorithm proceeds as follows:

- **Step 1:** Set $m = 0$ and some initial values $\theta^{(0)} = \{\Omega^{L(0)}, \Omega^{H(0)}, \mathbf{A}^{L(0)}, \mathbf{A}^{H(0)}\}$.

  Obtain the MLE: $(\widehat{B}^{L,MLE}, \widehat{B}^{H,MLE})$ using a Newton-type algorithm for optimization that minimizes the negative log-Likelihood function $(-\log \mathcal{L}_T^*)$ derived from (34) and (35) with a maximum of 1000 iterations. This is equivalent to maximizing

$\log \mathcal{L}_T^*$:

$$\arg \max_{B^L, B^H} \log \mathcal{L}_T^* = \frac{TK}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widehat{u}_t'\Omega_t^{-1}\widehat{u}_t$$
$$\text{where } \Omega_t = (1 - F(s_t, \gamma, c))B^L(B^L)' + F(s_t, \gamma, c)B^H(B^H)'$$

This is done for a given $F(s_t, \gamma, c)$. The starting values for the optimization above are the Cholesky decompositions: $\widetilde{B}^L = \text{chol}(\widetilde{\Omega}^L)$ and $\widetilde{B}^H = \text{chol}(\widetilde{\Omega}^H)$. These initial values take the OLS linear estimate of the covariance for the regime L and a similar matrix that has been 'randomly' modified to take a small departure for computational reasons:

$$\widetilde{\Omega}^L = \widehat{\Omega}_u \quad \text{and} \quad \widetilde{\Omega}^H = \widehat{\Omega}_u + \widehat{\lambda}_{\min}DD'$$

here the $\widehat{\lambda}_{\min}$ is the smallest eigenvalue of $\widehat{\Omega}_u$ and the matrices D contain random numbers drawn from a standard Normal distribution $N(0, 1)$. The initial values of the regime covariances can be constructed with the MLE estimates of B^j for $i = \{L, H\}$:

$$\Omega^{L(0)} = \widehat{B}^{L,MLE}(\widehat{B}^{L,MLE})' \quad \text{and} \quad \Omega^{H(0)} = \widehat{B}^{H,MLE}(\widehat{B}^{H,MLE})'$$

Ω_t is known for a given set $\{\Omega^{L(0)}, \Omega^{H(0)}\}$:

$$\widehat{\Omega}_t = (1 - F(s_t, \gamma, c))\Omega^{L(0)} + F(s_t, \gamma, c)\Omega^{L(0)}$$

Now the model is linear with respect to $\{\mathbf{A}^L, \mathbf{A}^H\}$:

$$\arg \max_{\mathbf{A}^L, \mathbf{A}^H} \log \mathcal{L}_T^* = \frac{TK}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\widehat{\Omega}_t| - \frac{1}{2}\sum_{t=1}^{T}u_t'\widehat{\Omega}_t^{-1}u_t$$

where

$$\begin{aligned}
u_t &= Y_t - \Big[(1 - F(s_t, \gamma, c))\mathbf{A}^L X_{t-1} + F(s_t, \gamma, c)\mathbf{A}^H X_{t-1}\Big] \\
&= Y_t - \begin{bmatrix}\mathbf{A}^L & \mathbf{A}^H\end{bmatrix}\begin{bmatrix}(1 - F(s_t, \gamma, c))X_{t-1} \\ F(s_t, \gamma, c)X_{t-1}\end{bmatrix} \\
&= Y_t - \begin{bmatrix}\mathbf{A}^L & \mathbf{A}^H\end{bmatrix} W_t
\end{aligned}$$

The Feasible WLS estimate of $[\mathbf{A}^L, \mathbf{A}^H]$ is given by:

$$\text{vec}\begin{bmatrix}\widehat{\mathbf{A}}^{L,WLS} & \widehat{\mathbf{A}}^{H,WLS}\end{bmatrix} = \sum_{t=1}^{T}\left(W_t W_t' \otimes \widehat{\Omega}_t^{-1}\right)^{-1}\left(\left(W_t \otimes \widehat{\Omega}_t^{-1}\right)Y_t\right) \tag{55}$$

Now $\mathbf{A}^{L(0)} = \widehat{\mathbf{A}}^{L,WLS}$ and $\mathbf{A}^{H(0)} = \widehat{\mathbf{A}}^{H,WLS}$ can be used as starting values for the Monte Carlo algorithm.

