UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Short Rate Modelling: A Data Driven Approach

by Karol Gellert

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Prof Erik Shlögl

University of Technology Sydney
Faculty of Business, Finance Discipline Group

November 2023

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Karol Gellert, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the Business/Finance Discipline Group at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.
**If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).*

**If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).*

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 09/03/2023

Short Rate Modelling: A Data Driven Approach

Karol Gellert

November 26, 2023

Contents

Chapter 1

Research Motivation

1.1 Foreword

The research presented in this thesis sets out to follow a data-motivated approach to modelling interest rate dynamics. The data informing model construction includes interest rate index prices namely the London Inter-Bank Offer Rate (LIBOR), Secured Overnight Funding Rate (SOFR), Effective Fed Funds Rate (EFFR) as well as corresponding futures contract prices and options on those futures. The initial goal of the research was to find a model that simultaneously reflects the empirical behaviour of the LIBOR index and risk-neutral dynamics inferred from corresponding futures and options on futures. The outcome of the research is a methodological contribution to empirical data analysis and a new modelling approach motivated by empirical observation of the new SOFR benchmark rate.

Initial data exploration suggested that the empirical behaviour of LIBOR evolves over time and is possibly subject to regular regime changes. This observation led to the development of a particle filter which can adapt and evolve an estimated posterior distribution to streaming data. The adaptation is achieved by enriching a standard particle filter with a genetic algorithm. The particle filter, extensively tested using simulated data, is able to rapidly detect regime changes including different stochastic volatility states.

Analysis of real data using the new particle filter as well as the emergence of the Secured Overnight Financing Rate (SOFR) as a replacement for LIBOR led to a change in understanding of what constitutes constructing an empirically driven model of financial data. The perhaps subtle insight is that financial data does not necessarily follow any particular model, rather it does exhibit persistent statistical features which can be replicated by models. In effect, it is possible to create a model which produces the same statistical properties as empirical data, without necessarily being able to find stable model parameters that statistically fit empirical data.

The primary feature observed in SOFR is jumps which coincide with Federal Open Market Committee (FOMC) changes to the Fed Funds policy target rate. At the same time forward rates associated with those rates evolve in a continuous diffusive fashion. Reconciling these two features within one modelling framework was the inspiration behind

the model proposed in this thesis.

The first versions of the proposed model are based on Gaussian dynamics but also facilitate model-free estimation, under certain assumptions, of forward states from futures. Analysis of the forward states suggests the presence of non-Gaussian higher-order moments, a statistical feature which motivates the addition of stochastic volatility to the model. More specifically, the model is extended with stochastic volatility dynamics. This extension also allows for calibration to interest rate options which imply excess leptokurtosis for risk-neutral dynamics, usually characterised as the implied volatility smile. The rest of this section outlines the motivation, background and original contributions of each phase of the research.

1.2 Particle Filter With Accelerated Adaptation

In an ideal world, using a well-specified model entails estimating the model parameters from historical data, and then applying the model with these parameters going forward, i.e. out of sample. Indeed, the bulk of the empirical academic literature in finance takes this approach. However, practitioners' use of models, in particular for the pricing and risk management of derivative financial products relative to observed prices for liquidly traded market instruments, typically tends to depart from this ideal. Primacy is accorded to model calibration over empirical consistency, i.e., choosing a set of liquidly traded market instruments (which may include liquidly traded derivatives) as calibration instruments, model parameters are determined so as to match model prices of these instruments as closely as possible to observed market prices at a given point in time. Once these market prices have changed, the model parameters (which were assumed to be constant, or at the most time–varying in a known deterministic fashion) are recalibrated, thereby contradicting the model assumptions. Legalising these parameter changes by expanding the state space (e.g. via regime–switching or stochastic volatility models) shifts, rather than resolves, the problem: for example in the case of stochastic volatility, volatility becomes a state variable rather than a model parameter and can evolve stochastically, but the parameters of the stochastic volatility process itself are assumed to be time–invariant. The limits of increasing model complexity are determined by a combination of mathematical tractability and the practicality of models. The result is a certain disparity between empirical research and how models are used in practice. Proposed in this light is a practically motivated methodology in the form of an adaptive particle filter which can rapidly detect discrepancies between the assumed model and data including parameter changes and model misspecification.

Particle filtering is a sequential Monte Carlo method which has become popular for its flexibility, wide applicability and ease of implementation. The origin of the particle filter is widely attributed to Gordon, Salmond and Smith (1993) and theirs has remained the most general filtering approach. It is an *online* filtering technique ideally suited for analysing streaming financial data in a live setting. It seeks to approximate the posterior distribution of latent (unobserved) dynamic states and/or model parameters by sets of discrete sample values, where these sampled values are called particles. For a more comprehensive introduction to particle filtering see Chen et al. (2003) for a general introduction and historical

perspective; Johannes and Polson (2009), Lopes and Tsay (2011), and Creal (2012) for reviews related to finance; Andrieu, Doucet, Singh and Tadic (2004), Cappé, Godsill and Moulines (2007), Chopin, Iacobucci, Marin, Mengersen, Robert, Ryder and Schäfer (2011), Kantas, Doucet, Singh and Maciejowski (2009), and Kantas, Doucet, Singh, Maciejowski and Chopin (2015) for parameter estimation techniques. The theoretical perspective is outlined by Del Moral and Doucet (2014) and covered in depth in Del Moral (2004, 2013).

Basic particle filter algorithms suffer from particle impoverishment, which can be broadly described as the increase in the number of zero-weighted particles as the number of observations increases, resulting in fewer particles available for the estimation of the posterior. A key distinguishing feature of most contemporary particle filters is the approach taken to deal with the problem of particle impoverishment and it continues to be a focus of effort from researchers. A variety of techniques have been proposed as a solution, the main approaches are the use of sufficient statistics as per Storvik (2002), Johannes and Polson (2007), Polson, Stroud and Müller (2008), and Carvalho, Johannes, Lopes and Polson (2010); maximising likelihood functions as per Andrieu, Doucet and Tadic (2005) and Yang, Xing, Shi and Pan (2008); and random perturbation or kernel methods as per West (1993a, 1993b), Liu and West (2001), Carvalho and Lopes (2007), Flury and Shephard (2009), and Smith and Hussain (2012).

The idea behind random perturbation, initially proposed by Gordon et al. (1993) in the context of the estimation of dynamic latent states, is that by introducing an artificial dynamic to the static parameters, the point estimates become slightly dispersed, effectively smoothing the posterior distribution and reducing the degeneracy problem. This comes at a cost of losing accuracy as the artificial dynamic embeds itself into the estimation. Motivated in part by this issue, Liu and West (2001) introduce a random kernel with shrinking variance, a mechanism which allows for a smoothed interim posterior, but where the dispersion reduces in tandem with the convergence of the posterior distribution. The method proposed by Flury and Shephard (2009) is another example of this approach, introducing a perturbation to the SIR filter just prior to the resampling stage such that new samples are drawn from an already smoothed distribution, avoiding damage to the asymptotic properties of the algorithm. A common theme of these approaches is the assumption that the parameters are fixed over the observation period. However, particle filters have also been applied to estimate regime–switching models, see for example Carvalho and Lopes (2007) and Bao, Chiarella and Kang (2012). To this end, in this thesis, the idea of random perturbation is adapted to a more general parameter detection filter, pursuing a similar objective as Nemeth, Fearnhead and Mihaylova (2014),[1] who develop a particle filter for the estimation of parameters subject to dynamic changes. This type of problem is usually motivated by tracking manoeuvring targets, perhaps an apt metaphor for a financial market model requiring repeated recalibration of model parameters as time moves on.

The contribution of this research to the literature is the introduction of a particle filter with accelerated adaptation designed for the situation where the subsequent data in online sequential filtering does not match the model posterior filtered based on data up to a current point in time. This covers cases of model misspecification as well as sudden regime changes

[1] See also Nemeth, Fearnhead, Mihaylova and Vorley (2012).

or rapidly changing parameters. The proposed filter is an extension of on-line methods for parameter estimation which achieve smoothing using random perturbation. The accelerated adaptation is achieved by introducing a dynamic to the random perturbation parameter, allowing particle–specific perturbation variance; this combines with re-selection to produce a genetic algorithm which allows for rapid adaptation to mismatching or changing dynamics in the data. The similarity between particle filtering and genetic algorithms has been noted before, see for example Smith and Hussain (2012), who use genetic algorithm mutation as a resampling step in the SIR filter. The particle filter is reinforced to detect and rapidly adapt to any discrepancies between the model and realised dynamic by exploiting random perturbations, in a sense taking the opposite direction of methods in the literature which seek to control random perturbation in order to remove biases in the estimation of parameters assumed to be *fixed*.

This approach leads to a useful indicator of when changes in model parameters are being signalled by the data. The effectiveness of this heuristic measure is based on the notion that in the case of perfect model specification no additional parameter learning would be required. This indicator can provide useful information for characterising the empirical underlying dynamics without using highly complex models (meaning models which assume stochastic state variables where the simpler model uses model parameters). This allows for the use of a more basic model implementation to gain insight into more complex models and to make data–driven choices on how the simpler models might most fruitfully be extended. For example, the indicator will behave quite distinctly for an unaccounted regime change in the dynamic as opposed to an unaccounted stochastic volatility dynamic.

The particle filter extended with accelerated adaptation is described in detail in Chapter 2. The remaining chapters focus on overnight rates, particularly SOFR, motivated by the emergence of SOFR as the new benchmark for US interest rates.

1.3 Reconciling discontinuous short rate and continuous forward rates

As the Secured Overnight Funding Rate (SOFR) is currently in the process of becoming the key Risk–Free Rate (RFR) benchmark in US dollars, interest rate term structure models need to be updated to reflect this. Historically, interest rate term structure modelling has been based on rates of substantially longer time to maturity than overnight, either directly as in the LIBOR Market Model,[2] or indirectly, in the sense that even models based on the continuously compounded short rate (i.e., with instantaneous maturity)[3] are typically calibrated to term rates of longer maturities, with any regard to a market overnight rate at best an afterthought. However, with SOFR this situation is reversed: The overnight rate now is the primary market observable, and term rates (i.e., interest rates for longer maturities) will be less readily available and therefore must be inferred (for example from derivatives prices).

[2]See Miltersen, Sandmann and Sondermann (1997), Brace, Gatarek and Musiela (1997) and Musiela and Rutkowski (1997).

[3]Of these, Hull and White (1990) is the most prominent example.

Thus the empirical idiosyncrasies of the overnight rate cannot be ignored when constructing interest rate term structure models in a SOFR–based world, and more than longer-term rates, these idiosyncrasies are driven by monetary policy. The dynamics of both SOFR and the closely related and more established Effective Fed Funds Rate (EFFR) are closely examined. Already, by simple inspection one sees that models, in which the short rate evolves as a diffusion, can no longer be justified by empirical data. Instead, the primary driver of the short rate is the piecewise flat behaviour of the Federal Open Market Committee (FOMC) policy target rate. Concurrently, it is observed that the forward rates associated with the policy target rate evolve in a more diffusive manner. Reconciliation of these two features is one of the main research contributions of the model proposed in this thesis.

Modelling the target rate may seem not quite reflective of reality since the FOMC sets a target range rather than a specific rate. However, documentary and empirical evidence are presented to show that the target rate continues to exist via the Interest on Excess Reserves (IOER). The IOER acting as the target rate is a deliberate strategy by the Federal Reserve, which has proven effective at keeping the EFFR near the policy target.

Prior to 2021, a prominent empirical feature of SOFR dynamics, and to a lesser degree EFFR dynamics is the occurrence of large spikes. The spikes tend to occur at predictable times on the last day of the month and particularly the end of quarter and end-of-year dates. Not all spikes occur on the last day of the month, such as the extreme spike in September 2019. An explanation provided by the Federal Reserve[4] for the September 2019 spike is that it occurred on a day on which large corporate tax receipts and Treasury bond expiries caused a sharp imbalance in demand and supply in the repo market. Both the reasons given occur on dates easily obtainable in advance, therefore arguably this spike also could be classified as occurring on a predictable date. Using a similar approach to modelling the target rates, the model is extended to allow for spikes occurring on known dates.

It is not the intent of this research to conduct an econometric study of EFFR or SOFR dynamics. However, as an illustration, the model is calibrated to Fed funds and SOFR futures to show that Fed funds futures anticipate target rate changes well, and SOFR futures anticipate end–of–month spikes in SOFR to some extent[5]. For the latter, this is in line with full econometric studies in the literature, see e.g. Krueger and Kuttner (1996), Robertson and Thornton (1997) and Fontaine (2016).

The literature refers to jumps with deterministic jump times as *stochastic discontinuities*, see for example Kim and Wright (2014), Keller-Ressel, Schmidt and Wardenga (2018), Fontana, Grbac, Gümbel and Schmidt (2020). The nomenclature reflects the treatment of discontinuities as extensions to an existing continuous stochastic model. The approach in the herein proposed model is distinctly different in that the discontinuity is the basis of the model for the short rate, while simultaneously the forward rates for maturities beyond the

[4]See Feds Notes link https://www.federalreserve.gov/econres/notes/feds-notes/what-happened-in-money-markets-in-september-2019-20200227.html.

[5]While it is possible to extend the research to SOFR swaps, this is not in the scope of this research. The interaction between forward term structure and FOMC meeting dates is best reflected in futures instruments.

next scheduled jump evolve as a continuous stochastic process.

Specific to SOFR, Heitfield and Park (2019) model forward rates using a step function, assuming that rates remain constant for all dates between FOMC meetings. This is a static approach for the purposes of calibrating a piecewise flat term structure. More recently, Andersen and Bang (2020) provide a SOFR–inspired general spike model to enable the extension of derivative pricing models to spikes. In the model proposed in Chapter 2 of the present dissertation, spikes at known times are included as a special case. However, the main focus of this research is on short rate discontinuities at known times.

Modelling discontinuities in the short rate is also attempted by Piazzesi (2001), Piazzesi (2005) and more recently by Backwell and Hayes (2022). In these references, the discontinuities are modelled as a jump process directly on the short rate. Piazzesi (2001) and Piazzesi (2005) are able to adapt the model to approximate jumps at known times by increasing the jump intensity around specific dates. Backwell and Hayes (2022) introduce dynamics of auxiliary state variables to model jumps in the continuously compounded short rate, i.e., modelling the size of jumps at known and at random times as well as the intensity of the random jump process. In contrast, in this dissertation the forward rate term structure is directly modelled, reflecting changing expectations of future policy target rates. This obviates the need for auxiliary state variables. Arguably, these market expectations evolve diffusively, the more so at times when policy rate changes are signalled well in advance by the central bank. The piecewise constant paths of the short rate are then a consequence of policy target rate changes becoming effective only at central bank meeting dates (regularly scheduled or convened at relatively short notice). All iterations of the model presented in this dissertation fit naturally in the general arbitrage–free framework of Heath, Jarrow and Morton (1992) (HJM) and lend themselves to principal component analysis of the forward term structure dynamics, where the (correlated) model factors can be directly mapped to central bank decision dates, or transformed (via rotation) into a set of independent HJM driving factors.

Thus, a key contribution of this research is to demonstrate that changing policy target rate expectations are the primary driver of empirical forward rate dynamics. This is directly reflected by the model, by design and as a result of the data-driven approach taken in this thesis. The link between the forward rate term structure and FOMC dates aligns the model directly with the fundamental economic behaviour of evolving policy rate change expectations. This alignment, upon the empirical estimation of the model, yields several novel insights linking interest rate modelling to the economic behaviour of interest rate markets as detailed in Chapters 3, 4 and 5.

Several other papers focus on adapting existing models to SOFR without considering discontinuities. These include Mercurio (2018) who uses a deterministic SOFR–OIS spread with a short rate model for the OIS. Lyashenko and Mercurio (2019) propose an extension to the LIBOR Market Model to accommodate the in–arrears setting nature of term rates related to SOFR and overnight benchmark rates in general. Skov and Skovmand (2021) show that a three–factor Gaussian arbitrage–free Nelson/Siegel model is well suited for the SOFR futures market, but they do not include the time series of SOFR itself in their estimation.

The empirical and economic motivation for the first iteration of the model, presented

in Chapter 3, is detailed in the next section.

1.4 Empirical Motivation

Monetary Policy and short rate Models

The first stochastic model of the short rate is attributed to Merton (1973), who employed a single–dimension Brownian motion as the driving dynamic. At least on cursory visual inspection, the empirical data at the time, see Figure 1.1, did not contradict the mathematically tractable Gaussian assumption of the model. The next major development came from Vasicek (1977), adding mean reversion, a strong empirical feature of rate dynamics. Modelling mean reversion also aligned with the notion of open market operations by the Federal Reserve trading desk managing the rate around the monetary policy target. Cox, Ingersoll and Ross (1985) (CIR) modified the dynamics of the continuously compounded short rate by scaling the volatility by the square root of the short rate, ensuring the non-negativity of interest rates. The next milestone in short rate modelling was an extension of the Vasicek model with time-dependent drift by Hull and White (1990), allowing the model to be fitted to an initial term structure of interest rates observed in the market — this was critical for use of the model to price interest rate derivatives. Heath, Jarrow and Morton (1992) developed the general framework into which all diffusion–based arbitrage–free interest rate term structure models must fit.

Starting around 2015, significant changes to the implementation of monetary policy have had a dramatic impact on the EFFR, resulting in a substantial divergence between its empirical behaviour and the dynamic assumptions of short rate models. The changes trace back to the 2008 financial crisis, prior to which monetary policy was administered primarily by direct intervention in the Fed Funds market to maintain the EFFR close to the target rate set by the FOMC. The approach relied on open market operations by the Federal Reserve trading desk resulting in the EFFR gravitating around the target rate with varying degrees of volatility.[6]

Open market operations are carried out by the Federal Reserve trading desk whose trading goal is to maintain the EFFR near the target rate. This involves monitoring the market and counteracting trades which move the EFFR away from the target, in essence, micro–managing market liquidity. The 2008 financial crisis included a crisis in liquidity and the ability of the Federal Reserve's trading desk to maintain the EFFR near the target rate significantly deteriorated. The trading desk did not have the means to counteract the dramatic drain in the supply of desperately demanded capital.

This was acknowledged by the Federal Reserve[7] as one of the factors considered when switching to a target range, initially set between 0 and 25 basis points. The Federal Reserve's strategy in response to the financial crisis centred around two key policies: near-zero interest rates and quantitative easing. The phases of quantitative easing became known as QE 1/2/3 and involved selling Treasury bonds and purchases of various credit risky assets[8]

[6] See Hilton (2005) for an analysis of factors impacting EFFR volatility related to open market operations.
[7] See Federal Open Market Committee (2000-2020) December 2008, page 9.
[8] Such as Agency Debt, Mortgage Backed Securities and Term Auction Facilities, see Binder (2010).

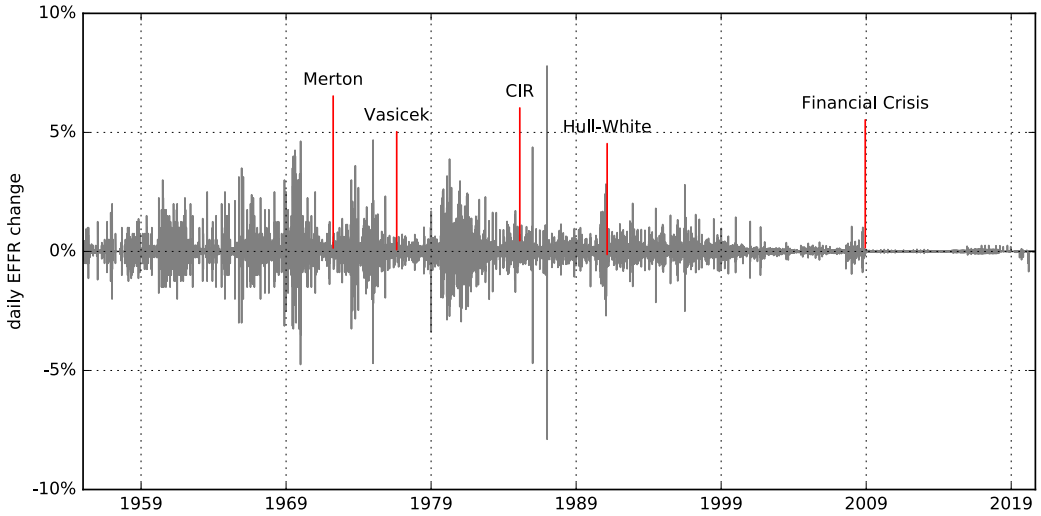

Figure 1.1: Empirical daily EFFR changes and the history of short rate models

in a bid to boost liquidity and credit conditions. The Federal Reserve's injection of liquidity resulted in an environment of elevated excess reserves[9]. By historical standards, the rise in excess reserves was extreme and without precedent. As can be seen in Figure 1.2, it increased from under $2 billion in September 2008 to $1 trillion by November 2009, before reaching a high of over $2.5 trillion in October 2015.

In October 2008, the Federal Reserve began paying IOER[10] to help control the EFFR in response to increasing excess reserves. It was thought at the time that the IOER should act as a lower bound for the EFFR since no institutions should want to lend below this rate. As such, effective from October 9 the IOER was set to 75 basis points, with the EFFR target rate at 150 basis points. In the following days, the EFFR was set well below the target rate, including some days below the IOER. On October 23, to lift rates closer to the target, IOER was increased to 110 basis points, in response EFFR rates increased but were still setting below the IOER. Other adjustments were made in November under the assumption of IOER acting as a lower bound. However, with EFFR persisting to settle well below the IOER it became clear the assumption was incorrect.

At the FOMC meeting immediately following the introduction of the IOER, it was noted that institutions not eligible to receive IOER were willing to sell (lend) funds at rates below the IOER.[11] However, it was not until December 2008, when together with the introduction of the target range, the IOER was set at the target range upper limit of 25 basis points in recognition that due to unique circumstances the IOER was acting as an upper bound for the EFFR. The large surpluses in excess reserves eliminated the demand for reserve loans. Instead, the Fed Funds rate was driven by Government Sponsored Institutions,

[9]Excess reserves are capital reserves held by financial institutions including banks in excess of the regulatory requirement

[10]See Federal Open Market Committee (2000-2020) October 2008, page 7.

[11]See Federal Open Market Committee (2000-2020) October 2008, page 2.

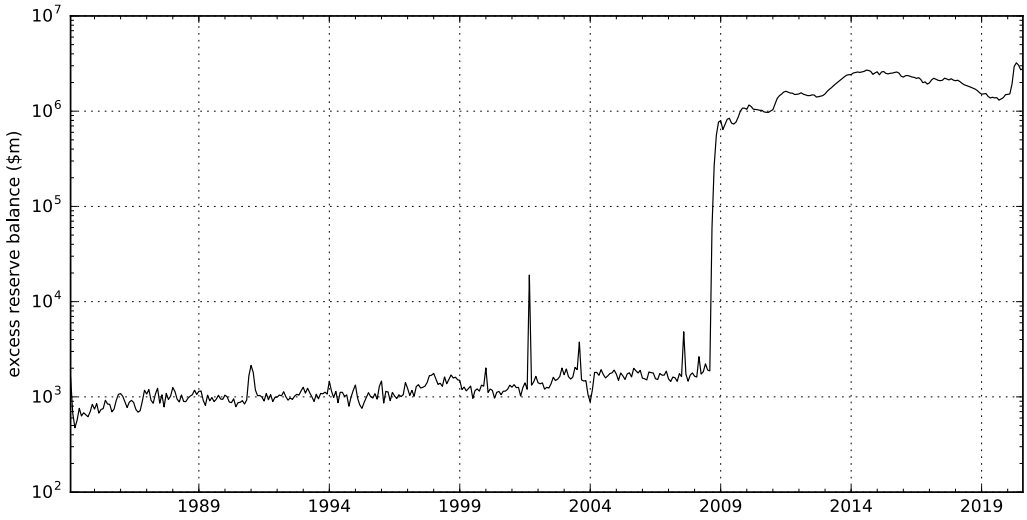

Figure 1.2: Excess reserves balance history

which do not earn interest on reserve balances, lending their excess reserves below the IOER to institutions that would then earn the difference between the Fed Funds rate and the IOER. In effect, by paying the IOER in a market flooded with liquidity, the Federal Reserve became the borrower, rather than the lender of last resort.

Plans for reversal of the post-financial crisis expansionary policy were formally laid out at the FOMC September 2014 meeting as the Policy Normalization Principles and Plans.[12] The aim of the normalisation strategy was to bring the EFFR back to normal levels and reduce the securities held by the Federal Reserve, thereby unwinding the excess reserves held by banks. Prior to the financial crisis, controlling the supply of reserves via open market operations was a key tool in controlling the Fed Funds rate. However, the Federal Reserve has adopted the view that with banks using reserves for liquidity more than prior to the crisis, it might be hard to predict demand for reserves and therefore open market operations would not be effective at precisely controlling the EFFR.[13] Instead, the new normal will constitute the Federal Reserve keeping excess reserves just large enough to remain on the flat part of the demand curve, a prerequisite condition for the use of the IOER to control the EFFR.

Thus the conditions in the Fed Funds market are dramatically different to when short rate models were first conceived. The flood of liquidity in excess reserves, by construction aimed at removing any supply–demand gradient, has removed most of the volatility from the short rate of interest, with changes in the short rate being mainly driven by changes in the IOER, leading to jumps at known times (the FOMC meeting dates). Forward rates are implied by traded market instruments. However, they continue to exhibit volatility as the evolution of market expectations of FOMC actions is priced into forward–looking instruments such as Fed Fund futures.

[12] See Federal Open Market Committee (2000-2020) September 2014, page 3.
[13] See Federal Open Market Committee (2000-2020) November 2018, page 3.

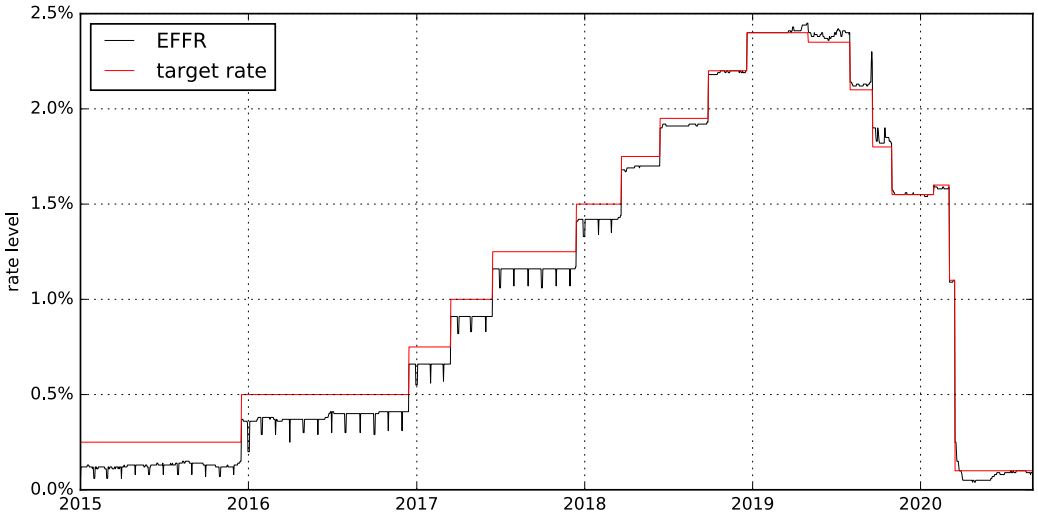

Figure 1.3: EFFR and FOMC target rate history

Effective Federal Funds Rate

In this section, the EFFR is examined by breaking it down into distinct components. A comparison of EFFR and the Fed Funds target rate since the beginning of 2015, see Figure 1.3, demonstrates the low volatility in deviations of EFFR from the target rate. The target rate therefore must be a major component of the EFFR dynamics. Another feature of Fed Funds empirical data in the earlier part of five years covered by Figure 1.3 are end–of–month downward spikes. These spikes used to occur as a result of regulations prescribing the last day of the month as a measurement day for reporting regulatory capital, resulting in a temporary imbalance in the demand–supply for excess reserve funds[14].

It is instructive to deconstruct the EFFR $r_E(t)$ rate into the two components, discontinuous at known times, and a residual such that:

$$r_E(t) = r_P(t) + \Delta r_Z(t) + \zeta(t) \tag{1.1}$$

The first component r_P, the policy target rate is directly observable as the IOER rate. The second component, Δr_Z the end–of–month spike, can be deduced from the data. Here any changes to the rate on the last trading day of the month regardless of magnitude are counted as spikes, a sufficient approach for the qualitative analysis in this section. The third component ζ captures any residual noise in addition to the first two components. The variance of the daily changes in each component, shown in Figure 1.4, is an indicator of the relative contribution of each component. It is clear that over the 5 years of data used to produce these results, the target rate is the main factor in EFFR dynamics, followed by the end–of–month spikes, with only a small contribution from the residual.

The existence of mean reversion in the residual is examined by finding an approximate Hurst exponent h for the time series of ζ. The Hurst exponent relates the variance of the

[14]See https://www.forbes.com/sites/jonhartley/2017/03/31/how-european-regulators-are-hindering-the-feds-ability-to-raise-interest-rates.

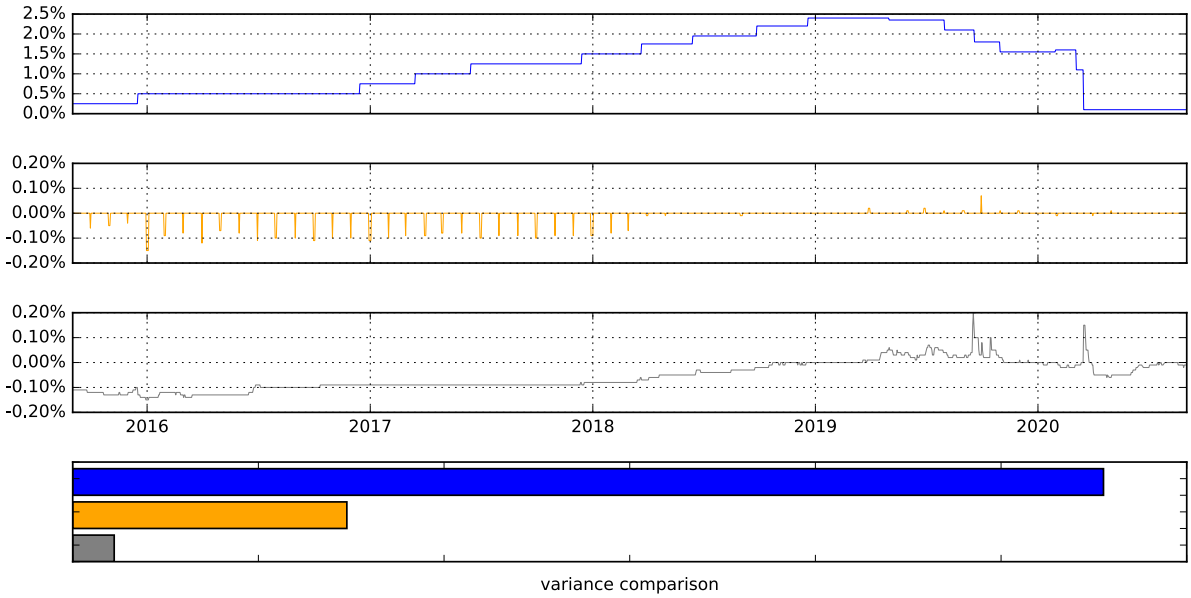

Figure 1.4: EFFR breakdown in vertical order (i) target rates (ii) end–of–month spikes (iii) residual (iv) variance contribution

lagged difference to the lag size as follows:

$$\mathrm{Var}\big[\zeta(t+\tau) - \zeta(t)\big] \sim \tau^{2h} \tag{1.2}$$

A Hurst exponent value of 0.5 indicates a Brownian motion, $h < 0.5$ indicates presence of mean reversion. For the residual noise time series, the estimate is $h = 0.31$, see Figure 1.5, indicating mean reversion.

In summary, the empirical characteristics of EFFR break down into the following components: piecewise flat target rates, followed by spikes occurring on known days and mean reverting residual noise. The correlation between the three components is close to zero, with the exception of a slightly negative correlation between the residual and the target rates. The negative correlation is due to a small lag between target rate changes and EFFR adapting the full magnitude of the change, temporarily changing the spread in the opposite direction to the target rate change.

Secured Overnight Funding Rate

Shortly following the well–publicised LIBOR manipulation scandals, the Financial Stability Board and Financial Stability Oversight Council highlighted one of the key problems related to the reference rate to be the decline in transactions underpinning LIBOR and the associated structural risks to the financial system.[15] As argued by Schrimpf and Sushko

[15] See The Alternative Reference Rates Committee (2018), page 1.

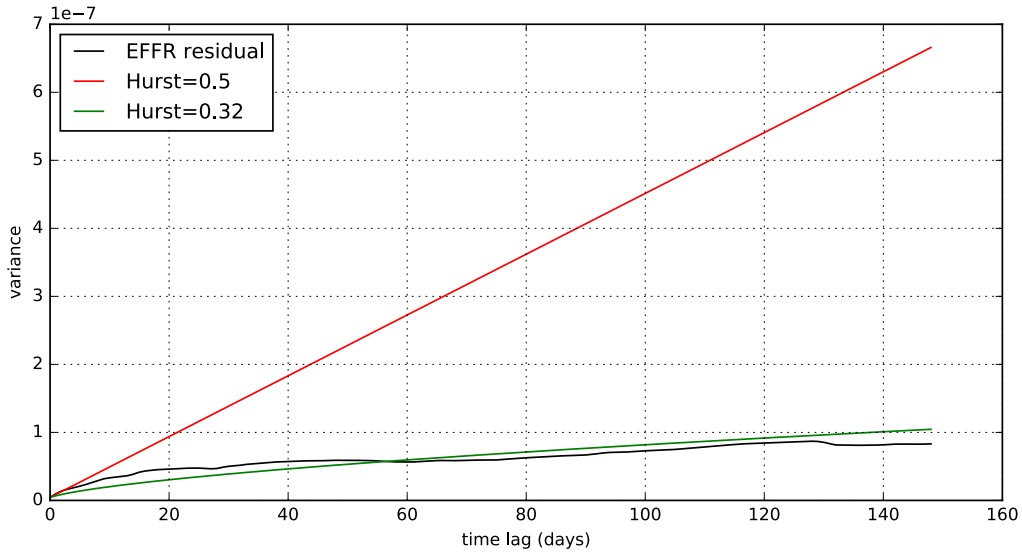

Figure 1.5: Variance for the difference of the residual noise time series for different lags, compared to theoretical Hurst exponent values

(2019), partly to blame for the decline in interbank term lending is the inflated excess reserves discussed in the previous section.[16] In response, the Federal Reserve convened the Alternative Reference Rates Committee (ARRC)[17] to explore alternative reference rates. In June 2017, the ARRC formally announced the Secured Overnight Financing Rate (SOFR) as the replacement for LIBOR. A key criterion for the choice was the large volume of transactions behind SOFR, translating to it being more representative of the bank's funding costs and less susceptible to manipulation. The calculation of SOFR is based on overnight repo transactions, which in 2017 averaged around $700b in daily transactions[18] (compared to less than $1b for US dollar LIBOR).

Official SOFR fixings have been calculated as far back as 2014 and can be seen in comparison to the target rate in Figure 1.6. Three features stand out, firstly SOFR appears to fluctuate around a stepwise function, suggesting that similarly to EFFR the Fed Funds target rate plays an important role in the SOFR dynamic. Another aspect is that SOFR is substantially more volatile than EFFR. A third feature is the prominence of spikes, most of which, similarly to EFFR, occur on the last trading day of the month. The end–of–month spikes are related to the measurement of dealers' balance sheet exposures at month–end for regulatory purposes. This single snapshot approach incentivises the management of exposures around reporting dates, which as explained in Schrimpf and Sushko (2019) has been resulting in increases in the SOFR rate on end–of–month dates. Therefore the main components of SOFR can be characterised as follows, see Figure 1.7:

[16] This suggests an interesting causal link between the financial crisis, the Federal Reserve response and the emergence of SOFR by linking the decline in LIBOR transactions to excess reserves.

[17] See https://www.newyorkfed.org/arrc.

[18] For details see The Alternative Reference Rates Committee (2018), page 7.

Figure 1.6: SOFR and FOMC target rate history

$$r_S(t) = r_P(t) + \Delta r_Z(t) + \Delta r_J(t) + \zeta_s(t) \tag{1.3}$$

Here $r_S(t)$ is the SOFR observation at time t, $r_P(t)$ the policy target rate, $\Delta r_Z(t)$ the end–of–month SOFR spikes, $\Delta r_J(t)$ spikes not occurring on end–of–month dates and $\zeta_s(t)$ the residual. The spikes not occurring on the last day of the trading month are the most prominent in terms of contribution to the net variance over the period. However, this is due to only one very large spike occurring in September 2019. This particular spike occurred on a day of large corporate tax payments and Treasury bond expiries, therefore it could be argued that the date was predictable. The next largest contribution comes from the end–of–month spike component, followed closely by the policy target rate component.

In contrast to EFFR, the contribution from the residual component is in the same order of magnitude as the target rate component as well as the end–of–month component. Using the same approach as for EFFR, the SOFR residual also exhibits strong mean reversion with an estimated Hurst parameter $h = 0.24$, see Figure 1.8. In summary, the components of SOFR mostly mirror the components of EFFR, but with different contributions to the overall variance.

The first iteration of the model introducing piecewise constant short rate and continuously diffusive forward rate dynamics is detailed in Chapter 3. A time-homogenous version, allowing empirical estimation of the model states, is presented in Chapter 4. The next section discusses the motivation behind a stochastic volatility extension to the model.

1.5 Extending the model with stochastic volatility

The first version of the model is set in an HJM framework with a piecewise constant volatility function to produce dynamics which are simultaneously piecewise constant for

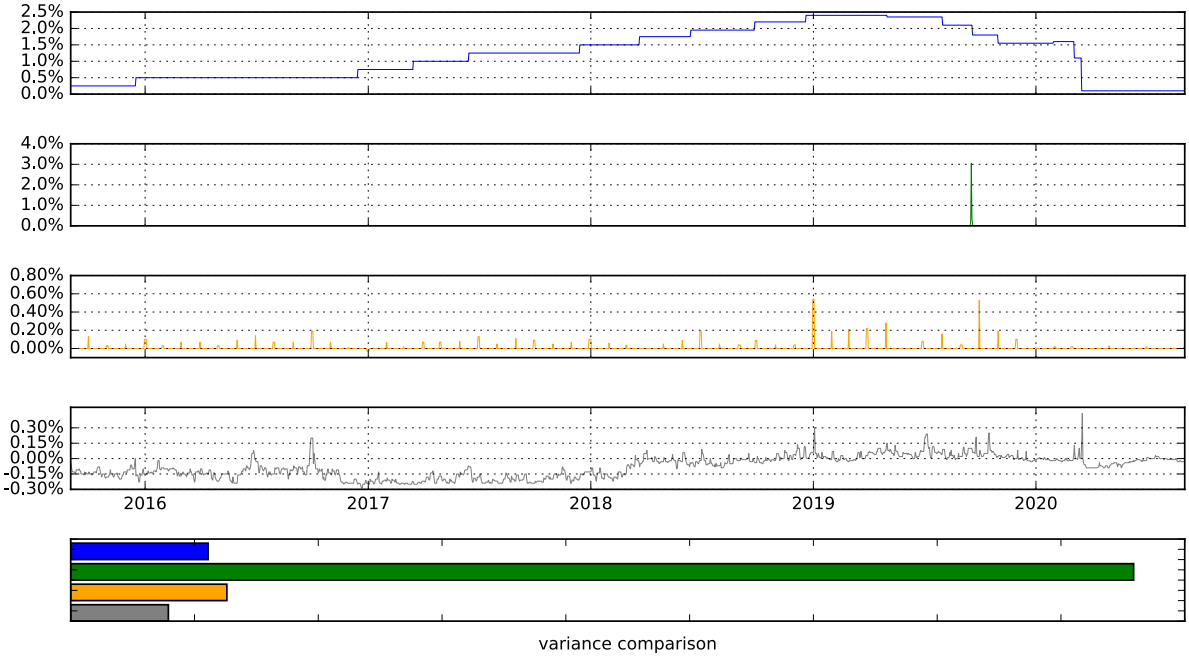

Figure 1.7: SOFR variance breakdown in vertical order (i) target rates (iii) non-eom spike (iii) end–of–month spikes (iv) residual (v) variance contribution

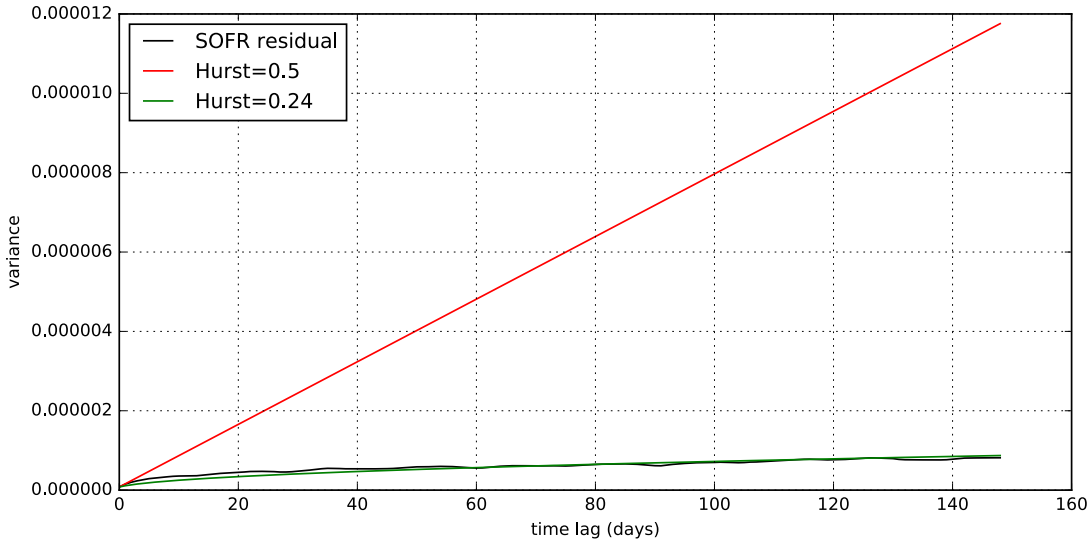

Figure 1.8: Variance for the difference of SOFR residual time series for different lags, compared to theoretical Hurst exponent values

the short rate and continuous for the forward rates. The approach is inspired directly by the empirical behaviour of overnight rates, namely the Fed Funds rate and SOFR. This initial version of the model focused on introducing the piecewise constant volatility and

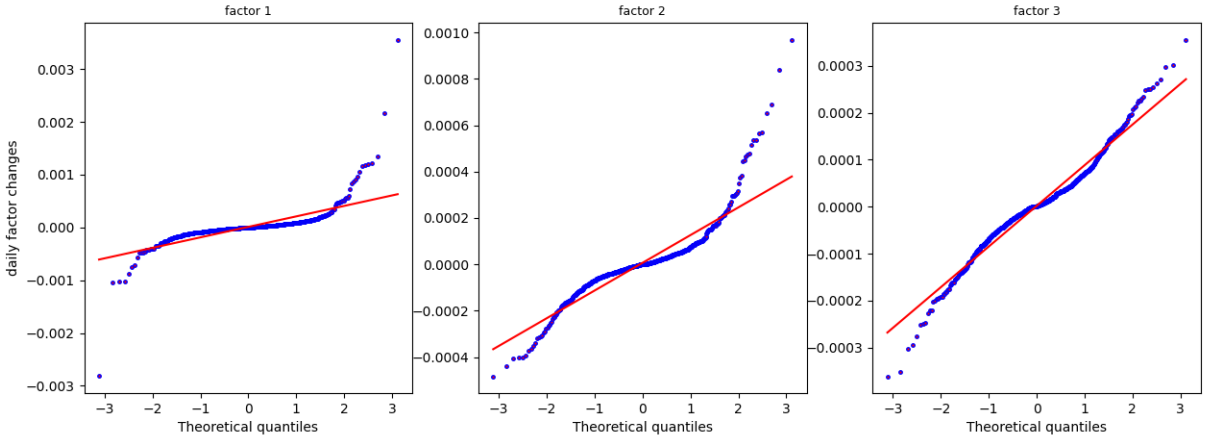

Figure 1.9: Empirical factor states(blue) compared with a normal distribution(red) quantile-quantile plots

did not include any consideration of the empirical dynamics of forward rates. However, forward rate states are derived from daily calibration of the piecewise constant structure to Fed Funds and SOFR futures.

To continue to build on the empirically inspired model the empirical dynamics of the forward rate states are examined. The empirical evidence as well as practical implications including calibration to options, both point to an extension with stochastic volatility and mean reversion. Additionally, an update of the analysis in the previous phase shows that the spikes and the overnight rate to target rate spread have become less relevant and are therefore excluded, resulting in this extension being focused entirely on stochastic volatility and mean reversion extension.

The forward rate states were derived from the initial version of the model assuming a piecewise flat structure between FOMC dates without any assumptions regarding the driving dynamics. The forward rate states are factorised using principle component analysis to produce the volatility parameter structure and empirical states of the driving factors. The time series of these factor states allows for empirical assessment of the factors driving forward rate state dynamics. This includes a measurement of the fourth moment (leptokurtosis) of daily changes in the factor state time series as shown in Figure 1.9.

The quantile-quantile (QQ) plots shown in Figure 1.9 compare the expected quantile values for a normal distribution (red line) against the empirical value (blue dots). All three factors exhibit clear leptokurtosis, with excess kurtosis of 63, 10 and 2 for the first three factors respectively. The presence of excess kurtosis can be caused by various underlying processes, stochastic volatility is a common and parsimonious modelling choice to reproduce this feature.

One of the consequences of linking the model to FOMC dates is that some of the empirical results have a direct economic interpretation. As detailed earlier, the first factor focuses on policy rate changes at the upcoming FOMC meeting while the higher factors tend

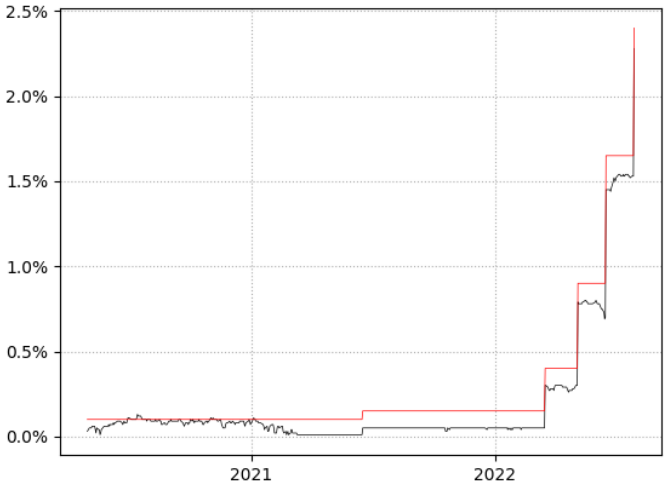

Figure 1.10: SOFR v FOMC target rate history May 2020 to August 2022

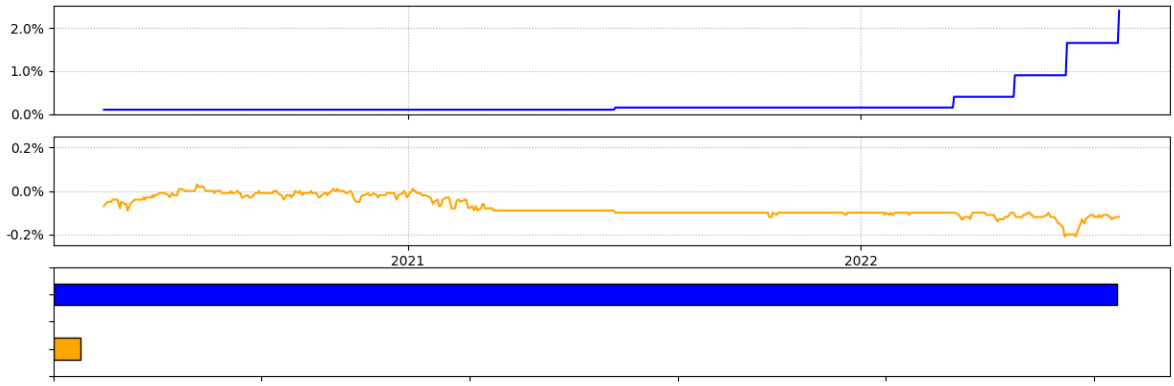

Figure 1.11: SOFR breakdown in vertical order (i) target rates (ii) SOFR-Target Rate spread (iii) variance contribution

to focus on FOMC meetings beyond the next one. Excess kurtosis is notably highest for the first factor, suggesting a link between high leptokurtosis and the next FOMC meeting. This is consistent with evidence from interest rate options which imply higher stochastic volatility for shorter expiry options. Anecdotally, interest rate market participants tend to focus on the next FOMC meeting date and the Federal Reserve tends to focus on managing the expectations related to the next FOMC date. This tends to make the expectations related to the next FOMC date most susceptible to news and changing economic circumstances: a feasible economic explanation for the highlighted excess kurtosis observation.

Another important aspect to consider is the calibration to interest rate options. In general calibration to interest rate options requires some degree of freedom to calibrate the skewness and convexity of implied volatilities for a range of strikes. Interest rate options also tend to imply a term structure of volatility, skewness and convexity for a range of expiries and forward terms. The initial versions of the model already provide the ability to control volatility term structure with different factors affecting different forward rate terms. Embedding stochastic volatility into each factor provides the ability to calibrate convexity and skewness[19] with some control of the term structure of those features inherited from the factor term structure of the original model.

Spikes have been a well-publicised feature of SOFR and reflect imbalances in the underlying overnight repo market. Spikes used to occur regularly at end of the month and occasionally on the non-end of the month dates. The initial analysis, based on data from Jan 2014 to May 2020, showed that the SOFR rate variance was largely dominated by spikes followed by an even contribution from the target rate and the SOFR spread to the target rate. A significant SOFR spike in September 2019 motivated the Federal Reserve to take action and effectively stabilise the future benchmark rate. Since that time as can be seen in Figure 1.10 spikes are no longer a feature of SOFR. The variance contribution now breaks down to changes in the target rate (around 99% of variance), with the remainder of the variance explained by the SOFR to target rate spread, see Figure 1.11. Consequently, the model in Chapter 5. focuses on modelling the target rate component.

The addition of stochastic volatility reflects empirical behaviour and is a critical component for interest rate option calibration, particularly for away from the money strikes. Mean reversion also reflects empirical behaviour but also controls the steepening of the forward curve resulting from the risk-neutral drift in HJM. Heston/Hull-White inspired stochastic volatility in an HJM framework fits very well into the piecewise continuous formulation in the initial version of the model. Having each factor in the model driven by stochastic volatility dynamics provides the ability to control the term structure of volatilities, stochastic volatilities and correlation, providing ample flexibility to calibrate to options across different strikes and expiries while at the same time reflecting key empirical features of the underlying forward rates.

The final phase of the research continues to take the data-driven approach and extends the model based on further consideration of both empirical time series data as well as cross-sectional calibration to interest rate options. It is particularly influenced by developments in the market and empirical data over the course of 2021 and 2022. Examining the empirical dynamics of the forward rate driving factors yields evidence of excess kurtosis, which motivates a stochastic volatility extension of the model. The empirical behaviour of SOFR is revisited to find that most recently it is driven almost exclusively by the Fed Funds target rate. In addition to stochastic volatility, mean reversion is identified as an important feature to include in the extended model.

The extended model endows each factor associated with target rate dynamics with its own stochastic volatility and mean reversion dynamic inspired by the Heston/Hull-White(HHW) model. In literature, Heston/Hull-White usually refers to Heston (1993)

[19] Skewness is impacted by the correlation of stochastic volatility to forward rate volatility.

stochastic volatility equity models with an interest rate driven by a Hull and White (1990) model, see for example Grzelak, Oosterlie and Weeran (2008). In this thesis, HHW refers to an interest rate term structure model along the lines of Hull and White (1990), where the deterministic volatility is replaced by a volatility process similar to the one used in Heston (1993). Using a multifactor HHW model within the HJM framework, the preceding chapter volatility functions are defined in such a way as to generate short rates which are constant between jump dates, while forward rate evolve diffusively (now with a stochastic volatility and mean reversion). It is demonstrated how this model has the flexibility required for the calibration to interest rate options.

Several papers consider various empirical aspects of SOFR. Skov and Skovmand (2021) propose a Nelson-Siegel model and demonstrate it can, to a certain degree of accuracy, reflect SOFR futures prices, albeit without taking into account the piecewise constant nature of the underlying SOFR rate. Andersen and Bang (2020) address SOFR spikes in their proposed model also without considering the SOFR rate as being driven by the piecewise constant target rate. Heitfield and Park (2019) consider the same piecewise term structure for cross-sectional calibration to futures only and without considering SOFR or SOFR forward rate dynamics. Backwell and Hayes (2022) (discussed in Section 1.3 above) conduct an empirical estimation of their model on Sterling (GBP) time series data, where the overnight rate benchmark is the Sterling Overnight Index Average (SONIA). The model presented in Chapter 5 considers the empirical behaviour of SOFR, SOFR forward rates implied from futures as well as cross-sectional calibration to both futures and options.

The proposed model performs well in cross-sectional calibration due to having a sufficient amount of variables which control various aspects of the model's behaviour. This allows the model to be calibrated across different maturities, underlying futures accrual periods and option strikes. This flexibility in the context of cross-sectional calibration is similar to prominent models deployed in practice. The SABR model, introduced by Hagan, Kumar, Lesniewski and Woodward (2002), can be calibrated to implied volatility convexity and skew across strikes. However, this approach generally requires a new calibration per option expiry (and swap/forward rate tenor). Short rate models, such as Hull and White (1990), are usually calibrated to only co-terminal swaptions chosen to match an underlying trade[20] and a singular strike per swaption. The BGM model introduced in Brace et al. (1997) is well suited for simultaneously calibrating to at the money swaptions across expiries and tenors. Most comparable in terms of the ability to calibrate across expiry, underlying tenor and strike are stochastic volatility extensions to BGM, see for example Piterbarg (2015) or Karlsson, Pilz and Schlögl (2017).

An analysis of the behaviour of options in the accrual period implied by the proposed model is also presented. Interest in this behaviour is mostly driven by the practicalities of adapting existing LIBOR-based modelling to SOFR and therefore requires casting option behaviour in the accrual period to the behaviour of the dynamics of partially set forward term rates. In terms of accrual period behaviour, under simplifying assumptions the model produces behaviour similar to the volatility decay proposed in Lyashenko and Mercurio (2019). However, the model presented in this thesis handles the case of partially set forwards

[20]For example call dates in a callable note.

naturally and also provides more granular insight into the decay characteristics of implied volatility within the accrual period, leading to arguably more realistic behaviour.

Additionally, the proposed model reveals a connection between forward rate empirical behaviour and short rate mean reversion. In the HJM framework, mean reversion is embedded a priori as a decay function of forward rate volatilities. The specification of the HJM volatility in the model proposed in Chapter 5 includes both a decay function and a piecewise constant component[21]. However, remarkably the piecewise component derived directly from empirical data without any shape restrictions closely resembles the decay function associated with mean reversion.

The rest of the thesis is organised as follows. Chapter 2 presents a novel version of the particle filter enhanced by a genetic algorithm which results in a rapid adaptation to regime changes. Chapter 3 introduces the model approach which reflects piecewise constant short rates and continuously diffusive forward rates. The next version of the model, presented in Chapter 4, is defined to be time-homogeneous to allow for empirical estimation of the model states. The model is extended with stochastic volatility in Chapter 5. Final remarks are given in Chapter 6.

[21] To impose a piecewise forward rate structure corresponding to FOMC meeting dates.

Chapter 2

Particle Filter with Accelerated Adaptation[1]

This chapter presents the construction of a particle filter, which incorporates elements inspired by genetic algorithms, to achieve accelerated adaptation of the estimated posterior distribution to changes in model parameters. Specifically, the filter is designed for the situation where the subsequent data in online sequential filtering does not match the model posterior filtered based on data up to a current point in time. The examples considered encompass parameter regime shifts and stochastic volatility. The filter adapts to regime shifts extremely rapidly and delivers a clear heuristic for distinguishing between regime shifts and stochastic volatility, even though the model dynamics assumed by the filter exhibit neither of those features.

The remainder of this chapter is organised as follows. Section 2.1 recalls the basic particle filter construction. Section 2.2 iteratively presents the evolution of the particle filter methodology based on the existing literature, culminating in the Liu and West (2001) filter, which forms the starting point of the filter with accelerated adaptation. Section 2.3 presents the particle filter incorporating additional elements inspired by genetic algorithms, adding these elements step by step and providing examples demonstrating their effectiveness. Section concludes.

2.1 Particle filter

2.1.1 General framework

Consider a sequence of Markovian discrete time states $x_{1:t} = \{x_1, ..., x_t\}$ and discrete observations $y_{1:t} = \{y_1, ..., y_t\}$. In general, particle filtering is concerned with the filtering

[1]This chapter is based on the paper Gellert and Schlögl (2021a) with Erik Schlögl contributing in a supervisory capacity.

problem characterised by state-space equations:[2]

$$x_{t+1} = f(x_t, d_t) \tag{2.1}$$
$$y_t = g(x_t, v_t) \tag{2.2}$$

with d_t and v_t as independent random sequences in the discrete time domain. The *transition density* of the state $p(x_{t+1}|x_t)$ and the observation *likelihood* $p(y_t|x_t)$ are obtained from equations (2.1) and (2.2). The objective of particle filtering is the sequential estimation of the posterior density $p(x_t|y_{1:t})$ which can be expressed in the recursive Bayesian setting:[3]

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \tag{2.3}$$

where $p(x_t|y_{1:t-1})$ is the *prior* and is determined by the integral:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \tag{2.4}$$

$p(y_t|y_{1:t-1})$ is the *evidence* and is determined by:

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t \tag{2.5}$$

The continuous posterior is approximated in the particle filter by a discrete random measure $\{x_t^{(i)}, \pi_t^{(i)}\}_{i=1}^N$ composed of sample values of the state $x_t^{(i)}$ with associated weights $\pi_t^{(i)}$, and N denoting the total number of particles[4] as follows:

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{x_t^{(i)}=x_t\}} \pi_t^{(i)} \tag{2.6}$$

where δ is the Dirac delta function.

2.1.2 Filtering for parameters with directly observed states

The general state-space framework is a combination of state and observation dynamics. The observation dynamics are useful in problems which require modelling of observation uncertainty, for example applications involving physical sensors. In empirical finance the observation usually consists of a directly observed price, which often is assumed not have any inherent observation uncertainty. It does not mean state uncertainty cannot be modelled but here unobserved states are artefacts of the modelling assumptions rather than physical quantities of the system — stochastic volatility is one such example.

In this chapter the focus is on estimation of only model parameters, in order to present the approach in the simplest setting where states are assumed to be directly observed — by

[2] Using notation from Chen et al. (2003).
[3] See Chen et al. (2003) for derivation.
[4] See Li, Bolic and Djuric (2015).

adapting the results in existing literature in the appropriate manner, the approach would be straightforward to extend to the estimation of latent state variables. Begin with a derivation of the recursive Bayesian framework below. Similarly to Chen et al. (2003), the last step of the derivation relies on the assumption that the states follow a first-order Markov process and therefore $p(x_t|x_{1:t-1}) = p(x_t|x_{t-1})$. Let θ represent the parameter set, the parameter posterior is

$$p(\theta|x_{1:t+1}) = \frac{p(x_{1:t+1}|\theta)p(\theta)}{p(x_{1:t+1})} \tag{2.7}$$

$$= \frac{p(x_{t+1}, x_{1:t}|\theta)p(\theta)}{p(x_{t+1}, x_{1:t})} \tag{2.8}$$

$$= \frac{p(x_{t+1}|x_{1:t}, \theta)p(x_{1:t}|\theta)p(\theta)}{p(x_{t+1}|x_{1:t})p(x_{1:t})} \tag{2.9}$$

$$= \frac{p(x_{t+1}|x_{1:t}, \theta)p(\theta|x_{1:t})}{p(x_{t+1}|x_{1:t})} \tag{2.10}$$

$$= \frac{p(x_{t+1}|x_{1:t}, \theta)p(\theta|x_{1:t})}{\int p(x_{t+1}|x_{1:t}, \theta)p(\theta|x_{1:t})d\theta} \tag{2.11}$$

$$= \frac{p(x_{t+1}|x_t, \theta)p(\theta|x_{1:t})}{\int p(x_{t+1}|x_t, \theta)p(\theta|x_{1:t})d\theta} \tag{2.12}$$

The above formulation establishes a recursive relationship between sequential posteriors. The particle approximation is based on the discretisation of $\theta^{(i)}$ as shown below. Introducing weight notation $\pi_{t+1}^{(i)} := p(\theta^{(i)}|x_{1:t+1})$ and the un-normalised weight $\hat{\pi}_{t+1}^{(i)}$:

$$\pi_{t+1}^{(i)} \approx \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta^{(i)})\pi_t^{(i)}}{\sum_{i=1}^{N} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta^{(i)})\pi_t^{(i)}} = \frac{\hat{\pi}_{t+1}^{(i)}}{\sum_{i=1}^{N} \hat{\pi}_{t+1}^{(i)}} \tag{2.13}$$

After initialisation a basic filtering algorithm consists of iterative application of two steps to calculate the above approximation. The *update* step is the calculation of $\hat{\pi}_{t+1}^{(i)}$ for all i, and the *normalisation* step obtains the posterior estimates $\pi_{t+1}^{(i)}$.

2.2 Implementation and numerical results

From this point, the chapter follows an iterative approach to demonstrate the evolution of particle filter methodology as it exists in the current literature. Each iteration consists of a definition of a particle filter algorithm, followed by a simulation study with a focus on deficiencies which are used to motivate the next innovation. Each incremental addition to the filter algorithm aims to resolve the deficiency found in the previous step. Thus this section focuses on already existing techniques, leading up to the next section which contains the main contributions of this chapter. This incremental process is initialised with the most basic algorithm and simulation model chosen to represent a minimal implementation

of a particle filter. The choice of a simple simulation model makes available a known posterior distribution, providing a benchmark for measuring the performance of the filter. The Kolmogorov–Smirnov statistic is used to measure performance. Although this is not a common choice, it is particularly useful for measuring performance against the benchmark, as well providing an intuitive measure for demonstrating particle impoverishment.

2.2.1 Preliminaries

Observation process

Begin with a basic Gaussian stochastic process, defined by the stochastic differential equation

$$dx_t = \sigma dW_t \tag{2.14}$$

where W_t denotes a standard Wiener process. Discrete observations used for the simulation study are generated using the Euler-Maruyama discretisation, i.e.

$$\Delta x_t = x_t - x_{t-1} = \sigma \Delta W_t \tag{2.15}$$

The particle filter presented in this section will be concerned with estimating the posterior of $p(\sigma_t|x_{1:t})$ given a set of observations $x_{1:t}$ generated by the above process. Note the use of subscript in σ_t to associate the estimate with data up to time t.
The transition density for this process is given by:

$$p(x_t|x_{t-1}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\Delta x_t{}^2}{2\sigma_t^2}} \tag{2.16}$$

Benchmark posterior

Finding the posterior of σ from observations generated by the above process is equivalent to finding the posterior distribution of the variance given a set of Gaussian increments Δx_t. An established result, based on Cochran's theorem[5], states that the distribution of σ_t^2 is obtained from the chi-square distribution with $n-1$ degrees of freedom; χ^2_{n-1} according to:

$$\frac{\hat{\sigma}_t^2 n}{\sigma_t^2} \sim \chi^2_{n-1} \tag{2.17}$$

where $\hat{\sigma}_t^2$ is the maximum likelihood estimator; $\hat{\sigma}_t^2 = \frac{\sum_t \Delta x_t^2}{n}$ with n the number of observations. The theoretical posterior can be written as:

$$p(\sigma_t|x_{1:t}) = \chi^2_{n-1}\left(\frac{\hat{\sigma}_t^2 n}{\sigma_t^2}\right) \tag{2.18}$$

with cumulative distribution function (CDF):

$$F(\sigma_t) = \int_0^{\sigma_t} \chi^2_{n-1}\left(\frac{\hat{\sigma}_t^2 n}{\sigma_t^2}\right) d\sigma_t \tag{2.19}$$

[5]See Cochran (1934).

Measuring performance

Assessing the performance of the particle filter in the presence of a theoretical benchmark amounts to measuring the distance between two posterior distributions with respect to increasing number of particles and increasing number of observations. The Kolmogorov-Smirnov (KS) statistic, which measures the maximum distance between CDFs, seems a natural choice. However, it is not commonly used in the literature, with Djuric and Míguez (2010) providing one of the few examples of usage related to the particle filter. Define the estimated CDF as

$$F^*(\sigma_t) = \sum_i \mathbb{I}_{(\sigma_t^{(i)} \leq \sigma_t)} \pi_t^{(i)}.$$

The KS statistic measures the maximum distance between the theoretical $F(\sigma_t)$ and estimated posterior $F^*(\sigma_t)$:

$$KS = \sup_{\sigma_t} |F^*(\sigma_t) - F(\sigma_t)| \tag{2.20}$$

It must be stressed that this approach is limited in use for this specific case due to the availability of a known posterior. Within this limitation, it is a simple and effective approach for demonstrating convergence, as well as demonstrating the issue of particle impoverishment; including the efficacy of the Liu and West filter for resolving it.

Convergence

Convergence has been the subject of significant research and a wide range of results exist in the literature. For detailed theoretical analysis see Chopin (2004), Del Moral and Doucet (2014), Douc and Moulines (2007) and Doucet and Johansen (2009). For a comprehensive survey of convergence results refer to Crisan and Doucet (2002). The present chapter employs numerical testing of expected convergence results as a means to assess particle filter performance. The numerical tests are based on the KS statistic relying on the basic assertion that convergence of the KS statistic implies convergence of the estimated posterior to the benchmark distribution:

$$\sup_{\sigma_t} |F^*(\sigma_t) - F(\sigma_t)| \xrightarrow[N\to\infty]{} 0 \implies F^*(\sigma_t) \xrightarrow[N\to\infty]{} F(\sigma_t) \tag{2.21}$$

Additionally, the KS statistic is used to assess convergence with respect to the number of observations, where it proves to be an effective indicator of a known issue with particle filters better known as particle impoverishment[6].

2.2.2 Basic Filter (SIS)

A basic filtering algorithm consists of initialisation followed by iterative application of two steps corresponding to (2.13). It is essentially an adaptation of the filter known as sequential importance sampling (SIS) for the detection of a static parameter. A filter for dynamic

[6] See Chen et al. (2003) p. 26 for an introduction to particle impoverishment.

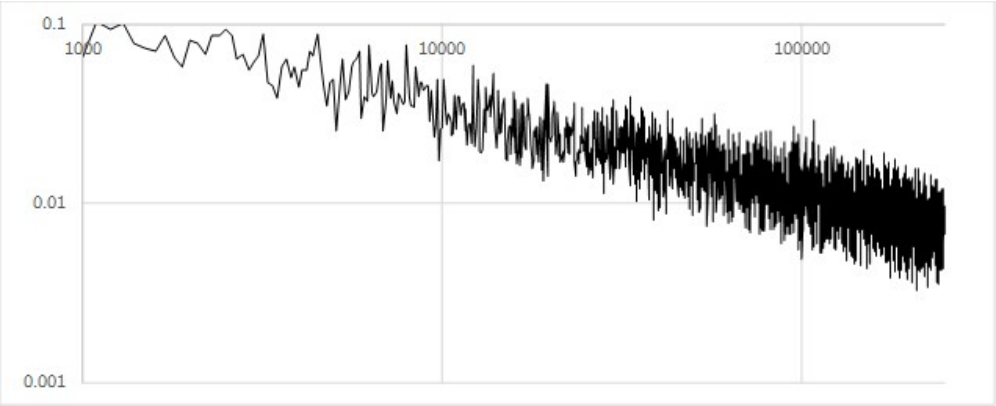

Figure 2.1: KS statistic for increasing number of particles (logarithmic scales)

state variables also would include a draw from the state variable transition kernel, see Chen et al. (2003). The *update* step is the calculation of $\hat{\pi}_{t+1}^{(i)}$ for all i, and the *normalisation* step obtains the posterior estimates $\pi_{t+1}^{(i)}$. The algorithm is defined as follows:

1: *Initilisation* For each particle; draw N particles $\sigma_0^{(i)} \sim U(a, b)$ and $\pi_0^{(i)} = \frac{1}{N}$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t | x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$

The performance of the filter is tested numerically by rerunning it with an increasing number of particles given a fixed set of observations, and recording the KS statistic for each run. The results are presented in Figure 2.1, showing the value of the KS statistic with respect to the number of particles, both in log-space. Although convergence is evident, the most prominent aspect of the results is instability as $N \to \infty$. The instability indicates that while there is an overall convergence trend, incremental increases in the number of particles result in significant noise in the KS statistic. To better understand the reason behind the instability, the theoretical (2.18) and the estimated posterior (2.6) PDF and CDF are compared visually for a small number of particles. The PDF comparison is shown in Figure 2.2. Each vertical line in the plot represents a particle weight $\pi_t^{(i)}$ and is shown against the theoretical posterior[7]. The scaled comparison demonstrates a very good correspondence between the estimated and theoretical shape of the PDF. It is also evident that the estimation points are unevenly distributed, reflecting the randomly initialised particle locations. The estimated and theoretical CDFs are compared in Figure 2.3. The KS statistic corresponds to the maximum vertical distance between the two plots. Intuitively it appears that the vertical distance between the CDFs is related to the horizontal distance between adjacent estimation points, identifiable in the CDF as the end points of each piecewise flat

[7]Note the estimate and theoretical posterior are shown at different scales as the scaling of the estimate depends on the number of particles, only converging to the scaling of the theoretical posterior as $N \to \infty$.

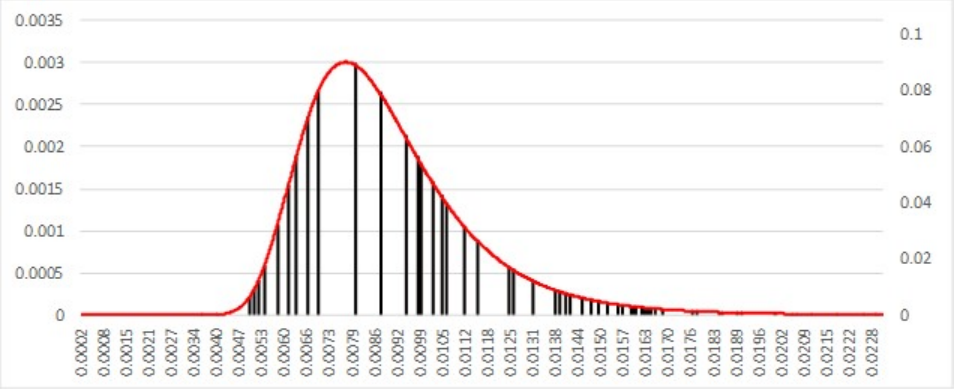

Figure 2.2: Theoretical (red, lhs) and estimated (black, rhs) σ (horizontal) posterior.

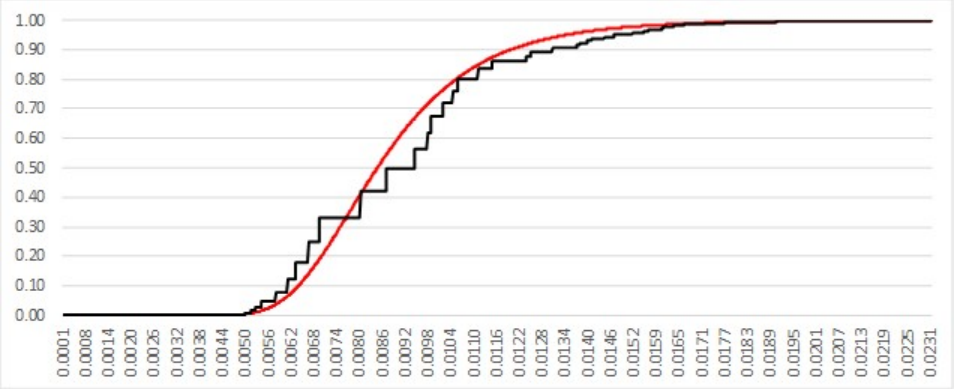

Figure 2.3: Theoretical (red) and estimated (black) σ (horizontal) posterior CDF.

interval. It follows that the maximum vertical distance, i.e the KS statistic, is related to the maximum distance between adjacent estimation points. The random initialisation of the estimation points means that the maximum distance between adjacent points does not decrease monotonically as $N \to \infty$, resulting in the instability observed in the KS statistic.

2.2.3 Equal Spacing

The above reasoning leads to a trivial improvement: if the point density locations are equally spaced the maximum distance between adjacent points will decrease monotonically as $N \to \infty$. Further, even spacing guarantees that this maximum distance is minimised for any given number of particles[8].The initialisation in the filtering algorithm is altered to reflect equally spaced estimation points:

[8]This does not guarantee the optimal estimation point distribution with respect to the KS statistic, but the improvement in performance is substantial.

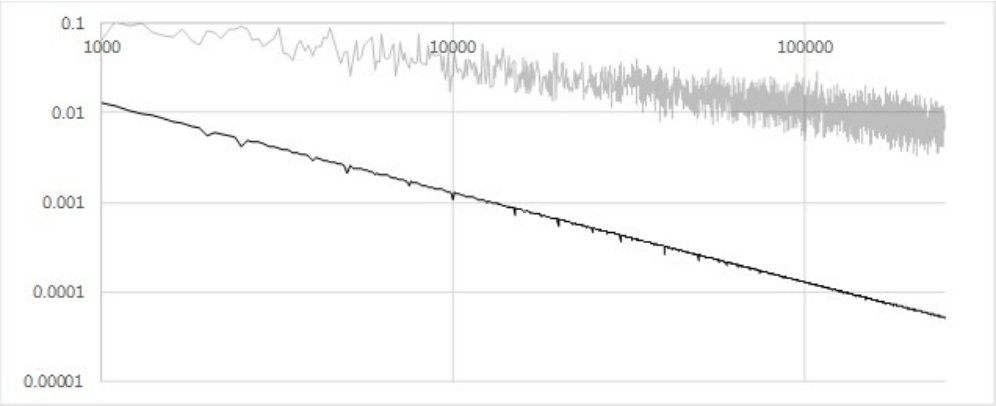

Figure 2.4: KS statistic or increasing number of particles (logarithmic scales), random (grey) vs equally spaced (black) distribution of estimation point densities

1: *Initilisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$ and $\pi_0^{(i)} = \frac{1}{N}$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t|x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$

The convergence test from the previous section is rerun with the above adjustment. The result, shown in Figure 2.4, clearly demonstrates that this simple change has resulted in a faster and more stable rate of convergence.[9] [10] The reason for the improvement is evident in the PDF Figure 2.5 and CDF Figure 2.6 and confirms the assertion from the previous section. The shape of the PDF is preserved at equally spaced intervals which allows closer and more consistent alignment between the CDFs, resulting in the elimination of noise from the convergence of the KS statistic.

A well known problem with the basic particle filter is that the number of particles with non-zero weights can only decrease with each iteration[11]. Zero weights occur when the estimated posterior probability at a particular estimation point falls below the smallest positive floating point number available for the computing machine on which the filter is implemented. To demonstrate the problem, the proportion of zero weighted particles is plotted against the number of observations in Figure 2.7. This *weight degeneration* is a problem particularly for detection of dynamic state variables where it essentially diminishes the sample domain. In this case the problem is largely solved by introducing a resampling step where the zero weight particles are replaced by sampling from the non-zero weight particles according to their relative estimated probabilities. For detection of static model

[9] This is somewhat related to research focusing on sequential Monte Carlo using quasi-random rather than pseudo-random draws, see for example Gerber and Chopin (2015).

[10] Notably the convergence is linear in log-space suggesting the form $\sup|F^*(\sigma) - F(\sigma)| \leq \frac{C}{N}$ echoing theoretical results for the convergence of mean square error, see Crisan and Doucet (2002), section V.

[11] See Chen et al. (2003) p. 26 for a good summary.

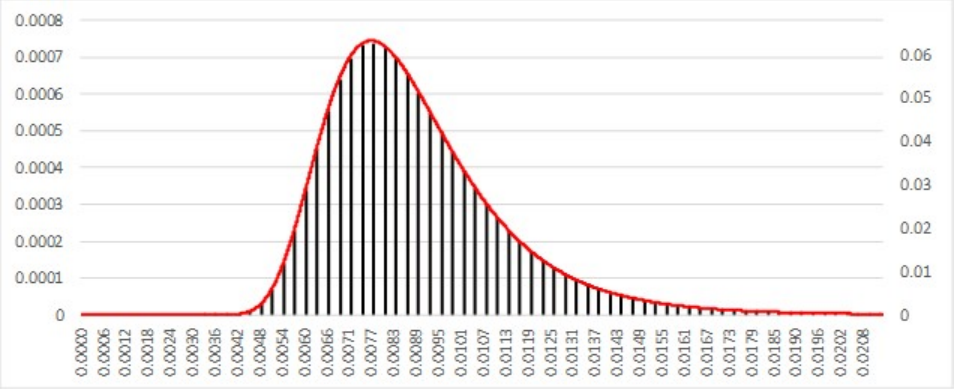

Figure 2.5: Theoretical (red, lhs) and estimated (black, rhs) σ (horizontal) posterior for equally spaced estimation points

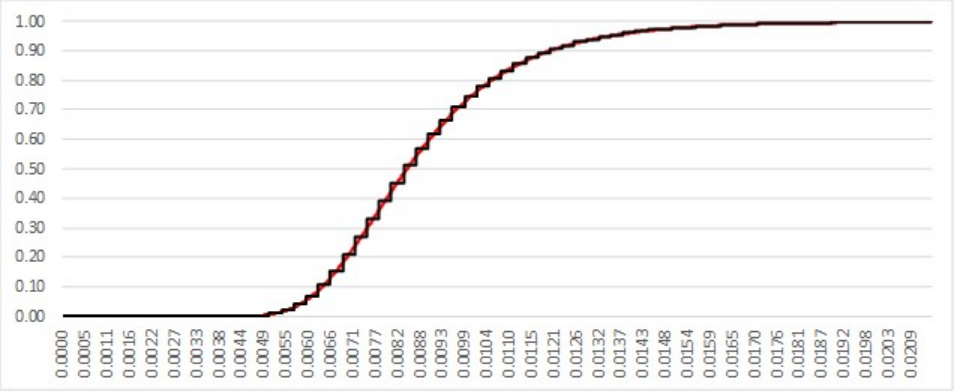

Figure 2.6: theoretical (red) and estimated (black) σ (horizontal) posterior CDF for equally spaced estimation points

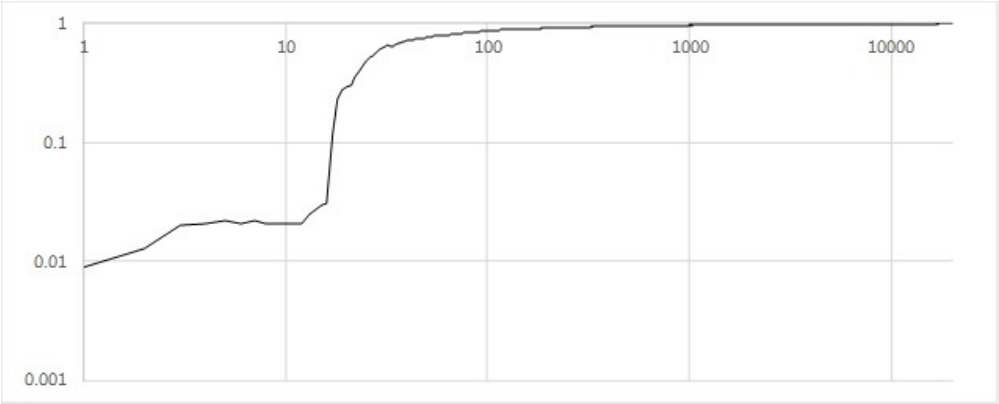

Figure 2.7: Proportion of zero weight particles with increasing number of observations for the SIS filter in log-space

parameters, simply resampling only removes the zero weight particles but does not resolve the sample impoverishment problem. Because the model parameters are assumed to be fixed, resampling simply concentrates more particles on the same (diminishing) number of estimation points. The resolution of this issue for static parameters will be demonstrated over the next two sections.

2.2.4 Resampling: SIR Filter

The first step to reducing particle impoverishment is to redistribute the particles according to the current posterior estimate, a technique called sample importance resampling (SIR). This redistribution of particles replicates higher probability particles and discards any with zero or very low weighting, thus resolving the problem highlighted in the previous section. Various methods for resampling have been proposed in literature. This chapter uses systematic resampling, which can be found in the survey analysis of Hol, Schon and Gustafsson (2006), where it is described as having the lowest discrepancy and reduced computation complexity without deterioration of the estimate. The goal of the resampling step is to transform the posterior distribution approximated by N particles of differing weights into one approximated by N particles of equal weight. The resampling begins by generating N ordered numbers:

$$u_k = \frac{(k-1)+\tilde{u}}{N} \text{ , with } \tilde{u} \sim U[0,1), \tag{2.22}$$

then reselecting the particles according to

$$\sigma_t^{(k)} = \sigma_t^{(i)} \text{ , with } i \text{ s.t. } u_k \in \left[\sum_{s=1}^{i-1} \pi_t^{(s)}, \sum_{s=1}^{i} \pi_t^{(s)} \right) \tag{2.23}$$

The filtering algorithm is adjusted to include the resampling step:

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$ and $\pi_0^{(i)} = \frac{1}{N}$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t | x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t | x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)} = \sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t | x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)} = \sigma_t\}}$$

As discussed, the above algorithm discards zero weighted particles (or, more accurately, is likely to discard particles of very low weight), resolving the problem posed in the previous section. However, for static model parameter estimation, the particle impoverishment persists, because resampling simply concentrates particles on the same estimation points. Because the parameters are assumed to be fixed, their value does not change as it would for dynamic state parameters. As the number of observations increases, the theoretical

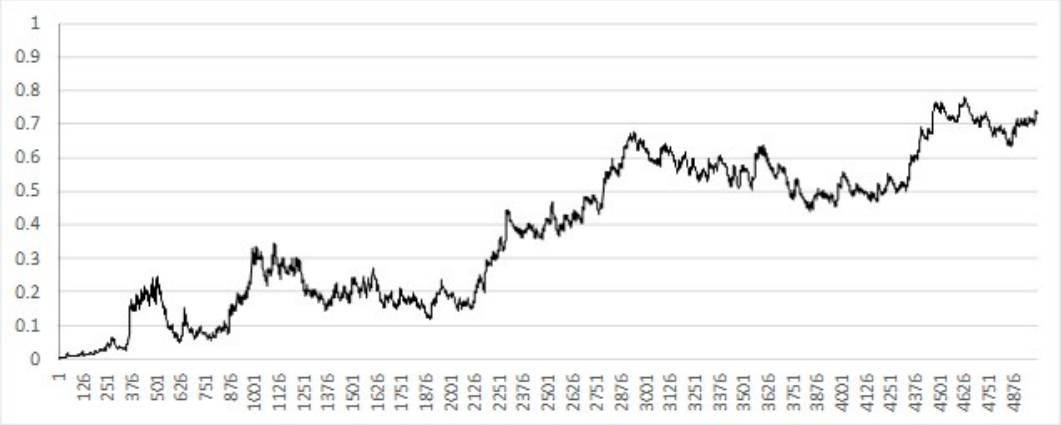

Figure 2.8: KS statistic for increasing number of observations for the SIR filter

posterior density becomes increasingly more concentrated around the maximum likelihood estimate, in the limit approaching the Dirac delta measure. The key intuition to understanding particle impoverishment is that the estimated posterior will concentrate the estimation density on a single point after a finite number of observations rather than in the limit: On each update of the weight $\pi_t^{(i)}$ the total weight becomes more and more concentrated on the parameter value $\theta^{(i)}$ of maximum likelihood given the observations, but for a finite number of observations the parameter value of maximum likelihood does not necessarily coincide with the "true" parameter of the data generating process, and for a finite number of particles, the discretisation of the parameter space by the $\theta^{(i)}$ will also mean that the best possible $\theta^{(i)}$ will not coincide with the exact "true" parameter. Therefore, in the limit, the theoretical and estimated posterior will both be concentrated on a single point at different locations, with the estimate reaching this state after a finite amount of observations. Once the estimate reaches this point the KS statistic is based on the one estimation point (denote as σ^*) and equals $\max(1 - F(\sigma^*), F(\sigma^*))$. As the theoretical posterior converges, its PDF narrows until the single remaining estimation point is outside its numerically significant domain; reflecting this the KS statistic approaches the maximum value of 1. This is demonstrated numerically by running the particle filter over a large number of observations and recording the KS statistic at each sequential estimate, as shown in Figure 2.8. The KS statistic approaching 1 as the number of observations increases is an indicator of theoretical and particle filter posteriors diverging from each other as a consequence of particle impoverishment. As an additional explanation, Figure 2.9 shows an example of a PDF for an impoverished state of the particle filter, where the posterior is estimated by just four particle locations. The example shows a state where the theoretical posterior has narrowed as it converges. However, the spacing between the estimation points has not changed. This will eventually lead to the posterior being estimated by a single point, which eventually will fall outside of the theoretical posterior. One of the many approaches motivated by the problem of particle impoverishment, initially proposed by Gordon et al. (1993), is to add small random perturbations to every particle at each iteration of the filtering algorithm,

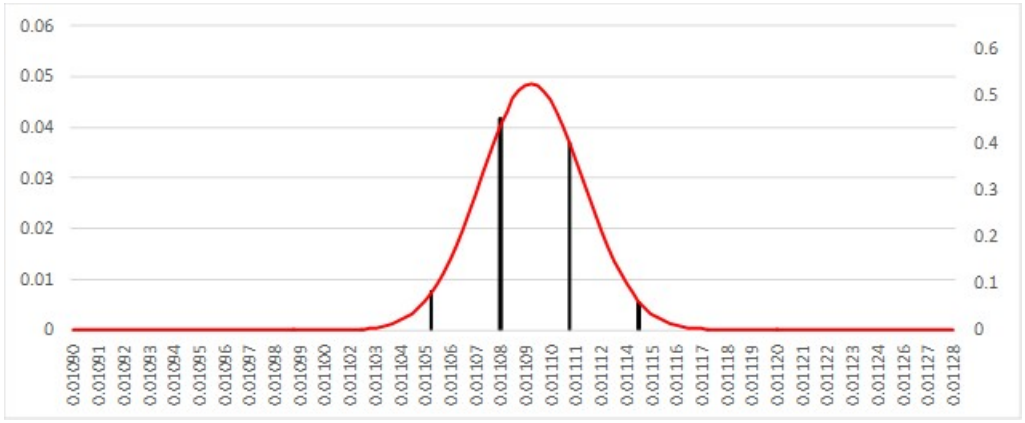

Figure 2.9: Theoretical (red, lhs) and estimated (black, rhs) σ (horizontal) posterior demonstrating particle impoverishment.

that is

$$\theta_{t+1} = \theta_t + \zeta_{t+1} \tag{2.24}$$

$$\zeta_{t+1} \sim N(0, W_{t+1}) \tag{2.25}$$

where W_{t+1} is a specified variance matrix. While this approach provides a framework for addressing particle impoverishment, it does so at the cost of accuracy to the posterior distribution. Any random perturbation to the fixed parameters introduces an artificial dynamic resulting in potential overdispersion of the parameter estimate. For example, if the variance W_{t+1} of the random perturbation is constant, the constant value becomes the minimum variance of the posterior estimate, i.e at some point the minimum variance becomes larger than the variance of the theoretical posterior, almost exactly the opposite effect to particle impoverishment. It is therefore desirable to have the perturbation variance shrink in line with the posterior convergence such that it always remains only a relatively small contributor to the estimation variance. One such approach which explicitly addresses overdispersion is proposed by Liu and West (2001), and the literature refers to this as the Liu and West filter.

2.2.5 Liu and West filter

To resolve the problem of over-dispersion, Liu and West (2001) put forward an approach using a kernel interpretation of the random perturbation proposed by Gordon et al. (1993). The idea of the kernel representation is that each parameter in the particle population exists as a density instead of a single point. The overdispersion is resolved by linking the variance of the kernel to the estimated posterior variance such that it shrinks proportionally to the convergence of the estimated posterior. The practical application within the filter algorithm is to draw the parameter from the kernel density for each particle at each iteration. The kernel is expressed as a normal density $\mathcal{N}(\sigma|m, S)$ with mean m and variance S and replaces

the Dirac delta density in equation (2.6):

$$p(\sigma_t|x_{1:t}) \approx \sum_{i=1}^{N} \pi_t^{(i)} \mathcal{N}(\sigma_t^{(i)}|m_t^{(i)}, h^2 V_t) \tag{2.26}$$

where V_t is the variance of the current posterior $V_t = \frac{1}{N}\sum_i (\sigma_t^{(i)} - \overline{\sigma}_t)^2$ and

$$m_t^{(i)} = c\sigma_t^{(i)} + (1-c)\overline{\sigma}_t \tag{2.27}$$

with $c = \sqrt{1-h^2}$ and $\overline{\sigma}_t$ the mean of the current posterior. The filtering algorithm now includes a kernel smoothing step where the posterior points are drawn from the kernel defined in eq. (2.26):

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$ and $\pi_0^{(i)} = \frac{1}{N}$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t|x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t|x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)}=\sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t|x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)}=\sigma_t\}}$$

2.4: *Kernel smoothing* For each particle apply $\sigma_t^{(i)} \sim \mathcal{N}(\sigma_t^{(i)}|m_t^{(i)}, h^2 V_t)$

Numerical results demonstrating the effectiveness of the Liu and West filter in reducing particle impoverishment are shown in Figure 2.10. The KS statistic for the Liu and West filter remains relatively constant as the number of observations increases, indicating that the filter estimate convergence with respect to the number of observations is well aligned with the theoretical posterior. The reason for the improvement is confirmed by a comparison of the estimated posterior to the theoretical PDF, shown in Figure 2.11. The example demonstrates the effectiveness of the kernel in estimating both the location and variance of the theoretical posterior.

2.3 Parameter learning and change detection

This section continues to extend the particle filter, introducing techniques representing the main research contribution of this chapter. The section begins by introducing a regime shift to the Gaussian reference model in order to pose a more difficult filtering problem and highlight the adaptive aspect of the Liu and West filter. It is shown that the adaptation is a result of the combination of random perturbation and re-selection, forming a genetic algorithm capable of adapting to parameter regime changes. After demonstrating the link between adaptation speed and the size of the random perturbation, an extension to the Liu and West filter which increases the kernel variance when required for adaptation is

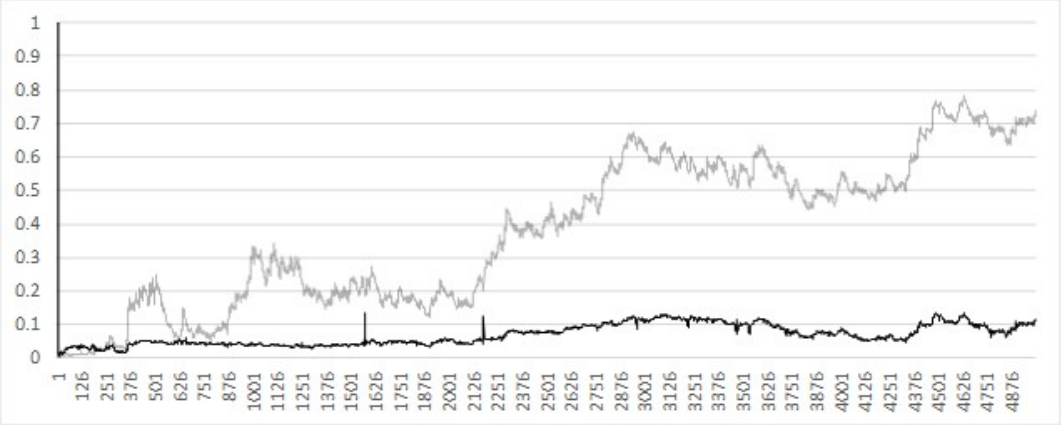

Figure 2.10: KS statistic for increasing number of observations SIR (grey) and Liu and West (black)

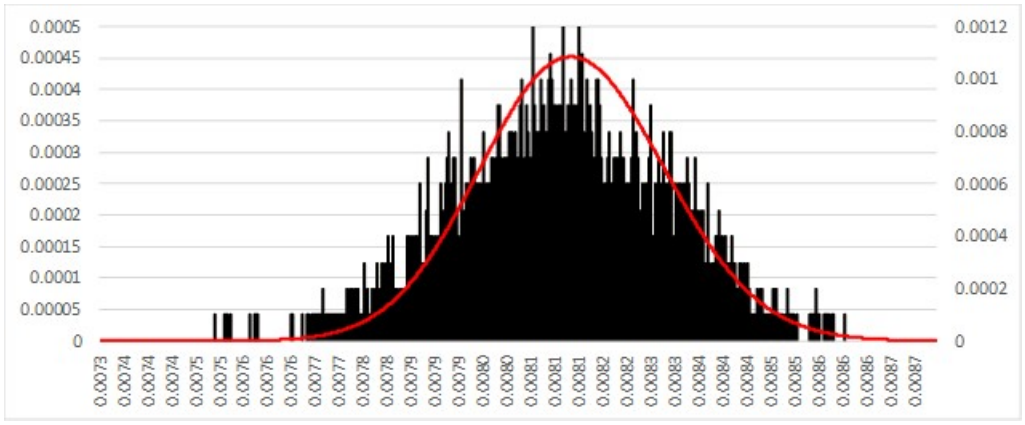

Figure 2.11: Theoretical (red, lhs) and estimated (black, rhs) posterior for σ (horizontal) in the Liu and West filter

proposed. This is achieved by exploiting the genetic algorithm embedded in the Liu and West filter, thus allowing the size of parameter random perturbation to evolve as part of the already existing process. The result is a filter capable of adapting to regime changes and converging to the Liu and West filter when adaptation is not required. The capacity of the filter to adapt to stochastic volatility is also illustrated. Finally, it is shown how measuring the average adaptation at each iteration can provide useful information which can be used to distinguish between different dynamics of the underlying data.