- **Step 2:** Consider $\Psi^{(m)} = \{B^{L(m)}, B^{H(m)}\}$ such that each covariance matrix can be constructed by: $\Omega^{j(m)} = B^{j(m)}(B^{j(m)})', j = \{L, H\}$. Draw a candidate vector $\Upsilon^{(m)}$ where: $\Upsilon^{(m)} = \Psi^{(m)} + \psi^{(m)}$. We are interested in $\Psi^{(m)}$ but the draw is constructed to include the noise from $\psi \sim i.i.d.N(0, \Sigma_\Psi)$ with variance given by $\Sigma_\Psi = \text{diag}(\Sigma_\Psi)$. The variance of ψ is initially set to $\Sigma_\Psi = 0.0025$ and adjusted on the basis of the acceptance rate of the algorithm. When the acceptance rate is too low and falls below 0.25, the variance is tightened to search for candidates closer to the priors. When the acceptance rate is too high and rises above 0.35 then the variance is increased to search further away from the priors. This way the algorithm ensures a high coverage of the distribution of considered candidates. These are the values chosen by Auerbach & Gorodnichenko (2012). Discard the draw $\Upsilon^{(m)}$ if any of the eigenvalues of $\Omega^{L(m)}$ and $\Omega^{H(m)}$ are equal to zero and repeat **Step 2**. If the eigenvalues are greater than zero, continue to **Step 3**. The reason for this is to ensure that the Cholesky decomposition is feasible.

- **Step 3:** Conditional on the convariances in $\Upsilon^{(m)}$, compute $\{\mathbf{A}^{L(0)}, \mathbf{A}^{H(0)}\}$ using the Feasible WLS method from (55). Let now $Q(\theta)$ be the criterion function to minimize which includes the negative log-Likelihood and penalty functions for each parameter within θ [32] to penalise for deviations from prior beliefs:

$$Q(\theta) = -\log \mathcal{L}_T^*(\theta) + \mathbf{1}'(\mathbf{A}^L - \underline{\mathbf{A}})\underline{\Sigma}_\mathbf{A}(\mathbf{A}^L - \underline{\mathbf{A}})'\mathbf{1} + \mathbf{1}'(\mathbf{A}^H - \underline{\mathbf{A}})\underline{\Sigma}_\mathbf{A}(\mathbf{A}^H - \underline{\mathbf{A}})'\mathbf{1}$$
$$+ \mathbf{1}'(\Omega^L - \underline{\Omega})\underline{\Sigma}_\Omega(\Omega^L - \underline{\Omega})'\mathbf{1} + \mathbf{1}'(\Omega^H - \underline{\Omega})\underline{\Sigma}_\Omega(\Omega^H - \underline{\Omega})'\mathbf{1}$$

  using the prior parameters given in Section 5.2.1 and the $(K \times 1)$ vector of 1s $\mathbf{1} = [1, \dots, 1]'$ that convert each penalty functions into a single number[33]. Now define the $(m+1)$th vector of covariances $\Psi^{(m+1)}$ to be equal to $\Upsilon^{(m)}$ with a higher probability[34] if the criterion $Q(\theta)$ evaluates the candidate vector as more suitable than than $\Psi^{(m)}$. Set the $m+1$ state of the chain as:

$$\Psi^{(m+1)} = \begin{cases} \Upsilon^{(m)} & \text{with probability} \quad P(\Psi, \psi) = \min\{1,\ \exp[Q(\Upsilon^{(m)}) - Q(\Psi^{(m)})]\} \\ \Psi^{(m)} & \text{with probability} \quad 1 - P(\Psi, \psi) \end{cases}$$