2.3.1 Regime shift

Consider a model where at time $t = t^*$ there is a change in volatility,

$$dx_t = (\sigma_1 \mathbb{I}_{t<t^*} + \sigma_2 \mathbb{I}_{t \geq t^*}) dW_t \tag{2.28}$$

Up to the time t^*, the above model is identical to (2.14) and the particle filter performs as demonstrated in the previous section. In the case of the Liu and West filter, the range of the posterior support will narrow in line with the converging theoretical posterior as the number of observations increases up to time t^*. At the point $t = t^*$, two situations are possible, the new value σ_2 could lie either inside or outside the range of the estimated posterior. In the case that it is inside, that is there are at least two particles such that $\sigma_t^{(i)} \leq \sigma_2 \leq \sigma_t^{(j)}$, the weights of the particles closest to σ_2 will start to increase and eventually the filter will converge to the new value. However, in order to develop and illustrate a genetic algorithm approach, this section will focus on the opposite case, where σ_2 is outside the range of the posterior. This situation will be labelled the adaptation phase.

In general, the posterior density, given enough observations, tends to converge around the parameter values set in the simulation used to generate the observations. However, this is not possible during the adaptation phase, since by definition the range of the posterior does not encompass the new parameter value. In this case, the posterior will converge to the point closest to the new value σ_2, located at the boundary of the existing posterior range. In the limit, all density will be focused on the single particle closest to σ_2, that is:

$$\pi_t^{(i)} \xrightarrow{p} 1 \text{ for } i \; s.t. \; |\sigma_2 - \sigma_t^{(i)}| = \inf_{1 \leq j \leq N} [|\sigma_2 - \sigma_t^{(j)}|] \tag{2.29}$$

The presence of random perturbation, in the form of the kernel used in the Liu and West filter, allows the posterior interval to expand and therefore decrease the distance of the interval boundary to σ_2. That is, there is a non-zero probability that the random perturbation results in at least one of the new particle locations falling outside the current estimation boundary. This is especially the case during the adaptation phase, where posterior density is accumulated at the boundary. This translates to

$$P(\inf_{1 \leq j \leq N} [|\sigma_2 - \widetilde{\sigma}_t^{(j)}|] < \inf_{1 \leq j \leq N} [|\sigma_2 - \sigma_t^{(j)}|]) > 0, \text{ where } \widetilde{\sigma}_t^{(j)} \sim \mathcal{N}(\sigma_t^{(j)} | m_t^{(j)}, h^2 V_t) \tag{2.30}$$

The random perturbation combines with the re-selection to form a genetic algorithm capable of adapting the posterior to the new value by expanding the posterior such that $\inf[|\sigma_2 - \sigma_t^{(j)}|] \to 0$, thereby allowing the posterior to shift towards σ_2. In the Liu and West filter, the kernel variance V_t is determined by the variance of the particles, which tends towards zero as the particles become increasingly concentrated around the boundary. The variance is prevented from reaching zero by the random expansion of the boundary, giving the posterior incremental space and preventing collapse to a single point. The net result is a situation where the two forces tend to balance out resulting in a relatively steady rate of boundary expansion towards σ_2. This is evident in Figure 2.12: after the regime change there is a slow and relatively constant change in the estimate in the direction of σ_2.

The adaptation demonstrated in this section stems from the combination of random perturbation via the kernel and re-selection, creating a type of genetic algorithm. Although adaptation is evident, it is very slow; the Liu and West filter was designed to smooth the posterior without causing overdispersion, and not to rapidly adapt to model parameter regime changes.

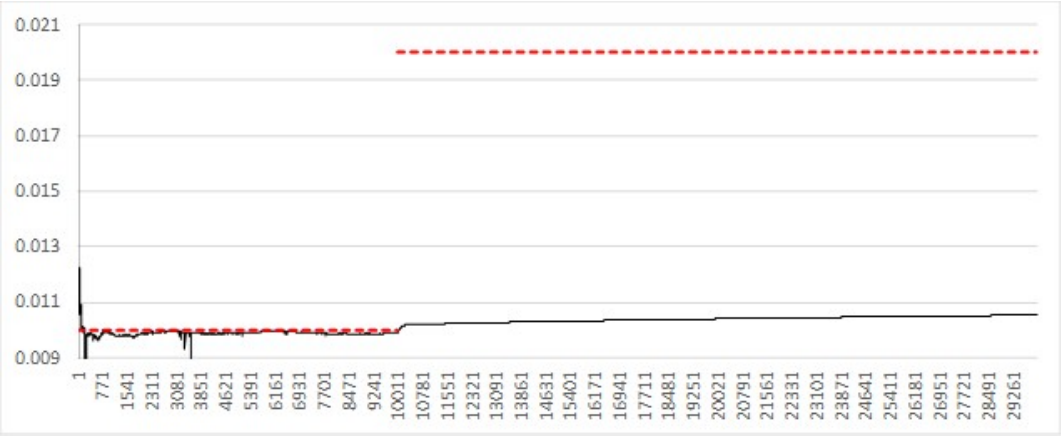

Figure 2.12: Estimated expected value (black) vs simulation input value (red) of σ_t with regime change after 10,000 steps for the Liu and West filter

2.3.2 Controlling the rate of adaptation

Equation (2.30) implies that the speed of adaptation is directly related to the speed of expansion of the posterior interval, which in turn is driven by the size of the kernel variance V_t. In the case of the Liu and West filter, the adaptation is slow since the variance of the kernel depends on the variance of the posterior and therefore shrinks in line with the convergence of the posterior. To illustrate the relationship between the size of the random kernel variance and adaptation speed, an additional noise term ϕ is introduced into the kernel used by Liu and West as follows:

$$p(\sigma_t|x_{1:t}) \approx \sum_{i=1}^{N} \pi_t^{(i)} \mathcal{N}(\sigma_t^{(i)}|m_t^{(i)}, h^2V_t + \phi) \tag{2.31}$$

The filtering algorithm becomes:

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$ and $\pi_0^{(i)} = \frac{1}{N}$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t|x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t|x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)}=\sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t|x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)}=\sigma_t\}}$$

2.4: *Kernel smoothing* For each particle apply $\sigma_t^{(i)} \sim \mathcal{N}(\sigma_t^{(i)}|m_t^{(i)}, h^2V_t + \phi)$

Figure 2.13 and Figure 2.14 show filtering results with the above change for varying levels of the noise term ϕ. The key aspect of the results is that increasing ϕ indeed increases the

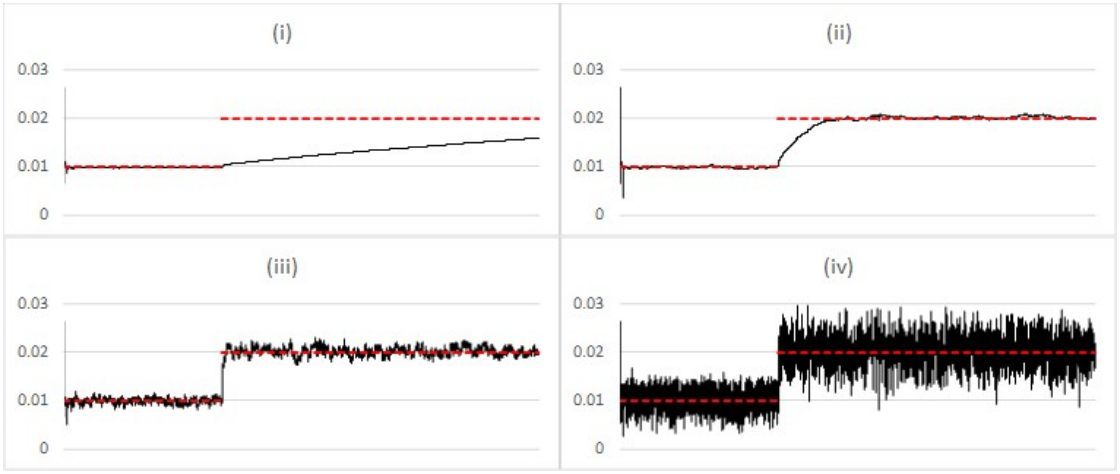

Figure 2.13: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ_t, when there is a regime change after 10,000 steps, using the Liu and West filter with additional noise, for different values of ϕ (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

adaptation speed, but at the cost of significant additional noise in the prediction.[12]

The regime shift, applied to the dynamics behind Figures 2.13 and 2.14, provides a test case with a large sudden change at a specific point in time. As another test case, Figure 2.15 and Figure 2.16 considers a basic stochastic volatility model. In contrast to a regime shift, changes in the model volatility are driven by a diffusion, testing the capability of the filter to detect continuous, rather than sudden discrete changes. The process is defined by the following system of SDEs:

$$dx_t = \alpha_t dW_{1,t} \tag{2.32}$$

$$d\alpha_t = \nu dW_{2,t} \tag{2.33}$$

where $W_{1,t}$ and $W_{2,t}$ denote independent standard Wiener processes.

The filter is applied to the stochastic volatility model without any alteration from the setup used to detect regime changes. As is apparent in particular in Figure 2.15(ii), some degree of additional noise seems to help in the detection of the underlying value of α. However, similarly to the regime shift, too much noise simply translates to a noisy estimate. These results also highlight the resemblance to a filter configured to detect only a stochastic volatility model, i.e., a filter set up to detect the state parameter α given a value of ν[13]. Each iteration would contain an additional step where each particle's $\alpha_{(i)}$ is updated according to the stochastic volatility dynamic. In this case, the additional noise parameter ϕ acts in a similar fashion to the stochastic volatility parameter ν. The difference in the approach highlights one of the motivating factors behind the method, the approach being presented

[12]Figure 2.12 results are equivalent to the above filter if one sets $\phi = 0$.

[13]See Bao et al. (2012), Casarin (2004) for examples of stochastic volatility model detection.

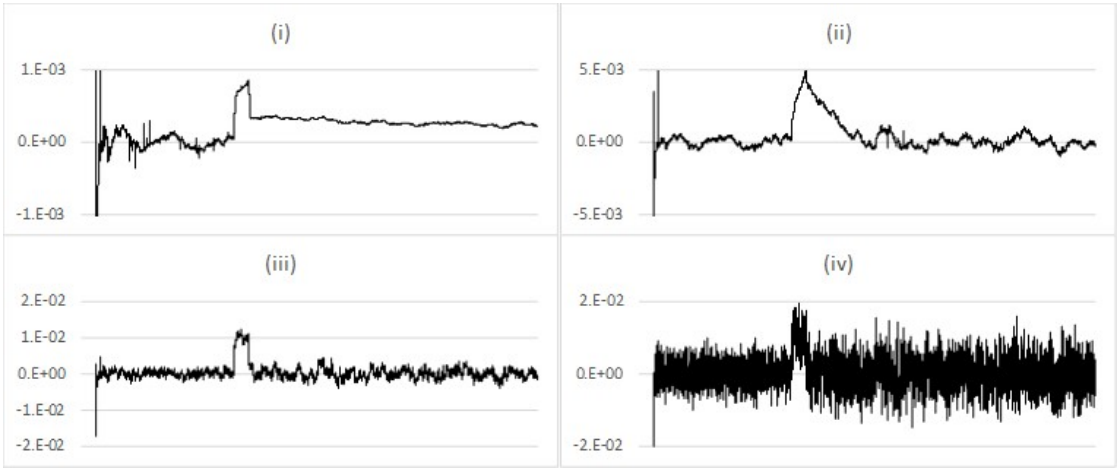

Figure 2.14: Comparing the change in estimated posterior expected value of σ_t, where there is a regime change after 10,000 steps, using the Liu and West filter with additional noise, for different values of ϕ (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

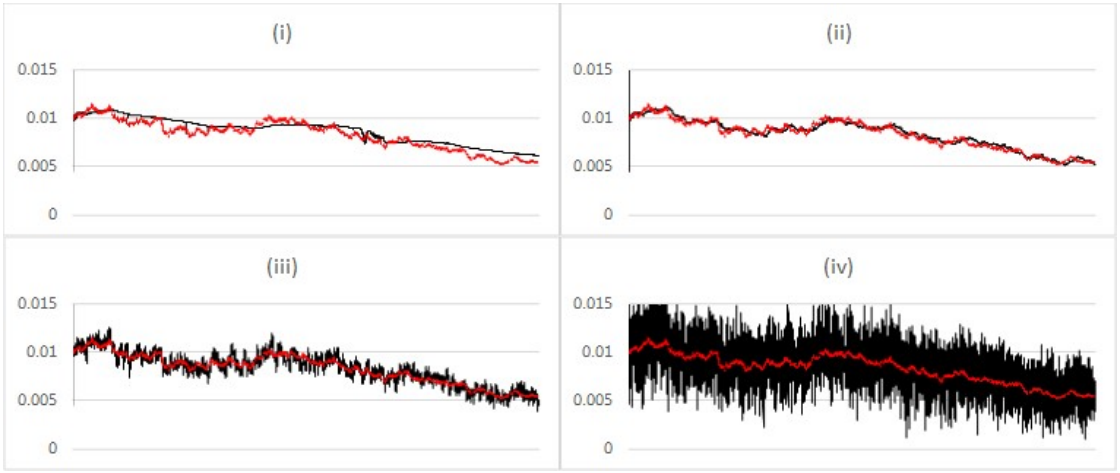

Figure 2.15: Comparing estimated posterior expected value (black) vs simulation input value (red) of α in the stochastic volatility model, using the Liu and West filter with additional noise, for different values of ϕ (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

does not have to assume prior knowledge of the underlying model, rather it can act as a gauge for empirical assessment of data and with limited modelling assumptions can suggest fruitful extensions toward more sophisticated models. In the regime shift example, the noise parameter ϕ improved the adaptation speed at the cost of prediction noise. A high level of ϕ is only desirable during the adaptation phase, at other times the ideal level of ϕ would be zero. For the stochastic volatility example clearly there is some optimal level of ϕ which achieves good filter performance without causing excessive noise. This is the motivation for a methodology for automatically selecting the level of ϕ based on the data, based on an examination of the behaviour of particles on the boundary of the posterior

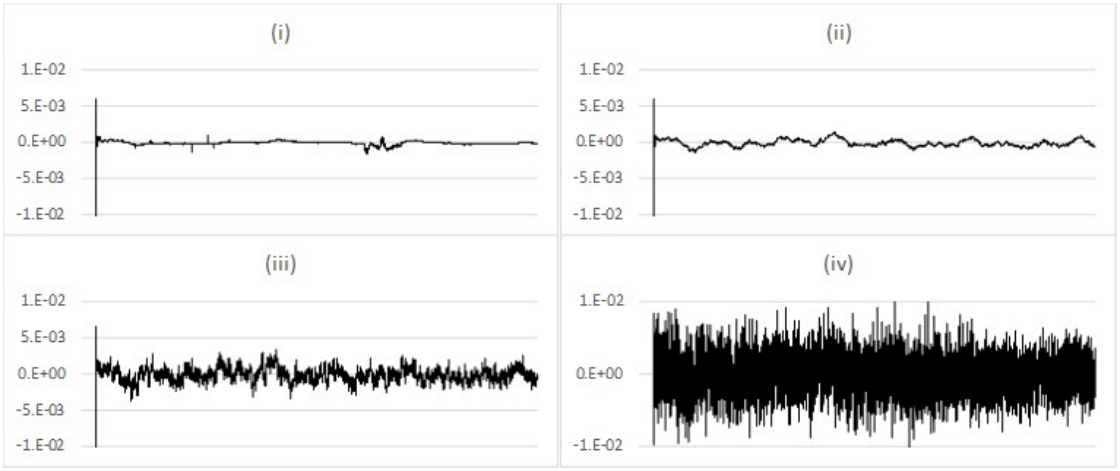

Figure 2.16: Comparing the change in estimated posterior expected value of α in the stochastic volatility model, using the Liu and West filter with additional noise, for different values of ϕ (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

during the adaptation phase, which will be considered next.

2.3.3 Applying selection to the rate of adaptation

Results from the previous section show that the adaptation of the filter after a regime change is driven by posterior boundary expansion resulting from random perturbation. During the adaptation phase, there is a persistent concentration of density around the boundary of the posterior closest to the new value. It is as though the particles seek to be as close as possible to the new value and are pushing the posterior in this direction. Therefore the behaviour of particles on the edge of the posterior should be quite different during the adaptation phase than at other times. It remains to quantify this difference and use it to enhance the performance of the filter.

One of the differences, already highlighted, is the concentration of posterior density around the boundary during the adaptation phase. This is examined numerically by measuring how much probability mass the update step moves into the pre-update tail of the posterior. The measurement is made by first, before the update step, finding the lowest σ_t^* such that

$$\sum_i \mathbb{I}_{(\sigma_t^{(i)} \geq \sigma_t^*)} \pi_t^{(i)} \leq p$$

when $\sigma_2 > \sup[\sigma_t^{(j)}]$ or the highest σ_t^* such that

$$\sum_i \mathbb{I}_{(\sigma_t^{(i)} \leq \sigma_t^*)} \pi_t^{(i)} \leq p$$

when $\sigma_2 < \inf[\sigma_t^{(j)}]$. Following the update step, compute the amount of probability mass which has moved beyond σ_t^*, i.e. into the tail, using either $\sum_i \mathbb{I}_{(\sigma_t^{(i)} \geq \sigma_t^*)} \hat{\pi}_{t+1}^{(i)}$ when $\sigma_2 > \sup[\sigma_t^{(j)}]$ or $\sum_i \mathbb{I}_{(\sigma_t^{(i)} \leq \sigma_t^*)} \hat{\pi}_{t+1}^{(i)}$ when $\sigma_2 < \inf[\sigma_t^{(j)}]$. If the new cumulative density is higher than p, it means that the density in the tail of the posterior has increased. If this measure is persistently high through cycles of weight update and re-selection, it is a strong indicator of regime change.

Indeed, Figure 2.17 generated with $p = 0.05$, reveals a notable increase in the weight associated with the edge particles during the adaption phase. In the first case, the measure persists at the maximum value of 1.0, reflecting the slow adaptation observed for this setting, where for a substantial number of update steps all probability mass is shifted beyond σ_t^* in each step (i.e., because of the choice of small ϕ, the posterior moves toward the new "true value" only in small increments). Consistent with the findings in the previous sections, the speed of adaptation depends on the size of ϕ at the cost of noise in the results.[14]

Another quantity to consider is the size of the dispersion of each particle from the application of the kernel. Define realised dispersion for each particle as the distance it has moved from the current location due to the application of the kernel, denoted as $|\Delta\sigma_t^{(i)}|$. Consider the situation where all particles are in the same location, i.e the posterior exists at one point. After the application of the kernel, it is obvious that the particles on the edge of the posterior will have the highest realised dispersion. In the opposite situation where the particles are very widely dispersed and the kernel variance is relatively small, the relative position of the particle after application of the kernel will have minimal relation with realised dispersion. Therefore the relation between realised dispersion and particle location at the posterior boundary depends on the existing level of dispersion and relative kernel variance. As already determined, during the adaptation phase the particles tend to be very concentrated at the boundary, therefore are closer to the situation where they are likely to exhibit a relation where particles located on the edges will tend to have higher realised dispersion.

The combination of higher density and realised dispersion at the edge of the posterior results in a selection bias of high realised dispersion particles during the adaptation phase. This is verified numerically by recording the total realised dispersion $\sum_i |\Delta\sigma_t^{(i)}|$ following each re-selection step. The results are shown in Figure 2.18 and show a similar pattern to the results in Figure 2.17, confirming the assertion.

The results so far have established a relationship between the value of ϕ and the adaptation speed, and a selection bias for particles with high realised dispersion during the adaptation phase. Realised dispersion is a function of ϕ, which so far has been kept constant, connecting the two results. Through its connection to realised dispersion, redefining

[14] Results for Figure 2.17 were generated with the identical filter configuration to the results shown in Figure 2.13.

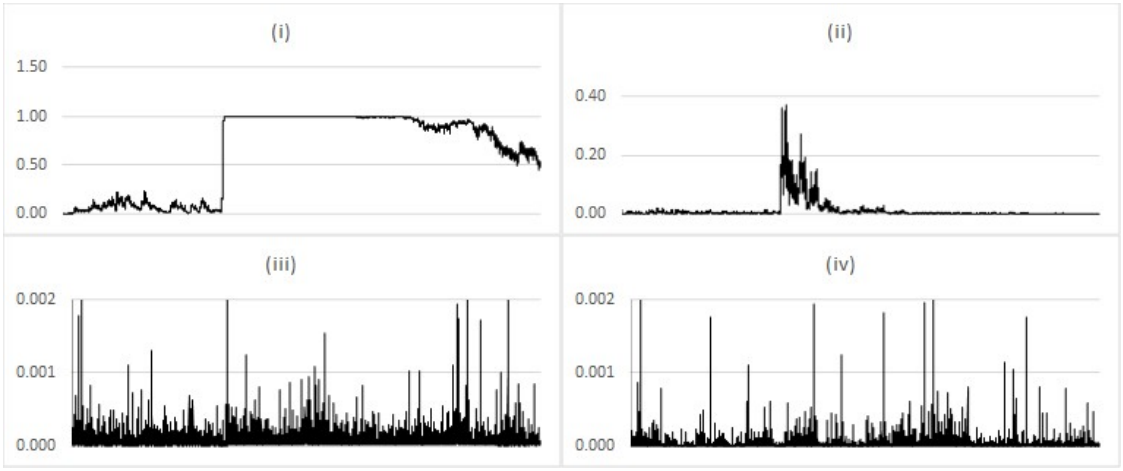

Figure 2.17: Probability mass shifted beyond the 5% quantile on the edge of the posterior with regime change after 10,000 steps, using the Liu and West filter with additional noise term ϕ equal to (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

ϕ to be non-constant will subject it to the same selection bias. To take advantage of this, define $\phi^{(i)}$ for each particle, initialised using $\phi^{(i)} \sim U(0, c)$. This way high values of ϕ leading to high dispersion will tend to be selected during the adaptation phase increasing adaptation speed. Conversely, low values of ϕ will tend to be selected when adaptation is not required, reducing noise. The filtering algorithm now becomes:

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$, $\pi_0^{(i)} = \frac{1}{N}$ and $\phi^{(i)} \sim U(0, c)$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t | x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t | x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)} = \sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t | x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)} = \sigma_t\}}$$

2.4: *Kernel smoothing* For each particle apply $\sigma_t^{(i)} \sim \mathcal{N}(\sigma_t^{(i)} | m_t^{(i)}, h^2 V_t + \phi^{(i)})$

The algorithm is tested with the initial distribution set such that the expected value of ϕ for each test is equivalent to the value set for the tests in the previous section. The results, shown in Figure 2.19 and Figure 2.20, when compared to Figures 2.13 and 2.14, reveal a significant reduction in noise coupled with an increase in the speed for charts (i) and (ii) but a decrease for charts (iii) and (iv). The reduction in noise results from a selection bias for low ϕ particles when not in the adaptation phase as discussed above. Conversely, the increase in adaptation speed for charts (i) and (ii) results from a selection bias towards higher ϕ particles during the adaptation phase. The slowdown in adaptation speed observed in charts (iii) and (iv) occurs because before the adaptation phase high ϕ particles

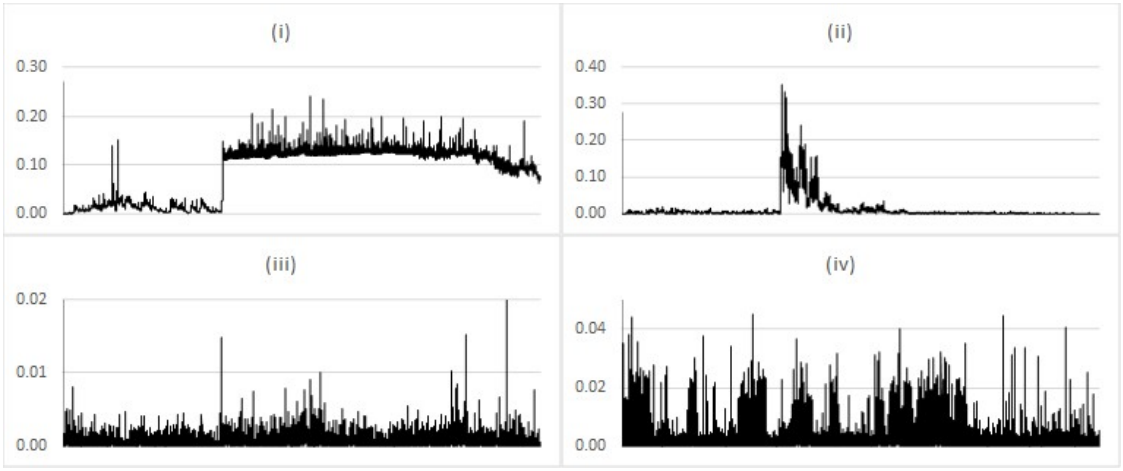

Figure 2.18: Realised dispersion after re-selection for simulated data with regime change after 10,000 steps, using the Liu and West filter with additional noise term ϕ equal to (i)$\frac{0.1}{N}$ (ii)$\frac{1.0}{N}$ (iii)$\frac{10.0}{N}$ (iv)$\frac{100.0}{N}$

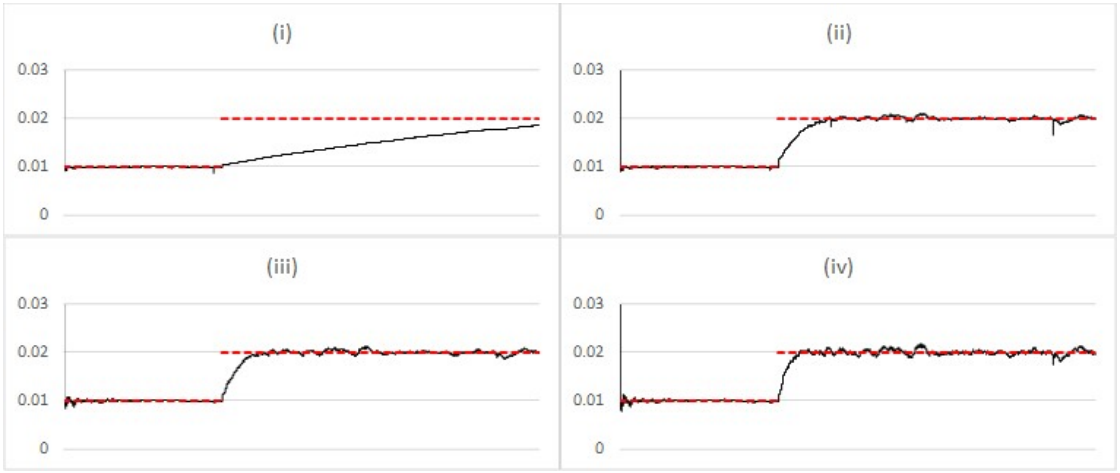

Figure 2.19: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ_t with regime change after 10,000 steps, using the Liu and West filter with additional noise parameter $\phi^{(i)} \sim U(0, c)$ with c set to (i)$\frac{0.2}{N}$ (ii)$\frac{2.0}{N}$ (iii)$\frac{20.0}{N}$ (iv)$\frac{200.0}{N}$

tend to be eliminated from the particle population by the selection process. The filter is also applied to the stochastic volatility model with results shown in Figure 2.21 and Figure 2.22. Similarly to the results for the regime change, there is an elimination of noise from the results. However, particularly for charts (ii), (iii) and (iv), the results are very similar to each other indicating that the selection process has converged on a similar level of α, highlighting the ability of the filter to find the correct level of additional noise corresponding to the constant stochastic volatility parameter.

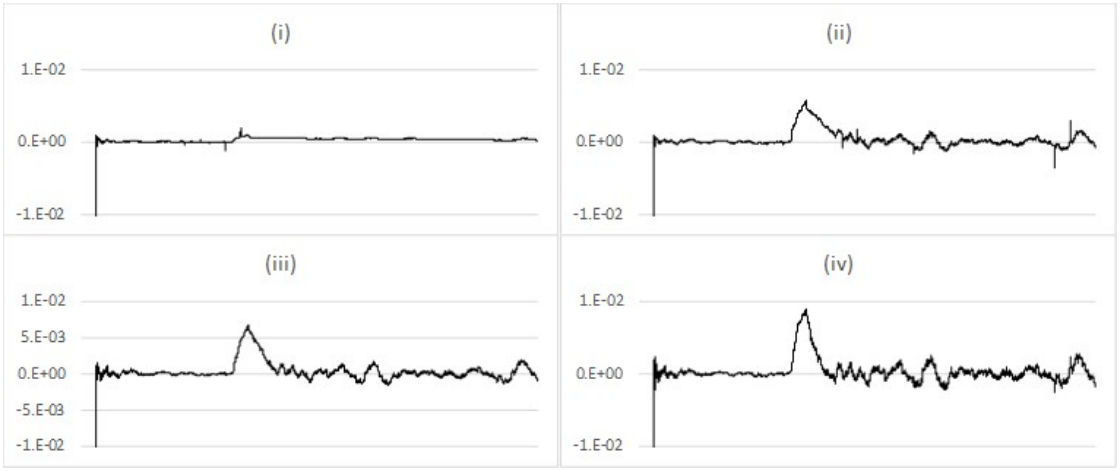

Figure 2.20: Comparing the change in estimated posterior expected value (black) vs simulation input value (red) of σ_t with regime change after 10,000 steps, using the Liu and West filter with additional noise parameter $\phi^{(i)} \sim U(0, c)$ with c set to (i)$\frac{0.2}{N}$ (ii)$\frac{2.0}{N}$ (iii)$\frac{20.0}{N}$ (iv)$\frac{200.0}{N}$

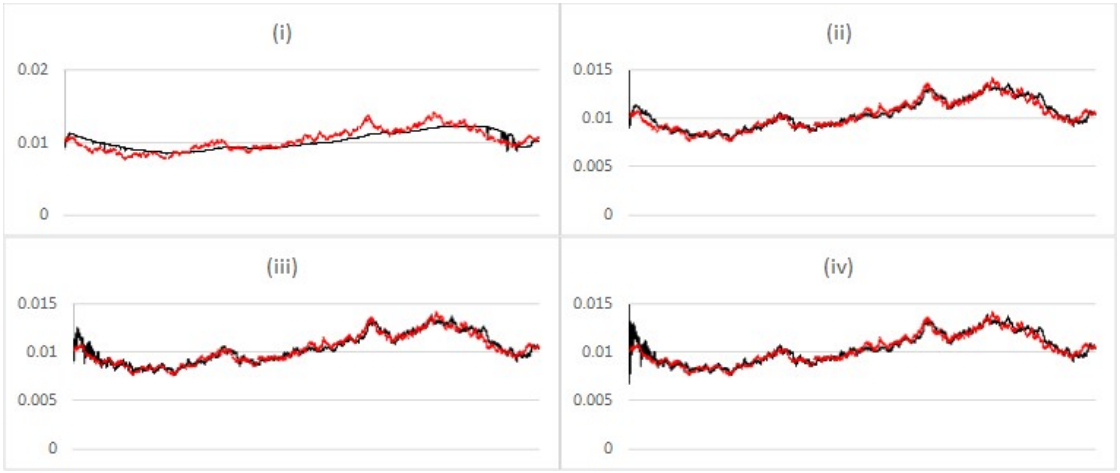

Figure 2.21: Comparing estimated expected value (black) vs simulation input value (red) of α in the stochastic volatility model for the Liu and West filter with additional noise parameter $\phi^{(i)} \sim U(0, c)$ with c set to (i)$\frac{0.2}{N}$ (ii)$\frac{2.0}{N}$ (iii)$\frac{20.0}{N}$ (iv)$\frac{200.0}{N}$

The above algorithm takes advantage of the existing selection process to increase the adaptation speed when required and reduce noise in the results when adaptation is not required. The detection of an increase in adaptation speed during the adaptation phase is now embedded in the algorithm via the selection of the noise term. However, the speed of adaptation remains relatively constant, bounded by the range of the initial distribution of ϕ, which can only shrink as a result of the selection process. The next section describes a method which overcomes this limitation and achieves accelerated adaptation.

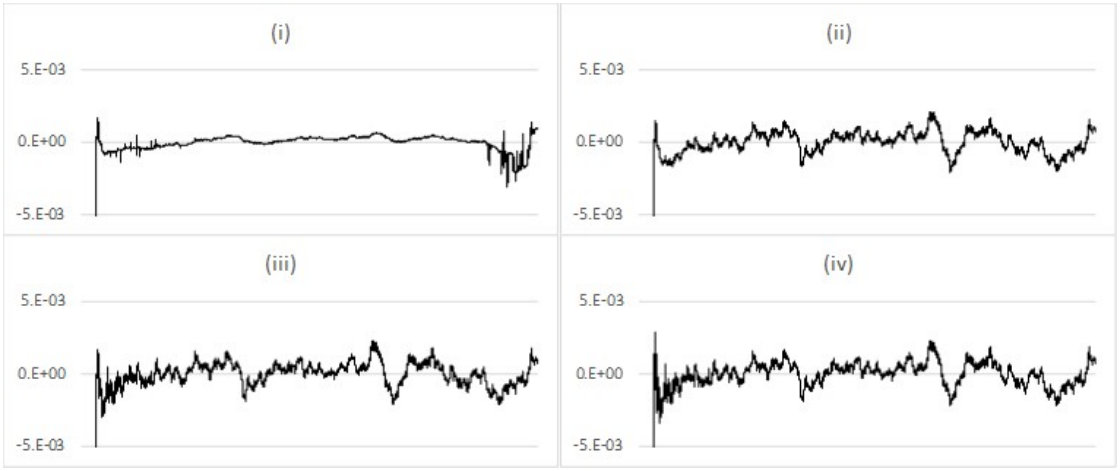

Figure 2.22: Comparing the change in estimated posterior expected value (black) vs simulation input value (red) of α in the stochastic volatility model, using the Liu and West filter with additional noise parameter $\phi^{(i)} \sim U(0, c)$ with c set to (i)$\frac{0.2}{N}$ (ii)$\frac{2.0}{N}$ (iii)$\frac{20.0}{N}$ (iv)$\frac{200.0}{N}$

2.3.4 Accelerated adaptation: selectively increasing the rate of adaptation

Adaptation in a particle filter is driven by a genetic algorithm resulting from a combination of selection and random perturbation. The speed of the adaptation is bounded by the size of the parameter ϕ, which sets the level of variance of the random perturbation via the smoothing kernel. To increase the speed of, or accelerate, the rate of adaptation, the parameter ϕ needs to constantly increase during the adaptation phase. The idea to allow ϕ itself to adapt this way, is to use the already existing genetic algorithm by subjecting ϕ to both selection and random perturbation. The effectiveness of selection on ϕ has already been demonstrated in the last section. In this section the genetic algorithm for ϕ is completed by adding a random perturbation; $\phi_{t+1}^{(i)} = \phi_t^{(i)} e^{\Delta\phi_t^{(i)}}$, where $\Delta\phi_t^{(i)} \sim \mathcal{N}(0, \gamma)$. To reflect this, the algorithm is altered as follows:

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$, $\pi_0^{(i)} = \frac{1}{N}$ and $\phi_0^{(i)} \sim U(0, c)$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t | x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t | x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)} = \sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t | x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)} = \sigma_t\}}$$

2.4: *Noise parameter perturbation* For each particle; $\phi_t^{(i)} = \phi_{t-1}^{(i)} e^{\Delta\phi_t^{(i)}}$ where $\Delta\phi_t^{(i)} \sim \mathcal{N}(0, \gamma)$
2.5: *Kernel smoothing* For each particle apply $\sigma_t^{(i)} \sim \mathcal{N}(\sigma_t^{(i)} | m_t^{(i)}, h^2 V_t + \phi^{(i)})$

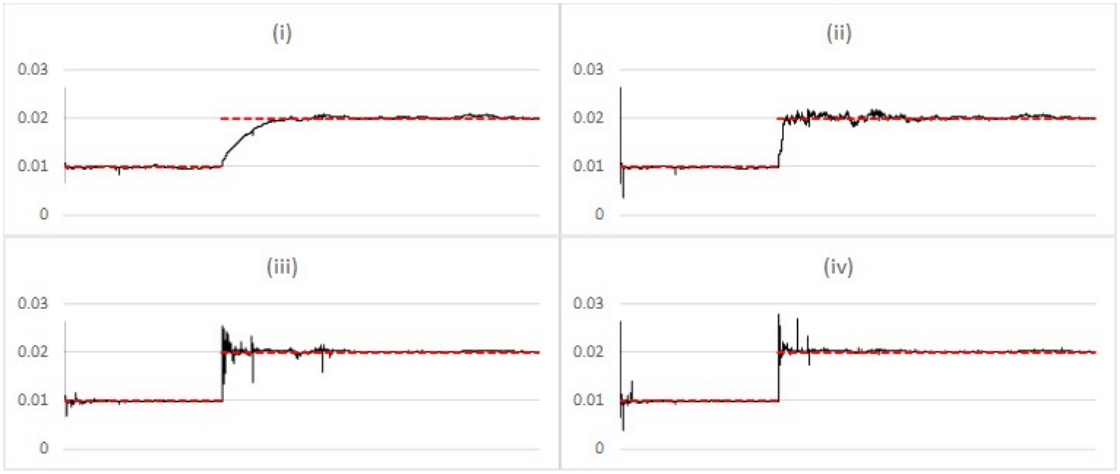

Figure 2.23: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ with regime change after 10,000 steps, using the Liu and West filter with learning for different values of γ (i)0.0001 (ii)0.001 (iii)0.01 (iv)0.1

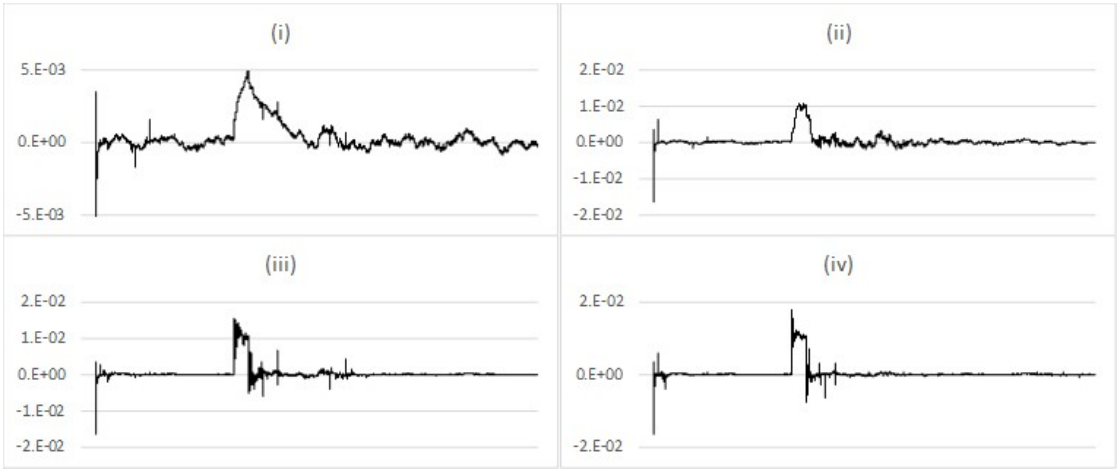

Figure 2.24: Comparing the change in estimated posterior expected value (black) vs simulation input value (red) of σ with regime change after 10,000 steps, using the Liu and West filter with learning for different values of γ (i)0.0001 (ii)0.001 (iii)0.01 (iv)0.1

The ability of the proposed approach to accelerate adaptation is demonstrated by adding noise parameter perturbation to the filter configuration used to produce the results shown in Figure 2.19, chart (ii). The results, shown in Figure 2.23 and Figure 2.24 for increasing values of γ, demonstrate a very effective acceleration of adaptation coupled with a significant reduction in noise compared to the implementation in the previous section. The adaptation of the filter to a regime change demonstrates the ability to rapidly increase the speed of adaptation when required. The same mechanism forces the additional noise to decrease when the adaptation phase is finished. The decrease in the noise factor can be further demonstrated with stochastic volatility model simulated data and the filter configured

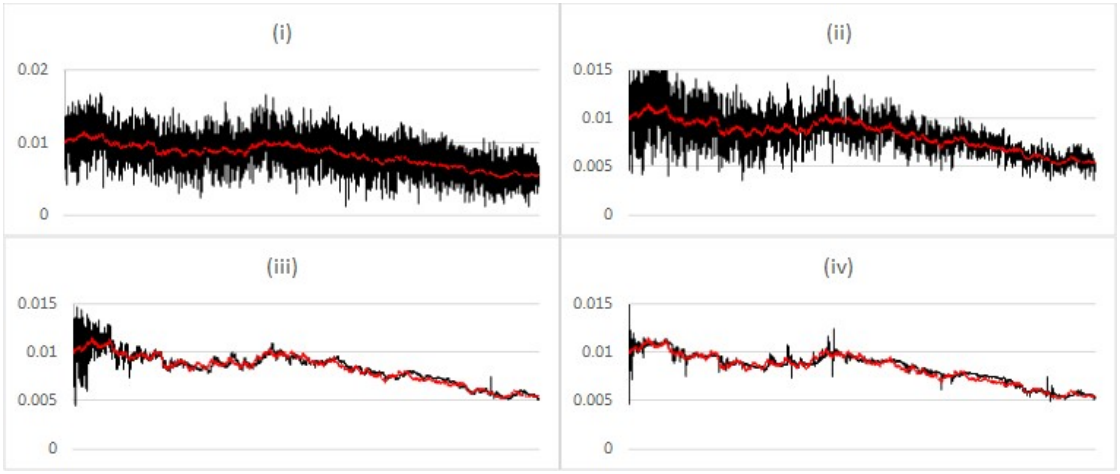

Figure 2.25: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ for stochastic volatility with learning for different values of γ (i)0.0001 (ii)0.001 (iii)0.01 (iv)0.1

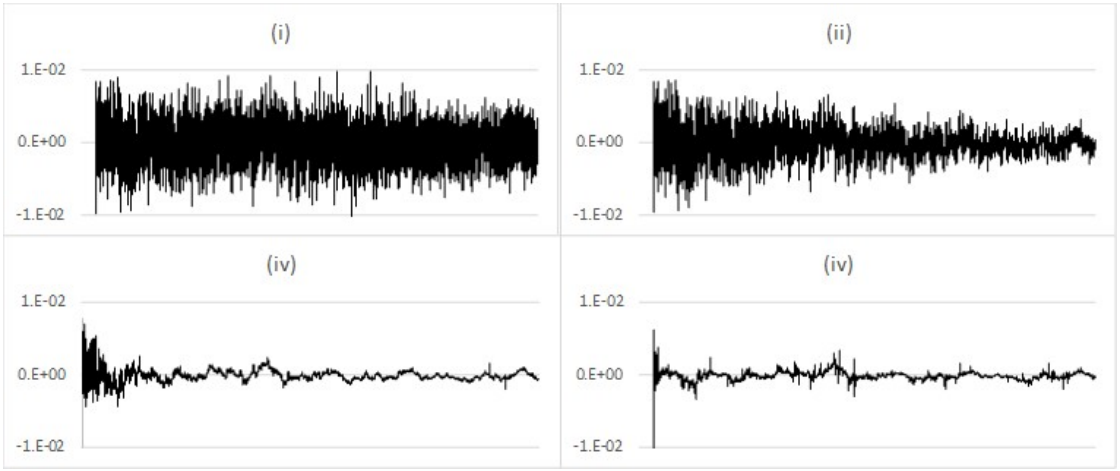

Figure 2.26: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ for stochastic volatility with learning for different values of γ (i)0.0001 (ii)0.001 (iii)0.01 (iv)0.1

with a very high starting point for ϕ as in Figure 2.15, chart (iv). The results in Figure 2.25 and Figure 2.26 demonstrate how the particle filter learns to reduce excess noise for increasing values of γ. It is also evident from the results, particularly Figure 2.24, that the reduction in noise post-adaptation phase tends to be slower than the initial increase. To speed up this reversal a dampening parameter is introduced in the next section.

2.3.5 Dampening the rate of adaptation

The post-adaptation learning parameter decrease tends to be slower than the adaptation increase because, relatively speaking, large observed changes are less likely to assume a low volatility than small observed changes assuming high volatility. Therefore, during adaptation low-noise particles are relatively less likely to survive than high-noise particles outside of the adaptation phase. This bias can be counteracted with the introduction of a dampening parameter in the form of a negative mean in the distribution used to perturb ϕ, that is $\Delta\phi_t^{(i)}$ in the learning step becomes $\Delta\phi_t^{(i)} \sim \mathcal{N}(-\kappa, \gamma)$. The addition of the dampening parameter also speeds up the convergence to the Liu and West filter when learning is not required, that is, in the idealised situation where the model assumed by the filter actually matches the observations. The filtering algorithm including the dampening factor becomes:

1: *Initialisation* For each particle; let $\sigma_0^{(i)} = \frac{(b-a)i}{N}$, $\pi_0^{(i)} = \frac{1}{N}$ and $\phi_0^{(i)} \sim U(0, c)$
2: Sequentially for each observation:
2.1: *Update* For each particle update weight $\hat{\pi}_t^{(i)} = \pi_t^{(i)} p(x_t | x_{t-1}, \sigma_t^{(i)})$
2.2: *Normalisation* For each particle $\pi_t^{(i)} = \frac{\hat{\pi}_t^{(i)}}{\sum \hat{\pi}_t^{(i)}}$
2.3: *Resampling* Generate a new set of particles:

$$p(\sigma_t | x_{1:t}) \approx \sum_{i=1}^{N} \delta_{\{\sigma_t^{(i)} = \sigma_t\}} \pi_t^{(i)} \xrightarrow[resample]{} p(\sigma_t | x_{1:t}) \approx \sum_{k=1}^{N} \frac{1}{N} \delta_{\{\sigma_t^{(k)} = \sigma_t\}}$$

2.4: *Noise parameter perturbation* For each particle; $\phi_t^{(i)} = \phi_{t-1}^{(i)} e^{\Delta\phi_t^{(i)}}$ where $\Delta\phi_t^{(i)} \sim \mathcal{N}(-\kappa, \gamma)$
2.5: *Kernel smoothing* For each particle apply $\sigma_t^{(i)} \sim \mathcal{N}(\sigma_t^{(i)} | m_t^{(i)}, h^2 V_t + \phi^{(i)})$

The impact of the dampening parameter was tested on the regime shift data with the particle filter configured with a high learning parameter used to generate the results in chart (iv) in Figure 2.23. The results for increasing values of the dampening parameter are shown in Figure 2.27 and Figure 2.28. Although the dampening parameter does indeed reduce estimate noise, it is important to note that the dampening parameter has to be set low enough so as not to completely offset the impact from perturbation. It may also be possible to evolve this parameter in the same manner as the noise parameter. However, this is not attempted in this research.