  where $Q(\Upsilon^{(m)}) = Q(\Upsilon^{(m)}, \mathbf{A}^{L(m)}, \mathbf{A}^{H(m)})$ evaluates the criterion for the candidate vector of parameters while $Q(\Psi^{(m)}) = Q(\Psi^{(m)}, \mathbf{A}^{L(m)}, \mathbf{A}^{H(m)})$ it at the current state of the Markov Chain. This means that the candidate could be accepted even if it yields a higher value of the criterion to minimize. This facilitates the coverage of the entire distribution by considering candidates that are further away from the optimum.

[32] The code also treats γ as one parameter to estimate and hence is included in $Q(\theta)$ with a penalty of $\left(\frac{(\gamma - \underline{\gamma})^2}{\underline{\Sigma}_\gamma}\right)$. Due to the low variance in the prior, γ will be chosen as the prior expected value $\gamma = 3.5$.

[33] This makes use of: $\mathbf{1}'X\mathbf{1}$ being equal to the sum of the elements within the $(K \times K)$ matrix X.

[34] Computationally, this is done by drawing z from the process $z \sim i.i.d.N(0, 1)$ and accepting the candidate vector $\Upsilon^{(m)}$ if $P(\Psi, \psi) > z$.

- **Step 4:** Set $m = m + 1$ and return to **Step 2**. Repeat the loop for $m = 1, \ldots, M$ where $M = 100000$ to ensure convergence towards the quasi-posterior distribution.

B Data Description and Sources

This Appendix briefly describes each variable from the data set included in the models. The data set includes US monthly data from 1980-01 to 2020-01 (both included). The variables are:

- Inflation (π_t): Inflation refers to the year-on-year growth rate of the non-seasonally adjusted Producer Price Index for all commodities. It therefore contains information from all commodities used as inputs or final goods in the production process. It is particularly affected by crude oil and energy prices. It serves to look into price inflation from the perspective of the producer, which means that it is a leading indicator of consumer price inflation. The data can be found in the FRED Economic Database as: Producer Price Index by Commodity, all commodities, not seasonally adjusted (PPIACO).

- Output Growth (x_t): Output growth refers to the year-on-year growth rate of the non-seasonally adjusted Industrial Production Index. This is a measure of output from the industrial sector. Once again, this looks into output from the producer's side, which is highly correlated with consumer's demand. Alternative measures like the Gross Domestic Product are only available in quarterly frequency, but both a highly correlated in the US. The data can be found in the FRED Economic Database as: Industrial Production: Total Index (IPB50001N).

- Interest Rate (i_t): Interest rate refers to the shadow Federal Funds Rate from Wu and Xia (2016). This index is adjusted to include unconventional monetary policies such as asset purchases and forward guidance, which are found to affect the economy in similar ways to changes in short term rates (Wu and Xia 2016, Borio and Zabai 2016). This implies that the shadow rate is allowed to fall below the Zero Lower Bound, better capturing the monetary policy state. The data can be found in Jing Cynthia Wu's website:
https://sites.google.com/view/jingcynthiawu/shadow-rates

- Bond Premium (bp_t): Bond premium refers to the Excess Bond Premium from Gilchrist and Zakrajšek (2011). This variable is constructed by design as the unpredictable component of credit spreads, therefore orthogonal to the rest of the economy by design. They interpret an increase in the excess bond premium as "reduction in the effective risk-bearing capacity of the financial sector" (Gilchrist and Zakrajšek 2011), or as an increase in the price of default risk. This unpredictable component is found to account for most the variation in credit spreads, and has a strong explanatory power in predicting macroeconomic downturns. The data can be found in the Federal Reserve Economic and Research Data:

https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/updating-the-recession-risk-and-the-excess-bond-premium-20161006.html