2.3.6 Average perturbation as a relative measure

The particle filter proposed in this chapter allows rapid detection of parameter changes by exploiting and enhancing the genetic algorithm aspect of a filter which includes random perturbation and selection. However, every random perturbation results in a deterioration of the quality of the posterior estimation, since the underlying assumption in the recursive calculation of the particle weights is that the parameters of each particle are fixed. Ideally, if the model assumption in the filter is reflected in the empirical data, the posterior estimation would not require any additional noise. This leads to the idea that the amount of additional

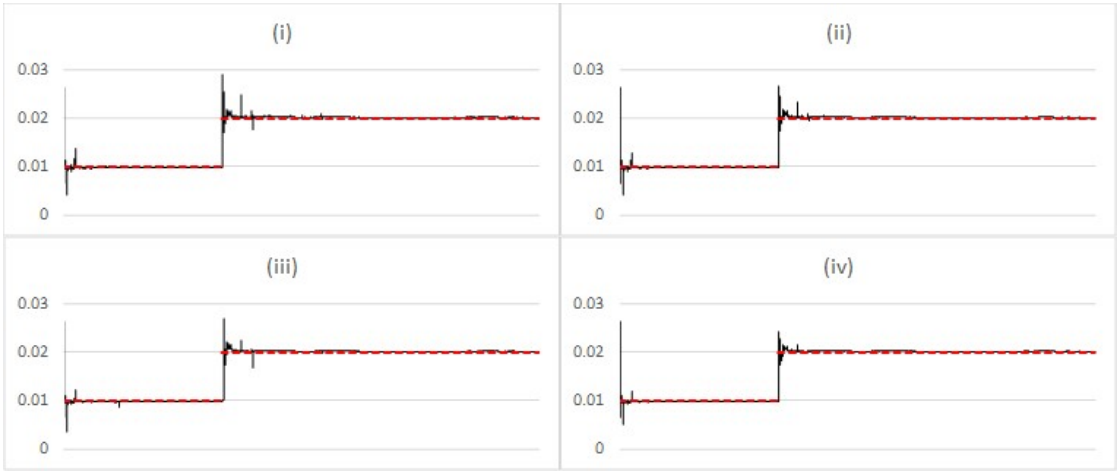

Figure 2.27: Comparing estimated posterior expected value (black) vs simulation input value (red) of σ with regime change after 10,000 steps, using the Liu and West filter with learning for different values of κ (i)0.01 (ii)0.02 (iii)0.03 (iv)0.04

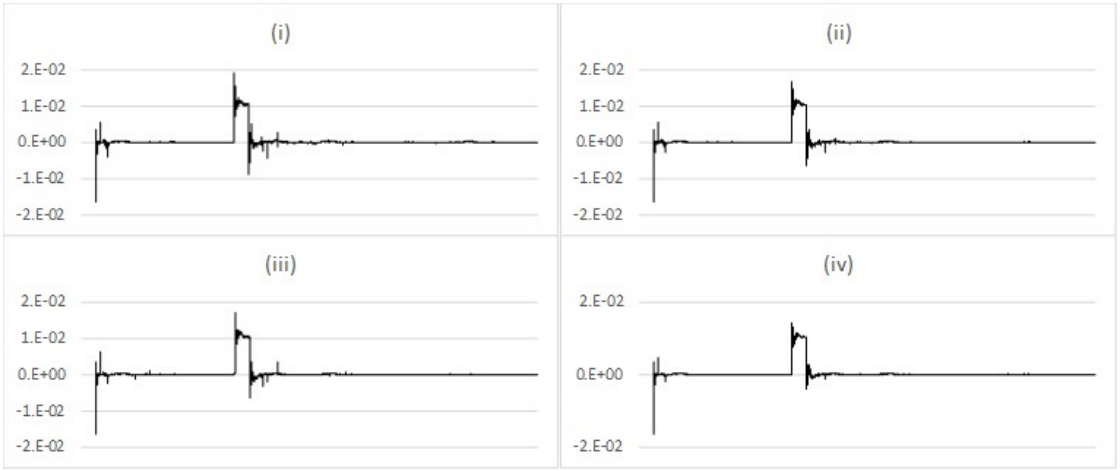

Figure 2.28: Comparing the change in estimated posterior expected value (black) vs simulation input value (red) of σ with regime change after 10,000 steps, using the Liu and West filter with learning for different values of κ (i)0.01 (ii)0.02 (iii)0.03 (iv)0.04

noise used by the filter can serve as an indicator of model adequacy, as well as distinguish between different dynamics present in the data set. Define this measure as the average of the ϕ parameter calculated at each iteration:

$$\frac{\sum_i \phi^{(i)}}{N}$$

The following offer some examples of how the behaviour of average ϕ can help to distinguish and identify the dynamics of the underlying data.

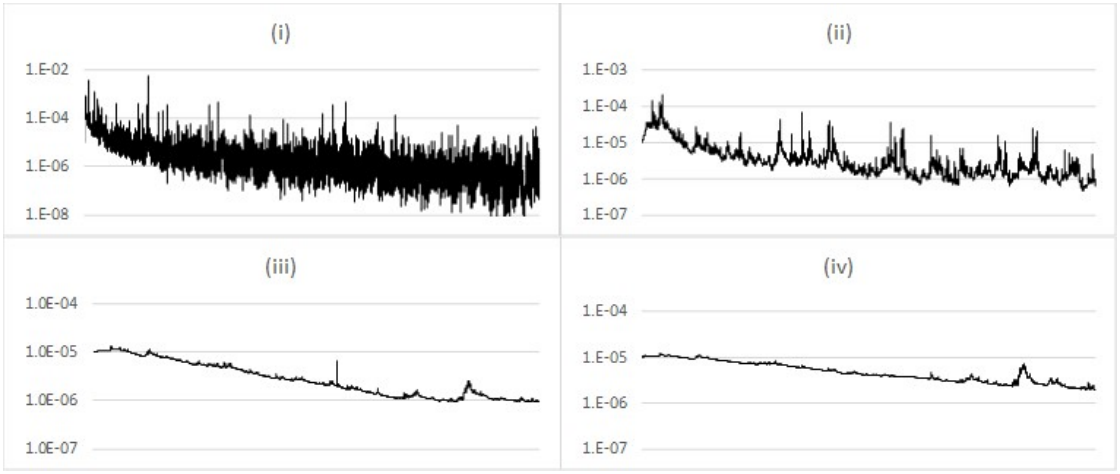

Figure 2.29: Comparison of average ϕ (log-scale) for the Gaussian model with different values of filter perturbation variance γ (i) 1.0 (ii) 0.01 (iii) 0.001 (iv) 0.0001

Gaussian process

The Gaussian process as the underlying dynamic demonstrates the behaviour of the measure when the data-generating process matches the assumption in the filter. As the estimated parameter posterior converges increasingly less perturbation is required, reflecting the correspondence between the filter assumption and underlying data. Simulation results shown in Figure 2.29 confirm the convergence of ϕ for different values of the perturbation variance parameter γ, expectedly high γ results in convergence noise highlighting the need for some implementation-specific tuning of this parameter.

Regime change

The regime change is marked by a sharp increase in ϕ, reflecting sudden adaptation to the new model state. Before and after the regime change the model is Gaussian and therefore the behaviour of ϕ is similar to the previous section. The rate of convergence of ϕ is slower for the higher σ following the regime change indicating a relationship between the rate of convergence of ϕ and σ. The results are shown in Figure 2.30 for varying levels of γ.

Stochastic volatility

If the filter assumption does not match the dynamics of the underlying data, ϕ will not tend to converge to zero. In the case of stochastic volatility, ϕ will tend towards a constant value reflecting the constantly changing volatility, with the level of this value indicating the level of stochasticity in the data. Figure 2.31 shows some examples of the behaviour of ϕ for varying levels of stochasticity in the data, using $\gamma = 0.001$ corresponding to chart (iii) in Figure 2.29.

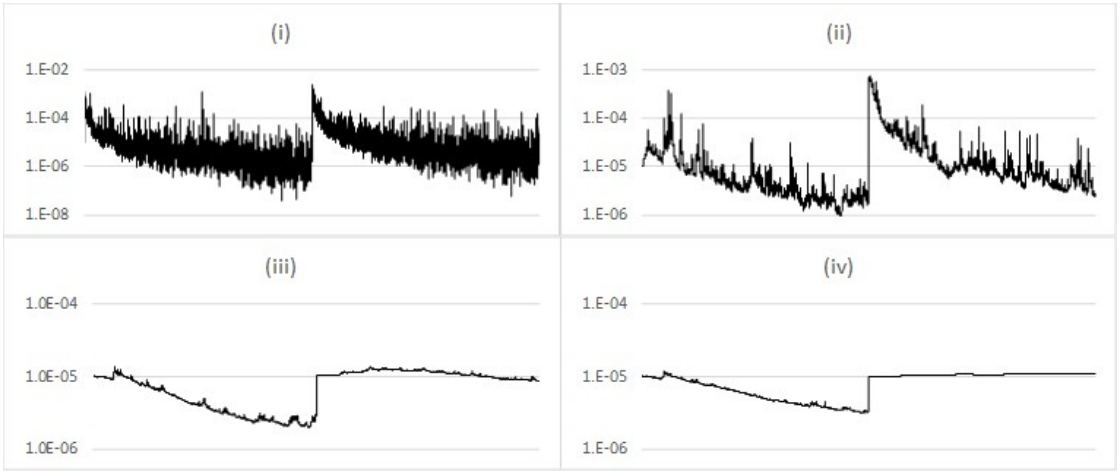

Figure 2.30: Comparison of average ϕ (log-scale) for the regime change model with different values of filter perturbation variance parameter γ (i) 1.0 (ii) 0.01 (iii) 0.001 (iv) 0.0001

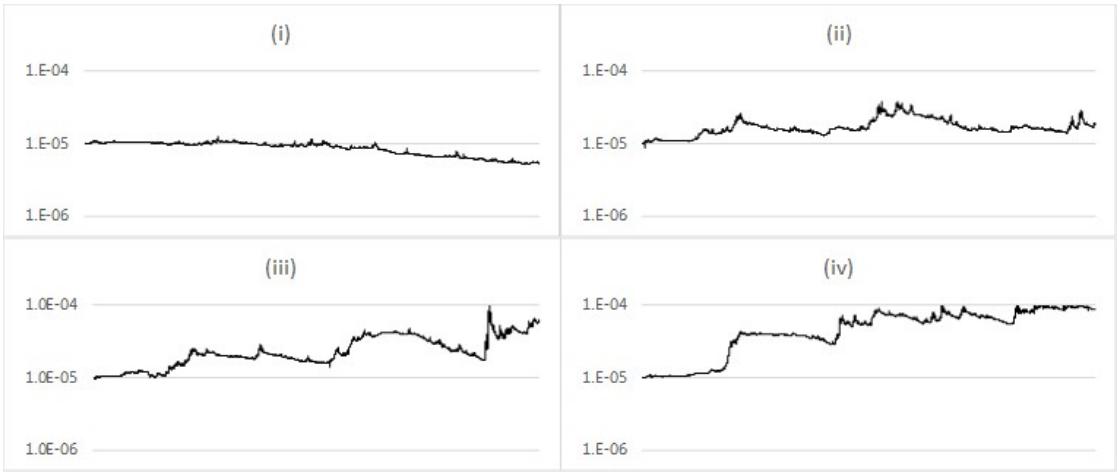

Figure 2.31: Comparison of average ϕ (log-scale) for the stochastic volatility model with different values of stochastic volatility ν (i) 0.1 (ii) 0.2 (iii) 0.3 (iv) 0.4

2.4 Conclusion

The methodology in this chapter was arrived at by first recognising that the random perturbation technique applied in a particle filter results in a genetic–type algorithm capable of adapting to changing parameters. At this point, an opposite direction to the approach of Liu and West (2001) was taken, instead of remediating the overdispersion caused by random perturbation, the random perturbation is allowed to freely evolve, enhancing the adaptive capability of the particle filter. The approach is highly adaptive when required and convergent conditional on the data matching modelling assumptions and no parameter changes. Given that the level of adaptability is governed by the variance of the random perturbation; the key insight of the approach is that an effective way of recognising the

level of required variance is to incorporate its selection into the already existing genetic algorithm framework. In terms of existing literature, it links particle filtering with genetic algorithms for parameter learning, resulting in a filtering algorithm particularly useful for parameter change detection and in the context of finance an effective on-line method for measuring volatility.

The motivation behind the particle filter was to create a tool which would inform modelling choices in a data-driven way. The original plan involved analysing the London Interbank Offer Rate(LIBOR) and its derivatives, mainly Eurodollar futures and options on those futures. However, throughout the development of the particle filter, there were increasing reports regarding the replacement of the LIBOR with SOFR. This prompted a re-focus on SOFR and short rates in general as reflected in the remaining chapters of this thesis.

Chapter 3

Reconciling Piecewise Constant Short Rates And Continuous Forward Rates[1]

This chapter introduces the first iteration of a model inspired by the empirical features outlined in Section 1.3. To reflect the empirical features, assume a three–component model driven by independent factors and construct it within the HJM framework. The three components include a step component to reflect the central bank target rate dynamics, a spike component for spikes occurring at known times and a continuous diffusion component for the residual noise.

The rest of the chapter is organised as follows. Section 3.1 presents the model for discontinuous short rates with continuous forward rates including both step and spike discontinuities. The model is presented within the Heath et al. (1992) (HJM) framework, and also includes an additional Gaussian diffusion to account for residual noise. Results from the calibration of the model to futures market data are presented in Section 3.2. Section 3.3 concludes.

3.1 Modelling Short Rates With Discontinuities At Known Times

Define a set of independent Brownian motions W comprising of subsets of Brownian motions W^P, W^Z, W^V related to the step, spike and continuous components respectively, where $W^P = [W_1^P, ..., W_m^P]$, $W^Z = [W_1^Z, ..., W_n^Z]$ and where $W = [W_1, ..., W_{m+n+1}] = [W_1^P, ..., W_m^P, W_1^Z, ..., W_n^Z, W^V]$. Under the spot risk–neutral measure, in the HJM frame-

[1]This chapter is based on the paper Gellert and Schlögl (2021b) with Erik Schlögl contributing in a supervisory capacity.

work:

$$f(t,T) = f(0,T) + \sum_{i=1}^{m+n+1} \int_0^t \sigma_i(u,T) \int_u^T \sigma_i(u,s) ds du + \sum_{i=1}^{m+n+1} \int_0^t \sigma_i(s,T) dW_i(s) \tag{3.1}$$

Define:

$$\sigma_i(t,T) = \mathbb{1}(i \leq m)\sigma_i^P(t,T) + \mathbb{1}(m < i \leq m+n)\sigma_{i-m}^Z(t,T) + \mathbb{1}(i = m+n+1)\sigma^V(t,T) \tag{3.2}$$

Therefore:

$$\begin{aligned} \sum_{i=1}^{m+n+1} \int_0^t \sigma_i(s,T) dW_i(s) = & \sum_{i=1}^{m} \int_0^t \sigma_i^P(s,T) dW_i^P(s) \\ & + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(s,T) dW_i^Z(s) \\ & + \int_0^t \sigma^V(s,T) dW^V(s) \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} \sum_{i=1}^{m+n+1} \int_0^t \sigma_i(u,T) \int_u^T \sigma_i(u,s) ds du = & \sum_{i=1}^{m} \int_0^t \sigma_i^P(u,T) \int_u^T \sigma_i^P(u,s) ds du \\ & + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(u,T) \int_u^T \sigma_i^Z(u,s) ds du \\ & + \int_0^t \sigma^V(u,T) \int_u^T \sigma^V(u,s) ds du \end{aligned} \tag{3.4}$$

therefore

$$f(t,T) = f^P(t,T) + f^Z(t,T) + f^V(t,T) \tag{3.5}$$

where

$$f^P(t,T) = f^P(0,T) + \sum_{i=1}^{m} \int_0^t \sigma_i^P(u,T) \int_u^T \sigma_i^P(u,s) ds du + \sum_{i=1}^{m} \int_0^t \sigma_i^P(s,T) dW_i^P(s) \tag{3.6}$$

$$f^Z(t,T) = f^Z(0,T) + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(u,T) \int_u^T \sigma_i^Z(u,s) ds du + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(s,T) dW_i^Z(s) \tag{3.7}$$

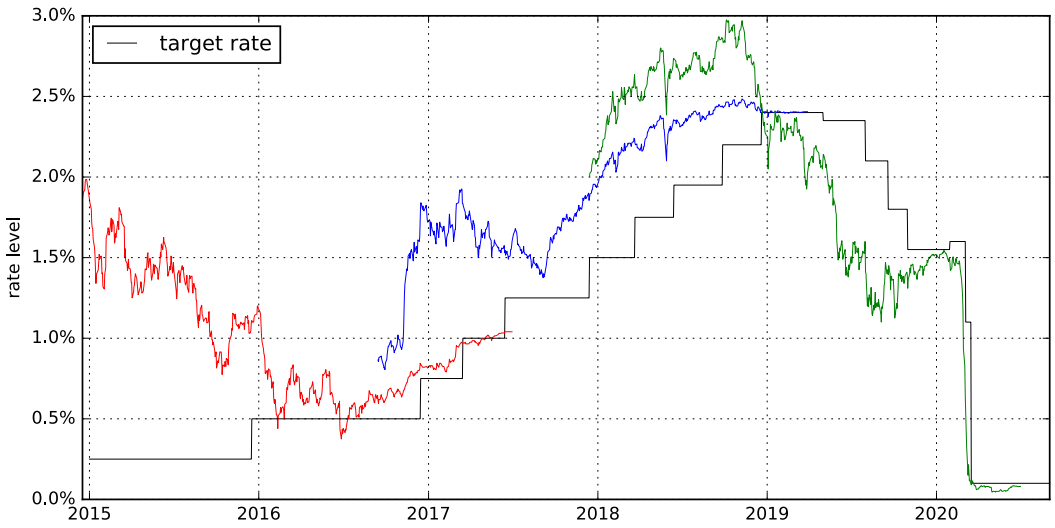

Figure 3.1: Target rate and various forward rates implied by specific 30–day Fed Funds futures

$$f^V(t,T) = f^V(0,T) + \int_0^t \sigma^V(u,T) \int_u^T \sigma^V(u,s) ds du + \int_0^t \sigma^V(s,T) dW^V(s) \tag{3.8}$$

similarly for the short rate:

$$r(t) = r^P(t) + r^Z(t) + r^V(t) \tag{3.9}$$

and zero coupon bonds:

$$B(t,T) = B^P(t,T) B^Z(t,T) B^V(t,T) \tag{3.10}$$

The modelling of each component is now discussed in more detail.

3.1.1 Target Rate Step Model

The main empirical feature of the target rate is that it is piecewise flat between the FOMC meeting dates at which a policy change has occurred. Most of the meetings are scheduled at least one year ahead of time with the exception of emergency meetings.[2]

Although forward target rates do not trade directly, the nature of their dynamics can be deduced from 30-day Fed Fund Futures which trade on the closely related EFFR rate. Figure 3.1 shows the historical target rate and various forward rates implied from specific futures contracts. The point at which the forward rates end and meet the target rate coincides with the expiry of the futures contracts.[3] In contrast to the overnight rate, the

[2]Since 2015 there have been 47 meetings (including 3 emergency meetings), of which 17 resulted in a target rate change.

[3]Futures without an FOMC date in the reference month were chosen such that the target rate is expected to be flat over the contract month and therefore the price of the futures reflects the expected target rate for that month plus a spread rather than reflecting two flat periods before and after the FOMC date.

dynamics of target rates are more diffusive and do not jump at deterministic dates. Jumps conceivably could occur on unexpected dates, reflecting sudden large changes in market sentiment. However, these are not captured by this model. The main contribution of this model is that having observed that empirically the overnight rate (EFFR or SOFR) follows dynamics determined primarily by jumps at known times, but forward rates follow primarily diffusive dynamics, it reconciles these two (naively contradictory) observations.

An interpretation of the forward rates deduced from Fed fund futures is that they reflect the expectations of prospective FOMC target rate changes. The diffusive dynamics of forward rates then reflect the nature of the changes in those expectations. From this perspective, the expectations corresponding to each scheduled FOMC meeting are not independent of each other. In some circumstances, for example, a change in the overall Federal Reserve monetary policy stance, they will be positively correlated. In other cases, where for example the aggregated change to the target rate over some period of time is anticipated but the timing is less certain, the expectations may be negatively correlated to each other as the expected timing but not the net outcome evolves.

Therefore the target rate model is motivated by the following empirical features. The target rate represented by the short rate $r^P(t)$ must be piecewise flat with respect to t. The forward rate with maturity T evolves diffusively with respect to t until the FOMC meeting immediately preceding maturity T, reflecting the expectations of any FOMC policy target rate change. A model is constructed which reconciles these features, reflecting both the discontinuous nature of the short rate and diffusively evolving forward rates.

Forward Rates

Construct a model such that the forward rates are driven by the evolution of expectations associated with FOMC target rate changes, where the target rate change for each scheduled meeting date evolves under its own dynamic. The forward rate $f(t,T)$ dynamics under the empirical measure can be written as follows:

$$f^P(t,T) = f^P(0,T) + \alpha(t,T) + \sum_{i=1}^{m} \int_0^t \xi_i(s,T) dZ_i(s) \tag{3.11}$$

[4] where $f^P(0,T)$ is the initial term structure of forward rates, $\alpha(t,T)$ a deterministic drift and $dZ_i(s)$ the Wiener increment corresponding to the i_{th} FOMC date with correlation $dZ_i(t)dZ_j(t) = \rho_{i,j}dt$. The volatility term is defined as follows:

$$\xi_i(t,T) = \xi_i \mathbb{1}(t < x_i)\mathbb{1}(T \geq x_i) \tag{3.12}$$

where x_i denotes the i^{th} FOMC meeting date. The intuition behind this construction is that each stochastic component corresponds to an FOMC date and any changes to the target rate are carried forward from that date. The indicator function $\mathbb{1}(T \geq x_i)$ ensures that the i^{th} factor is only applied to forwards with maturities greater or equal to x_i. For any maturities prior to the first meeting date $T < x_1$, therefore $\mathbb{1}(T \geq x_i) = 0, \forall i \geq 1$, thus

[4] For clarity throughout this paper indexed variables in the form ξ_i are constants.

ensuring no diffusion for forward rates with maturities prior to the first FOMC meeting date. The indicator function $\mathbb{1}(t < x_i)$ terminates the diffusion from the i^{th} stochastic component on the corresponding FOMC date.[5] Solving the integral, see Appendix (A.2), yields:

$$f^P(t,T) = f^P(0,T) + \alpha(t,T) + \sum_{i=1}^{m} \xi_i \mathbb{1}(T \geq x_i) Z_i(t \wedge x_i) \tag{3.13}$$

To demonstrate the behaviour of the model with an example, let $x_2 < T < x_3$ and $t < x_1$:

$$f^P(t,T) = f^P(0,T) + \alpha(t,T) + \xi_1 Z_1(t) + \xi_2 Z_2(t) \tag{3.14}$$

Here one sees that both stochastic components, corresponding to FOMC dates x_1 and x_2, impact the forward rate up to time t. Any stochastic components beyond x_2 do not apply since the forward rate matures prior to x_3. Now let $x_1 < t < x_2$:

$$f^P(t,T) = f^P(0,T) + \alpha(t,T) + \xi_1 Z_1(x_1) + \xi_2 Z_2(t) \tag{3.15}$$

In this case the first stochastic component terminates at x_1, prior to t. That is, the expectations of the target rate change at time x_1 evolve diffusively only up until this date.

Short Rates

These forward rate dynamics create the piecewise dynamic in the short rate, which can be derived from (3.11) by setting $r(t) = f(t,t)$:

$$r^P(t) = f^P(t,t) = f^P(0,t) + \alpha(t,t) + \sum_{i=1}^{m} \int_0^t \xi_i(s,t) dZ_i(s) \tag{3.16}$$

Solving the integral, see Appendix (A.3), yields:

$$r^P(t) = f^P(0,t) + \alpha(t,t) + \sum_{i=1}^{m} \xi_i \mathbb{1}(t \geq x_i) Z_i(x_i) \tag{3.17}$$

From this it is evident that the short rate has no diffusion up until the first FOMC date at which point it picks up all the diffusion from the forward rate accumulated up until this point in time. To illustrate this, for $t < x_1$:

$$r^P(t) = f^P(0,t) + \alpha(t,t) \tag{3.18}$$

for $x_1 < t < x_2$:

$$r^P(t) = f^P(0,t) + \alpha(t,t) + \xi_1 Z_1(x_1) \tag{3.19}$$

[5] By construction the diffusion coefficient is constant relative to each forward rate period, however the total diffusion for any forward rate is an aggregate of these diffusive forward periods and therefore the stochasticity of each forward rate is a function of the number of FOMC dates from t to the forward date and is thus a function of t.

for $x_2 < t < x_3$:

$$r^P(t) = f^P(0,t) + \alpha(t,t) + \xi_1 Z_1(x_1) + \xi_2 Z_2(x_2) \tag{3.20}$$

In general, the accumulated diffusion for the forward rates creates discontinuities in the short rate on FOMC dates, reflecting the empirical behaviour for the target rate and the associated forward rates.

Decomposition to Independent Components

The model can be easily transformed to independent components which will make it consistent with the HJM framework, thus facilitating the derivation of risk–neutral dynamics. Define Σ to be the covariance matrix of the vector $dZ = [dZ_1, ..., dZ_m]$. To transform the system to independent components, find a transformation matrix γ, such that $\Sigma = \gamma\gamma^T$ which is applied using $dZ = \gamma dW^P$, to result in a vector of uncorrelated Wiener increments $dW^P = [dW_1^P, ..., dW_m^P]$. Therefore:

$$dZ_i = \sum_{j=1}^{m} \gamma_{i,j} dW_j^P \tag{3.21}$$

Rewrite the forward rate dynamics with respect to the uncorrelated components:

$$\begin{aligned}
\sum_{i=1}^{m} \int_0^t \xi_i(s,T) dZ_i(s) &= \sum_{i=1}^{m} \int_0^t \sigma_i \mathbb{1}(s < x_i) \mathbb{1}(T \geq x_i) dZ_i(s) \\
&= \sum_{i=1}^{m} \int_0^t \xi_i \mathbb{1}(s < x_i) \mathbb{1}(T \geq x_i) \sum_{j=1}^{m} \gamma_{i,j} dW_j^P(s) \\
&= \sum_{j=1}^{m} \int_0^t \sigma_j^P(s,T) dW_j^P(s)
\end{aligned} \tag{3.22}$$

where

$$\sigma_j^P(t,T) = \sum_{i=1}^{m} \xi_i \gamma_{i,j} \mathbb{1}(t < x_i) \mathbb{1}(T \geq x_i) \tag{3.23}$$

It is also worth noting that this transformation lends itself to principal components analysis (PCA). Thus, rather than requiring as many driving stochastic components as there are FOMC meeting dates, the model can be driven by a smaller number of independent components. The transformation is thus a factor reduction approach that even in a reduced factor form retains the key modelling property of piecewise short rates and continuously diffusive forward rates. That is in this form the short rate continues to pick up the accumulated forward rate diffusion only on FOMC dates resulting in a piecewise constant short rate.

Forward Rates Under the Spot Risk–Neutral Measure

It is now possible to formulate the risk neutral dynamics by using the result from HJM. Under the spot risk–neutral measure:

$$f^P(t,T) = f^P(0,T) + \sum_{j=1}^{m} \int_0^t \sigma_j^P(u,T) \int_u^T \sigma_j^P(u,s) ds du + \sum_{j=1}^{m} \int_0^t \sigma_j^P(s,T) dW_j^P(s) \quad (3.24)$$

Therefore, see (A.4) and (A.11):

$$\begin{aligned} f^P(t,T) = f^P(0,T) &+ \sum_{j=1}^{m} \sum_{q=1}^{m} \sum_{i=1}^{m} \xi_q \xi_i \gamma_{q,j} \gamma_{i,j} \mathbb{1}(T \geq x_{q \vee i})(T - x_i)[t \wedge x_q \wedge x_i] \\ &+ \sum_{j=1}^{m} \sum_{i=1}^{m} \xi_i \gamma_{i,j} \mathbb{1}(T \geq x_i) W_j^P(t \wedge x_i) \end{aligned} \quad (3.25)$$

Short Rates Under the Spot Risk–Neutral Measure

Short rate dynamics can are obtained as follows:

$$r^P(t) = f^P(0,t) + \sum_{j=1}^{m} \int_0^t \sigma_j^P(u,t) \int_u^t \sigma_j^P(u,s) ds du + \sum_{j=1}^{m} \int_0^t \sigma_j^P(s,t) dW_j^P(s) \quad (3.26)$$

Therefore, see (A.5) and (A.12):

$$\begin{aligned} r^P(t) = &\underbrace{f^P(0,t) + \sum_{j=1}^{m} \sum_{q=1}^{m} \sum_{i=1}^{m} \xi_q \xi_i \gamma_{q,j} \gamma_{i,j} \mathbb{1}(t \geq x_{q \vee i})(t - x_i)[x_q \wedge x_i]}_{\text{deterministic term (**)}} \\ &+ \underbrace{\sum_{j=1}^{m} \sum_{i=1}^{m} \xi_i \gamma_{i,j} \mathbb{1}(t \geq x_i) W_j^P(x_i)}_{\text{stochastic term (*)}} \end{aligned} \quad (3.27)$$

The stochastic term (*) follows piecewise constant dynamics, jumping almost surely at each x_i.[6] Because at present, only the target rate is modelled, the paths of $r^P(t)$ should be constant between FOMC meeting dates. This implies that the deterministic term (**) should not depend on t, i.e., the dependence on t of the triple sum must cancel against the dependence on t of the initial term structure $f^P(0,t)$.[7] When considering a time horizon

[6] At this point, one might object that in reality, rates do not jump at every FOMC meeting date. However, one could argue that this is because target rates are only updated in discrete increments. The model could be extended to reflect this, but as a first approximation, we'll accept the implication of a continuous distribution of jump sizes, with jumps occurring at every FOMC meeting date.

[7] Note that the term $(t - x_i)$ appearing in the triple sum reflects a feature of a classical Gaussian term structure model without mean reversion (as noted, for example, in Schlögl and Sommer (1998)), that the term structure of forward rates endogenously steepens ever more (see also (3.25) above) as time passes — this can be avoided by introducing mean reversion.

of two years or less (as per the empirical section of this chapter), the triple sum in (**) is practically flat in t, so this is consistent with an initial term structure $f^P(0,t)$ which is approximately constant between FOMC meeting dates. Note that (3.27) implies that if the paths of $r^P(t)$ are required to be constant between FOMC meeting dates, it is not correct to arbitrarily choose an interpolation method for the initial term structure. In particular, requiring piecewise constant paths of $r^P(t)$ precludes applying the popular Nelson/Siegel interpolation to the initial term structure.[8]

Bond Prices

Bond prices can be written as follows:

$$B^P(t,T) = \exp\Big(-\int_t^T f^P(t,s)ds \Big) = \frac{B^P(0,T)}{B^P(0,t)} \exp\Big(a(t,T) + b(t,T) \Big) \tag{3.28}$$

where

$$\begin{aligned} a(t,T) &= -\int_t^T \sum_{j=1}^m \sum_{q=1}^m \sum_{i=1}^m \xi_q \xi_i \gamma_{q,j} \gamma_{i,j} \mathbb{1}(s \geq x_{q\vee i})(s - x_i)[t \wedge x_q \wedge x_i] ds \\ &= -\sum_{j=1}^m \sum_{q=1}^m \sum_{i=1}^m \xi_q \xi_i \gamma_{q,j} \gamma_{i,j} [t \wedge x_q \wedge x_i] \int_t^T \mathbb{1}(s \geq x_{q\vee i})(s - x_i) ds \\ &= -\sum_{j=1}^m \sum_{q=1}^m \sum_{i=1}^m \xi_q \xi_i \gamma_{q,j} \gamma_{i,j} [t \wedge x_q \wedge x_i][I_1 - I_2] \end{aligned} \tag{3.29}$$

where

$$\begin{aligned} I_1 &= \int_0^T \mathbb{1}(s \geq x_{q\vee i})(s - x_i) ds \\ &= \int_{x_{q\vee i}}^T (s - x_i) ds = \mathbb{1}(T \geq x_{q\vee i})[T(\frac{T}{2} - x_i) - x_{q\vee i}(\frac{x_{q\vee i}}{2} - x_i)] \end{aligned} \tag{3.30}$$

$$I_2 = \int_0^t \mathbb{1}(s \geq x_{q\vee i})(s - x_i) ds = \mathbb{1}(t \geq x_{q\vee i})[t(\frac{t}{2} - x_i) - x_{q\vee i}(\frac{x_{q\vee i}}{2} - x_i)] \tag{3.31}$$

[8] Skov and Skovmand (2021) show that a three–factor Gaussian arbitrage–free Nelson/Siegel model is well suited for the SOFR futures market, but they do not include the time series of SOFR itself in their estimation, i.e., their objective is not to match the SOFR dynamics, which have a substantial piecewise flat component.

$$
\begin{aligned}
b(t,T) &= -\int_t^T \sum_{j=1}^m \sum_{i=1}^m \xi_i \gamma_{i,j} \mathbb{1}(s \geq x_i) W_j^P(t \wedge x_i) ds \\
&= -\sum_{j=1}^m \sum_{i=1}^m \xi_i \gamma_{i,j} W_j^P(t \wedge x_i) \int_t^T \mathbb{1}(s \geq x_i) ds \\
&= -\sum_{j=1}^m \sum_{i=1}^m \xi_i \gamma_{i,j} W_j^P(t \wedge x_i) \mathbb{1}(T \geq x_i)[T - (t \vee x_i)]
\end{aligned}
\tag{3.32}
$$

Note that zero coupon bond prices are exponential affine functions of the $W_j^P(t \wedge x_i)$. However, unlike in classical Gauss/Markov HJM term structure models, here the entire term structure cannot be represented as an exponential affine function of n factors.

3.1.2 Known Spike Time Model

In addition to the step-wise behaviour reflecting FOMC decisions modelled above, spikes in the short rate have been a prominent feature of EFFR and particularly SOFR dynamics until 2021. In this section, the approach is adapted to also model the occurrence of spikes at known dates.

Spiked Forward Rates

The model for the spike component f^Z of forward rates is constructed such that the spike on each spike date z_i is driven by its own independent factor. This allows the use the HJM result to formulate f^Z under the risk neutral measure:

$$
f^Z(t,T) = f^Z(0,T) + \sum_{i=1}^n \int_0^t \sigma_i^Z(u,T) \int_u^T \sigma_i^Z(u,s) ds du + \sum_{i=1}^n \int_0^t \sigma_i^Z(s,T) dW_i^Z(s) \tag{3.33}
$$

Assume that when spikes occur, they impact a fixed period h_i starting from time z_i.[9] Let $H_i = [z_i, z_i + h_i]$, the volatility function is defined as follows:

$$
\sigma_i^Z(t,T) = \sigma_i^Z \mathbb{1}(t < z_i) \mathbb{1}(T \in H_i) \tag{3.34}
$$

Therefore, see (A.6) and (A.14) :

$$
f^Z(t,T) = f^Z(0,T) + \sum_{i=1}^n \left(\sigma_i^Z\right)^2 \mathbb{1}(T \in H_i)(T - z_i)[t \wedge z_i] + \sum_{i=1}^n \sigma_i^Z \mathbb{1}(T \in H_i) W_i(t \wedge z_i) \tag{3.35}
$$

[9] Usually this period of time would be equivalent to 1 day but could be more if for example it is a SOFR rate set on a Friday, therefore applying for compounding and averaging payoff calculations over the weekend.

To demonstrate the behaviour of the model with an example, let $T \in H_2$ and $t < z_1$:

$$f^Z(t,T) = f^Z(0,T) + \left(\sigma_2^Z\right)^2 (T - z_2)t + \sigma_2^Z W_2(t) \tag{3.36}$$

Here the interpretation is that the f^Z only evolve when $T \in H_i$ up to the minimum of time t or z_i, the beginning of the period H_i.

Spiked Short Rates

By this construction, the spike component of the short rate follows the spiked trajectory:

$$r^Z(t) = f^Z(t,t) = f^Z(0,t) + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(u,t) \int_u^t \sigma_i^Z(u,s) ds du + \sum_{i=1}^{n} \int_0^t \sigma_i^Z(s,t) dW_i^Z(s) \tag{3.37}$$

Therefore, see (A.7) and (A.15):

$$r^Z(t) = f^Z(0,t) + \sum_{i=1}^{n} \left(\sigma_i^Z\right)^2 \mathbb{1}(t \in H_i)(t - z_i)z_i + \sum_{i=1}^{n} \sigma_i^Z \mathbb{1}(t \in H_i) W_i(z_i) \tag{3.38}$$

From this it is evident that the short rate is deterministic until the spike interval over which a spike applies, with a magnitude which includes the associated forward rate diffusion accumulated up to the beginning of the interval. For example let $t \in H_2$:

$$r^Z(t) = f^Z(0,t) + \left(\sigma_2^Z\right)^2 (t - z_2)z_2 + \sigma_2^Z W_2(z_2) \tag{3.39}$$

Spiked Bond Prices

The spike component of the bond prices can be written as follows:

$$B^Z(t,T) = \exp\left(- \int_t^T f^Z(t,s) ds \right) = \frac{B^Z(0,T)}{B^Z(0,t)} \exp\Big(a(t,T) + b(t,T) \Big) \tag{3.40}$$

$$\begin{aligned} a(t,T) &= - \int_t^T \sum_{i=1}^{n} \left(\sigma_i^Z\right)^2 \mathbb{1}(s \in H_i)(s - z_i)[t \wedge z_i] ds \\ &= -\left(\sigma_i^Z\right)^2 [t \wedge z_i] \int_t^T \mathbb{1}(s \in H_i)(s - z_i) ds \\ &= -\left(\sigma_i^Z\right)^2 [t \wedge z_i][I_1 - I_2] \end{aligned} \tag{3.41}$$

where

$$I_1 = \int_0^T \mathbb{1}(s \in H_i)(s - z_i) ds = \mathbb{1}(T \geq z_i) \left[\frac{h_i^2}{2} \wedge \frac{(T - z_i)^2}{2} \right] \tag{3.42}$$

and

$$I_2 = \int_0^t \mathbb{1}(s \in H_i)(s - z_i)ds = \mathbb{1}(t \geq z_i)\left[\frac{h_i^2}{2} \wedge \frac{(t - z_i)^2}{2}\right] \tag{3.43}$$

and

$$\begin{aligned} b(t,T) &= -\int_t^T \sum_{i=1}^n \sigma_i^Z W_i(t \wedge z_i)\mathbb{1}(s \in H_i)ds \\ &= -\sum_{i=1}^n \sigma_i^Z W_i(t \wedge z_i)\int_t^T \mathbb{1}(s \in H_i)ds \\ &= -\sum_{i=1}^n \sigma_i^Z W_i(t \wedge z_i)([(T - z_i) \wedge (T - t) \wedge h_i \wedge (z_i + h_i - t)] \vee 0) \end{aligned} \tag{3.44}$$

3.1.3 Modelling the Diffusive Residual

An empirical feature of the residual noise component of both EFFR and SOFR is mean reversion. Since the initial bond term structure is most naturally contained in the initial target rate term structure, the mean-reverting model proposed in Vasicek (1977) should be sufficient to model the noise component of short rates. The model is presented based on the results shown in Carmona (2007). The dynamics of the diffusive residual are given by:

$$dr^V(t) = (\theta - \beta r^V(t))dt + \sigma^V dW^V(t) \tag{3.45}$$

The solution is given by:

$$r^V(t) = e^{-\beta t}r^V(0) + (1 - e^{-\beta t})\frac{\theta}{\beta} + \sigma^V \int_0^t e^{-\kappa(t-s)}dW^V(s) \tag{3.46}$$

With forward rates:

$$f^V(t,T) = r^V(t)e^{-\beta(T-t)} + \frac{\theta}{\beta}\left(1 - e^{-\beta(T-t)}\right) - \frac{\theta^2}{2\beta^2}\left(1 - e^{-\beta(T-t)}\right)^2 \tag{3.47}$$

The diffusive residual component of the zero coupon bond price is given by:

$$B^V(t,T) = a(t,T)e^{b(t,T)r(0)} \tag{3.48}$$

with

$$b(t,T) = -\frac{1 - e^{-\beta(T-t)}}{\beta} \tag{3.49}$$

and

$$a(t,T) = \frac{4\theta\beta - 3\sigma^2}{4\beta^3} + \frac{(\sigma^V)^2 - 2\alpha\beta}{2\beta^2}T + \frac{(\sigma^V)^2 - \alpha\beta}{\beta^3}e^{-\beta T} - \frac{(\sigma^V)^2}{4\beta^3}e^{-2\beta T} \tag{3.50}$$

3.2 Calibration to Futures Contracts

This section presents results calibrating the model to Fed Funds and SOFR futures data. Fed Fund futures are used to calibrate the target rate term structure, which is then used as the basis for calibration to SOFR futures, from which the term structure of forward rates related to end–of–month spikes is inferred. The time series of calibrated EFFR and SOFR forward rates are used to examine how well the market anticipates FOMC policy target rate changes as well as end–of–month spikes. The time series of SOFR forward rates is then used to compare the forward-looking SOFR term rates to LIBOR.

3.2.1 30 day Fed Funds futures

Fed Funds futures contracts[10] are based on the arithmetic average of the EFFR, denoted r_E over the specified contract month. Define m as the number of months from the current trading month ($m = 0$), $\tau_{m,i}$:= as the date corresponding to day i in month m with n_m denoting the total days in month m.

Define the futures contract index for reference month m at time t as $\widetilde{F}_m(t)$, the value of a single contract is \$4,167 $\times$ $\widetilde{F}_m(t)$. The terminal value of the contract is determined as $\widetilde{F}_m(\tau_{m,n_m}) = 100 - R_m$ where R_m is the arithmetic average of the daily EFFR fixing during the contract month, settled on the first business day after the final fixing date.

Defining $R_m := \frac{100}{n_m} \sum_{i=1}^{n_m} r_E(\tau_{m,i})$, the terminal payoff is:

$$\widetilde{F}_m(\tau_{m,n_m}) = 100 - R_m = 100\bigg(1 - \frac{1}{n_m}\sum_{i=1}^{n_m} r_E(\tau_{m,i})\bigg)$$

Using the generic futures pricing theorem,[11] the expected value at t of the futures contract index $\widetilde{F}_m$ under the spot risk neutral measure is:

$$F_m(t) = E_t[\widetilde{F}_m(\tau_{m,n_m})] = 100\bigg(1 - \frac{1}{n_m}\sum_{i=1}^{n_m} E_t[r_E(\tau_{m,i})]\bigg) \tag{3.51}$$

The current futures contract continues to trade during the observation month, therefore the valuation needs to account for already observed values of r_E:

$$F_0(t) = 100\bigg(1 - \frac{1}{n_0}\bigg(\sum_{i=1}^{n_0} \mathbf{1}_{(t>\tau_{0,i})} r_E(\tau_{0,i}) + \sum_{i=1}^{n_0} \mathbf{1}_{(t\leq\tau_{0,i})} E_t[r_E(\tau_{0,i})]\bigg)\bigg) \tag{3.52}$$

Calibration

Fed Funds futures contracts are available for each calendar month approximately 3 years ahead of time. However, the liquidity beyond 1 year deteriorates and therefore the calibration is limited to the first 12 contracts. The availability of contracts for each calendar

[10] Source:https://www.cmegroup.com/markets/interest-rates/stirs/30-day-federal-fund.contractSpecs.html.

[11] See Cox, Ingersoll and Ross (1981).

month makes the Fed Funds futures particularly useful for extracting information regarding expected target rate changes, which are scheduled 8 times per year and never twice in the same month. Calibrating the expected policy target rate jumps from Fed Funds futures is performed by making the following assumptions.

Firstly the initial term structure of $f(0,T)$ is assumed to be piecewise flat between FOMC meeting dates. This aligns the initial term structure to the driving factors of the target rate model and therefore the daily changes in the calibrated $f(0,T)$ vector provides an empirical estimate for the dynamics of $f(t,T)$. To simplify the calibration, it is assumed that the impact of the drift component is negligible, particularly if the calibration is used to obtain the empirical dynamics of the forward rate based on daily increments obtained from the calibration. The spikes are a secondary component of EFFR empirical dynamics and are ignored in the calibration. Additionally, a constant spread between EFFR and the target rate is calibrated, which is equivalent to assuming zero volatility in the Gaussian residual noise component of the model.

Observable market prices exist in the form of the current bid and offer and the last observed price, which reflects a trade at either the bid or the offer levels at the time of the transaction. At any given time the true market state is at some point between the bid and offer prices. Closing prices which are recorded at the end of each day's trading session also reflect either the bid or the offer. Therefore the closing price could be either the offer, inferring that the bid is one price fluctuation below the closing price, or conversely inferring that the offer is one price fluctuation above the closing price. Based on this reasoning, a minimum price fluctuation size tolerance is embedded in the calibration error $e_m(t)$ for month m:

$$e_m(t) = (|F_m(t) - \widetilde{F}_m(t)| - h_m)^+$$

Where the minimum fluctuation of the index for month m as h_m with $h_0 = 0.0025$ and $h_m = 0.005$ for $m \neq 0$. The error bounds result in better solution stability, and less subject to bid-ask fluctuations in the cross-sectional and longitudinal data. The calibration is performed using a genetic algorithm approach based on the method developed in Chapter 2

Fed Fund futures implied forward rates as a predictor of FOMC decisions

To analyse the dynamics of the stepwise model forward rates, the calibration is performed on daily data in the period from January 2015 to September 2020. Additionally, the agreement between actual target rate changes and the corresponding change inferred from the initial term structure of calibrated forward rates is measured. This demonstrates how well the futures market was able to predict target rate changes in the test period. It is also a good indicator of the ability of the model to translate futures data into a meaningful term structure of anticipated target rate changes.

An R-squared is calculated between actual target rate changes $\Delta r^P(x_i)$ and corresponding initial forward rate term structure inferred changes $f^P(0,x_i) - f^P(0,x_i - h)$, grouped by the number of days in the forward rate term, that is the number of days between the calibration date corresponding to $t = 0$ and x_i. The results in Figure 3.2 show the R-squared for an increasing number of days between x_i and the calibration date. For comparison, the

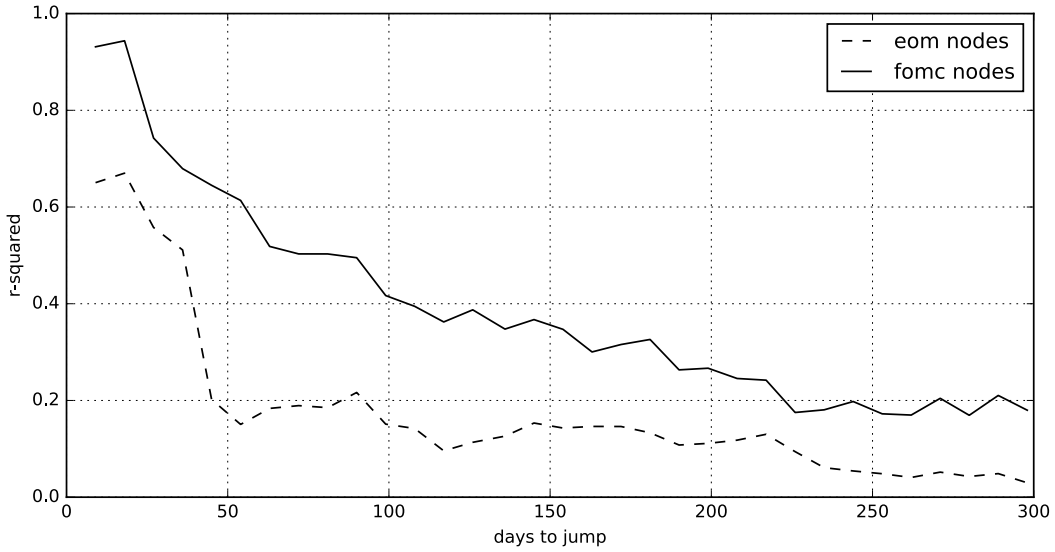

Figure 3.2: R-squared of EFFR realised vs forward rates for different forward periods

same calculation is shown with the same piecewise flat assumption but with discontinuity dates naively set to coincide with futures contract maturities (dashed line).

The results show a clear correspondence between actual and anticipated target rate changes. The correspondence deteriorates as the forward term increases but still shows evidence of some anticipation for terms over 200 days. The results comprise a mixture of good long-term anticipation of rate increases and poor anticipation of rate decreases. This can be attributed to the well-communicated and regular increases in the target rate during the normalisation phase following near-zero target rates. The rapid drop in target rates at the beginning of 2020 was not expected by the market, excluding this period would substantially improve the R-squared results.

3.2.2 SOFR Futures

SOFR futures are available in monthly and quarterly contract period lengths. The SOFR 1M futures contracts[12] are defined to reflect the specification of the Fed Funds 30 day futures with SOFR replacing the EFFR as the reference rate. Therefore the pricing formulas described in the previous section also apply to SOFR 1M futures.

In contrast to the monthly contracts, the final payoff of the SOFR 3M futures contracts[13] compounds SOFR, denoted by r_s, over IMM quarterly dates,[14] aligning the dates of the contracts to the LIBOR–referenced quarterly Eurodollar futures. Define q as the number of IMM quarters from the current trading quarter ($q = 0$), $\tau^*_{q,i} :=$ as the date corresponding

[12]Source:https://www.cmegroup.com/education/brochures-and-handbooks/sofr-futures-contract-specifications.html.

[13]Source:https://www.cmegroup.com/education/brochures-and-handbooks/sofr-futures-contract-specifications.html.

[14]Third Wednesday of March, June, September and December.

to day i in quarter q with n_q denoting the total days in quarter q. The SOFR 3M futures contract terminal payoff is:

$$\widetilde{F}_q^{s3}(\tau_{q,n_q}) = 100 - R_q^{s3}$$

where R^{s3} is based on SOFR compounded over the reference quarter:

$$R^{s3} = 100 \times \frac{360}{n_q}\left[\prod_{i=1}^{n_q}\left\{\mathbf{1}_{(\tau_{q,i}\in\mathbf{b})}\left(1+\frac{d_i r_s(\tau_{q,i})}{360}\right)\right\} - 1\right]$$

where $\mathbf{b}$ is the set of US government securities business days and d_i the number of days the rate $r_s(\tau_{q,i})$ applies.[15] Using two approximations which greatly simplify calculations but have an insubstantial numerical impact, see section 4.3.5 for numerical justification, the expected value at t of the futures contract index $\widetilde{F}_q^{s3}$ under the spot risk neutral measure is:

$$F_q^{s3}(t) = E_t[\widetilde{F}_q^{s3}(\tau_{q,n_q})] = 100\left(1 - \frac{360}{n_q}\left[\prod_{i=1}^{n_q}\left\{\mathbf{1}_{(\tau_{q,i}\in\mathbf{b})}\left(1+\frac{d_i E_t[r_s(\tau_{q,i})]}{360}\right)\right\} - 1\right]\right) \tag{3.53}$$

The current futures continues to trade during the observation quarter, therefore the valuation needs to account for already observed values of r_s:

$$F_0^{s3}(t) = E_t[\widetilde{F}_q^{s3}(\tau_{0,n_0})] = 100\left(1 - \frac{360}{n_0}\left[\prod_{i=1}^{n_0}\left\{\mathbf{1}_{(\tau_{0,i}\in\mathbf{b})}\left(1+\frac{d_i r_s^*(\tau_{0,i})}{360}\right)\right\} - 1\right]\right) \tag{3.54}$$

where $r_s^*(\tau_{0,i}) = \mathbf{1}_{(t>\tau_{0,i})} r_s(\tau_{0,i}) + \mathbf{1}_{(t\leq\tau_{0,i})} E_t[r_s(\tau_{0,i})]$

Calibration

Similarly to Fed Funds futures, SOFR 1M futures are available for each calendar month with liquidity approximately 1 year ahead and SOFR 3M futures are available between quarterly IMM dates approximately 2 years ahead of expiry. Calibrating the target rate term structure to Fed Fund futures allows the use of SOFR futures to extract information regarding the expected SOFR end–of–month spikes. To calibrate the spike component of the model to SOFR futures, similar assumptions in the case of Fed Fund futures are taken.

The SOFR term structure is assumed to consist of the target rate term structure obtained from Fed Fund futures, an end–of–month spike specific to the SOFR rate and a SOFR-specific spread. The drift component of the spike is ignored assuming it has a negligible effect on the inferred spike forward dynamics. The spread is assumed constant for all forwards, which is equivalent to the assumption of zero volatility for the noise component. The treatment related to the bid-ask spread is applied in the same way as for Fed Fund futures.

[15] d_i is equal to one plus the number of consecutive business days immediately following $\tau_{q,i}$.

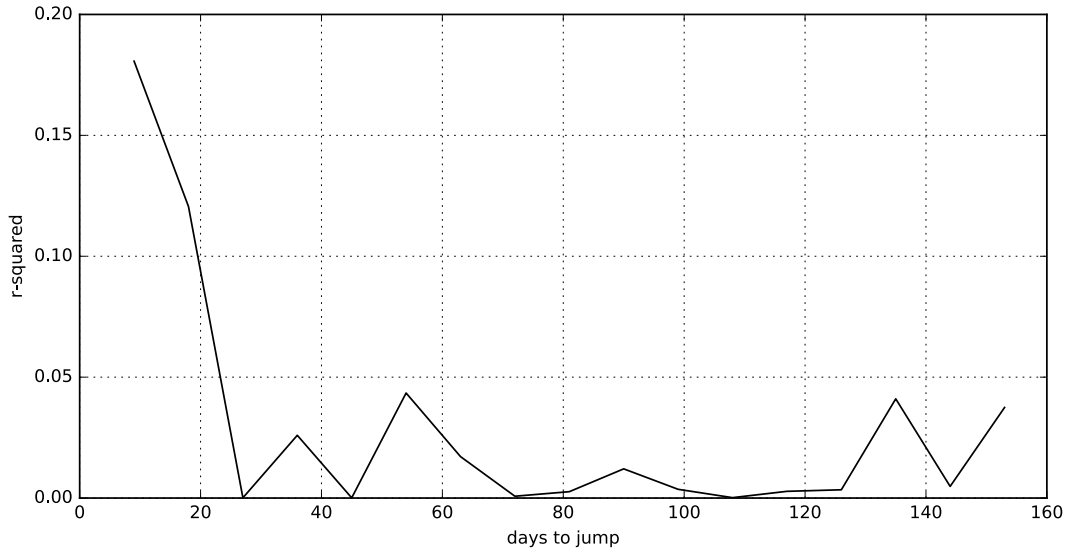

Figure 3.3: R-squared of SOFR realised vs forward rates for different forward periods

SOFR futures implied forwards as a predictor of SOFR spikes

Calibration is performed for all available SOFR futures data since the commencement of trading in June 2018. The agreement between expected SOFR spikes and actual spikes is measured by calculating the R-squared between end–of–month changes in the SOFR rate $\Delta r(z_i)$ and the corresponding forward spike $f^Z(0, z_i) - f^Z(0, z_i - h)$. The comparison is grouped by the number of days between the calibration date corresponding to $t = 0$ and z_i. The results in Figure 3.3 show the R-squared for an increasing number of days between z_i and the calibration date.

The results reveal some evidence of short-term anticipation of spikes close to the spike date. This is particularly true for the last trading day of the futures contract because the trading activity in the repo market from which the day's SOFR rate is calculated occurs simultaneously with trading in the futures market. The contrast to the high R-squared for target rate jumps anticipated by Fed Funds futures comes from the fact that FOMC target rate changes are communicated well ahead of time, particularly for rate increases, while the SOFR spikes depend on liquidity conditions, which are only anticipated in a short time frame, if at all. However, the most negative impact on the results is not a lack of anticipation of spikes, rather it is the over-anticipation of spikes, particularly when spikes do not occur.

3.2.3 Term rate dynamics

One of the approaches considered as the replacement for the LIBOR indexation of loan terms is a rate based on retrospectively compounding SOFR over the same term, see Figure 3.4 for a historical comparison. Both rates appear to follow the same underlying trend, this is related to the target rate term structure, which underlies all interest rates. LIBOR also

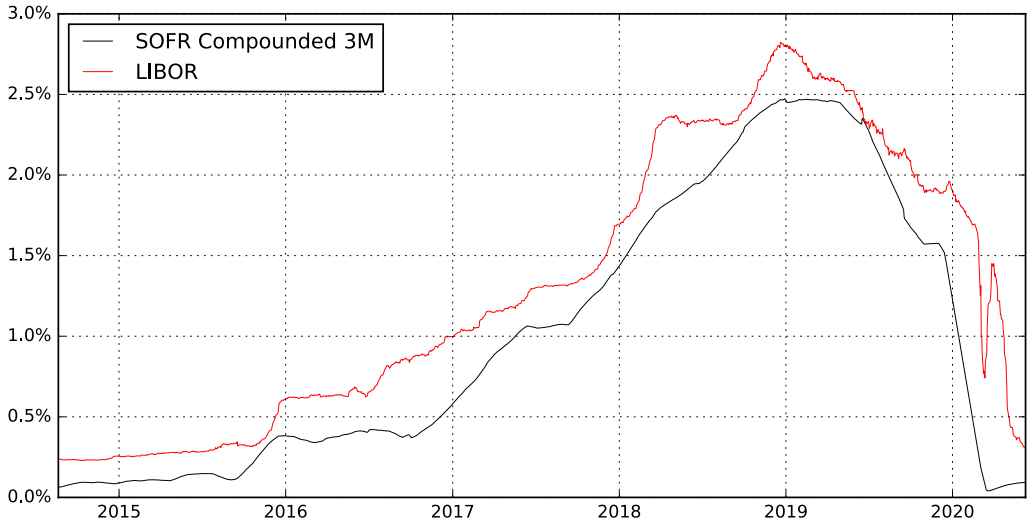

Figure 3.4: SOFR 3m rolling compounded rate compared to LIBOR

exhibits considerably more volatility. This is because the SOFR compounded rate is a rolling compounding calculation of already set rates, with only one new rate rolled in the calculation on each day. LIBOR, on the other hand, is a forward-looking term rate and is not subject to the volatility reduction from rolling compounding. The two rates, therefore, are not really comparable, which highlights one aspect of substantial problems with any proposal to replace LIBOR with a compounded SOFR.[16]

The calibration presented in the previous section enables a more analogous comparison of LIBOR and the SOFR forward-looking spot term rate. The SOFR term rates are calculated according to the compounding formula used to calculate SOFR 3M futures terminal payoff, using the daily forward rates obtained from the calibration.

The calculated spot SOFR 3M term rate is shown in Figure 3.5 in comparison to spot LIBOR. The rates are well correlated, approximately 50% of the LIBOR variance can be attributed to the SOFR 3M term rate. The impact of SOFR spikes dissipates over a 3-month compounding period, instead, the term rate is mostly driven by the target rate term structure. In turn, this shows that a significant proportion of LIBOR dynamics is driven by the target rate term structure exposed in the modelling framework. From this perspective, one can think of LIBOR trading at a spread to the term rates implied from the target rate term structure. One would expect this spread to be partly due to credit risk, but not entirely, since the term rate extracted from SOFR futures is not a "true" term rate in the sense that market participants could actually borrow at this rate[17] — one would therefore expect this spread also to include a "funding liquidity risk" component analogous to the one

[16]Other problems include the disconnect due to credit risk between SOFR and the cost of funding of private–sector banks, see Berndt, Duffie and Zhu (2020).

[17]If one takes into account a borrower's risk of not being able to refinance roll–over borrowing at (a constant spread to) a benchmark rate, this gives rise to additional basis spreads as observed in the market, see Alfeus, Grasselli and Schlögl (2020).

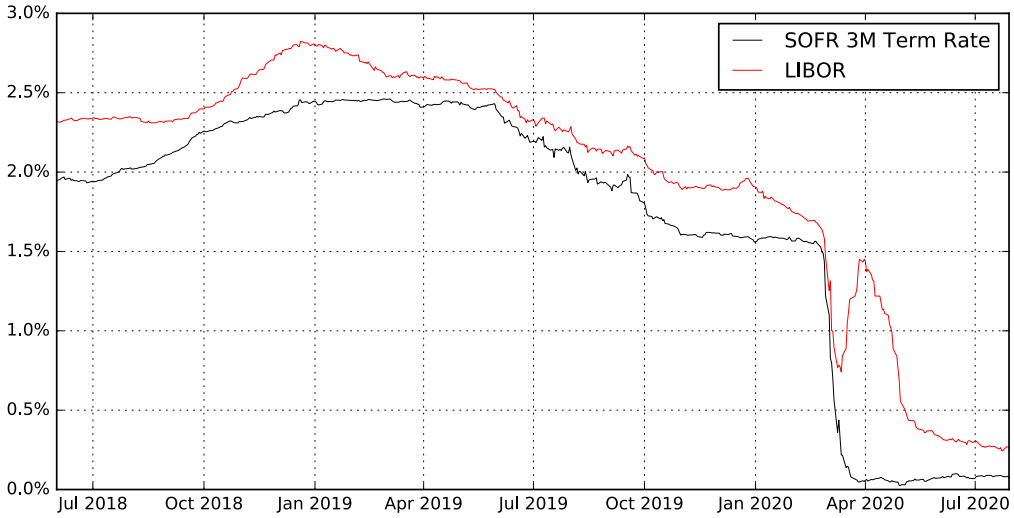

Figure 3.5: Spot SOFR 3m Term Rate vs LIBOR

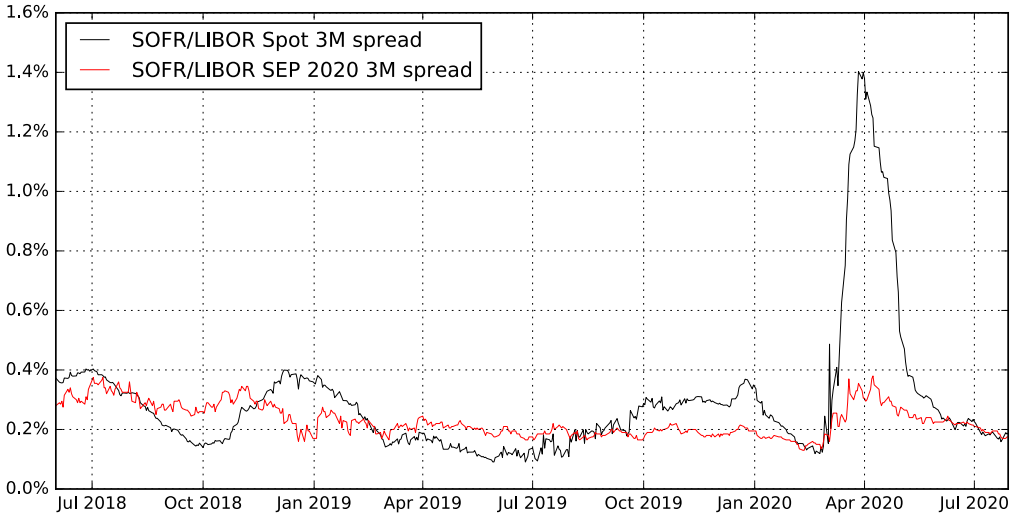

Figure 3.6: SOFR 3m Term Rate/LIBOR spread Spot vs Sep-2020 3M Term Rate

found in the LIBOR/OIS spread by Backwell, Macrina, Schlögl and Skovmand (2019).

It is also interesting to compare the spot and forward LIBOR to SOFR spread. As shown in Figure 3.6, the spread in the forward rates appears more stable, especially during the market turmoil in February and March of 2020. This is also in contrast to the large instability exhibited by the repo rates during the financial crisis of 2008, see Andersen and Bang (2020) for details. This is most likely due to Federal Reserve increasing operations in the repo market as a response to the September 2019 spike, which also appears to have eliminated end-of-month spikes.[18]

[18]See Federal Open Market Committee (2000-2020) September 2019 page 5.

3.3 Conclusion

The model introduced in this chapter reflects key empirical features present in short rate data. Calibration to futures showed that the piecewise structure imposed by the model induces better predictability of FOMC policy rate changes. This indicates the model is capturing significant information contained in derivative prices. However, the model in this form cannot be properly estimated since each factor relates to a specific FOMC date, therefore is not relevant in a general sense. The next iteration of the model presented in the next chapter overcomes this problem by defining the HJM volatility in a time-homogenous way.

Chapter 4

Adjusting For Time Homogeneity[1]

This chapter introduces a time-homogeneous version of the model introduced in the previous chapter. This allows the estimation of forward rate driving factors and their empirical states. The connection to FOMC dates allows for an intuitive economic interpretation of the model and empirical results. Comparison to the existing literature on the ability of the model to replicate historical futures prices is also provided in this chapter.

The rest of the chapter is organised as follows. The time-homogeneous version of the model is presented in Section 4.1. The approach to empirical analysis is detailed in section 4.2 and empirical results are presented in section 4.3. Section 4.4 concludes.

4.1 Model

4.1.1 Forward Rates

The model for forward rates is constructed such that they are driven by the evolution of expectations associated with FOMC target rate changes, where the target rate change for each scheduled meeting date evolves under its own dynamic. The forward rate $f(t,T)$ dynamics under the empirical measure can be written as follows:

$$f(t,T) = f(0,T) + \alpha(t,T) + \sum_{i=1}^{n} \int_0^t \xi_i(s,T) dZ_i(s) \tag{4.1}$$

where $f(0,T)$ is the initial term structure of forward rates, $\alpha(t,T)$ a deterministic drift and $dZ_i(s)$ the Wiener increment corresponding to the i^{th} driving Brownian motion with correlation $dZ_i(t)dZ_j(t) = \rho_{i,j}dt$. The volatility term is defined as follows:

$$\xi_i(t,T) = \xi_i \mathbb{1}(i \leq \mathcal{A}_{t,T}) \tag{4.2}$$

where $\mathcal{A}_{t,T}$ reflects the number of meeting dates between t and T:

$$\mathcal{A}_{t,T} := \big|\{x_1, ..., x_m | t < x_i \leq T\}\big| \tag{4.3}$$

[1]This chapter is based on a forthcoming revision of the paper Gellert and Schlögl (2021b) with Erik Schlögl contributing in a supervisory capacity.

where x_i denotes the i^{th} FOMC meeting date. The intuition behind this construction is that each stochastic component corresponds to evolving expectations related to an FOMC policy decision based on its order from the current state time.[2] The indicator function $\mathbb{1}(i \leq \mathcal{A}_{t,T})$ ensures that the i^{th} factor is only applied to forwards impacted by the relevant policy decision. Solving the integral yields:

$$\begin{aligned}\int_0^t \xi_i \mathbb{1}(i \leq \mathcal{A}_{s,T}) dZ_i(s) &= \xi_i \begin{cases} 0, & i > \mathcal{A}_{0,T} \\ Z_i(x_{\bar{i}(T)}), & i > \mathcal{A}_{t,T}, i \leq \mathcal{A}_{0,T} \\ Z_i(t), & i \leq \mathcal{A}_{t,T}, i \leq \mathcal{A}_{0,T} \end{cases} \\ &= \xi_i \mathbb{1}(i \leq \mathcal{A}_{0,T}) Z_i(t \wedge x_{\bar{i}(T)})\end{aligned} \tag{4.4}$$

where $\bar{i}(T) = \mathcal{A}_{0,T} - i + 1$. Therefore:

$$f(t,T) = f(0,T) + \alpha(t,T) + \sum_{i=1}^{n} \xi_i \mathbb{1}(i \leq \mathcal{A}_{0,T}) Z_i(t \wedge x_{\bar{i}(T)}) \tag{4.5}$$

To demonstrate the behaviour of the model with an example, let $x_2 < T < x_3$ and $t < x_1$:

$$f(t,T) = f(0,T) + \alpha(t,T) + \xi_1 Z_1(t) + \xi_2 Z_2(t) \tag{4.6}$$

Here one sees that both stochastic components, corresponding to FOMC dates x_1 and x_2, impact the forward rate up to time t. Any stochastic components beyond x_2 do not apply since the forward rate matures prior to x_3. Now let $x_1 < t < x_2$:

$$f(t,T) = f(0,T) + \alpha(t,T) + \xi_1 Z_1(t) + \xi_2 Z_2(x_1) \tag{4.7}$$

In this case both stochastic components apply to the period from $t = 0$ to $t = x_1$ since in that interval there are two meetings dates until forward time T. Once the first meeting occurs at $t = x_1$ the second stochastic component no longer applies since there is only one remaining meeting to T, the first component continues until time t. This reflects the dependency of the volatility function on the number of meetings dates between the current and forward time, which is what makes this version of the model time homogeneous in contrast to the version introduced in the previous chapter.

4.1.2 Short Rates

These forward rate dynamics create the piecewise dynamic in the short rate, which can be derived from (4.1) by setting $r(t) = f(t,t)$:

$$r(t) = f(t,t) = f(0,t) + \alpha(t,t) + \sum_{i=1}^{n} \int_0^t \xi_i(s,t) dZ_i(s) \tag{4.8}$$

[2]This is distinct from the previous chapter where each stochastic component was related to specific FOMC date. The approach shown in this chapter makes the model time homogeneous (thus allowing empirical estimation) and also more consistent with empirical behaviour where forward rate dynamics are a function of cardinality of meeting dates between the forward date and t.

Solving the integral yields:

$$\int_0^t \xi_i \mathbb{1}(i \leq \mathcal{A}_{s,t}) dZ_i(s) = \xi_i \begin{cases} 0, & i > \mathcal{A}_{0,t} \\ Z_i(x_{\bar{i}(t)}), & i \leq \mathcal{A}_{0,t} \end{cases} \\ = \xi_i \mathbb{1}(i \leq \mathcal{A}_{0,t}) Z_i(x_{\bar{i}(t)}) \tag{4.9}$$

Therefore:

$$r(t) = f(0,t) + \alpha(t,t) + \sum_{i=1}^{n} \xi_i \mathbb{1}(i \leq \mathcal{A}_{0,t}) Z_i(x_{\bar{i}(t)}) \tag{4.10}$$

From this it is evident that the short rate has no diffusion up until the first FOMC date at which point it picks up all the diffusion from the forward rate accumulated up until this point in time. To illustrate this, for $t < x_1$:

$$r(t) = f(0,t) + \alpha(t,t) \tag{4.11}$$

for $x_1 < t < x_2$:

$$r(t) = f(0,t) + \alpha(t,t) + \xi_1 Z_1(x_1) \tag{4.12}$$

for $x_2 < t < x_3$:

$$r(t) = f(0,t) + \alpha(t,t) + \xi_1 Z_1(x_2) + \xi_2 Z_2(x_1) \tag{4.13}$$

In general, the accumulated diffusion for the forward rates creates discontinuities in the short rate on FOMC dates, reflecting the empirical behaviour for the target rate and the associated forward rates.

4.1.3 Decomposition to Independent Components

The model can be easily transformed to independent components which will make it consistent with the HJM framework, thus facilitating the derivation of risk–neutral dynamics. Define Σ to be the covariance matrix of the vector $dZ = [dZ_1, ..., dZ_n]$. To transform the system to independent components, find a transformation matrix γ, such that $\Sigma = \gamma\gamma^T$ which is applied using $dZ = \gamma dW$, to result in a vector of uncorrelated Wiener increments $dW = [dW_1, ..., dW_n]$. Therefore:

$$\xi_i dZ_i = \sum_{j=1}^{n} \sigma_j \gamma_{i,j} dW_j \tag{4.14}$$

Rewrite the forward rate dynamics with respect to the uncorrelated components:

$$\begin{aligned}\sum_{i=1}^{n}\int_0^t \xi_i(s,T)dZ_i(s) &= \sum_{i=1}^{n}\int_0^t \xi_i \mathbb{1}(i \le \mathcal{A}_{t,T})dZ_i(s) \\ &= \sum_{i=1}^{n}\int_0^t \mathbb{1}(i \le \mathcal{A}_{t,T})\sum_{j=1}^{n}\sigma_j\gamma_{i,j}dW_j(s) \\ &= \sum_{j=1}^{n}\int_0^t \sigma_j(s,T)dW_j(s)\end{aligned} \tag{4.15}$$

where

$$\sigma_j(t,T) = \sigma_j \sum_{i=1}^{n}\gamma_{i,j}\mathbb{1}(i \le \mathcal{A}_{t,T}) \tag{4.16}$$

4.1.4 Forward Rates Under the Spot Risk–Neutral Measure

Formulate the risk neutral dynamics by using the result from HJM. Under the spot risk–neutral measure:

$$f(t,T) = f(0,T) + \sum_{j=1}^{n}\int_0^t \sigma_j(u,T)\int_u^T \sigma_j(u,s)dsdu + \sum_{j=1}^{n}\int_0^t \sigma_j(s,T)dW_j(s) \tag{4.17}$$

Similarly to (4.5) for the stochastic component and using (4.16) for the drift component:

$$\begin{aligned}f(t,T) - f(0,T) = &\sum_{j=1}^{n}\sum_{q=1}^{n}\sum_{i=1}^{n}\sigma_j^2\gamma_{q,j}\gamma_{i,j}\int_0^t \mathbb{1}(q \le \mathcal{A}_{u,T})\int_u^T \mathbb{1}(i \le \mathcal{A}_{u,s})dsdu \\ &+ \sum_{j=1}^{n}\sum_{i=1}^{n}\sigma_j\gamma_{i,j}\mathbb{1}(i \le \mathcal{A}_{0,T})W_j(t \wedge x_{\bar{i}(T)})\end{aligned} \tag{4.18}$$

Where:

$$\int_0^t \mathbb{1}(q \le \mathcal{A}_{u,T})\int_u^T \mathbb{1}(i \le \mathcal{A}_{u,s})dsdu = \int_0^t \mathbb{1}(q \le \mathcal{A}_{u,T})(T - x_{\hat{i}(u)})^+du \tag{4.19}$$

where $\hat{i}(u) = \mathcal{A}_{0,u} + i$. The integral $\int_0^t \mathbb{1}(q \le \mathcal{A}_{u,T})(T - x_{\hat{i}(u)})^+du$ is an integral over a simple piecewise constant function. While it is possible to solve the integral it is best left in this form. The solution expression appears over complicated since it needs to account for distinct logical combinations of t, T, q and i.

4.1.5 Short Rates Under the Spot Risk–Neutral Measure

Short rate dynamics can are obtained as follows:

$$r(t) = f(0,t) + \sum_{j=1}^{n} \int_0^t \sigma_j(u,t) \int_u^t \sigma_j(u,s) ds du + \sum_{j=1}^{n} \int_0^t \sigma_j(s,t) dW_j(s) \tag{4.20}$$

Similarly to (4.10) and (4.18):

$$\begin{aligned} r(t) - f(0,t) = & \underbrace{\sum_{j=1}^{n} \sum_{q=1}^{n} \sum_{i=1}^{n} \sigma_j^2 \gamma_{q,j} \gamma_{i,j} \int_0^t \mathbb{1}(q \leq \mathcal{A}_{u,t})(t - x_{\hat{i}(u)})^+ du}_{\text{deterministic term (**)}} \\ & + \underbrace{\sum_{j=1}^{n} \sum_{i=1}^{n} \sigma_j \gamma_{i,j} \mathbb{1}(i \leq \mathcal{A}_{0,t}) W_j(x_{\bar{i}(t)})}_{\text{stochastic term (*)}} \end{aligned} \tag{4.21}$$

4.1.6 Bond Prices

Bond prices can be written as follows:

$$B(t,T) = \exp\Big(- \int_t^T f(t,s) ds \Big) = \frac{B(0,T)}{B(0,t)} \exp\Big(a(t,T) + b(t,T) \Big) \tag{4.22}$$

where

$$a(t,T) = - \sum_{j=1}^{n} \sum_{q=1}^{n} \sum_{i=1}^{n} \sigma_j^2 \gamma_{q,j} \gamma_{i,j} \int_t^T \int_0^t \mathbb{1}(q \leq \mathcal{A}_{u,s})(s - x_{\hat{i}(u)})^+ du ds \tag{4.23}$$

$$b(t,T) = - \sum_{j=1}^{n} \sum_{i=1}^{n} \sigma_j \gamma_{i,j} \int_t^T \mathbb{1}(i \leq \mathcal{A}_{0,s}) W_j(t \wedge x_{\bar{i}(s)}) ds \tag{4.24}$$

4.2 Empirical approach

The previous section presented a forward rate model where a piecewise continuous forward rate structure induces discontinuities in the SOFR rate. Assuming Gaussian dynamics the HJM framework was used to derive a closed-form solution for forward rate dynamics. Estimation of the model requires the vector $\gamma_{i,j}$ and volatility term σ_j for each factor j. Aided by the intuition of this framework, in this section, the Gaussian dynamics assumption is relaxed to demonstrate an approach which allows a model agnostic estimation of the

vectors $\gamma_{i,j}$ and their empirical states. This allows close examination of the empirical behaviour of the model factors to inform further development of the model to resemble the empirical dynamics of the observed states.

SOFR forward rate term structures are not directly observable but can be inferred from the relevant futures contract prices. SOFR futures have been trading since May 2018 and consist of monthly and quarterly contracts to align them with respectively 30 day Fed Fund futures and Eurodollar Libor-based contracts. The contracts are available up to a forward period limited to approximately 2 years covered by liquidly traded futures contracts. The period of observation used for this section begins in June 2018 to capture the first full month of trading through to June 2021 consisting of 757 trading dates for which SOFR futures end-of-day settlement prices are available. The notation used in this section reflects the discrete nature of the empirical observations, SOFR futures payoffs and aligns with the discrete forward rate term structure of the model.

Define the observation period as a set of discrete dates $(t_0, ..., t_n)$ corresponding to trading days for SOFR futures. A set of SOFR futures settlement prices consisting of 1m and 3m futures, denoted as $F_j^1(t_a)$ and $F_j^3(t_a)$ respectively, is observed on each date. The subscript j indicates the position of the contract maturity, e.g. $F_3^1(t_a)$ is the third maturing 1m contract from t_a.

Denote the corresponding model price as $\hat{F}_j^1(t)$ and $\hat{F}_j^3(t)$. Using the generic futures pricing theorem, where E_β denotes the expectation under the spot risk neutral measure:

$$\hat{F}_j^1(t_a) = E_\beta\left[\hat{F}_j^1(\tau_j^1(t_a))\big|\mathcal{F}_{t_a}\right] = 100 - \frac{100}{D_j^1(t_a)}\left[S_j^1(t_a) + G_j^1(t_a)\right] \tag{4.25}$$

where $\tau_j^1(t_a)$ is the terminal date and $D_j^1(t_a)$ is the number of calendar dates in the reference period of the contract. $S_j^1(t_a)$ accounts for the accrued SOFR fixings if the contract is trading during the reference period:

$$S_j^1(t_a) = \sum_{i=n_j^1(t_a)}^{(t_a^*-1)\wedge N_j^1(t_a)} r(t_i)d_i \tag{4.26}$$

where t_a^* is the index of time t_a, $n_j^1(t_a)$ is the index of the first date in the reference period and $N_j^1(t_a)$ is the index of the last date in the reference period. d_i is the number of calendar days the rate at time t_i applies to in order to include accrual over non trading dates. The upper limit for the sum reflects that the SOFR is published the day following its reference date. $G_j^1(t_a)$ represents the sum of the SOFR forward rates relevant to the reference period:

$$G_j^1(t_a) = E_\beta\left[G_j^1(\tau_j^1(t_a))\big|\mathcal{F}_{t_a}\right] = \sum_{i=t_a^*\vee n_j^1(t_a)}^{N_j^1(t_a)} E_\beta\left[r(t_i)\big|\mathcal{F}_{t_a}\right]d_i \tag{4.27}$$

where

$$E_\beta\Big[r(t_i)\big|\mathcal{F}_{t_a}\Big] = E_\beta\Big[f^P(t_a,t_k) + \sum_{j=1}^{n}\int_{t_a}^{t_k}\sigma_j^P(u,t)\int_{u}^{t_k}\sigma_j^P(u,s)dsdu + \sum_{j=1}^{n}\int_{t_a}^{t_k}\sigma_j^P(s,t_k)dW_j^P(s)\Big|\mathcal{F}_{t_a}\Big] \tag{4.28}$$

$$= f^P(t_a,t_k) + \sum_{j=1}^{n}\int_{t_a}^{t_k}\sigma_j^P(u,t)\int_{u}^{t_k}\sigma_j^P(u,s)dsdu \tag{4.29}$$

The price of the quarterly SOFR futures is based on the compounding payoff defined for the contract:

$$\hat{F}_j^3(t_a) = 100 - 100\Big[\mathcal{S}_j^3(t_a)\mathcal{G}_j^3(t_a)\Big]\frac{360}{D_j^3(t_a)} \tag{4.30}$$

$\mathcal{S}_j^3(t_a)$ accounts for the compounded SOFR fixings if the contract is trading during the reference period:

$$\mathcal{S}_j^3(t_a) = \prod_{i=n_j^3(t_a)}^{(t_a^*-1)\wedge N_j^3(t_a)}\left(1+\frac{r(t_i)d_i}{360}\right) \tag{4.31}$$

$\mathcal{G}_j^3(t_a)$ represents the compounding of the SOFR forward rates relevant to the reference period:

$$\mathcal{G}_j^3(t_a) = E_\beta\Big[\prod_{i=t_a^*\vee n_j^3(t_a)}^{N_j^3(t_a)}\left(1+r(t_i)\frac{d_i}{360}\right)\Big|\mathcal{F}_{t_a}\Big] \tag{4.32}$$

The following two approximations greatly simplify calculations but have an insubstantial numerical impact, see Section 4.3.5 for numerical justification.

$$\mathcal{G}_j^3(t_a) \approx \prod_{i=t_a^*\vee n_j^3(t_a)}^{N_j^3(t_a)}\left(1+E_\beta\Big[r(t_i)\big|\mathcal{F}_{t_a}\Big]\frac{d_i}{360}\right) \tag{4.33}$$

Furthermore, the convexity correction due to the distinction between expectations under the spot and forward measures is ignored by setting:

$$E_\beta\Big[r(t_i)\big|\mathcal{F}_{t_a}\Big] \approx f(t_a,t_i) \tag{4.34}$$

For each day in the observation period assume forward rates are piecewise constant between FOMC dates and solve:

$$\mathbf{f}(t_a) = \underset{\mathbf{f}(t_a)}{\arg\min}\, O(t_a) \tag{4.35}$$

where $\mathbf{f}(t_a)$ is the vector of forward rates $f(t_a, t_i)$ which are piecewise constant between FOMC dates. The objective function is defined as the sum of square errors between the price given $\mathbf{f}(t_a)$ and the market price for monthly and quarterly futures:

$$O(t_a) = \sum_j (\hat{F}_j^1(t_a) - F_j^1(t_a))^2 + \sum_j (\hat{F}_j^3(t_a) - F_j^3(t_a))^2 \tag{4.36}$$

The vector $\mathbf{f}(t_a)$ can instead be expressed as a step function of the forward rate:

$$f(t_a, T) = f_0(t_a, T) + \sum_{i=1}^{n} v_i(t_a) \mathbb{1}(i \leq \mathcal{A}_{t_a,T}) \tag{4.37}$$

from which a vector of discrete forward rate levels $(f_0(t_a), ..., f_n(t_a))$ can be extracted, where:

$$f_k(t_a) = f_0(t_a) + \sum_{i=1}^{k} v_i(t_a) \tag{4.38}$$

v_i correspond to changes in FOMC policy rate change expectations which form the basis of the model. To obtain the empirical dynamics observe the changes in v_i. For any dates t_a not immediately following an FOMC meeting:

$$\Delta v_i(t_a) = v_i(t_a) - v_i(t_{a-1}), t_{a-1} \notin \{x_0, ..., x_n\} \tag{4.39}$$

For dates following an FOMC meeting the effect of rolling the FOMC meeting index needs to be considered:

$$\Delta v_i(t_a) = v_i(t_a) - v_{i+1}(t_{a-1}), t_{a-1} \in \{x_0, ..., x_n\} \tag{4.40}$$

v_{n+1} is by definition not observed, therefore truncate the estimate to $n-1$ FOMC dates. Define a matrix of $\Delta v_i(t_a)$ observations.

$$\mathbf{V} = \begin{bmatrix} \Delta v_1(t_1) & \dots & \Delta v_{n-1}(t_1) \\ \vdots & \ddots & \\ \Delta v_1(t_m) & \dots & \Delta v_{n-1}(t_m) \end{bmatrix} \in \mathbb{R}^{m \times (n-1)} \tag{4.41}$$

The matrix $\mathbf{V}$ can be factorised using principal component decomposition:

$$\mathbf{S} = \mathbf{V}\mathbf{W} \tag{4.42}$$

where $\mathbf{W} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a matrix of column wise eigenvectors of the matrix $\mathbf{V}^T\mathbf{V}$. The eigenvectors represent a new basis which factorises the matrix $\mathbf{V}$ into $n-1$ independent factors. The matrix $\mathbf{S} \in \mathbb{R}^{m\times(n-1)}$ denotes the empirical states of the independent factors.

In order to estimate a reduced factor model, truncate the matrices such that $\mathbf{W}^* = \mathbf{W}[\{1, ..., n-1\}, \{1, ..., b\}] \in \mathbb{R}^{(n-1)\times b}$ and $\mathbf{S}^* = \mathbf{S}[\{1, ..., m\}, \{1, ..., b\}] \in \mathbb{R}^{m\times b}$. The matrix

$\mathbf{V}^* = \mathbf{S}^*(\mathbf{W}^*)^T$ represents changes in FOMC policy rate change expectations corresponding to the reduced factor truncation. Let $s_j(t_a)$ be the (a, j) element of matrix $\mathbf{S}^*$ and let $w_{i,j}$ be the (i, j) element of matrix $\mathbf{W}$. Then:

$$\Delta v_i^*(t_a) = \sum_{j=1}^{b} s_j(t_a) w_{i,j} \tag{4.43}$$

From which the truncated jump states can be obtained as follows:

$$v_i^*(t_a) = \Delta v_i^*(t_a) + \begin{cases} v_i^*(t_{a-1}) & , t_{a-1} \notin \{x_0, ..., x_n\} \\ v_{i+1}^*(t_{a-1}) & , t_{a-1} \in \{x_0, ..., x_n\} \end{cases} \tag{4.44}$$

where the initial state is obtained from the calibration, i.e. $v_i^*(t_0) = v_i(t_0)$, therefore:

$$v_i^*(t_a) = v_i(t_0) + \sum_{a^*=1}^{a} \Delta v_{i^*}^*(t_{a^*}) = v_i(t_0) + \sum_{a^*=1}^{a} \sum_{j=1}^{b} s_j(t_{a^*}) w_{i,j} \tag{4.45}$$

where $i^* = i + \mathcal{A}_{t_{a^*}, t_a}$ denoting the number of FOMC meetings between t_{a^*} and t_a. Therefore the truncated forwards rates can be written as follows:

$$\begin{aligned} f^*(t_a, T) &= f_0(t_a, T) + \sum_{i=1}^{n} \left(v_i(t_0) + \sum_{a^*=1}^{a} \sum_{j=1}^{b} s_j(t_{a^*}) w_{i,j} \right) \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) \\ &= f_0(t_a, T) + \sum_{i=1}^{n} \left(v_i(t_0) + \sum_{a^*=1}^{a} \sum_{j=1}^{b} s_j(t_{a^*}) w_{i,j} \right) \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) \\ &= f_0(t_a, T) + \sum_{i=1}^{n} v_i(t_0) \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) + \sum_{i=1}^{n} \sum_{a^*=1}^{a} \sum_{j=1}^{b} s_j(t_{a^*}) w_{i,j} \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) \\ &= f_0(t_a, T) + \sum_{i=1}^{n} v_i(t_0) \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) + \sum_{j=1}^{b} \sum_{i=1}^{n} w_{i,j} \mathbb{1}(i \leq \mathcal{A}_{t_a, T}) \sum_{a^*=1}^{a} s_j(t_{a^*}) \end{aligned} \tag{4.46}$$

Taking the increment over time, for $t_{a-1} \notin \{x_0, ..., x_n\}$:

$$\begin{aligned}
\Delta f^*(t_a, T) &= f^*(t_a, T) - f^*(t_{a-1}, T) \\
&= f_0(t_a, T) + \sum_{i=1}^{n} v_i(t_0)\mathbb{1}(i \leq \mathcal{A}_{t_a,T}) + \sum_{j=1}^{b}\sum_{i=1}^{n} w_{i,j}\mathbb{1}(i \leq \mathcal{A}_{t_a,T}) \sum_{a^*=1}^{a} s_j(t_{a^*}) \\
&- f_0(t_{a-1}, T) - \sum_{i=1}^{n} v_i(t_0)\mathbb{1}(i \leq \mathcal{A}_{t_{a-1},T}) - \sum_{j=1}^{b}\sum_{i=1}^{n} w_{i,j}\mathbb{1}(i \leq \mathcal{A}_{t_{a-1},T}) \sum_{a^*=1}^{a-1} s_j(t_{a^*}) \\
&= \sum_{j=1}^{b}\sum_{i=1}^{n} w_{i,j}\mathbb{1}(i \leq \mathcal{A}_{t_a,T})\Big(\sum_{a^*=1}^{a} s_j(t_{a^*}) - \sum_{a^*=1}^{a-1} s_j(t_{a^*})\Big) \\
&= \sum_{j=1}^{b}\sum_{i=1}^{n} w_{i,j}\mathbb{1}(i \leq \mathcal{A}_{t_a,T}) s_j(t_a)
\end{aligned} \tag{4.47}$$

The above equation connects the empirical results to the model as follows. First write the forwards rates without the drift component for a reduced factor model:

$$f^P(t, T) = f^P(0, T) + \sum_{j=1}^{b}\sum_{i=1}^{n} \gamma_{i,j}\mathbb{1}(i \leq \mathcal{A}_{0,T})\sigma_j W_j^P(t \wedge x_{\bar{i}(T)}) \tag{4.48}$$

Taking the increment between t_{a-1} and t_a in the case where $\mathcal{A}_{t_{a-1},t_a} = 0$

$$\begin{aligned}
\Delta f^P(t_a, T) &= f^P(t_a, T) - f^P(t_{a-1}, T) \\
&= f^P(0, T) + \sum_{j=1}^{n}\sum_{i=1}^{n} \gamma_{i,j}\mathbb{1}(i \leq \mathcal{A}_{0,T})\sigma_j W_j^P(t_a \wedge x_{\bar{i}(T)}) \\
&- f^P(0, T) - \sum_{j=1}^{n}\sum_{i=1}^{n} \gamma_{i,j}\mathbb{1}(i \leq \mathcal{A}_{0,T})\sigma_j W_j^P(t_{a-1} \wedge x_{\bar{i}(T)}) \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} \gamma_{i,j}\mathbb{1}(i \leq \mathcal{A}_{0,T})\sigma_j \Big(W_j^P(t_a \wedge x_{\bar{i}(T)}) - W_j^P(t_{a-1} \wedge x_{\bar{i}(T)})\Big)
\end{aligned} \tag{4.49}$$

Now:

$$W_j^P(t_a \wedge x_{\bar{i}(T)}) - W_j^P(t_{a-1} \wedge x_{\bar{i}(T)}) = \begin{cases} W_j^P(t_a) - W_j^P(t_{a-1}) & , t_{a-1} < x_{\bar{i}(T)} \\ 0 & , t_a > x_{\bar{i}(T)} \end{cases} \tag{4.50}$$

therefore:

$$W_j^P(t_a \wedge x_{\bar{i}(T)}) - W_j^P(t_{a-1} \wedge x_{\bar{i}(T)}) = \mathbb{1}(i \leq \mathcal{A}_{t_a,T})\big(W_j^P(t_a) - W_j^P(t_{a-1})\big) \tag{4.51}$$

Let $\Delta W_j^P(t_a) = W_j^P(t_a) - W_j^P(t_{a-1})$:

$$\Delta f^P(t_a, T) = \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_{i,j} \mathbb{1}(i \leq \mathcal{A}_{t_a,T}) \sigma_j \Delta W_j^P(t_a) \tag{4.52}$$

Comparing equations Eq.(4.47) and Eq.(4.52) it is apparent that the empirical results are connected to the model with $w_{i,j} = \gamma_{i,j}$ and $s_j(t_a) = \sigma_j \Delta W_j^P(t_a)$.

4.3 Empirical Results

4.3.1 Factor decomposition

The estimation described in the previous section produces an empirically based decomposition into the piecewise forward rate structures corresponding to the modelling approach proposed in this chapter. The decomposition informs a dimension reduction effected by removing factors which do not significantly impact estimated dynamics. In order to choose the appropriate level of factor reduction the relationship between the number of factors and the total root mean square error (RMSE) is examined. Define the total RMSE as:

$$e(b) = \sqrt{\sum_{a=1}^{m} \frac{O(t_a)}{m}} \tag{4.53}$$

where b is the number of factors used to obtain the forward rates from the calibration. Figure 4.1 suggests that over the sample period factors 8 and above are redundant. The HJM framework allows for the inclusion of any number of factors, therefore the number of factors becomes a choice between modelling accuracy and parsimony.

The estimated λ vectors reflect the empirical dynamics of policy rate expectations. Therefore they offer an interesting economic interpretation of the driving dynamics of forward rates, especially those with short maturities. Figure 4.2 shows the λ vectors corresponding to the top three factors. The factors reflect the general level across the term structure (factor 1), the gradient (factor 2) and curvature (factor 3). The general shape of these vectors aligns with their parametrised counterparts in the Nelson-Siegel model (see Nelson and Siegel (1987)).

The modelling setup proposed in this chapter interprets forward rates as an accumulation of expected policy rate changes. Therefore the decomposition weight vectors can be cast as factorised policy rate change expectations, where the resulting λ vectors correspond directly to factors in the proposed model. An interesting insight emerges regarding the dynamics of forward rates. As shown in Figure 4.2, the λ vector corresponding to factor 1 is concentrated mostly on the first element, revealing that parallel changes in the forward rate term structure are equivalent to changing expectations regarding the next FOMC policy rate change.