C Additional Figures

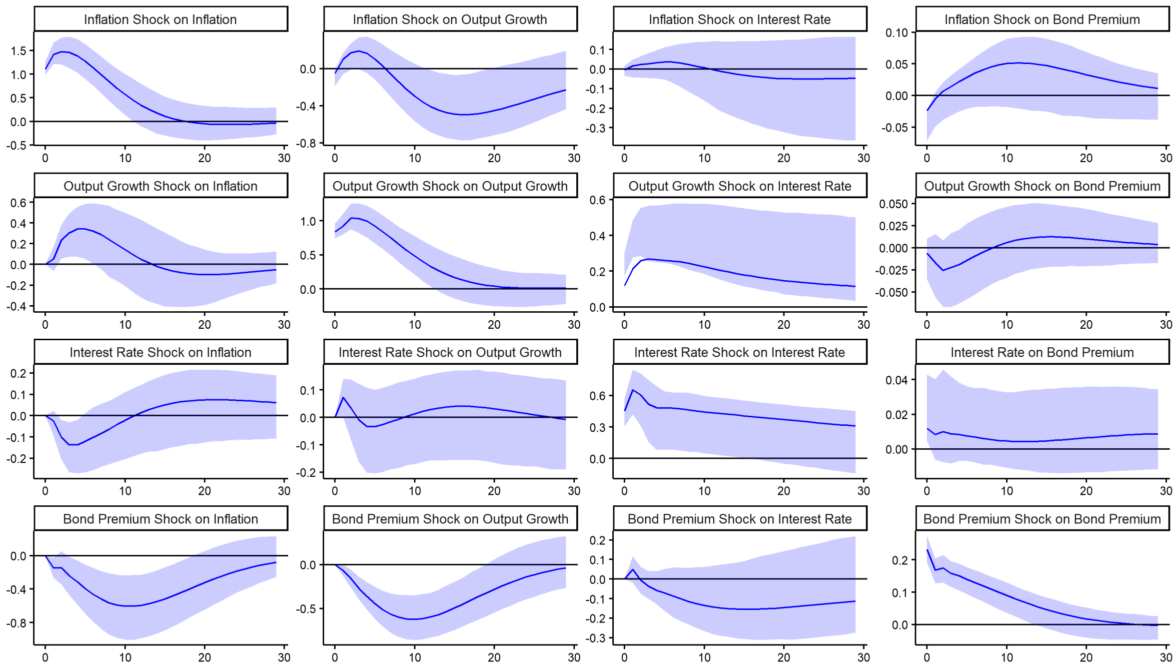


Figure 12: Impulse responses with Cholesky identification for every shock. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws.