Although these results are based on the short end of the term structure, this connection can be conceptually extended to longer terms. Parallel changes have been long known to

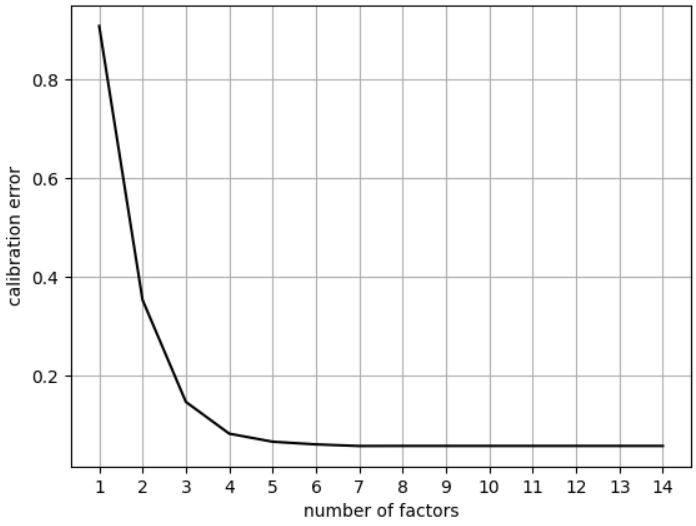

Figure 4.1: Total calibration RMSE given number of factors

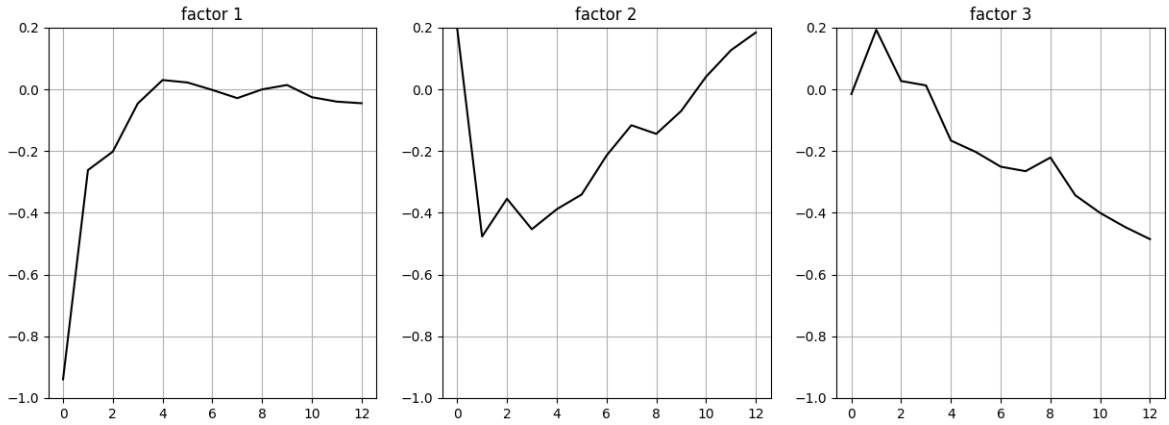

Figure 4.2: λ vectors for the first three factors

be the primary driver of term structure dynamics, these results reveal that parallel forward curve changes are directly related to changing expectations related to the first FOMC date. The connection between FOMC policy rate expectations and parallel changes in the forward curve is a key insight stemming from this modelling approach.

The second factor is similarly concentrated on the first element but with a distinctive change in the opposite direction for the remaining values. The magnitude of the first element is smaller than the sum of the remaining values which gives rise to an opposite change in the forward rates between the first and last forward with a smooth transition

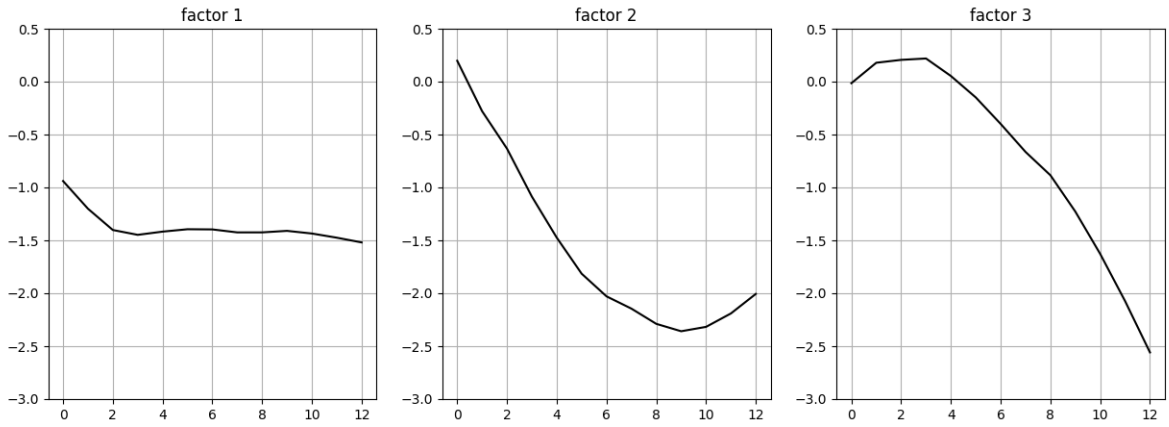

Figure 4.3: Forward rate level vectors for the first three factors

in between. For the forward curve, this is a gradient change, which again is typical for the second factor in term rate dynamics. Although the interpretation is that this captures dynamics where the expected aggregate changes in policy rates do not change, the timing of those changes does. For example, the expectations of a rate rise increase for the next meeting. However, the aggregate expected level does not change therefore the expectations of rate rises decrease for subsequent meetings. An equivalent interpretation is that the second factor represents a negative correlation between the expectations related to the next policy rate change with the remaining term structure. Again this is a key insight, connecting FOMC policy rate change expectations and a long-known shape of the second factor for term structure changes.

The third factor, which appears as a curvature change in the forward rate term structure, is actually quite similar to the second factor. It also represents a negative correlation involving the next policy rate change. However, this time focusing mostly on the relationship with the second FOMC meeting rather than the remaining term structure. Similarly to factor 2, the interpretation is related to the aggregate outcome over the next two meetings remaining fairly constant while allowing for uncertainty regarding the timing of the change. The insight from this interpretation can be extended even further. The top three factors, which cover most of the forward rate dynamics, are all related to the next policy rate change and the resulting implications for the expectations of subsequent policy rate changes.

4.3.2 Calibration performance

A notable feature of the calibration error shown in Figure 4.1 is that it does not converge to zero, meaning that even a full-factor model is not able to perfectly calibrate to futures prices on a cross-sectional basis. Futures are readily tradable basic derivative instruments and as such, it is an arbitrage requirement that they are fully reconciled by pricing models. In general, cross-sectional calibration performance depends on the number and nature of

instrument	M0	M1	M2	M3	M4	M5	M6	Q0	Q1	Q2	Q3	Q4
1 factor	1.2	1.7	6.6	12.8	19.6	26.0	30.7	7.2	22.9	36.3	47.4	68.6
2 factor	1.1	1.4	1.5	1.4	2.2	2.6	2.6	0.8	2.1	1.9	5.0	10.6
3 factor	1.1	1.4	1.5	1.3	1.6	1.5	1.3	0.8	1.0	1.0	1.1	2.1
Skov & Skovmand	2.9	3.1	3.3	2.6	2.1	1.6	1.6	0.9	1.3	1.8	0.9	1.8

Figure 4.4: RMSE table for monthly(M) and quarterly(Q) contracts, comparing results between different number of factors for the proposed model and results published in Skov and Skovmand (2021)

factors since they translate directly to degrees of freedom in the calibration. The results indicate that there are still residual errors which could be addressed by the inclusion of spikes as per the previous chapter. However, the aim is to first determine the set of primary factors that explain as much of the cross-sectional fit, state dynamics and short-rate dynamics as possible. From this perspective and for a more granular understanding define the calibration error per instrument as a root mean square error (RMSE):

$$e_j^x(b) = \sqrt{\frac{\sum_a (\hat{F}_j^x(t_a) - F_j^x(t_a))^2}{m}} \tag{4.54}$$

The RMSE results for the first three factors in the model are shown Figure 4.4. A distinct feature of SOFR futures is the existence of both monthly and quarterly futures with overlapping reference periods. An approach which would yield zero calibration error would be to choose interpolation points on each of the futures maturity dates, thus lining up the calibration factors directly to the instruments. This approach has a much-reduced ability to predict FOMC policy rate changes [3] and therefore does not align well with empirical short-rate dynamics.

FOMC meetings occur eight times per year, therefore the choice of piecewise flat regions between FOMC dates results in fewer degrees of freedom than there are instruments even in the full factor model. This could be mitigated with the addition of further factors such as the spike and a mean-reverting spread factor proposed in the previous chapter. A factor for end-of-month spikes would provide an extra degree of freedom aligned with each monthly future. However, the main purpose of this section is to compare to published results by Skov and Skovmand (2021).

To allow the comparison the results are produced over the same period and instruments as the aforementioned paper. The results produced by Skov and Skovmand (2021) show a significant deterioration for the first three monthly contracts, which does not seem to occur in the model proposed in this chapter. This could be explained by as well as highlighting a key difference between the two models. Both models are three-factor models and the state vectors show similar characteristics. However, in the Nelson-Siegel model used by Skov and Skovmand (2021), the forward structures are continuous as opposed to piecewise flat as proposed in this model. It is observed as part of this empirical research that short expiry

[3] As shown in section 3.2.1.

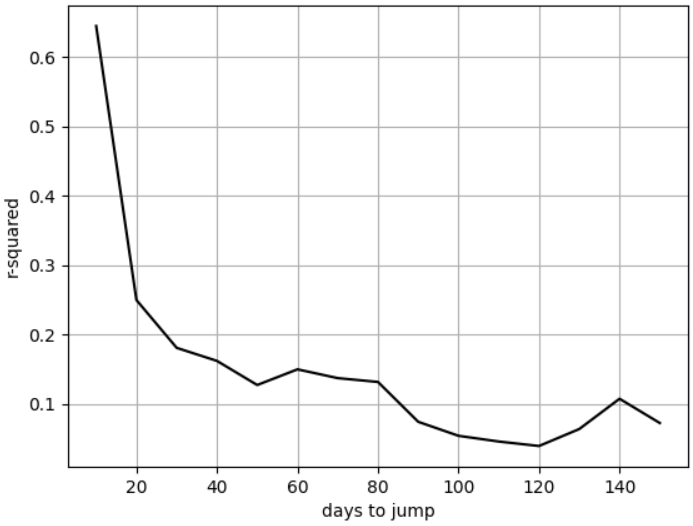

Figure 4.5: R-squared of FOMC policy rate change vs forward rates for different forward periods

futures are particularly sensitive to the forward rate structure, with the proposed piecewise structure being able to better align with the front monthly futures.

4.3.3 Policy rate change prediction

In this section, the analysis discussed in section 3.2.1 is repeated for the calibration performed in this chapter. However, this time the information is extracted directly from SOFR futures. Similarly to section 3.2.1, the r-squared is calculated between expected and actual policy rate changes for different forward rate periods. However, this time focusing on forward rates derived directly from SOFR futures. As shown in Figure 4.5 there is a strong indication that policy rate changes are increasingly well anticipated as the length of the forward rate period decreases. Additionally, there is a distinction between well-communicated policy rate increases and the more sudden nature of policy rate decreases, see Figure 4.6.

The expectations reflected in the model are directly related to FOMC policy rate changes which tend to be well communicated ahead of time. One exception is in the case of crisis situations where the Federal Reserve has suddenly sharply decreased the policy rate without much advance warning. Therefore there should be increasingly good agreement between the actual and expected policy rate changes as the forward period to the next FOMC meeting decreases. This agreement should be better for rate increases which were well communicated than the decreases which were associated with emergency Federal Reserve action.

The drop in correspondence at the very short end of the forward rate period for policy rate increases can also be attributed to noise in the SOFR rate which is reflected in short forward rates in the form of a spread to the policy target rate, making it difficult to precisely disentangle changes in spread and changes to policy rate change expectations.

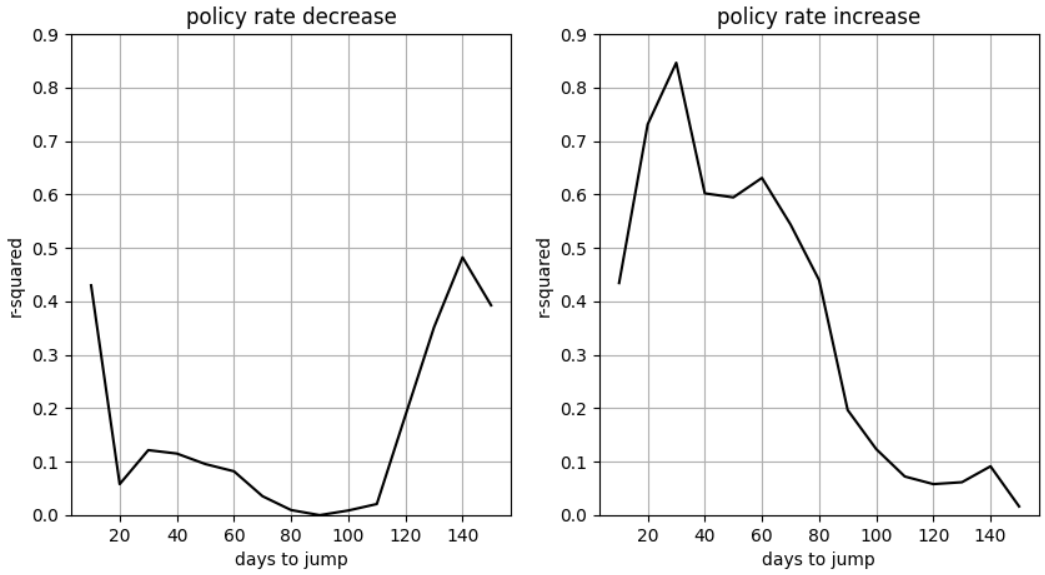

Figure 4.6: R-squared of FOMC policy rate change vs forward rates for different forward periods split by up and down changes

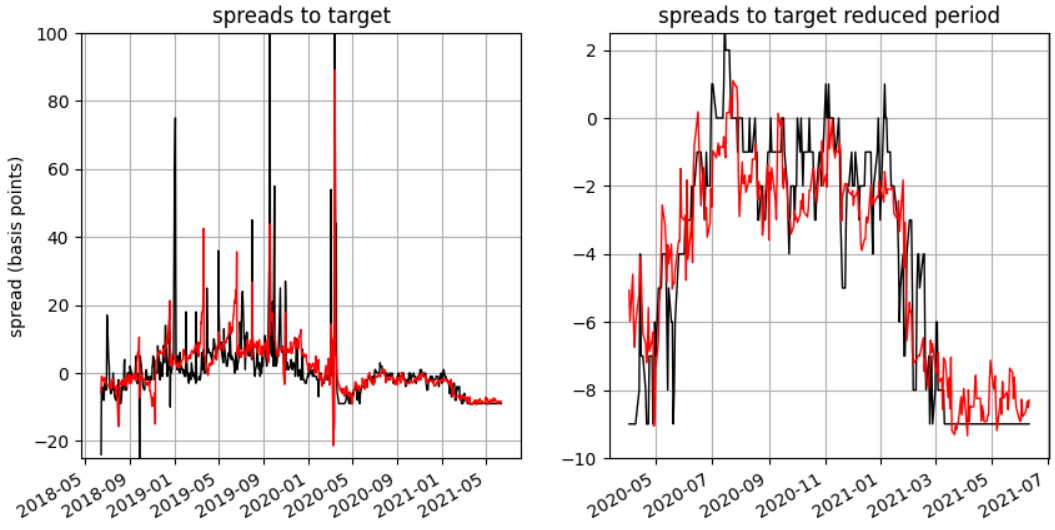

Figure 4.7: comparison of actual (black) and model(red) SOFR to the target rate spread between 2018-2021(LHS) and between May-2020 and July 2021(RHS)

4.3.4 SOFR spread from calibration

Another aspect captured by the calibration is the SOFR spread to the target rate. This is reflected in the forward rate before the first FOMC meeting date, corresponding to the

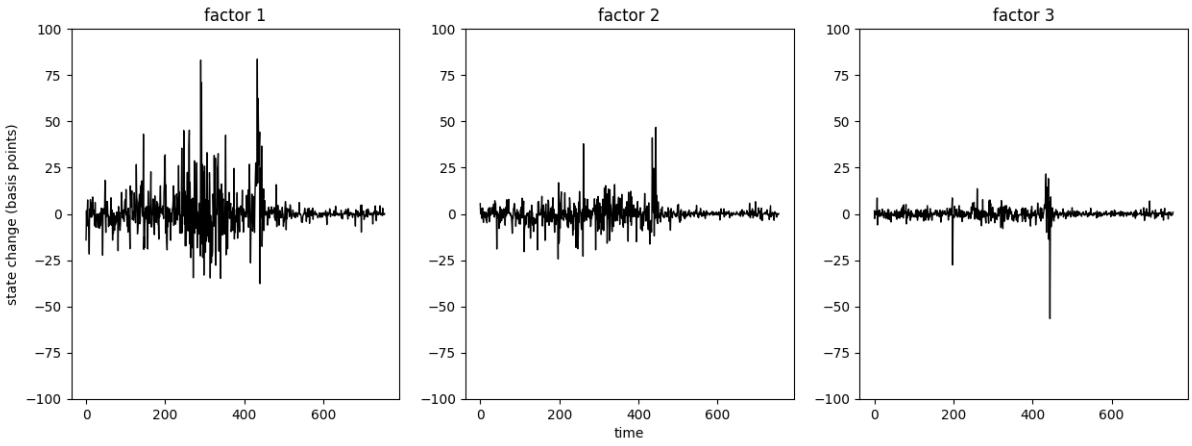

Figure 4.8: Daily changes in the states of estimated factors

first term in Eq.(4.37). Since spikes are not included, in some cases the presence of spikes and expectations of the spike is also reflected in the SOFR spread and the calibration, particularly in the period when spikes are present. For the SOFR spread spikes results in an elevated spread on the day and a few days following the spike. As shown in the previous chapter there is evidence that the expectation of spikes is embedded in futures prices and it is possible to extract this expectation in calibration. However, the focus of this chapter is the estimation of the main driving dynamics of SOFR forward term structure and therefore spike estimation is not repeated.

The very large spike in September 2019 resulted in active intervention by the Federal Reserve to stabilise SOFR. Since then they have mostly disappeared with one exception in March 2020. The calibrated spread corresponds very well to the actual spread in the post-spike period, see Figure 4.7. In the period where spikes are present, it follows the spread reasonably well but there are discrepancies related to the spikes. These results highlight that the dynamics of the forward rate maturing before the first FOMC date are driven by the SOFR spread. The correspondence between the actual spread and the calibrated spread shows that the piecewise structure facilitates the accurate extraction of this information from futures.

4.3.5 Risk neutral drift estimate

The empirical results presented in this chapter relied on two key simplifying assumptions, mostly motivated by computational speed, allowing empirical analysis across a broader range of data. The calibration was performed with zero drift, assuming it is negligible over the test period. The other assumption relates to Eq. (4.32), which approximates the expectation of the compounded rates by compounding the expected rates, violating Jensen's inequality. The error resulting from the two approximations can be quantified by the adjustment which would be required to obtain the strictly correct price.

The adjustments to a set of futures used for the calibration are calculated using Monte

	Jensen			Jensen + drift		
contract	adjustment(bp)	95% ci (2m paths)		adjustment(bp)	95% ci (2m paths)	
M0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
M1	-0.0042	-0.0199	0.0116	0.0085	-0.0072	0.0243
M2	0.0110	-0.0103	0.0322	-0.0133	-0.0345	0.0080
M3	0.0059	-0.0218	0.0337	0.0124	-0.0154	0.0401
M4	-0.0084	-0.0452	0.0284	0.0000	-0.0369	0.0368
M5	-0.0028	-0.0424	0.0367	-0.0373	-0.0768	0.0022
M6	-0.0221	-0.0674	0.0232	-0.0101	-0.0554	0.0351
M7	0.0521	-0.0008	0.1049	-0.0078	-0.0606	0.0449
M8	-0.0401	-0.0952	0.0149	-0.0986	-0.1537	-0.0436
M9	0.0339	-0.0268	0.0946	-0.0248	-0.0855	0.0359
M10	0.0397	-0.0260	0.1054	-0.0778	-0.1434	-0.0121
M11	0.0167	-0.0518	0.0853	-0.0301	-0.0987	0.0384
M12	0.0044	-0.0685	0.0773	-0.1067	-0.1796	-0.0338
Q0	-0.0018	-0.0102	0.0066	0.0049	-0.0036	0.0133
Q1	0.0018	-0.0296	0.0332	-0.0367	-0.0681	-0.0052
Q2	-0.0034	-0.0518	0.0450	-0.0280	-0.0764	0.0203
Q3	-0.0645	-0.1266	-0.0023	-0.0857	-0.1478	-0.0235
Q4	-0.0798	-0.1553	-0.0043	-0.1991	-0.2747	-0.1235
Q5	-0.0633	-0.1528	0.0262	-0.3142	-0.4037	-0.2246
Q6	-0.0857	-0.1854	0.0140	-0.3932	-0.4929	-0.2935
Q7	-0.2121	-0.3214	-0.1028	-0.5277	-0.6369	-0.4185

Table 4.1: Jensen inequality and drift Impact of assumptions on futures

Carlo simulation using volatilities estimated from the empirical states. The results, showing both the Jensen inequality adjustment and the drift adjustment are presented in Table 4.1. Jensen's inequality should only impact quarterly futures and indeed for all the monthly futures, the zero adjustments is within the 95% confidence interval. The maximum adjustment amounts to around half a basis point for the eighth quarterly contract.

The adjustments are relatively small(and economically insignificant), particularly for the contracts used to obtain the results in section 4.3.2, with the maximum adjustment for the fifth quarterly contract at around 0.2 basis points. Including the adjustments in the calibration would result in a slight change in the calibration state therefore the RMSE reported in section 4.3.2 would increase by less than the adjustment if at all. The adjustments are stable over time and therefore the change in states from the adjustments is also stable over time. Therefore the adjustments would have a minimal contribution to the daily change states. Since the covariance used for the PCA reduction is estimated from the change in states, the adjustments are not expected to change the covariance estimate or the reduced factor model.

4.4 Conclusion

The reformulated model presented in this chapter allowed for a time-homogeneous estimation of the driving factors and their states. This allows an analysis of the factor state dynamics. This analysis is used to motivate a stochastic volatility extension to the model which is presented in the next chapter.

Chapter 5

Stochastic Volatility Extension[1]

5.1 Introduction

This chapter introduces the final version of the model proposed in this thesis. The model includes a stochastic volatility extension, where volatility follows dynamics inspired by Heston (1993). It constitutes a tractable multi-factor, stochastic volatility model which incorporates piecewise constant short-rate dynamics and diffusive forward-rate dynamics introduced in previous chapters. Calibrating to prices for options on SOFR futures, a good fit to the market is achieved across available maturities and strikes in a single, consistent model. The model also provides novel insights into SOFR term rate behaviour (and implied volatilities) within the SOFR term rate accrual periods, as well as into empirical mean reversion dynamics. The empirical motivation for this Chapter was outlined in Section 1.5.

The rest of the chapter is organised as follows. The stochastic volatility version of the model is introduced in Section 5.2. The resulting term rate dynamics in the context of accrual period behaviour, mean reversion and option calibration are examined in Section 5.3. Cross-sectional calibration results are presented in Section 5.4. Section 5.5 concludes the chapter.

5.2 Model

5.2.1 HJM with a piecewise volatility function

The model introduced in Chapter 3, produces dynamics where the short rate is constant between specified dates while the forward rates evolve as a continuous diffusion. This is achieved within the HJM framework by specifying a volatility function which results in discontinuities in the short rate. The change in the short rate at the point in time when the discontinuity occurs reflects the accumulated diffusion of the corresponding forward rates. This reflects empirical behaviour where changes in the short rate occur at known times due

[1]This chapter is based on the paper Brace, Gellert and Schlögl (2022) with Erik Schlögl contributing in a supervisory capacity and Alan Brace contributing notes on which Section 5.2.2 introducing the Heston stochastic volatility model, is based.

to FOMC policy rate changes while the expected value of those changes continuously over time. The model begins with the standard HJM result for forward rate dynamics with N factors:

$$f(t,T) = f(0,T) + \sum_{j=1}^{N} \int_0^t \sigma_j(u,T) \int_u^T \sigma_j(u,s) ds du + \sum_{j=1}^{N} \int_0^t \sigma_j(s,T) dW_j(s) \tag{5.1}$$

Define $\sigma_j(t,T)$ as a piecewise constant function between FOMC meeting dates:

$$\sigma_j(t,T) = \sigma_j \sum_{i=1}^{n} \gamma_{i,j} \mathbb{1}(i \leq \mathcal{A}_{t,T}) \tag{5.2}$$

where n is the total number of meetings dates and $\mathcal{A}_{t,T}$ reflects the number of meeting dates between t and T:

$$\mathcal{A}_{t,T} := \big|\{x_1, ..., x_m | t < x_i \leq T\}\big| \tag{5.3}$$

σ_j and $\gamma_{i,j}$ scales the variance of each factor. σ_j allows control of the overall level of variance and is the key variable used in calibration to option prices. $\gamma_{i,j}$ scales the volatility based on the number of FOMC meeting dates between t and T. It can be empirically derived to reflect the covariance structure between forward rates.[2] Solving the stochastic integral yields:

$$\begin{aligned} f^P(t,T) - f^P(0,T) = &\sum_{j=1}^{n} \sum_{q=1}^{n} \sum_{i=1}^{n} \sigma_j^2 \gamma_{q,j} \gamma_{i,j} \int_0^t \mathbb{1}(q \leq \mathcal{A}_{u,T}) \int_u^T \mathbb{1}(i \leq \mathcal{A}_{u,s}) ds du \\ &+ \sum_{j=1}^{n} \sum_{i=1}^{n} \sigma_j \gamma_{i,j} \mathbb{1}(i \leq \mathcal{A}_{0,T}) W_j^P(t \wedge x_{\bar{\imath}(T)}) \end{aligned} \tag{5.4}$$

where $\bar{\imath}(T) = \mathcal{A}_{0,T} - i + 1$. The solution reveals that the total variance is an increasing function of the number of meeting dates between 0 and T, up to the minimum of t and the last meeting date before T. This implies that the variance of the forward rate is zero if the forward date occurs prior to the next meeting date.

5.2.2 HJM with a piecewise stochastic volatility function

The stochastic volatility extension of the model in Chapter 4 proposed in this chapter draws inspiration from an approach endowing a Hull and White (1990) interest rate terms structure model with stochastic volatility. The quasi-Gaussian HHW, presented in this section, builds on the Gaussian Hull-White model by adding a Heston stochastic volatility component. Start with a 1-factor quasi-Gaussian model (QG1) model with HJM defined volatility:

$$\sigma(t,T) = \chi(t)\phi(T) \tag{5.5}$$

[2] This is derived in Chapter 4 using principal component analysis and is in fact given by the eigenvectors stemming from the analysis.

where $\chi(t)$ is generally stochastic, and integrate from 0 to t the SDE for the instantaneous forward $f(t,T)$ under the spot measure, resulting in

$$df(t,T) = F(t,T)dt + \sigma(t,T)dW(t) \text{ where } F(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du \tag{5.6}$$

$$\implies f(t,T) - f(0,T) = \int_0^t F(s,T)ds + \phi(T)\int_0^t \chi(s)dW(s) \tag{5.7}$$

Then set $T = t$ and differentiate with respect to t to express the spot $r(t)$ in the form

$$\begin{aligned} r(t) - f(0,t) = x(t) &= \int_0^t F(s,t)ds + \phi(t)\int_0^t \chi(s)dW(s), x(0) = 0 \\ dx(t) &= \frac{d}{dt}\Big\{\int_0^t F(s,t)ds\Big\}dt + \phi'(t)\int_0^t \chi(s)dW(s) + \phi(t)\chi(t)dW(t) \\ &= \frac{d}{dt}\Big[\int_0^t F(s,t)ds\Big]dt + \frac{\phi'(t)}{\phi(t)}\Big[x(t) - \int_0^t F(s,t)ds\Big]dt + \sigma(t,t)dW(t) \end{aligned} \tag{5.8}$$

Define

$$\phi(T) = \exp\Big(-\int_0^T \lambda(v)dv\Big) \implies \frac{\phi'(t)}{\phi(t)} = -\lambda(t) \tag{5.9}$$

$$\chi(t) = \sigma(t)\exp\Big(\int_0^t \lambda(v)dv\Big) \implies \sigma(t,T) = \chi(t)\phi(T) = \sigma(t)\exp\Big(-\int_t^T \lambda(v)dv\Big) \tag{5.10}$$

therefore

$$F(t,T) = \sigma^2(t)\exp\Big(-\int_t^T \lambda(v)dv\Big)\int_t^T \exp\Big(-\int_t^u \lambda(v)dv\Big)du, F(t,t) = 0 \tag{5.11}$$

Hence σ inherits the stochasticity of $\chi, \sigma(t,t) = \sigma(t)$ and the SDE changes to

$$dx(t) = \Big\{\frac{d}{dt}\Big[\int_0^t F(s,t)ds\Big] + \lambda(t)\int_0^t F(s,t)ds\Big\}dt - \lambda(t)x(t)dt + \sigma(t)dW(t) \tag{5.12}$$

in which the part of the drift term involving $F(t,T)$ simplifies to

$$\begin{aligned}\Phi(t) &= F(t,t) + \int_0^t \frac{\partial}{\partial T} F(s,T)\Big|_{T=t} ds + \lambda(t) \int_0^t F(s,t) ds \\ &= \int_0^t \sigma^2(s) \exp\Big(-2 \int_s^t \lambda(v) dv \Big) ds = \int_0^t \sigma^2(s,t) ds\end{aligned} \tag{5.13}$$

The volatility $\sigma(.)$ is made stochastic by incorporating a Heston process $v(.)$ in it:

$$\sigma(t) \to \sigma(t)\sqrt{v(t)} \tag{5.14}$$

which constitutes an affine system under the HJM spot measure:

$$\begin{aligned} dx(t) &= [\Phi(t) - \lambda(t)x(t)]dt + \sigma(t)\sqrt{v(t)}dW(t), x(0) = 0 \\ d\Phi(t) &= [\sigma^2(t)v(t) - 2\lambda(t)\Phi(t)]dt, \Phi(0) = 0 \\ dv(t) &= \theta(t)(1 - v(t))dt + \alpha(t)\sqrt{v(t)}dU(t), v(0) = 1 \end{aligned} \tag{5.15}$$

with

$$\langle dW(.), dU(.) \rangle (t) = \rho dt \tag{5.16}$$

Bond price dynamics can be written as follows (see Appendix B for derivation):

$$B(t,T) = \exp\Big(-\int_t^T f(t,u)du \Big) = \frac{B(0,T)}{B(0,t)} \exp\Big(-\Lambda(t,T)y(t) - \frac{1}{2}\Phi(t)\Lambda^2(t,T) \Big) \tag{5.17}$$

where:

$$\Lambda(t,T) = \int_t^T \exp\Big(-\int_t^u \lambda(v)dv \Big) du \tag{5.18}$$

$$\Phi(t) = \int_0^t \sigma^2(s) \exp\Big(-2\int_s^t \lambda(v)dv \Big) ds \tag{5.19}$$

5.2.3 Piecewise Heston HJM

The idea behind the stochastic volatility extension of the model is that each factor evolves with its own independent Heston stochastic volatility. That is, each factor in the piecewise HJM model is extended in the same manner as HW in the HHW model. Thereby, the model proposed inherits the piecewise constant structure but is instead driven by Heston rather than Gaussian dynamics.

This set-up provides ample flexibility to calibrate to volatility term structure (since each factor uniquely impacts different aspects of the forward rate term structure), as well as option implied volatility skew and smile across different expiries. The level of flexibility is regulated by the choice of the number of factors and the time dependence of the HHW variables. Notably in the absence of the indicator function and for $n = 1$, the model collapses to the HHW.

Starting with the standard HJM multifactor framework:

$$f(t,T) = f(0,T) + \sum_{j=1}^{N} \int_0^t \sigma_j(u,T) \int_u^T \sigma_j(u,s) ds du + \sum_{j=1}^{N} \int_0^t \sigma_j(s,T) dW_j(s) \tag{5.20}$$

Define the HJM volatility as follows:

$$\sigma_j(t,T) = \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(t,T)\}} \chi_j(t) \phi_j(T) \gamma_{i,j} \tag{5.21}$$

where

$$\phi_j(T) = \exp\left(- \int_0^T \lambda_j(s) ds \right) \tag{5.22}$$

and

$$\chi_j(t) = \sigma_j(t) \sqrt{v_j(t)} \exp\left(\int_0^t \lambda_j(s) ds \right) \tag{5.23}$$

$v(t)$ evolves with a Heston dynamic:

$$dv(t) = \theta(t)(1 - v(t))dt + \alpha(t)\sqrt{v(t)} dU(t), v(0) = 1 \tag{5.24}$$

with

$$\langle dW_j(.), dU_j(.) \rangle (t) = \rho_j dt \tag{5.25}$$

and

$$\langle dW_i(.), dU_j(.) \rangle (t) = 0, \text{ for } i \neq j \tag{5.26}$$

The bond price dynamics for a single factor[3] can be written as (See Appendix B for

[3]The multifactor expression is trivial but notationally complicated.

derivation.):

$$B(t,T) = \exp\Big(- \int_t^T f(t,u)du \Big) \tag{5.27}$$

$$= \frac{B(0,T)}{B(0,t)} \exp\Big(- \sum_{b=0}^{\eta(t)-2} \Lambda_{x_{\eta(T)-1}}(x_{b+1},T) y_{\eta(T)-1}(x_{b+1}) - \Lambda_{x_{\eta(T)-1}}(t,T) y_{\eta(T)-1}(t) \tag{5.28}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \sum_{b=0}^{\eta(t)-2} \Lambda_{x_k}(x_{b+1},x_{k+1}) y_k(x_{b+1}) - \sum_{k=\eta(t)}^{\eta(T)-2} \Lambda_{x_k}(t,x_{k+1}) y_k(t) \tag{5.29}$$

$$- \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{\eta(T)-1-b} \sum_{j=1}^{\eta(T)-1-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \{ \Lambda^2(x_{b+1},T) - \Lambda^2(x_{b+1}, x_{\eta(T)-1}) \} \tag{5.30}$$

$$- \frac{1}{2} \sum_{i=1}^{\eta(T)-\eta(t)} \sum_{j=1}^{\eta(T)-\eta(t)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \{ \Lambda^2(t,T) - \Lambda^2(t, x_{\eta(T)-1}) \} \tag{5.31}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{k-b} \sum_{j=1}^{k-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \{ \Lambda^2(x_{b+1}, x_{k+1}) - \Lambda^2(x_{b+1}, x_k) \} \tag{5.32}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{i=1}^{(k-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \{ \Lambda^2(t, x_{k+1}) - \Lambda^2(t, x_k) \} \Big) \tag{5.33}$$

5.2.4 SOFR term rates

The LIBOR to SOFR transition imposes on the market a change from rates set for a longer-term (usually 3 months) to rates with an effective term of 1 business day. Transitioning to daily frequency for derivative instruments would not be desirable for many reasons, including burdening the system with a large increase in transaction volumes to settle daily flows. Instead, the market is adopting an approach where instruments are still defined with longer term rates. However, those term rates are now calculated using either compounding or averaging of SOFR over the term. These rates are better known as term SOFR.

A LIBOR term would be defined by the start date T_i and an end date T_k of the period over which it applies. A SOFR term for the corresponding dates is defined as a set of discrete dates $\{T_i, , , T_k\}$ on which SOFR is observed. The most common definition of term SOFR is based on compounding over the term (usually 3m):

$$S(T_i, T_k) = \tau_{i,k} \Big[\prod_{j=i}^{k} (1 + s(T_j)\delta_j) - 1 \Big] \tag{5.34}$$

where $\tau_{i,k}$ is the year fraction of the term length and $s(t)$ is the SOFR observed set for T_j. δ_j is the year fraction of the period between T_j and T_{j+1} to account for days on which SOFR

is not observed (weekends and holidays). Note that for the empirical results presented in this section it is assumed that the daily SOFR rate is approximated by the continuous short rate $r(t)$.

5.2.5 Pricing Futures

Define a 3M SOFR futures contract $F(T_i, T_k)$ with accrual period starting at T_i and ending at T_k, with payoff measurable at T_k:

$$F(T_i, T_k) = 100\bigg(1 - S(T_i, T_k)\bigg) \tag{5.35}$$

where $\delta_{i,k}$ is the year fraction between T_i and T_k. Using the generic futures pricing theorem the expected value at t under the spot risk neutral measure β of the futures contract is:

$$F(t, T_i, T_k) = E_\beta\bigg[F(T_i, T_k)|\mathcal{F}_t\bigg] \tag{5.36}$$

5.2.6 Pricing Options on Futures

Options on 3M SOFR futures exist for a variety of strikes and expiries. Although they are specified with American style exercise, here they are used to approximate European style implied volatilities as is common in practice. The impact of the American exercise is not assessed in this chapter, instead, the options are used to demonstrate the ability of the model to calibrate to a variety of strikes and expiries. The value of a call option at time t, expiring at $T_e < T_i$ with strike K on the futures contract is given by the expectation under the spot risk-neutral measure as:

$$C(t, T_e, F(T_i, T_k), K) = E_\beta\bigg[\frac{1}{\beta(T_e)}(F(T_i, T_k) - K)^+|\mathcal{F}_t\bigg] \tag{5.37}$$

5.2.7 Monte Carlo Simulation

As an initial proof of concept, particularly the ability of the model to calibrate to SOFR options, the following simulation is implemented. Rewriting to isolate stochastic components from Eq.(5.20) requiring Monte Carlo simulation:

$$\int_0^t \sigma_j(s,T)dW_j(s) = \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(t,T)\}}\gamma_{i,j}\phi_j(T) \int_0^t \sigma_j(s)\sqrt{v_j(s)}\exp\bigg(\int_0^s \lambda_j(q)dq\bigg)dW_j(s) \tag{5.38}$$

Setting constant parameters $\sigma_j(s) = \sigma_j$ and $\lambda_j(q) = \lambda_j$:

$$\int_0^t \sigma_j(s)\sqrt{v_j(s)}\exp\left(\int_0^s \lambda_j(q)dq\right)dW_j(s) = \sigma_j \int_0^t \sqrt{v_j(s)}e^{s\lambda_j}dW_j(s) \tag{5.39}$$

The stochastic component is approximated as follows:

$$\int_0^t \sqrt{v_j(s)}e^{s\lambda_j}dW_j(s) \approx \frac{1}{N}\sum_{p=1}^{N} g(t) \tag{5.40}$$

where

$$g_j(t) = \int_0^t \sqrt{v_j(s)}e^{s\lambda_j}dW_j(s) \tag{5.41}$$

calculated with Euler discretisation:

$$\Delta g_j(s) = \sqrt{v_j(s)}e^{s\lambda_j}\Delta W_j(s) \tag{5.42}$$

where $\Delta W_j(s) \sim N(0, \sqrt{\Delta t})$, $v_j(s) = v_j(s - \Delta t) + \Delta v_j(s)$ and:

$$\Delta v_j(s) = \theta(1 - v_j(s - \Delta t))\Delta t + \alpha\sqrt{v_j(s - \Delta t)}\Delta U_j(s), v_j(0) = 1 \tag{5.43}$$

where $\Delta U_j(s) \sim N(0, \sqrt{\Delta t})$ and $\langle \Delta W_j(.), \Delta U_j(.)\rangle(s) = \rho_j \Delta t$

5.3 Term rate dynamics

5.3.1 Factor Sensitivities

Calibration to interest rate options requires the model to fit the first four moments of terminal distributions across different expiries and tenors. The moments of a terminal distribution at a specific expiry are usually characterised in terms of implied volatilities across different strikes. In this representation, the first moment corresponds to a horizontal shift in the implied volatilities (across strikes), the second moment a vertical shift (across all implied volatilities), the third moment a gradient shift and the fourth moment a shift in convexity.

Using a model calibration on 10-June-2022 to the first four quarterly SOFR options including all available strikes, the sensitivity of implied volatilities to the model parameters is examined. Starting with σ, see Figure 5.1, it is apparent that changing this variable results in a parallel shift in the implied volatilities, thereby controlling the second moment. The table in Figure 5.1, shows how different factors impact different expiries with factor 1 focusing on short expiries, while factors 2 and 3 increasingly focus on the longer expiries, providing calibration flexibility across the term structure.

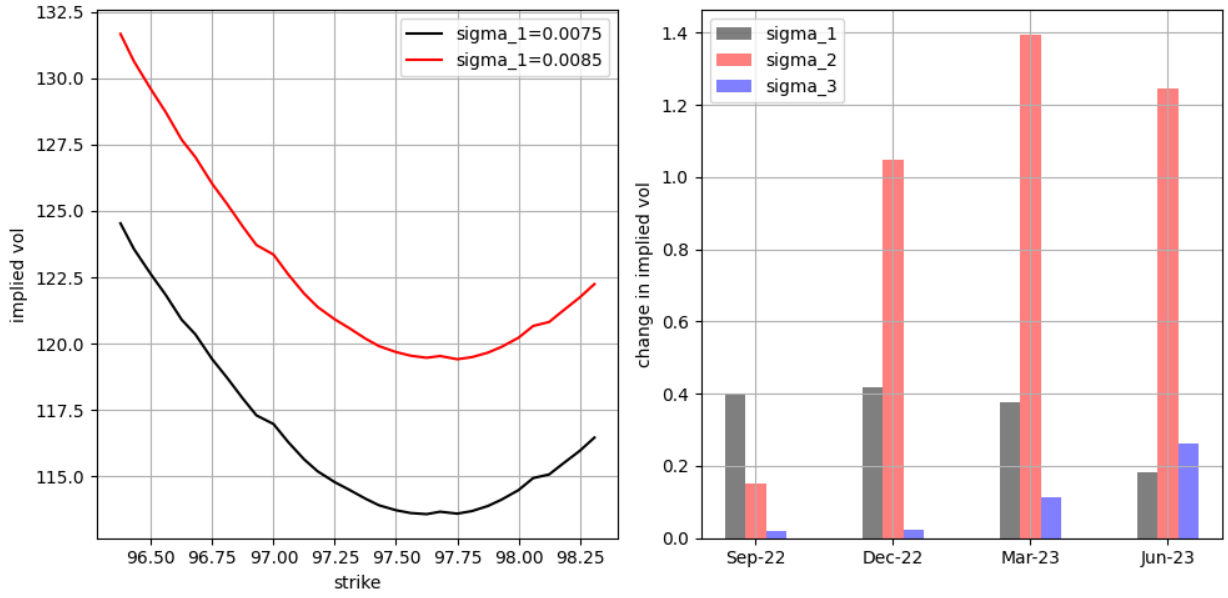

Figure 5.1: (LHS)Implied volatility for different values of σ.(RHS) ATM implied volatility sensitivity across contracts to changes in σ

The α parameter determines the level of stochastic volatility in the model and is associated with the fourth moment. As shown in Figure 5.2, changing the α parameter results in a change in convexity as well the level of volatilities. Control of just convexity, without changing ATM volatilities is possible by combining offsetting changes in the σ parameter. As per the σ parameter, the table in Figure 5.2 shows a different impact on convexity from different factors across expiries enabling the model to calibrate to different stochastic volatility term structures.

The ρ parameter determines the correlation between the stochasticity of the volatility and the forward rates. As can be seen in Figure 5.3, changing the ρ parameter results in a gradient change in implied volatilities corresponding to a change in the third moment. Similarly to the other parameters, the impact on implied volatility skewness varies from different factors across expiries allowing the model to calibrate to different correlation term structures.

λ and θ are two variables associated with mean reversion. The λ parameter controls the mean reversion of the forward rates while the θ parameter controls the mean reversion of the stochastic volatility level. From an implied volatility perspective, as shown in Figure 5.4, the mean reversion parameters work in reverse to their corresponding volatility parameters. The λ parameters offset the impact from σ and result in a parallel change in implied volatility in the opposite sign to the change in the parameter. The θ parameter reverses the α parameter and therefore results in both a level and convexity change in the implied volatilities.

The model proposed has the flexibility to attempt simultaneous calibration to implied volatilities across both strikes and expiries. Additional flexibility for calibration comes from

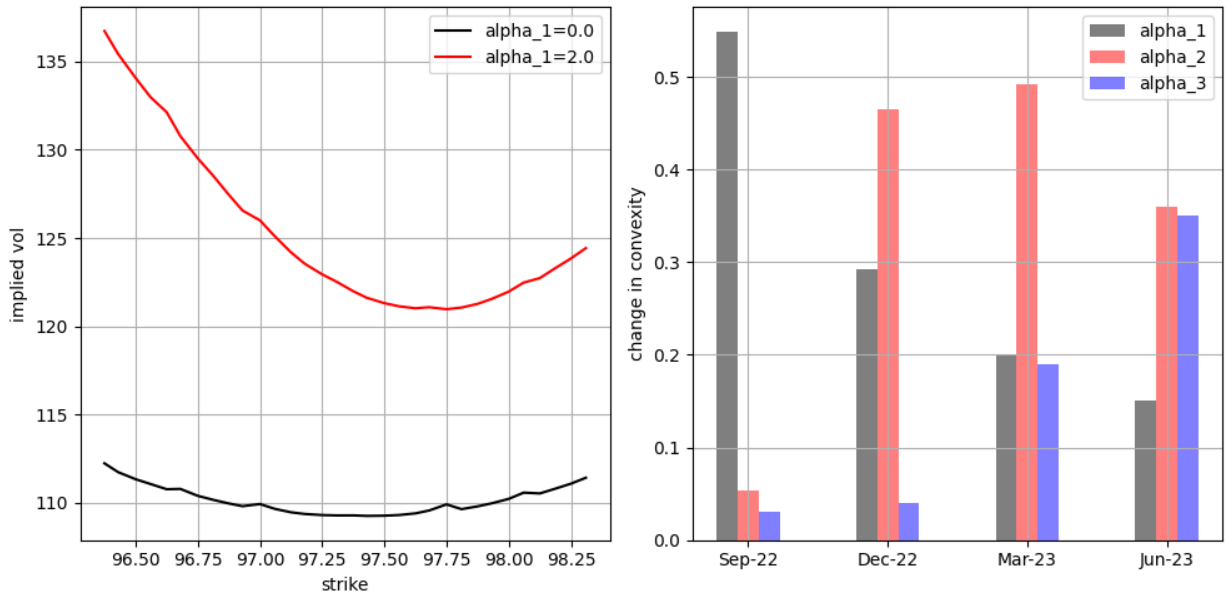

Figure 5.2: (LHS) Implied volatility for different values of α. (RHS) Implied volatility convexity (measured as the second derivative of implied volatility as a function of strike) sensitivity across contracts to changes in α

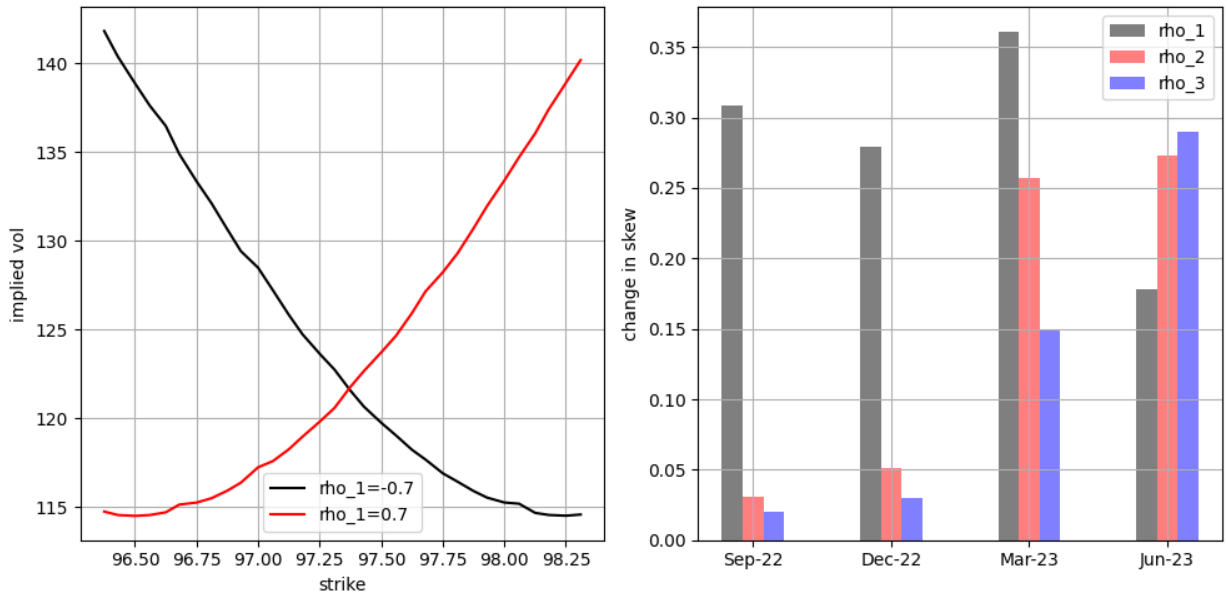

Figure 5.3: (LHS) Implied volatility for different values of ρ. (RHS) Implied volatility skew (measured as the first derivative of implied volatility as a function of strike) sensitivity across contracts to changes in ρ

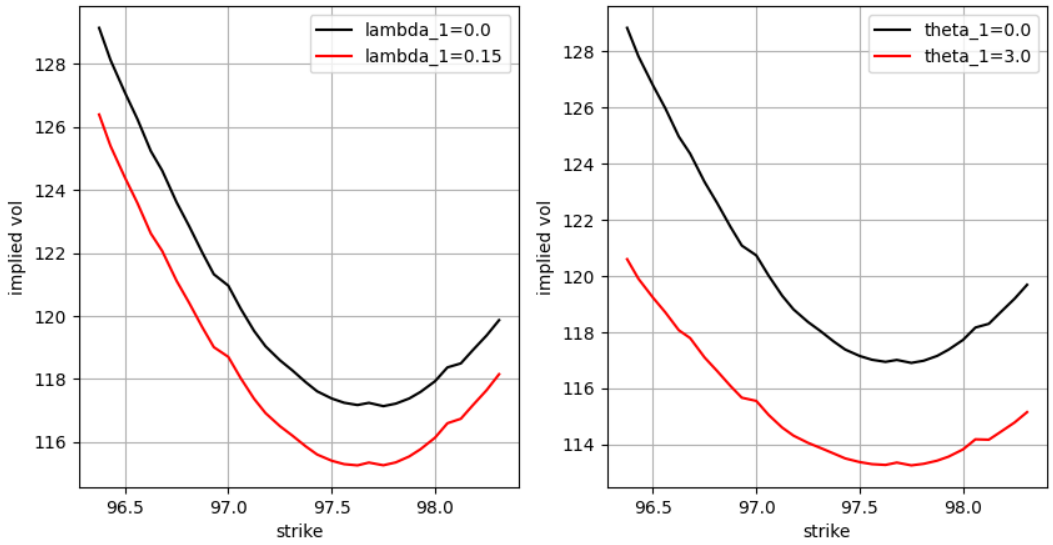

Figure 5.4: (LHS) Implied volatility for different values of λ (RHS) implied volatility for different values of θ

the ability to define the variables as functions of time. In the next section, the model's ability to calibrate to options on SOFR futures is demonstrated.

5.3.2 Accrual period

A prevalent approach in the LIBOR to SOFR transition, as reflected in the literature, see Lyashenko and Mercurio (2019), is the adaptation of existing LIBOR-based modelling to SOFR. A trivial but highly practical problem stemming from this approach is the behaviour of options in the accrual period of the SOFR term rate, i.e. for term forwards $S(T_i, T_k)$ at time $T_i < t \leq T_k$. This occurs when the expiry of the option is set past the beginning of the accrual period.

Examples of impacted options are in arrears SOFR caps and exchange-traded options on 1M SOFR futures[4]. Options on averaging and compounding SOFR term rates can be considered as average rate options on the short rate. However, existing LIBOR-based pricing models simply treat SOFR term rates like a LIBOR term rate set in arrears (i.e., observed at the end of its accrual period). These models do not have an in-built capacity to cater for the compounding or averaging aspect of SOFR. This means that these models need to be adapted (in an arbitrary fashion) for the behaviour of SOFR within the accrual period. This is done by making the implied volatility a function of the accrual period. Hence, in the LIBOR-based models, SOFR essentially continues to be treated like LIBOR,

[4]Options on 3M futures expire prior to the accrual period, hence are not impacted by accrual period behaviour.

where accrual period behaviour is absorbed into the implied volatility behaviour. To this effect, Lyashenko and Mercurio (2019) suggest that implied volatility is a linearly decaying function of the accrual time.

The model proposed in this thesis handles the case of partially set forwards naturally and also provides an alternative insight into the decay characteristics of implied volatility within the accrual period. The model presented is a factor model of the daily SOFR forward rates. Therefore the entire term structure of daily SOFR is available at any forward simulation point without additional simulation cost. The appropriate dynamics are embedded in the partially set term forwards by evolving each SOFR forward rate up to its observation/accrual time. As shown in Figure 5.5, setting a constant sigma and removing the indicator functions in the HJM volatility function results in a linearly decaying implied volatility. Although the result is consistent with Lyashenko and Mercurio (2019), it is directly reflective of model behaviour rather than derived to improve the accrual period behaviour of term structure models.

The model proposed in this thesis assumes that all forward rate volatility is driven by FOMC meetings and therefore there is zero volatility in the period between the end of the accrual period and the last FOMC meeting within the accrual period. The insight from this modelling set-up, as shown in Figure 5.5, results in an accelerating decay in implied volatility to zero at the final meeting date before the end of the accrual period. This analysis does ignore the volatility of the SOFR to policy target rate spread. However, empirical spread variance is relatively small suggesting the implication on the accrual period dynamics is likely to be accurate.

This dynamic can also be tested against the behaviour of options on 1M SOFR which began listing in early 2022. These options expire just before the end of the accrual period and therefore one should observe a drop in implied volatility as the expiry rolls over an FOMC date. However, at this point there is no liquidity in this market. Therefore prices are not subject to price discovery and as such testing is not pursued for now.

5.3.3 Mean reversion

Mean reversion is embedded in the model in the definition of $\sigma_j(t,T)$ which can be rewritten as follows:

$$\sigma_j(t,T) = \sigma_j(t)\sqrt{v_j(t)}\exp\bigg(-\int_t^T \lambda_j(s)ds\bigg)\sum_{i=1}^{n}\mathbb{I}_{\{i\leq\mathcal{A}(t,T)\}}\gamma_{i,j} \tag{5.44}$$

The mean reversion is reflected in the term $\exp\bigg(-\int_t^T \lambda_j(s)ds\bigg)$ and is a function of $(T-t)$. The $\gamma_{i,j}$ vector scales the volatility function based on the number of FOMC meetings between t and T and as such has an inherent dependence on $T-t$. Therefore for a given $\lambda_j(s)$ function, it is possible to define $\gamma_{i,j}$ such that:

$$\sum_{i=1}^{n}\mathbb{I}_{\{i\leq\mathcal{A}(t,T)\}}\gamma_{i,j} \approx \exp\bigg(-\int_t^T \lambda_j(s)ds\bigg) \tag{5.45}$$

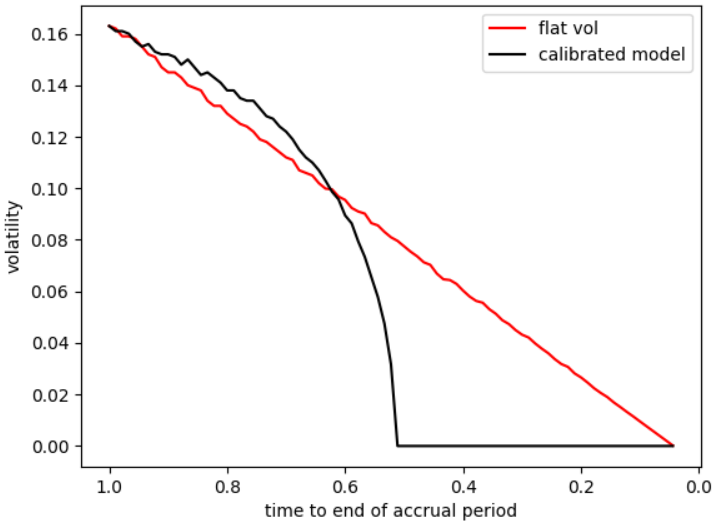

Figure 5.5: Accrual period term volatility comparing a flat volatility assumption to the calibrated model

That is, it is possible to set $\lambda_j = 0$ and instead mimic mean-reverting dynamics with the appropriate choice of $\gamma_{i,j}$. In Chapter 4, $\gamma_{i,j}$ is derived from the PCA of forward states implied from SOFR and Fed Funds futures. The states are derived assuming a piecewise flat structure between FOMC dates without any assumptions regarding the driving dynamics, in turn allowing for empirical assessment of the state dynamics. The first factor, i.e for $j = 1$, explains a large proportion (around 80%) of the forward state variance. It has a clear economic interpretation of focusing forward rate dynamics on the changing expectations related to the change in policy rate at the FOMC date immediately following t. This in itself is a powerful and intuitively agreeable insight; forward rate dynamics are largely driven by changing expectations of the next move in the policy rate. However, critical to the mean-reverting behaviour is that $\gamma_{i,j}$ has the opposite sign between $\gamma_{1,1}$ and $\gamma_{i,1}$ for $i > 1$.

Inspection of the empirically derived $\gamma_{i,j}$ vector for $j = 1$, reveals that it is now possible to choose $\lambda_j(s)$ such that:

$$\exp\left(-\int_t^T \lambda_j(s)ds \right) \approx \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(t,T)\}} \gamma_{i,j} \tag{5.46}$$

by setting:

$$\lambda_j(s) = \begin{cases} 0.9, & s - t < 0.5 \\ 0.08, & \text{otherwise} \end{cases} \tag{5.47}$$

which results in the comparison shown in Figure 5.6, demonstrating it how it is possible to replicate the calibrated $\gamma_{i,j}$ vector with an appropriate choice of $\lambda_j(s)$. It is clear that most

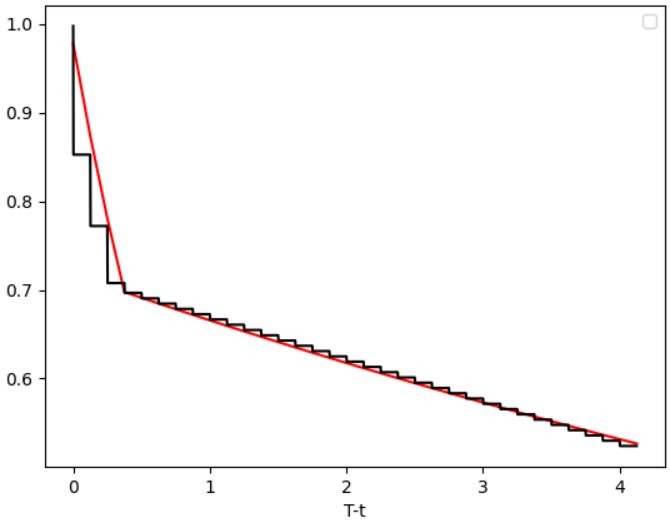

Figure 5.6: Comparison of $\exp\Big(-\int_t^T \lambda_j(s)ds\Big)$(red) and $\sum_{i=1}^{n} \mathbb{I}_{\{i\leq\mathcal{A}(t,T)\}}\gamma_{i,j}$(black)

of the difference stems from the continuous and piecewise definitions, but both approaches are very similar in terms of embedding dynamics which can be interpreted as mean reversion.

It is evident that the empirically derived $\gamma_{i,j}$ imposes mean-reverting dynamics in the model. Therefore, the estimation of the proposed model creates an implicit connection between forward rate dynamics driven by the next policy rate change and the mean-reverting behaviour of the short rate. That is mean reversion is detectable in the correlation structure of the expectation of FOMC policy rate changes.

5.4 Calibration to options

The ability to calibrate to cross-sectional option data is an important feature of interest rate models. Although it violates the time consistency of the model, it is standard industry practice to recalibrate interest rate models daily to vanilla instruments such as swaptions and caps. Although calibration using options on futures is less common, options on SOFR futures have been some of the first SOFR-related option instruments to trade since the inception of the new benchmark. It is also the only SOFR-related option traded directly on an exchange, meaning that the price information is widely available for research purposes.

At the time of writing, most of the market liquidity in options on SOFR futures is concentrated on the front four options on 3M SOFR futures. This set of options is coincident with the empirical research in the previous chapters, which focused on the underlying 3M SOFR futures. Arguably, shorter expiry interest rate options are the most difficult to calibrate due to steep and highly variable term structures in implied volatilities and implied kurtosis as is evident in the data set used for this section, making it a good test set to assess

the model's calibration capability.

For the calibration, take the γ parameters from the empirical estimation performed in Chapter 4. The remaining parameters σ, α, λ, θ and ρ are calibrated to option prices. In the calibration, σ controls the general level of volatility, the mean-reversion parameter λ gives some control of volatility levels across expiries. α controls the level of kurtosis, the stochastic volatility mean reversion parameter α gives some control of kurtosis across expiries. The correlation parameter ρ controls the implied volatility skew.

Each of the model calibration parameters can be defined as a function of time. Combined with the ability to choose the number of HJM factors, this provides significant flexibility in the model for calibration. Begin by performing the calibration with the parameters constant across time before adding time-dependent parameters. The results are presented as a comparison of normal volatilities implied from the bid/offer prices based on settlement price information on the 10-June-2022 with the calibrated model 5% confidence interval based on simulation results.

The calibration results shown in Figure 5.7, show that the model can calibrate general volatility levels, skew and convexity. However, it does not exactly match market-implied volatilities. A main feature of market implied volatilities is the sharply declining convexity as a function of expiry. Another feature is the term structure in skew slightly declining as a function of expiry. As can be seen in Table 5.1, with constant parameters the calibration focuses on α_0, which is the stochastic volatility parameter associated with the first factor. This understates the convexity on the first expiry and overstates for the longest expiry, thus effectively freezing factor 2 and 3 stochastic volatility (α_1, α_2) at zero. These results suggest the introduction of time-dependent stochastic volatility parameters.

Define the parameters of the stochastic volatility as a function of t, piecewise constant between the option expiry dates. As shown in Figure 5.8 this change provides enough flexibility across different expiries to result in a large improvement in calibration performance. With the added time dependence flexibility, the first stochastic volatility parameter α_0 has increased for short expiries and decreased for longer expiries to improve the results obtained with constant parameters.

This example demonstrates the flexibility of the model. Based on the calibration with constant parameters, one can make an informed choice concerning which parameters could be made time-dependent to benefit the calibration. Only three of the fifteen available parameters were changed to achieve substantially better calibration results, albeit for a limited set of calibration instruments. This same approach could be repeated for a larger set of more traditional calibration instruments, making other parameters time-dependent or increasing the number the factors if required.

5.5 Conclusion

The model proposed in this chapter is an outcome of a data-driven approach focused on the new SOFR benchmark. Primarily it accounts for the piecewise constant nature of SOFR, an ability inherited from the previous version of the model. Analysis of the factor dynamics implied by the piecewise constant model led to a stochastic volatility extension

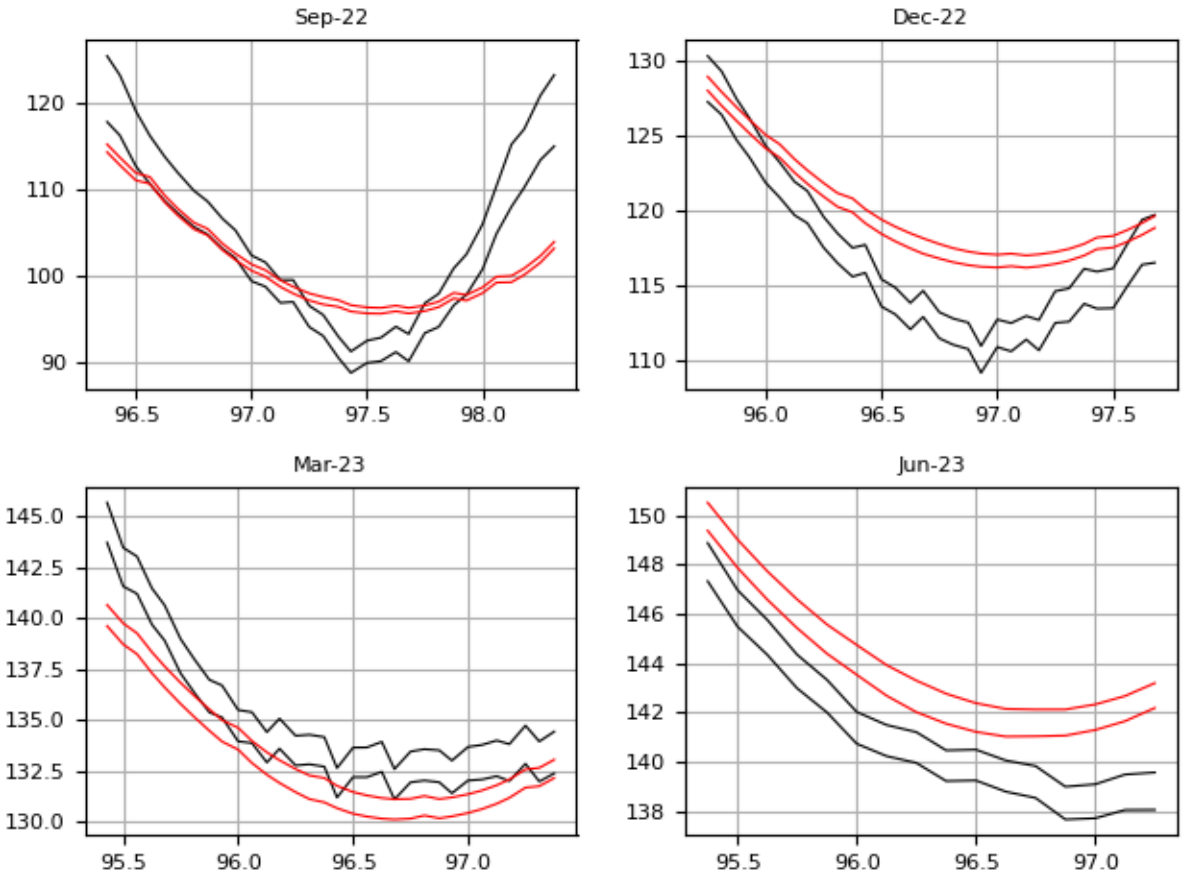

Figure 5.7: Calibration results using constant model parameter for options expiring on 4 different dates. Comparing market prices bid and offer (black) and calibrated model 95% Monte Carlo confidence interval (red)

utilising Heston/Hull-White inspired stochastic volatility dynamics for each HJM factor. These empirically inspired features also allowed the model to be calibrated to interest rate options across different expiries, forward times and strikes.

Final remarks and a summary of the thesis are provided in the next section.

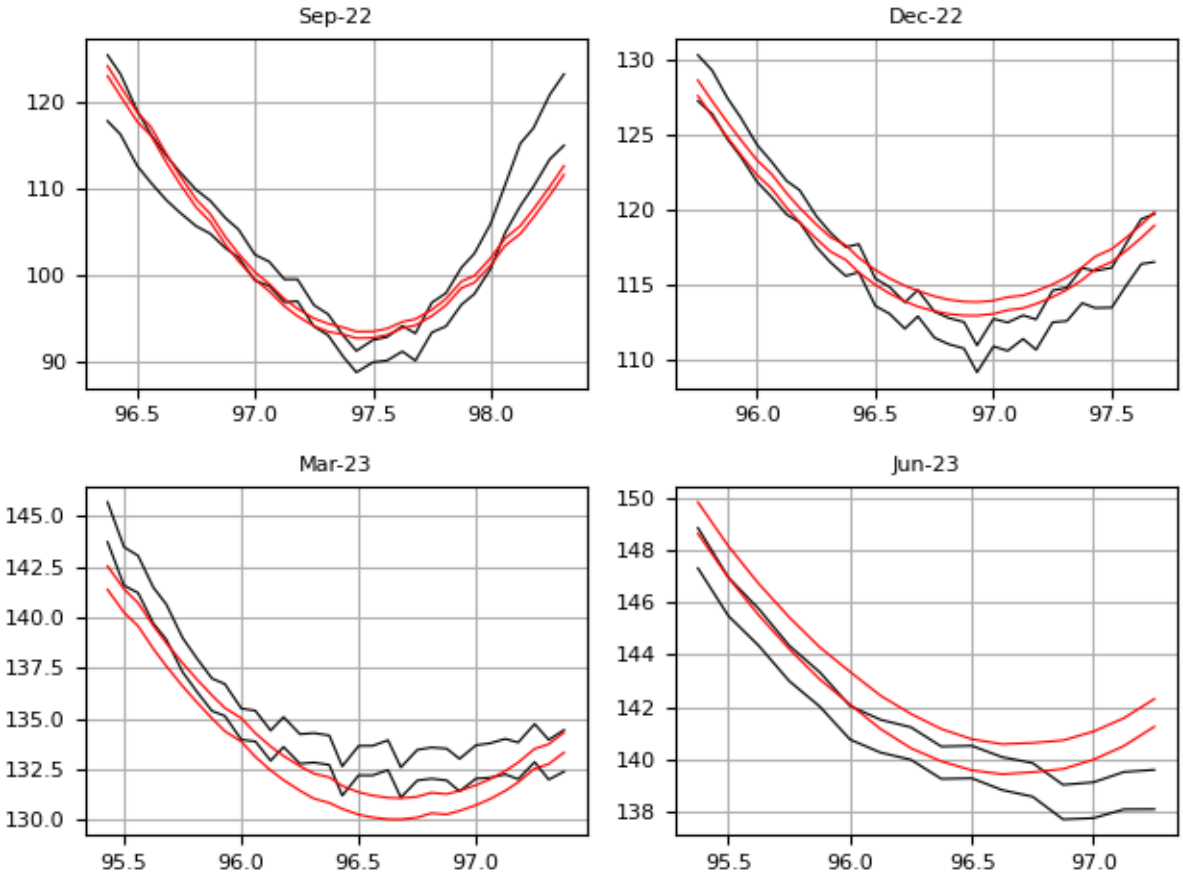

Figure 5.8: Calibration results using time dependent model parameters for options expiring on 4 different dates. Comparing market prices bid and offer (black) and calibrated model 95% Monte Carlo confidence interval (red)

parameters	constant	time dependent
σ_1	0.0081	0.00663
σ_2	0.006	0.00587
σ_3	0.0041	0.00447
λ_1	0.01	0.02
λ_2	0.0	0.004
λ_3	0.16	0.35
α_1	1.57	[3.142, 1.35, 3.2, 0.86]
α_2	0.82	[0.76, 0.66, 0.6, 0.22]
α_3	0.0	[3.0, 0.6, 0.5, 4.1]
ρ_1	-0.2	-0.14
ρ_2	0.0	-0.025
ρ_3	0.0	-0.83
θ_1	0.0	0.1
θ_2	0.0	0.0
θ_3	0.0	11.0

Table 5.1: Calibrated models using showing constant and time dependant parameters

Chapter 6

Final Summary

The research presented in this thesis began with the development of a particle filter capable of adapting to regime changes and detecting stochastic volatility levels. The methodology was arrived at by first recognising that the random perturbation technique applied in a particle filter results in a genetic–type algorithm capable of adapting to changing parameters. At this point, an opposite direction to the approach of (Liu and West 2001) was taken, instead of remediating the over-dispersion caused by random perturbation, the random perturbation is allowed to freely evolve, enhancing the adaptive capability of the particle filter. The approach is highly adaptive when required and convergent conditional on the data matching modelling assumptions and no parameter changes. Given that the level of adaptability is governed by the variance of the random perturbation; the key insight of the approach is that an effective way of recognising the level of required variance is to incorporate its selection into the already existing genetic algorithm framework. In terms of existing literature, it links particle filtering with genetic algorithms for parameter learning, resulting in a filtering algorithm particularly useful for parameter change detection and in the context of finance an effective on-line method for measuring volatility.

The motivation behind the particle filter was to create a tool which would inform modelling choices in a data-driven way. The original plan involved analysing the LIBOR rate and its derivatives, mainly Eurodollar futures and options on those futures. Exploratory analysis, employing the particle filter, combined with the emergence of SOFR as the new benchmark rate led to a refinement in the understanding of data-driven research in finance. Financial data does not necessarily follow any particular model, instead, it has particular statistical features which can be replicated by models. Therefore the research on which this thesis is based takes the approach of identifying empirical features in SOFR and proposing a model whose primary features reflect those identified in empirical data.

The key empirical behaviour motivating the first version of the model is the piecewise constant trajectory of short rates. Simultaneously, forward rates are found to behave in a mostly diffusive manner. Incorporating this structure into a model and calibrating this model to SOFR futures yielded empirical state dynamics. The state dynamics based on daily calibration exhibit clear leptokurtosis, motivating a stochastic volatility extension of the model. Consequently, stochastic volatility with the inherent correlation between the

volatility and state factors allows for effective calibration to options on SOFR futures.

The piecewise behaviour of short rates can be attributed to the setting of the policy target rate by the FOMC and the intervention by the Federal Reserve which forces both EFFR and SOFR to closely follow the target rate. This is a contemporary aspect of short-rate behaviour emerging following the financial crisis in 2009, before which the short rate could be observed to evolve in a much more diffusive manner. Although initially, SOFR did exhibit significant noise, including pronounced spikes, since 2021 has also predominantly followed the target rate.

On the other hand, the behaviour of SOFR forward rates cannot be directly observed, instead, it is inferred from SOFR futures. The observed diffusive behaviour of forward rates led to the following key insight. While short rates follow the target rate and are therefore mostly piecewise constant, the market expectations related to those changes evolve in a diffusive manner. Forward rates reflect market expectations of short rates in the future and therefore inherit the continuously diffusive behaviour. The model presented in this paper is based on the idea that the expectations related to each FOMC date are each driven by a separate factor. In this modelling setup, the forward rate term structure is cast into a correlated system of FOMC date-specific diffusive factors, which naturally yields piecewise constant short rates. The first iteration of the model aims to succinctly demonstrate this concept.

The first iteration of the model achieves the desired behaviour by defining the HJM volatility with indicator functions which act to assign each HJM volatility to a specific FOMC date. Each forward rate segment between FOMC dates evolves separately. As a result, discontinuities are created at the segment endpoints, i.e. the FOMC dates. The short rate inherits the discontinuous structure from this system. An embedded assumption of this approach is that on the FOMC date the short rate is fully determined by the forward rate structure, i.e. all information about the move is embedded in the forward rate. A justification for this assumption is related to the timing of the information transfer from the actual policy target rate change: when the new policy target rate is announced the affected short rate does not happen until the next day, therefore at announcement time the market is still trading the effected rate as a forward rate.

The next iteration of the model is an adaption to facilitate empirical estimation. The indicator functions are changed such that each factor is related to FOMC dates based on their order from the current state time. This allowed for time-consistent estimation as empirical dynamics can be related to each other regardless of the time frame. This adaptation does not change the nature of the model, rather it can be considered a remapping of the relationship between factors and FOMC dates.

The final version of the model was motivated by the observation of leptokurtosis in the estimated factor states. Leptokurtosis can be produced by many modelling approaches, the choice of stochastic volatility presents as the most tractable and pragmatic choice. This facilitates the calibration to options across different expiries, forward rate periods and strikes. The ability of the combined model to calibrate to options is demonstrated on options on 3M SOFR futures.

The model presented aims to replicate only three features of interest rates: piecewise constant short rates, continuously diffusive forward rates and leptokurtosis in the forward

rate process. However, linking the model directly to FOMC dates[1] has yielded several key results and insights, ranging from economic interpretation of the estimated model factors, the ability to extract predictive information content related to FOMC policy rate changes from market data, calibration to options and insight regarding accrual period behaviour and mean reversion.

The factorisation of the calibrated model using PCA yields a structure where each factor element corresponds to the order of FOMC meeting dates. Similarly to the more common approach of factorising the forward rates directly, the first factor accounts for around 80% of the system variance. The characteristic of the first factor is that is dominated by the first element which corresponds to the next (relative to current state time) FOMC date. Therefore the dominant element in the dominant factor relates directly to the upcoming FOMC date. In the model, forward rates are obtained through an aggregation of policy rate changes up to the forward time, therefore a factor with a dominant first element corresponds to a parallel shift in the forward rate term structure. A parallel shift is known to be the primary factor in a PCA decomposition of forward rates. Therefore the model presented in this paper links the predominant factor known in interest rates to expectations related to the next FOMC meeting.

Another key result is the ability of the model to extract predictive information regarding policy rate changes from SOFR and Fed Funds futures. Predictive information embedded in market-traded instruments is not surprising, part of the FOMC strategy is to carefully manage market expectations. By modelling those expectations directly, the model extracts this information resulting in a high degree of prediction accuracy. The model itself does not perform the prediction, rather it extracts the predictive information already embedded in futures contracts. Capturing this information should be important for pricing and managing the risk of derivative instruments. If features of a model do not match well to empirical behaviour, implied prices and risk management reflect model-induced artefacts which deviate from empirical reality.

The calibration has also been shown to reflect the empirical target rate to SOFR spread without this spread being used as an input into the calibration algorithm. Similarly to predicting FOMC policy rate changes, this means that the information regarding the spread is extracted from futures prices by the calibration. This result constitutes further evidence that the modelling setup, specifically the embedded forward rate structure is successful at extracting information from market instruments reflecting empirical data of the underlying SOFR.

Another more direct aspect of calibrating to futures is the ability of the model to recover the actual futures prices. The model proposed in this paper does give up some calibration accuracy in exchange for extracting more accurate empirical information. However, for a three-factor model, it compares favourably to the recently published research by Skov and Skovmand (2021), particularly for short-expiry futures. It would be possible to reduce the calibration error to futures further with the addition of spike factors.

In the context of cross-sectional calibration, stochastic volatility allows the model to calibrate to skewness and convexity across strikes, a prominent feature in interest rate options

[1] Arguably the most important set of dates in the interest rate calendar.

prices. Time-dependent parameters inherited from Heston/Hull-White and dependence on forward time inherited from the model proposed in Chapter 3 and Chapter 4 combine to provide flexibility in the calendar time and forward time dimension. Hence, the model definition produces a large degree of flexibility for calibration to interest rate options. This was demonstrated on options on SOFR futures. However, the model could be adapted to other calibration instruments such as caps and swaptions.

The data-driven approach has resulted in a model directly connected to FOMC meeting dates, the most important regular economic event in US interest rate options markets. Arguably, other significant economic events and data funnel into FOMC policy rate decisions as well as expectations of the decisions reflected in derivative pricing. The connection to FOMC meeting dates is made through a component in the HJM volatility function inducing piecewise short rates, a key feature of the model. The calibration of this feature to a history of SOFR futures prices has revealed a connection between interest rate mean reversion and FOMC policy rate expectations.

The primary driver of SOFR futures prices are changing expectations related to the next FOMC meeting which in turn tends to be negatively correlated with changes in expectations to subsequent meetings creating variance decay as a function of forward time. Historically, the Federal Reserve in managing economic cycles acts to mean revert interest rates. The market expects the Federal Reserve to continue to act this way and using our modelling set up this expectation is detectable in the evolution of SOFR futures prices.

The approach in this thesis has been to allow empirical data to inform the modelling choices. The approach aimed to identify and create a model which reflects key empirical features of SOFR. The result is a model which has yielded more than the original goal. The model accommodates cross-sectional calibration arguably better than leading industry models in addition to providing genuine economic insights related to the evolution of interest rates.

Bibliography

Alfeus, M., Grasselli, M. and Schlögl, E.: 2020, A Consistent Stochastic Model of the Term Structure of Interest Rates for Multiple Tenors, *Journal of Economic Dynamics and Control* **114**.

Andersen, L. B. G. and Bang, D. R. A.: 2020, Spike Modelling for Interest Rate Derivatives with an Application to SOFR Caplets. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3700446.

Andrieu, C., Doucet, A., Singh, S. S. and Tadic, V. B.: 2004, Particle Methods for Change Detection, System Identification, and Control, *Proceedings of the IEEE* **92**(3), 423–438.

Andrieu, C., Doucet, A. and Tadic, V. B.: 2005, On-Line Parameter Estimation in General State-Space Models, *44th IEEE Conference on Decision and Control*, IEEE, pp. 332–337.

Backwell, A. and Hayes, J.: 2022, Expected and unexpected jumps in the overnight rate: Consistent management of the libor transition, *Journal of Banking and Finance* **145**.

Backwell, A., Macrina, A., Schlögl, E. and Skovmand, D.: 2019, Term Rates, Multicurve Term Structures and Overnight Rate Benchmarks: a Roll–Over Risk Approach. Available at SSRN: https://ssrn.com/abstract=3399680.

Bao, Y., Chiarella, C. and Kang, B.: 2012, Particle Filters for Markov Switching Stochastic Volatility Models, *Technical report*, Quantitative Finance Research Centre, University of Technology, Sydney.

Berndt, A., Duffie, D. and Zhu, Y.: 2020, Across–the–Curve Credit Spread Indices, *Technical report*, Australian National University and Stanford University.

Binder, A. S.: 2010, Quantitative Easing: Entrance and Exit Strategies, *Federal Reserve Bank of St. Louis Review* **92**(6), 465–79.

Brace, A., Gatarek, D. and Musiela, M.: 1997, The market model of interest rate dynamics, *Mathematical Finance* **7**(2), 127–155.

Brace, A., Gellert, K. and Schlögl, E.: 2022, Sofr term structure dynamics - discontinuous short rates and stochastic volatility forward rates. Available at SSRN: https://ssrn.com/abstract=4270811.

Cappé, O., Godsill, S. J. and Moulines, E.: 2007, An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo, *Proceedings of the IEEE* **95**(5), 899–924.

Carmona, R. A.: 2007, *HJM: A unified approach to dynamic models for fixed income, credit and equity markets*, Springer, Berlin, Heidelberg.

Carvalho, C. M., Johannes, M. S., Lopes, H. F. and Polson, N. G.: 2010, Particle Learning and Smoothing, *Statistical Science* **25**(1), 88–106.

Carvalho, C. M. and Lopes, H. F.: 2007, Simulation-based sequential analysis of Markov switching stochastic volatility models, *Computational Statistics & Data Analysis* **51**(9), 4526–4542.

Casarin, R.: 2004, Bayesian Monte Carlo Filtering For Stochastic Volatility Models, *Technical report*, University Paris Dauphine.

Chen, Z. et al.: 2003, Bayesian filtering: From Kalman Filters to Particle Filters, and Beyond, *Statistics* **182**(1), 1–69.

Chopin, N., Iacobucci, A., Marin, J.-M., Mengersen, K., Robert, C. P., Ryder, R. and Schäfer, C.: 2011, On Particle Learning, *Bayesian Statistics 9* .

Chopin, N. x.: 2004, Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference, *The Annals of Statistics* **32**(6), 2385–2411.

Cochran, W. G.: 1934, The Distribution of Quadratic Forms in a Normal System, with Applications to the Analysis of Covariance, *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 30, Cambridge University Press, pp. 178–191.

Cox, J. C., Ingersoll, J. E. and Ross, S. A.: 1981, The Relation Between Forward Prices and Futures Prices, *Journal of Financial Economics* **9**, 321–346.

Cox, J. C., Ingersoll, J. E. and Ross, S. A.: 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* **53**(2), 385–407.

Creal, D.: 2012, A survey of sequential Monte Carlo methods for economics and finance, *Econometric reviews* **31**(3), 245–296.

Crisan, D. and Doucet, A.: 2002, A Survey of Convergence Results on Particle Filtering Methods for Practitioners, *IEEE Transactions on signal processing* **50**(3), 736–746.

Del Moral, P.: 2004, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, New York: Springer-Verlag.

Del Moral, P.: 2013, *Mean Field Simulation for Monte Carlo Integration*, Chapman-Hall/CRC Press.

Del Moral, P. and Doucet, A.: 2014, Particle Methods: An introduction with Applications, *ESAIM: Proceedings*, Vol. 44, EDP Sciences, pp. 1–46.

Djuric, P. M. and Míguez, J.: 2010, Assessment of Nonlinear Dynamic Models by Kolmogorov–Smirnov Statistics, *IEEE Transactions on Signal Processing* **58**(10), 5069–5079.

Douc, R. and Moulines, E.: 2007, Limit Theorems for Weighted Samples with Applications to Sequential Monte Carlo Methods, *The Annals of Statistics* **36**, 2344–2376.

Doucet, A. and Johansen, A. M.: 2009, A tutorial on Particle Filtering and Smoothing: Fifteen years later, *Handbook of Nonlinear Filtering* **12**(656-704), 3.

Federal Open Market Committee: 2000-2020, Minutes of the Federal Open Market Committee.

Flury, T. and Shephard, N.: 2009, Learning and Filtering via Simulation: Smoothly Jittered Particle Filters, *Technical report*, Department of Economics Discussion Paper Series, University of Oxford.

Fontaine, J.-S.: 2016, What fed funds futures tell us about monetary policy uncertainty, *Technical Report 2016–61*, Bank of Canada Staff Working Paper.

Fontana, C., Grbac, Z., Gümbel, S. and Schmidt, T.: 2020, Term structure modelling for multiple curves with stochastic discontinuities, *Finance and Stochastics* **24**(2), 465–511.

Gellert, K. and Schlögl, E.: 2021a, Parameter learning and change detection using a particle filter with accelerated adaptation, *Risks* **9**(12).

Gellert, K. and Schlögl, E.: 2021b, Short rate dynamics: A fed funds and sofr perspective. Available at SSRN: https://ssrn.com/abstract=3763589.

Gerber, M. and Chopin, N.: 2015, Sequential Quasi-Monte Carlo, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(3), 509–579.

Gordon, N. J., Salmond, D. J. and Smith, A. F.: 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F*, Vol. 140, IET, pp. 107–113.

Grzelak, L., Oosterlie, C. W. and Weeran, S. V.: 2008, Extension of stochastic volatility equity models with the hull white interest rate process, *Quantitative Finance* **12**(1).

Hagan, P. S., Kumar, D., Lesniewski, A. S. and Woodward, D. E.: 2002, Managing smile risk, *The Best of Wilmott* **1**, 249–296.

Heath, D., Jarrow, R. and Morton, A.: 1992, Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation, *Econometrica* **60**(1), 77–105.

Heitfield, E. and Park, Y.-H.: 2019, Inferring term rates from SOFR futures prices. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3352598.

Heston, S. L.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* **6**, 327–343.

Hilton, S.: 2005, Trends in Federal Funds Rate Volatility, Federal Reserve Bank of New York, *Current Issues in Economics and Finance* **11**(7).

Hol, J. D., Schon, T. B. and Gustafsson, F.: 2006, On Resampling Algorithms for Particle Filters, *Nonlinear Statistical Signal Processing Workshop*, IEEE, pp. 79–82.

Hull, J. and White, A.: 1990, Pricing interest–rate derivative securities, *Review of Financial Studies* **3**(4), 573–592.

Johannes, M. and Polson, N.: 2009, Particle filtering, *Handbook of Financial Time Series* pp. 1015–1029.

Johannes, M. S. and Polson, N.: 2007, Particle Filtering and Parameter Learning, *Technical report*, Chicago Business School.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J. and Chopin, N.: 2015, On Particle Methods for Parameter Estimation in State-Space Models, *Statistical Science* **30**(3), 328–351.

Kantas, N., Doucet, A., Singh, S. S. and Maciejowski, J. M.: 2009, An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models, *IFAC Proceedings Volumes* **42**(10), 774–785.

Karlsson, P., Pilz, K. and Schlögl, E.: 2017, Calibrating a market model with stochastic volatility to commodity and interest rate risk, *Quantitative Finance* **17**(6).

Keller-Ressel, M., Schmidt, T. and Wardenga, R.: 2018, Affine processes beyond stochastic continuity, *Annals of Applied Probability* **29**(6), 3387–3437.

Kim, D. H. and Wright, J. H.: 2014, Jumps in bond yields at known times, *National Bureau of Economic Research* (No. w20711).

Krueger, J. T. and Kuttner, K. N.: 1996, The fed funds futures rate as a predictor of federal reserve policy, *The Journal of Futures Markets* **16**(8), 865–879.

Li, T., Bolic, M. and Djuric, P. M.: 2015, Resampling Methods for Particle Filtering: Classification, implementation, and strategies, *IEEE Signal Processing Magazine* **32**(3), 70–86.

Liu, J. and West, M.: 2001, Combined parameter and state estimation in simulation-based filtering, *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, pp. 197–223.

Lopes, H. F. and Tsay, R. S.: 2011, Particle Filters and Bayesian Inference in Financial Econometrics, *Journal of Forecasting* **30**(1), 168–209.

Lyashenko, A. and Mercurio, F.: 2019, Looking forward to backward-looking rates: A modeling framework for term rates replacing libor. Available at SSRN: https://ssrn.com/abstract=3330240.

Mercurio, F.: 2018, A simple multi-curve model for pricing SOFR futures and other derivatives. Available at SSRN: https://ssrn.com/abstract=3225872.

Merton, R. C.: 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* pp. 141–183.

Miltersen, K. R., Sandmann, K. and Sondermann, D.: 1997, Closed form solutions for term structure derivatives with log-normal interest rates, *The Journal of Finance* **52**(1), 409–430.

Musiela, M. and Rutkowski, M.: 1997, Continuous–time term structure models: A forward measure approach, *Finance and Stochastics* **1**(4).

Nelson, C. and Siegel, A. F.: 1987, Parsimonious modelling of yield curves, *Journal of Business* **60**, 473–489.

Nemeth, C., Fearnhead, P. and Mihaylova, L.: 2014, Sequential Monte Carlo Methods for State and Parameter Estimation in Abruptly Changing Environments, *IEEE Transactions on Signal Processing* **62**(5), 1245–1255.

Nemeth, C., Fearnhead, P., Mihaylova, L. and Vorley, D.: 2012, Particle Learning Methods for State and Parameter Estimation, *9th Data Fusion and Target Tracking Conference*, IET.

Piazzesi, M.: 2001, An econometric model of the yield curve with macroeconomic jump effects, *Technical report*, National Bureau of Economic Research.

Piazzesi, M.: 2005, Bond yields and the federal reserve, *Journal of Political Economy* **113**(2).

Piterbarg, V.: 2015, *A stochastic volatility forward Libor model with a term structure of volatility smiles*. Available at SSRN: https://ssrn.com/abstract=359001.

Polson, N. G., Stroud, J. R. and Müller, P.: 2008, Practical Filtering with Sequential Parameter Learning, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(2), 413–428.

Robertson, J. C. and Thornton, D. L.: 1997, Using federal funds futures rates to predict federal reserve actions, *Federal Reserve Bank of St. Louis Review* pp. 45–53.

Schlögl, E. and Sommer, D.: 1998, Factor models and the shape of the term structure, *The Journal of Financial Engineering* **7**(1), 79–88.

Schrimpf, A. and Sushko, V.: 2019, Beyond LIBOR: a primer on the new reference rates,[w:] international banking and financial market developments, *BIS Quarterly Review* .

Skov, J. B. and Skovmand, D.: 2021, Dynamic Term Structure Models for SOFR Futures, *Journal of Futures Markets* **41**(10).

Smith, R. and Hussain, M. S.: 2012, Genetic Algorithm Sequential Monte Carlo Methods For Stochastic Volatility And Parameter Estimation, *Proceedings of the World Congress on Engineering*, Vol. 1.

Storvik, G.: 2002, Particle filters for state space models with the presence of unknown static parameters, *IEEE Transactions on signal Processing* **50**(2), 281–289.

The Alternative Reference Rates Committee: 2018, Second report.

Vasicek, O.: 1977, An equilibrium characterization of the term structure, *Journal Of Financial Economics* **5**(2), 177–188.

West, M.: 1993a, Approximating Posterior Distributions by Mixtures, *Journal of the Royal Statistical Society Series B* **55**(2), 409–422.

West, M.: 1993b, Mixture models, Monte Carlo, Bayesian Updating and Dynamic Models, *Computing Science and Statistics*, Vol. 24, pp. 325–333.

Yang, X., Xing, K., Shi, K. and Pan, Q.: 2008, Joint State and Parameter Estimation in Particle Filtering and Stochastic Optimization, *Journal of Control Theory and Applications* **6**(2), 215–220.