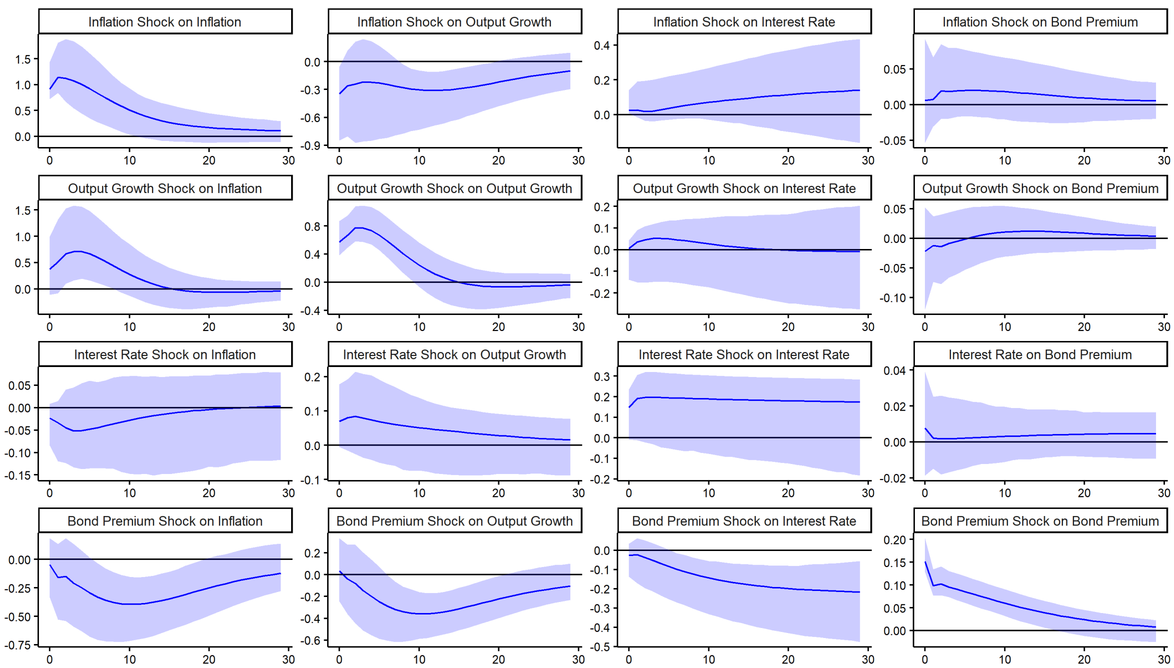


Figure 13: Impulse responses with heteroskedasticity identification for every shock. Impulse responses with heteroskedasticity identification for every shock. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws.

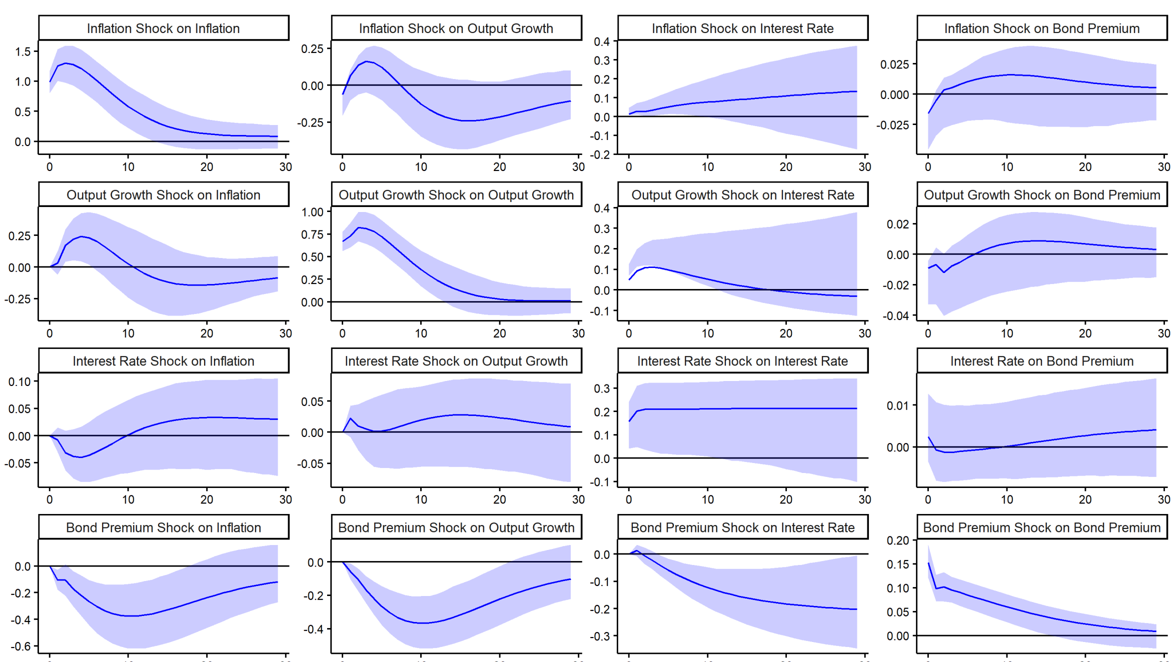


Figure 14: Impulse responses with heteroskedasticity identification and Cholesky over-identifying restrictions for every shock. 90% Hall's confidence interval using a fixed design Wild Bootstrap with 2000 draws.