Appendix A

For a Brownian motion $W(t)$:

$$\int_0^t \mathbb{1}(s < x) dW(s) = W(t \wedge x) \tag{A.1}$$

Therefore we have the following solutions to various stochastic integrals appearing in this paper:

$$\int_0^t \xi_i \mathbb{1}(s < x_i) \mathbb{1}(T \geq x_i) dZ_i(s) = \xi_i \mathbb{1}(T \geq x_i) Z_i(t \wedge x_i) \tag{A.2}$$

$$\int_0^t \xi_i \mathbb{1}(s < x_i) \mathbb{1}(t \geq x_i) dZ_i(s) = \xi_i \mathbb{1}(t \geq x_i) Z_i(x_i) \tag{A.3}$$

$$\int_0^t \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(s < x_i) \mathbb{1}(T \geq x_i) dW_j^P(s) = \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(T \geq x_i) W_j^P(t \wedge x_i) \tag{A.4}$$

$$\int_0^t \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(s < x_i) \mathbb{1}(t \geq x_i) dW_j^P(s) = \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(t \geq x_i) W_j^P(x_i) \tag{A.5}$$

$$\int_0^t \sigma_i^Z \mathbb{1}(s < z_i) \mathbb{1}(T \in H_i) dW_i^Z(s) = \sigma_i^Z \mathbb{1}(T \in H_i) W_i^Z(t \wedge z_i) \tag{A.6}$$

$$\int_0^t \sigma_i^Z \mathbb{1}(s < z_i) \mathbb{1}(t \in H_i) dW_i^Z(s) = \sigma_i^Z \mathbb{1}(t \in H_i) W_i^Z(z_i) \tag{A.7}$$

Solving the drift term $\int_0^t \sigma_j^P(u,T) \int_u^T \sigma_j^P(u,s)dsdu$, we have:

$$\begin{aligned}\int_u^T \sigma_j^P(u,s)ds &= \int_u^T \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(u < x_i)\mathbb{1}(s \geq x_i)ds \\ &= \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(u < x_i) \int_u^T \mathbb{1}(s \geq x_i)ds\end{aligned} \tag{A.8}$$

where:

$$\int_u^T \mathbb{1}(s \geq x_i)ds = \begin{cases} 0, & T < x_i \\ T - x_i, & u < x_i, T \geq x_i \\ T - u, & u \geq x_i \end{cases} \tag{A.9}$$

therefore:

$$\int_u^T \sigma_j^P(u,s)ds = \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(u < x_i)\mathbb{1}(T \geq x_i)(T - x_i) \tag{A.10}$$

therefore:

$$\begin{aligned}&\int_0^t \sigma_j^P(u,T) \int_u^T \sigma_j^P(u,s)dsdu \\ &= \int_0^t \sum_{q=1}^n \xi_q \lambda_{q,j} \mathbb{1}(u < x_q)\mathbb{1}(T \geq x_q) \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(u < x_i)\mathbb{1}(T \geq x_i)(T - x_i)du \\ &= \sum_{q=1}^n \xi_q \lambda_{q,j} \mathbb{1}(T \geq x_q) \sum_{i=1}^n \xi_i \lambda_{i,j} \mathbb{1}(T \geq x_i)(T - x_i) \int_0^t \mathbb{1}(u < x_q)\mathbb{1}(u < x_i)du \\ &= \sum_{q=1}^n \sum_{i=1}^n \xi_q \xi_i \lambda_{q,j} \lambda_{i,j} \mathbb{1}(T \geq x_{q \vee i})(T - x_i)[t \wedge x_q \wedge x_i]\end{aligned} \tag{A.11}$$

Similarly:

$$\int_0^t \sigma_j^P(u,t) \int_u^t \sigma_j^P(u,s)dsdu = \sum_{q=1}^n \sum_{i=1}^n \xi_q \xi_i \lambda_{q,j} \lambda_{i,j} \mathbb{1}(t \geq x_{q \vee i})(t - x_i)[x_q \wedge x_i] \tag{A.12}$$

Solving $\int_0^t \sigma_i^Z(u,T) \int_u^T \sigma_i^Z(u,s) ds du$:

$$\begin{aligned} \int_u^T \sigma_i^Z(u,s) ds &= \int_u^T \sigma_i^Z \mathbb{1}(u < z_i) \mathbb{1}(s \in H_i) ds \\ &= \sigma_i^Z \mathbb{1}(u < z_i) \int_u^T \mathbb{1}(s \in H_i) ds \\ &= \sigma_i^Z \mathbb{1}(u < z_i) \mathbb{1}(T \geq z_i)[h_i \wedge (T - z_i)] \end{aligned} \tag{A.13}$$

therefore

$$\begin{aligned} &\int_0^t \sigma_i^Z(u,T) \int_u^T \sigma_i^Z(u,s) ds du \\ &= \int_0^t \sigma_i^Z \mathbb{1}(u < z_i) \mathbb{1}_{H_i}(T) \sigma_i^Z \mathbb{1}(u < z_i) \mathbb{1}(T \geq z_i)[h_i \wedge (T - z_i)] du \\ &= \left(\sigma_i^Z\right)^2 \mathbb{1}(T \in H_i)[h_i \wedge (T - z_i)] \int_0^t \mathbb{1}(u < z_i) du \\ &= \left(\sigma_i^Z\right)^2 \mathbb{1}(T \in H_i)(T - z_i)[t \wedge z_i] \end{aligned} \tag{A.14}$$

Similarly:

$$\int_0^t \sigma_i^Z(u,t) \int_u^t \sigma_i^Z(u,s) ds du = \left(\sigma_i^Z\right)^2 \mathbb{1}(t \in H_i)(t - z_i) z_i \tag{A.15}$$

Appendix B

B.1 Single dimensional case

Define the following:

$$\sigma(t,T) = \chi(t)\phi(T) \tag{B.1}$$

$$\phi(T) = \exp\Big(- \int_0^T \lambda(v)dv \Big) \tag{B.2}$$

$$\chi(t) = \sigma(t)\exp\Big(\int_0^t \lambda(v)dv \Big) \tag{B.3}$$

$$\Lambda(t,T) = \int_t^T \exp\Big(- \int_t^u \lambda(v)dv \Big) du \tag{B.4}$$

$$\Phi(t) = \int_0^t \sigma^2(s)\exp\Big(- 2\int_s^t \lambda(v)dv \Big) ds \tag{B.5}$$

HJM result:

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T) \int_s^T \sigma(s,u) du ds + \int_0^t \sigma(s,T) dW(s) \tag{B.6}$$

$$= f(0,T) + \int_0^t \chi(s)\phi(T) \int_s^T \chi(s)\phi(u) du ds + \int_0^t \chi(s)\phi(T) dW(s) \tag{B.7}$$

$$= f(0,T) + \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) \tag{B.8}$$

$$\times \int_s^T \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^u \lambda(v) dv \Big) du ds \tag{B.9}$$

$$+ \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) dW(s) \tag{B.10}$$

$$= f(0,T) + \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \int_s^T \exp\Big(- \int_s^u \lambda(v) dv \Big) du ds \tag{B.11}$$

$$+ \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.12}$$

$$= f(0,T) + \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda(s,T) ds \tag{B.13}$$

$$+ \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.14}$$

let

$$y(t) = \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda(s,t) ds \tag{B.15}$$

$$+ \exp\Big(\int_t^T \lambda(v) dv \Big) \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.16}$$

substitute

$$\int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.17}$$

$$= \frac{y(t) - \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda(s,t) ds}{\exp\Big(\int_t^T \lambda(v) dv \Big)} \tag{B.18}$$

$$f(t,T) = f(0,T) + \int_0^t \sigma^2(s)\exp\Big(- \int_s^T \lambda(v)dv \Big)\Lambda(s,T)ds \tag{B.19}$$

$$+ \int_0^t \sigma(s)\exp\Big(- \int_s^T \lambda(v)dv \Big)dW(s) \tag{B.20}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv \Big)y(t) \tag{B.21}$$

$$+ \int_0^t \sigma^2(s)\exp\Big(- \int_s^T \lambda(v)dv \Big)\Lambda(s,T)ds \tag{B.22}$$

$$- \exp\Big(- \int_t^T \lambda(v)dv \Big) \int_0^t \sigma^2(s)\exp\Big(- \int_s^t \lambda(v)dv \Big)\Lambda(s,t)ds \tag{B.23}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv \Big)y(t) + \exp\Big(- \int_t^T \lambda(v)dv \Big) \tag{B.24}$$

$$\times \int_0^t \sigma^2(s)\Big\{ \exp\Big(\int_t^T \lambda(v)dv \Big)\exp\Big(- \int_s^T \lambda(v)dv \Big)\Lambda(s,T) \tag{B.25}$$

$$- \exp\Big(- \int_s^t \lambda(v)dv \Big)\Lambda(s,t) \Big\}ds \tag{B.26}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv \Big)y(t) + \exp\Big(- \int_t^T \lambda(v)dv \Big) \tag{B.27}$$

$$\times \int_0^t \sigma^2(s)\exp\Big(- \int_s^t \lambda(v)dv \Big)\Big\{ \Lambda(s,T) - \Lambda(s,t) \Big\}ds \tag{B.28}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv \Big)y(t) + \Lambda(t,T)\exp\Big(- \int_t^T \lambda(v)dv \Big) \tag{B.29}$$

$$\times \int_0^t \sigma^2(s)\exp\Big(- 2\int_s^t \lambda(v)dv \Big)ds \tag{B.30}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv \Big)y(t) + \Phi(t)\Lambda(t,T)\exp\Big(- \int_t^T \lambda(v)dv \Big) \tag{B.31}$$

therefore:

$$\int_t^T f(t,u)du \tag{B.32}$$

$$= \int_t^T \Big(f(0,u) + \exp\Big(- \int_t^u \lambda(v)dv \Big) y(t) \tag{B.33}$$

$$+ \Phi_j(t)\Lambda(t,u)\exp\Big(- \int_t^u \lambda(v)dv \Big) \Big) du \tag{B.34}$$

$$= \int_t^T f(0,u)du + y(t) \int_t^T \exp\Big(- \int_t^u \lambda(v)dv \Big) du \tag{B.35}$$

$$+ \Phi(t) \int_t^T \Lambda(t,u)\exp\Big(- \int_t^u \lambda(v)dv \Big) du \tag{B.36}$$

$$= \int_t^T f(0,u)du + \Lambda(t,T)y(t) + \Phi(t) \int_t^T \Lambda(t,u)d\Lambda(t,u) \tag{B.37}$$

$$= \int_t^T f(0,u)du + \Lambda(t,T)y(t) + \frac{1}{2}\Phi(t)\Lambda^2(t,T) \tag{B.38}$$

Therefore, bond price:

$$B(t,T) = \exp\Big(- \int_t^T f(t,u)du \Big) = \frac{B(0,T)}{B(0,t)} \exp\Big(- \Lambda(t,T)y(t) - \frac{1}{2}\Phi(t)\Lambda^2(t,T) \Big) \tag{B.39}$$

B.2 Single dimensional case with piecewise continuous short rate

Define the following:

$$\sigma(t,T) = \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(t,T)\}} \chi(t)\phi(T)\gamma_i \tag{B.40}$$

$$\phi(T) = \exp\Big(- \int_0^T \lambda(v)dv \Big) \tag{B.41}$$

$$\chi(t) = \sigma(t)\exp\Big(\int_0^t \lambda(v)dv \Big) \tag{B.42}$$

$$\Lambda(t,T) = \int_t^T \exp\Big(- \int_t^u \lambda(v)dv \Big) du \tag{B.43}$$

$$\Lambda_a(t,T) = \int_a^T \exp\Big(- \int_t^u \lambda(v)dv \Big) du \tag{B.44}$$

$$\Phi(t) = \int_0^t \sigma^2(s)\exp\Big(- 2\int_s^t \lambda(v)dv \Big) ds \tag{B.45}$$

B.2.1 Trivial case $t < T < x_1$

HJM result:

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T) \int_t^T \sigma(s,u) du ds + \int_0^t \sigma(s,T) dW(s) = f(0,T) \tag{B.46}$$

$$\int_t^T f(t,u) du = \int_t^T f(0,u) du \tag{B.47}$$

Therefore, bond price:

$$B(t,T) = \exp\Big(- \int_t^T f(t,u) du \Big) = \frac{B(0,T)}{B(0,t)} \tag{B.48}$$

B.2.2 Basic case $t < x_1 < T < x_2$

$$\int_t^T f(t,u) du = \int_t^{x_1} f(t,u) du + \int_{x_1}^T f(t,u) du \tag{B.49}$$

$$= \int_t^{x_1} f(0,u) du + \int_{x_1}^T f(t,u) du \tag{B.50}$$

To solve $\int_{x_1}^T f(t,u) du$, restrict $T \in [x1, x2]$ and $t < x_1$, HJM result:

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T) \int_s^T \sigma(s,u) du ds + \int_0^t \sigma(s,T) dW(s) \tag{B.51}$$

$$= f(0,T) + \int_0^t \sum_{i=1}^n \mathbb{I}_{\{i \leq \mathcal{A}(s,T)\}} \chi(s) \phi(T) \gamma_i \tag{B.52}$$

$$\times \int_s^T \sum_{j=1}^n \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \chi(s) \phi(u) \gamma_j du ds \tag{B.53}$$

$$+ \int_0^t \sum_{i=1}^n \mathbb{I}_{\{i \leq \mathcal{A}(s,T)\}} \chi(s) \phi(T) \gamma_i dW(s) \tag{B.54}$$

$$= f(0,T) + \int_0^t \chi(s) \phi(T) \gamma_1 \int_s^T \mathbb{I}_{\{s<x_1\}} \mathbb{I}_{\{u>x_1\}} \chi(s) \phi(u) \gamma_1 du ds \tag{B.55}$$

$$+ \int_0^t \chi(s) \phi(T) \gamma_1 dW(s) \tag{B.56}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \chi(s) \phi(T) \mathbb{I}_{\{s<x_1\}} \int_s^T \mathbb{I}_{\{u>x_1\}} \chi(s) \phi(u) du ds \tag{B.57}$$

$$+ \gamma_1 \int_0^t \chi(s) \phi(T) dW(s) \tag{B.58}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \chi(s) \phi(T) \int_s^T \mathbb{I}_{\{u>x_1\}} \chi(s) \phi(u) du ds \tag{B.59}$$

$$+ \gamma_1 \int_0^t \chi(s) \phi(T) dW(s) \tag{B.60}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \chi(s) \phi(T) \int_{x_1}^T \chi(s) \phi(u) du ds + \gamma_1 \int_0^t \chi(s) \phi(T) dW(s) \tag{B.61}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) \tag{B.62}$$

$$\times \int_{x_1}^T \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^u \lambda(v) dv \Big) du ds \tag{B.63}$$

$$+ \gamma_1 \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) dW(s) \tag{B.64}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \int_{x_1}^T \exp\Big(- \int_s^u \lambda(v) dv \Big) du ds \tag{B.65}$$

$$+ \gamma_1 \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.66}$$

$$= f(0,T) + \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda_{x_1}(s,T) ds \tag{B.67}$$

$$+ \gamma_1 \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.68}$$

let

$$y(t) = \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_1}(s,t) ds \tag{B.69}$$

$$+ \exp\Big(\int_t^T \lambda(v) dv \Big) \gamma_1 \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.70}$$

substitute

$$\gamma_1 \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) = \frac{y(t) - \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_1}(s,t) ds}{\exp\Big(\int_t^T \lambda(v) dv \Big)} \tag{B.71}$$

$$f(t,T) = f(0,T) + \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda_{x_1}(s,T) ds \tag{B.72}$$

$$+ \gamma_1 \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.73}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v) dv \Big) y(t) \tag{B.74}$$

$$+ \gamma_1^2 \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda_{x_1}(s,T) ds \tag{B.75}$$

$$- \gamma_1^2 \exp\Big(- \int_t^T \lambda(v) dv \Big) \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_1}(s,t) ds \tag{B.76}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v) dv \Big) y(t) + \gamma_1^2 \exp\Big(- \int_t^T \lambda(v) dv \Big) \tag{B.77}$$

$$\times \int_0^t \sigma^2(s) \Big\{ \exp\Big(\int_t^T \lambda(v) dv \Big) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda_{x_1}(s,T) \tag{B.78}$$

$$- \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_1}(s,t) \Big\} ds \tag{B.79}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v) dv \Big) y(t) + \gamma_1^2 \exp\Big(- \int_t^T \lambda(v) dv \Big) \tag{B.80}$$

$$\times \int_0^t \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Big\{ \Lambda_{x_1}(s,T) - \Lambda_{x_1}(s,t) \Big\} ds \tag{B.81}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v) dv \Big) y(t) \tag{B.82}$$

$$+ \gamma_1^2 \Lambda(t,T) \exp\Big(- \int_t^T \lambda(v) dv \Big) \int_0^t \sigma^2(s) \exp\Big(- 2 \int_s^t \lambda(v) dv \Big) ds \tag{B.83}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v) dv \Big) y(t) + \gamma_1^2 \Phi(t) \Lambda(t,T) \exp\Big(- \int_t^T \lambda(v) dv \Big) \tag{B.84}$$

therefore:

$$\int_{x_1}^{T} f(t,u)du = \int_{x_1}^{T} \Big(f(0,u) + \exp\Big(-\int_t^u \lambda(v)dv \Big) y(t) \tag{B.85}$$

$$+ \gamma_1^2 \Phi(t) \Lambda(t,u) \exp\Big(-\int_t^u \lambda(v)dv \Big) \Big) du \tag{B.86}$$

$$= \int_{x_1}^{T} f(0,u)du + y(t) \int_{x_1}^{T} \exp\Big(-\int_t^u \lambda(v)dv \Big) du \tag{B.87}$$

$$+ \gamma_1^2 \Phi(t) \int_{x_1}^{T} \Lambda(t,u) \exp\Big(-\int_t^u \lambda(v)dv \Big) du \tag{B.88}$$

$$= \int_{x_1}^{T} f(0,u)du + \Lambda_{x_1}(t,T)y(t) + \gamma_1^2 \Phi(t) \int_{x_1}^{T} \Lambda(t,u) d\Lambda(t,u) \tag{B.89}$$

$$= \int_{x_1}^{T} f(0,u)du + \Lambda_{x_1}(t,T)y(t) + \frac{1}{2}\gamma_1^2 \Phi(t) \{ \Lambda^2(t,T) - \Lambda^2(t,x_1) \} \tag{B.90}$$

therefore:

$$\int_t^{x_1} f(t,u)du + \int_{x_1}^{T} f(t,u)du = \int_t^{T} f(0,u)du + \Lambda_{x_1}(t,T)y(t) \tag{B.91}$$

$$+ \frac{1}{2}\gamma_1^2 \Phi(t) \{ \Lambda^2(t,T) - \Lambda^2(t,x_1) \} \tag{B.92}$$

Therefore, bond price:

$$B(t,T) = \exp\Big(-\int_t^T f(t,u)du \Big) \tag{B.93}$$

$$= \frac{B(0,T)}{B(0,t)} \exp\Big(-\Lambda_{x_1}(t,T)y(t) - \frac{1}{2}\gamma_1^2 \Phi(t) \{ \Lambda^2(t,T) - \Lambda^2(t,x_1) \} \Big) \tag{B.94}$$

B.2.3 More general case $t < x_1 < T$

In general we will need $\int_{x_a}^{T} f(t,u)du$ where $T < x_{a+1}$, therefore restrict $T \in [x_a, x_{a+1}]$ and $t < x_1$, HJM result:

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T) \int_s^T \sigma(s,u)duds + \int_0^t \sigma(s,T)dW(s) \tag{B.95}$$

$$= f(0,T) + \int_0^t \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(s,T)\}} \chi(s)\phi(T)\gamma_i \tag{B.96}$$

$$\times \int_s^T \sum_{j=1}^{n} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \chi(s)\phi(u)\gamma_j duds \tag{B.97}$$

$$+ \int_0^t \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(s,T)\}} \chi(s)\phi(T)\gamma_i dW(s) \tag{B.98}$$

$$= f(0,T) + \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i \int_s^T \sum_{j=1}^{n} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \chi(s)\phi(u)\gamma_j duds \tag{B.99}$$

$$+ \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i dW(s) \tag{B.100}$$

Now

$$\int_s^T \sum_{j=1}^{n} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \chi(s)\phi(u)\gamma_j du = \chi(s) \sum_{j=1}^{n} \gamma_j \int_s^T \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du \tag{B.101}$$

$$\int_s^T \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du = \int_s^{x_1} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du + \sum_{k=1}^{a-1} \int_{x_k}^{x_{k+1}} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du \tag{B.102}$$

$$+ \int_{x_a}^{T} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du \tag{B.103}$$

$$= \sum_{k=1}^{a-1} \int_{x_k}^{x_{k+1}} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du + \int_{x_a}^{T} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u)du \tag{B.104}$$

Therefore:

$$\sum_{j=1}^{n} \gamma_j \int_s^T \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u) du = \sum_{j=1}^{n} \gamma_j \sum_{k=1}^{a-1} \int_{x_k}^{x_{k+1}} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u) du \tag{B.105}$$

$$+ \sum_{j=1}^{n} \gamma_j \int_{x_a}^{T} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \phi(u) du \tag{B.106}$$

$$= \sum_{j=1}^{a-1} \gamma_j \int_{x_j}^{x_a} \phi(u) du + \sum_{j=1}^{a} \gamma_j \int_{x_a}^{T} \phi(u) du \tag{B.107}$$

$$= \sum_{j=1}^{a-1} \gamma_j \int_{x_j}^{T} \phi(u) du + \gamma_a \int_{x_a}^{T} \phi(u) du \tag{B.108}$$

$$= \sum_{j=1}^{a} \gamma_j \int_{x_j}^{T} \phi(u) du \tag{B.109}$$

Therefore:

$$f(t,T) = f(0,T) + \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i \int_s^T \sum_{j=1}^{n} \mathbb{I}_{\{j \leq \mathcal{A}(s,u)\}} \chi(s)\phi(u)\gamma_j du ds \tag{B.110}$$

$$+ \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i dW(s) \tag{B.111}$$

$$= f(0,T) + \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i \chi(s) \sum_{j=1}^{a} \gamma_j \int_{x_j}^T \phi(u) du ds \tag{B.112}$$

$$+ \int_0^t \chi(s)\phi(T) \sum_{i=1}^{a} \gamma_i dW(s) \tag{B.113}$$

$$= f(0,T) + \sum_{i=1}^{a} \sum_{j=1}^{a} \gamma_i \gamma_j \int_0^t \chi(s)\phi(T) \int_{x_j}^T \chi(s)\phi(u) du ds \tag{B.114}$$

$$+ \sum_{i=1}^{a} \gamma_i \int_0^t \chi(s)\phi(T) dW(s) \tag{B.115}$$

$$= f(0,T) + \sum_{i=1}^{a} \sum_{j=1}^{a} \gamma_i \gamma_j \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) \tag{B.116}$$

$$\times \int_{x_j}^T \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^u \lambda(v) dv \Big) du ds \tag{B.117}$$

$$+ \sum_{i=1}^{a} \gamma_i \int_0^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) dW(s) \tag{B.118}$$

$$= f(0,T) + \sum_{i=1}^{a} \sum_{j=1}^{a} \gamma_i \gamma_j \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \tag{B.119}$$

$$\times \int_{x_j}^T \exp\Big(- \int_s^u \lambda(v) dv \Big) du ds \tag{B.120}$$

$$+ \sum_{i=1}^{a} \gamma_i \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.121}$$

$$= f(0,T) + \sum_{i=1}^{a} \sum_{j=1}^{a} \gamma_i \gamma_j \int_0^t \sigma^2(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) \Lambda_{x_j}(s,T) ds \tag{B.122}$$

$$+ \sum_{i=1}^{a} \gamma_i \int_0^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.123}$$

let

$$y_a(t) = \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\int_0^t \sigma^2(s)\exp\Big(-\int_s^t \lambda(v)dv\Big)\Lambda_{x_j}(s,t)ds \tag{B.124}$$

$$+\exp\Big(\int_t^T \lambda(v)dv\Big)\sum_{i=1}^{a}\gamma_i\int_0^t \sigma(s)\exp\Big(-\int_s^T \lambda(v)dv\Big)dW(s) \tag{B.125}$$

substitute

$$\sum_{i=1}^{a}\gamma_i\int_0^t \sigma(s)\exp\Big(-\int_s^T \lambda(v)dv\Big)dW(s) \tag{B.126}$$

$$= \frac{y_a(t) - \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\int_0^t \sigma^2(s)\exp\Big(-\int_s^t \lambda(v)dv\Big)\Lambda_{x_j}(s,t)ds}{\exp\Big(\int_t^T \lambda(v)dv\Big)} \tag{B.127}$$

$$f(t,T) = f(0,T) + \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j \int_0^t \sigma^2(s)\exp\Big(- \int_s^T \lambda(v)dv\Big)\Lambda_{x_j}(s,T)ds \tag{B.128}$$

$$+ \sum_{i=1}^{a}\gamma_i \int_0^t \sigma(s)\exp\Big(- \int_s^T \lambda(v)dv\Big)dW(s) \tag{B.129}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv\Big)y_a(t) \tag{B.130}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j \int_0^t \sigma^2(s)\exp\Big(- \int_s^T \lambda(v)dv\Big)\Lambda_{x_j}(s,T)ds \tag{B.131}$$

$$- \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\exp\Big(- \int_t^T \lambda(v)dv\Big) \int_0^t \sigma^2(s)\exp\Big(- \int_s^t \lambda(v)dv\Big)\Lambda_{x_j}(s,t)ds \tag{B.132}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv\Big)y(t) + \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\exp\Big(- \int_t^T \lambda(v)dv\Big) \tag{B.133}$$

$$\times \int_0^t \sigma^2(s)\Big\{\exp\Big(\int_t^T \lambda(v)dv\Big)\exp\Big(- \int_s^T \lambda(v)dv\Big)\Lambda_{x_j}(s,T) \tag{B.134}$$

$$- \exp\Big(- \int_s^t \lambda(v)dv\Big)\Lambda_{x_j}(s,t)\Big\}ds \tag{B.135}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv\Big)y_a(t) + \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\exp\Big(- \int_t^T \lambda(v)dv\Big) \tag{B.136}$$

$$\times \int_0^t \sigma^2(s)\exp\Big(- \int_s^t \lambda(v)dv\Big)\Big\{\Lambda_{x_j}(s,T) - \Lambda_{x_j}(s,t)\Big\}ds \tag{B.137}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv\Big)y_a(t) \tag{B.138}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\Lambda(t,T)\exp\Big(- \int_t^T \lambda(v)dv\Big) \int_0^t \sigma^2(s)\exp\Big(- 2\int_s^t \lambda(v)dv\Big)ds \tag{B.139}$$

$$= f(0,T) + \exp\Big(- \int_t^T \lambda(v)dv\Big)y_a(t) \tag{B.140}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{a}\gamma_i\gamma_j\Phi(t)\Lambda(t,T)\exp\Big(- \int_t^T \lambda(v)dv\Big) \tag{B.141}$$

therefore:

$$\int_{x_a}^{T} f(t,u)du = \int_{x_a}^{T} \Big(f(0,u) + \exp\Big(-\int_t^u \lambda(v)dv \Big) y_a(t) \tag{B.142}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{a} \gamma_i\gamma_j \Phi(t)\Lambda(t,u)\exp\Big(-\int_t^u \lambda(v)dv \Big)\Big) du \tag{B.143}$$

$$= \int_{x_a}^{T} f(0,u)du + y_a(t)\int_{x_a}^{T} \exp\Big(-\int_t^u \lambda(v)dv \Big) du \tag{B.144}$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{a} \gamma_i\gamma_j \Phi(t) \int_{x_a}^{T} \Lambda(t,u)\exp\Big(-\int_t^u \lambda(v)dv \Big) du \tag{B.145}$$

$$= \int_{x_a}^{T} f(0,u)du + \Lambda_{x_a}(t,T)y_a(t) + \sum_{i=1}^{a}\sum_{j=1}^{a} \gamma_i\gamma_j \Phi(t) \int_{x_a}^{T} \Lambda(t,u)d\Lambda(t,u) \tag{B.146}$$

$$= \int_{x_a}^{T} f(0,u)du + \Lambda_{x_a}(t,T)y_a(t) + \frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{a} \gamma_i\gamma_j \Phi(t)\{\Lambda^2(t,T) - \Lambda^2(t,x_a)\} \tag{B.147}$$

define $\eta(t) = \min\{k | x_k > t\}$, now:

$$\int_{t}^{T} f(t,u)du = \int_{t}^{x_{\eta(t)}} f(t,u)du + \int_{x_{\eta(T)-1}}^{T} f(t,u)du + \sum_{k=\eta(t)}^{\eta(T)-2} \int_{x_k}^{x_{k+1}} f(t,u)du \tag{B.148}$$

$$= \int_{t}^{T} f(0,u)du + \Lambda_{x_{\eta(T)-1}}(t,T)y_{x_{\eta(T)-1}}(t) \tag{B.149}$$

$$+ \frac{1}{2}\sum_{i=1}^{x_{\eta(T)-1}}\sum_{j=1}^{x_{\eta(T)-1}} \gamma_i\gamma_j \Phi(t)\{\Lambda^2(t,T) - \Lambda^2(t,x_{\eta(T)-1})\} \tag{B.150}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \Big(\Lambda_{x_k}(t,x_{k+1})y_k(t) + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \gamma_i\gamma_j \Phi(t)\{\Lambda^2(t,x_{k+1}) - \Lambda^2(t,x_k)\}\Big) \tag{B.151}$$

Therefore, bond price:

$$B(t,T) = \exp\Big(- \int_t^T f(t,u)du \Big) \tag{B.152}$$

$$= \frac{B(0,T)}{B(0,t)} \exp\Big(\Lambda_{x_{\eta(T)-1}}(t,T) y_{x_{\eta(T)-1}}(t) \tag{B.153}$$

$$+ \frac{1}{2} \sum_{i=1}^{x_{\eta(T)-1}} \sum_{j=1}^{x_{\eta(T)-1}} \gamma_i \gamma_j \Phi(t) \{ \Lambda^2(t,T) - \Lambda^2(t, x_{\eta(T)-1}) \} \tag{B.154}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \Big(\Lambda_{x_k}(t, x_{k+1}) y_k(t) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \gamma_i \gamma_j \Phi(t) \{ \Lambda^2(t, x_{k+1}) - \Lambda^2(t, x_k) \} \Big) \Big) \tag{B.155}$$

B.2.4 General case $t < T$

In general we will need $\int_{x_a}^T f(t,u)du$ where $T < x_{a+1}$, therefore restrict $T \in [x_a, x_{a+1}]$ and $t < T$. Define $\eta(t) = \min\{b | x_b \geq t\}$, now HJM result:

$$f(t,T) = f(0,T) + \int_0^t \sigma(s,T) \int_s^T \sigma(s,u) du ds + \int_0^t \sigma(s,T) dW(s) \tag{B.156}$$

$$= f(0,T) + \int_{x_{\eta(t)-1}}^t \sigma(s,T) \int_s^T \sigma(s,u) du ds + \sum_{b=0}^{\eta(t)-2} \int_{x_b}^{x_{b+1}} \sigma(s,T) \int_s^T \sigma(s,u) du ds \tag{B.157}$$

$$+ \int_{x_{\eta(t)-1}}^t \sigma(s,T) dW(s) + \sum_{b=0}^{\eta(t)-2} \int_{x_b}^{x_{b+1}} \sigma(s,T) dW(s) \tag{B.158}$$

therefore in general we need to solve $\int_{x_b}^t \sigma(s,T) \int_s^T \sigma(s,u) du ds$ and $\int_{x_b}^t \sigma(s,T) dW(s)$ where $t \in [x_b, x_{b+1}]$. Now:

$$\int_{x_b}^t \sigma(s,T) dW(s) = \int_{x_b}^t \sum_{i=1}^{n} \mathbb{I}_{\{i \leq \mathcal{A}(s,T)\}} \chi(s) \phi(T) \gamma_i dW(s) \tag{B.159}$$

$$= \int_{x_b}^t \chi(s) \phi(T) \sum_{i=1}^{a-b} \gamma_i dW(s) \tag{B.160}$$

$$= \sum_{i=1}^{a-b} \gamma_i \int_{x_b}^t \sigma(s) \exp\Big(\int_0^s \lambda(v) dv \Big) \exp\Big(- \int_0^T \lambda(v) dv \Big) dW(s) \tag{B.161}$$

$$= \sum_{i=1}^{a-b} \gamma_i \int_{x_b}^t \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.162}$$

Also:

$$\int_{x_b}^{t}\sigma(s,T)\int_{s}^{T}\sigma(s,u)duds=\int_{x_b}^{t}\sum_{i=1}^{n}\mathbb{I}_{\{i\leq\mathcal{A}(s,T)\}}\chi(s)\phi(T)\gamma_i\int_{s}^{T}\sum_{j=1}^{n}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\chi(s)\phi(u)\gamma_j duds \tag{B.163}$$

$$=\int_{x_b}^{t}\chi(s)\phi(T)\sum_{i=1}^{a-b}\gamma_i\int_{s}^{T}\sum_{j=1}^{n}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\chi(s)\phi(u)\gamma_j duds \tag{B.164}$$

$$=\int_{x_b}^{t}\chi(s)\phi(T)\sum_{i=1}^{a-b}\gamma_i\chi(s)\sum_{j=1}^{n}\gamma_j\int_{s}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)duds \tag{B.165}$$

Now, implicitly with $s\in[x_b,t]$:

$$\int_{s}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du=\int_{s}^{x_1}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du+\sum_{k=1}^{a-1}\int_{x_k}^{x_{k+1}}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du \tag{B.166}$$

$$+\int_{x_a}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du \tag{B.167}$$

$$=\sum_{k=b+1}^{a-1}\int_{x_k}^{x_{k+1}}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du+\int_{x_a}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du \tag{B.168}$$

Therefore:

$$\sum_{j=1}^{n}\gamma_j\int_{s}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du=\sum_{j=1}^{n}\gamma_j\sum_{k=b+1}^{a-1}\int_{x_k}^{x_{k+1}}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du \tag{B.169}$$

$$+\sum_{j=1}^{n}\gamma_j\int_{x_a}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)du \tag{B.170}$$

$$=\sum_{j=1}^{a-b-1}\gamma_j\int_{x_{b+j}}^{x_a}\phi(u)du+\sum_{j=1}^{a-b}\gamma_j\int_{x_a}^{T}\phi(u)du \tag{B.171}$$

$$=\sum_{j=1}^{a-b-1}\gamma_j\int_{x_{b+j}}^{T}\phi(u)du+\gamma_{a-b}\int_{x_a}^{T}\phi(u)du \tag{B.172}$$

$$=\sum_{j=1}^{a-b}\gamma_j\int_{x_{b+j}}^{T}\phi(u)du \tag{B.173}$$

Therefore:

$$\int_{x_b}^{t}\sigma(s,T)\int_{s}^{T}\sigma(s,u)duds=\int_{x_b}^{t}\chi(s)\phi(T)\sum_{i=1}^{a-b}\gamma_i\chi(s)\sum_{j=1}^{n}\gamma_j\int_{s}^{T}\mathbb{I}_{\{j\leq\mathcal{A}(s,u)\}}\phi(u)duds \tag{B.174}$$

$$=\int_{x_b}^{t}\chi(s)\phi(T)\sum_{i=1}^{a-b}\gamma_i\sum_{j=1}^{a-b}\gamma_j\int_{x_{b+j}}^{T}\chi(s)\phi(u)duds \tag{B.175}$$

$$=\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{t}\chi(s)\phi(T)\int_{x_{b+j}}^{T}\chi(s)\phi(u)duds \tag{B.176}$$

$$=\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{t}\sigma(s)\exp\Big(\int_0^s\lambda(v)dv\Big)\exp\Big(-\int_0^T\lambda(v)dv\Big) \tag{B.177}$$

$$\times\int_{x_{b+j}}^{T}\sigma(s)\exp\Big(\int_0^s\lambda(v)dv\Big)\exp\Big(-\int_0^u\lambda(v)dv\Big)duds \tag{B.178}$$

$$=\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{t}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\int_{x_{b+j}}^{T}\exp\Big(-\int_s^u\lambda(v)dv\Big)duds \tag{B.179}$$

$$=\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{t}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,T)ds \tag{B.180}$$

Rewrite (B.156):

$$f(t,T)=f(0,T)+\int_0^t\sigma(s,T)\int_s^T\sigma(s,u)duds+\int_0^t\sigma(s,T)dW(s) \tag{B.181}$$

$$=f(0,T)+\sum_{b=0}^{\eta(t)-2}\left\{\int_{x_b}^{x_{b+1}}\sigma(s,T)\int_s^T\sigma(s,u)duds+\int_{x_b}^{x_{b+1}}\sigma(s,T)dW(s)\right\} \tag{B.182}$$

$$+\int_{x_{\eta(t)-1}}^{t}\sigma(s,T)\int_s^T\sigma(s,u)duds+\int_{x_{\eta(t)-1}}^{t}\sigma(s,T)dW(s) \tag{B.183}$$

$$=f(0,T)+\sum_{b=0}^{\eta(t)-2}\left\{\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{x_{b+1}}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,T)ds\right. \tag{B.184}$$

$$\left.+\sum_{i=1}^{a-b}\gamma_i\int_{x_b}^{x_{b+1}}\sigma(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)dW(s)\right\} \tag{B.185}$$

$$+\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,T)ds \tag{B.186}$$

$$+\sum_{i=1}^{a-\eta(t)+1}\gamma_i\int_{x_{\eta(t)-1}}^{t}\sigma(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)dW(s) \tag{B.187}$$

let

$$y_a(t) = \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \int_{x_{\eta(t)-1}}^{t} \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_{(\eta(t)-1+j)}}(s,t) ds \tag{B.188}$$

$$+ \exp\Big(\int_t^T \lambda(v) dv \Big) \sum_{i=1}^{a-\eta(t)+1} \gamma_i \int_{x_{\eta(t)-1}}^{t} \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.189}$$

substitute

$$\sum_{i=1}^{a-\eta(t)+1} \gamma_i \int_{x_{\eta(t)-1}}^{t} \sigma(s) \exp\Big(- \int_s^T \lambda(v) dv \Big) dW(s) \tag{B.190}$$

$$= \frac{y_a(t) - \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \int_{x_{\eta(t)-1}}^{t} \sigma^2(s) \exp\Big(- \int_s^t \lambda(v) dv \Big) \Lambda_{x_{(\eta(t)-1+j)}}(s,t) ds}{\exp\Big(\int_t^T \lambda(v) dv \Big)} \tag{B.191}$$

Define

$$\Phi(u,t) = \int_u^t \sigma^2(s) \exp\Big(- 2 \int_s^t \lambda(v) dv \Big) ds \tag{B.192}$$

therefore

$$\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{x_{b+1}}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,T)ds \tag{B.193}$$

$$+\sum_{i=1}^{a-b}\gamma_i\int_{x_b}^{x_{b+1}}\sigma(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)dW(s) \tag{B.194}$$

$$=\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big)y_a(x_{b+1}) \tag{B.195}$$

$$+\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\int_{x_b}^{x_{b+1}}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,T)ds \tag{B.196}$$

$$-\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big)\int_{x_b}^{x_{b+1}}\sigma^2(s)\exp\Big(-\int_s^{x_{b+1}}\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,x_{b+1})ds \tag{B.197}$$

$$=\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big)y_a(x_{b+1})+\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\Big\{\gamma_i\gamma_j\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big) \tag{B.198}$$

$$\times\Big[\int_{x_b}^{x_{b+1}}\sigma^2(s)\Big\{\exp\Big(\int_{x_{b+1}}^T\lambda(v)dv\Big)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,T) \tag{B.199}$$

$$-\exp\Big(-\int_s^{x_{b+1}}\lambda(v)dv\Big)\Lambda_{x_{b+j}}(s,x_{b+1})\Big\}ds\Big]\Big\} \tag{B.200}$$

$$=\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big)y_a(x_{b+1})+\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\Lambda(x_{b+1},T)\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big) \tag{B.201}$$

$$\times\int_{x_b}^{x_{b+1}}\sigma^2(s)\exp\Big(-2\int_s^{x_{b+1}}\lambda(v)dv\Big)ds \tag{B.202}$$

$$=\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big)y_a(x_{b+1})+\sum_{i=1}^{a-b}\sum_{j=1}^{a-b}\gamma_i\gamma_j\Phi(x_b,x_{b+1})\Lambda(x_{b+1},T)\exp\Big(-\int_{x_{b+1}}^T\lambda(v)dv\Big) \tag{B.203}$$

and

$$\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,T)ds \tag{B.204}$$

$$+\sum_{i=1}^{a-\eta(t)+1}\gamma_i\int_{x_{\eta(t)+1}}^{t}\sigma(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)dW(s) \tag{B.205}$$

$$=\exp\Big(-\int_t^T\lambda(v)dv\Big)y_a(t) \tag{B.206}$$

$$+\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,T)ds \tag{B.207}$$

$$-\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j \tag{B.208}$$

$$\times\exp\Big(-\int_t^T\lambda(v)dv\Big)\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\exp\Big(-\int_s^t\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,t)ds \tag{B.209}$$

$$=\exp\Big(-\int_t^T\lambda(v)dv\Big)y_a(t)+\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\Big\{\gamma_i\gamma_j\exp\Big(-\int_t^T\lambda(v)dv\Big) \tag{B.210}$$

$$\times\Big[\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\Big\{\exp\Big(\int_t^T\lambda(v)dv\Big)\exp\Big(-\int_s^T\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,T) \tag{B.211}$$

$$-\exp\Big(-\int_s^t\lambda(v)dv\Big)\Lambda_{x_{(\eta(t)-1+j)}}(s,t)\Big\}ds\Big]\Big\} \tag{B.212}$$

$$=\exp\Big(-\int_t^T\lambda(v)dv\Big)y_a(t)+\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j\Lambda(t,T)\exp\Big(-\int_t^T\lambda(v)dv\Big) \tag{B.213}$$

$$\times\int_{x_{\eta(t)-1}}^{t}\sigma^2(s)\exp\Big(-2\int_s^t\lambda(v)dv\Big)ds \tag{B.214}$$

$$=\exp\Big(-\int_t^T\lambda(v)dv\Big)y_a(t)+\sum_{i=1}^{(a-\eta(t)+1)}\sum_{j=1}^{(a-\eta(t)+1)}\gamma_i\gamma_j\Phi(x_{\eta(t)-1},t)\Lambda(t,T)\exp\Big(-\int_t^T\lambda(v)dv\Big) \tag{B.215}$$

Therefore:

$$f(t,T) = f(0,T) + \sum_{b=0}^{\eta(t)-2} \Bigg\{ \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \int_{x_b}^{x_{b+1}} \sigma^2(s) \exp\Bigg(- \int_s^T \lambda(v) dv \Bigg) \Lambda_{x_{b+j}}(s,T) ds \tag{B.216}$$

$$+ \sum_{i=1}^{a-b} \gamma_i \int_{x_b}^{x_{b+1}} \sigma(s) \exp\Bigg(- \int_s^T \lambda(v) dv \Bigg) dW(s) \Bigg\} \tag{B.217}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \int_{x_{\eta(t)-1}}^{t} \sigma^2(s) \exp\Bigg(- \int_s^T \lambda(v) dv \Bigg) \Lambda_{x_{(\eta(t)-1+j)}}(s,T) ds \tag{B.218}$$

$$+ \sum_{i=1}^{a-\eta(t)+1} \gamma_i \int_{x_{\eta(t)-1}}^{t} \sigma(s) \exp\Bigg(- \int_s^T \lambda(v) dv \Bigg) dW(s) \tag{B.219}$$

$$= f(0,T) + \sum_{b=0}^{\eta(t)-2} \Bigg\{ \exp\Bigg(- \int_{x_{b+1}}^T \lambda(v) dv \Bigg) y_a(x_{b+1}) \tag{B.220}$$

$$+ \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \Lambda(x_{b+1}, T) \exp\Bigg(- \int_{x_{b+1}}^T \lambda(v) dv \Bigg) \Bigg\} \tag{B.221}$$

$$+ \exp\Bigg(- \int_t^T \lambda(v) dv \Bigg) y_a(t) \tag{B.222}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \Lambda(t,T) \exp\Bigg(- \int_t^T \lambda(v) dv \Bigg) \tag{B.223}$$

$$= f(0,T) + \sum_{b=0}^{\eta(t)-2} \exp\Bigg(- \int_{x_{b+1}}^T \lambda(v) dv \Bigg) y_a(x_{b+1}) + \exp\Bigg(- \int_t^T \lambda(v) dv \Bigg) y_a(t) \tag{B.224}$$

$$+ \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \Lambda(x_{b+1}, T) \exp\Bigg(- \int_{x_{b+1}}^T \lambda(v) dv \Bigg) \tag{B.225}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \Lambda(t,T) \exp\Bigg(- \int_t^T \lambda(v) dv \Bigg) \tag{B.226}$$

therefore:

$$\int_{x_a}^{T} f(t,u)du = \int_{x_a}^{T} \Big(f(0,u) + \sum_{b=0}^{\eta(t)-2} \exp\Big(-\int_{x_{b+1}}^{u} \lambda(v)dv \Big) y_a(x_{b+1}) \tag{B.227}$$

$$+ \exp\Big(-\int_{t}^{u} \lambda(v)dv \Big) y_a(t) \tag{B.228}$$

$$+ \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \Lambda(x_{b+1}, u) \exp\Big(-\int_{x_{b+1}}^{u} \lambda(v)dv \Big) \tag{B.229}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \Lambda(t,u) \exp\Big(-\int_{t}^{u} \lambda(v)dv \Big) \Big) du \tag{B.230}$$

$$= \int_{x_a}^{T} f(0,u)du + \sum_{b=0}^{\eta(t)-2} y_a(x_{b+1}) \int_{x_a}^{T} \exp\Big(-\int_{x_{b+1}}^{u} \lambda(v)dv \Big) du \tag{B.231}$$

$$+ y_a(t) \int_{x_a}^{T} \exp\Big(-\int_{t}^{u} \lambda(v)dv \Big) du \tag{B.232}$$

$$+ \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \int_{x_a}^{T} \Lambda(x_{b+1}, u) \exp\Big(-\int_{x_{b+1}}^{u} \lambda(v)dv \Big) du \tag{B.233}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \int_{x_a}^{T} \Lambda(t,u) \exp\Big(-\int_{t}^{u} \lambda(v)dv \Big) du \tag{B.234}$$

$$= \int_{x_a}^{T} f(0,u)du + \sum_{b=0}^{\eta(t)-2} \Lambda_{x_a}(x_{b+1}, T) y_a(x_{b+1}) + \Lambda_{x_a}(t,T) y_a(t) \tag{B.235}$$

$$+ \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \int_{x_a}^{T} \Lambda(x_{b+1}, u) d\Lambda(x_{b+1}, u) \tag{B.236}$$

$$+ \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \int_{x_a}^{T} \Lambda(t,u) d\Lambda(t,u) \tag{B.237}$$

$$= \int_{x_a}^{T} f(0,u)du + \sum_{b=0}^{\eta(t)-2} \Lambda_{x_a}(x_{b+1}, T) y_a(x_{b+1}) + \Lambda_{x_a}(t,T) y_a(t) \tag{B.238}$$

$$+ \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{a-b} \sum_{j=1}^{a-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \{ \Lambda^2(x_{b+1}, T) - \Lambda^2(x_{b+1}, x_a) \} \tag{B.239}$$

$$+ \frac{1}{2} \sum_{i=1}^{(a-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \{ \Lambda^2(t, T) - \Lambda^2(t, x_a) \} \tag{B.240}$$

now

$$\int_t^T f(t,u)du = \int_t^{x_{\eta(t)}} f(t,u)du + \int_{x_{\eta(T)-1}}^T f(t,u)du + \sum_{k=\eta(t)}^{\eta(T)-2} \int_{x_k}^{x_{k+1}} f(t,u)du \tag{B.241}$$

$$= \int_t^T f(0,u)du + \sum_{b=0}^{\eta(t)-2} \Lambda_{x_{\eta(T)-1}}(x_{b+1},T)y_{\eta(T)-1}(x_{b+1}) + \Lambda_{x_{\eta(T)-1}}(t,T)y_{\eta(T)-1}(t) \tag{B.242}$$

$$+ \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{\eta(T)-1-b} \sum_{j=1}^{\eta(T)-1-b} \gamma_i\gamma_j\Phi(x_b,x_{b+1})\{\Lambda^2(x_{b+1},T) - \Lambda^2(x_{b+1},x_{\eta(T)-1})\} \tag{B.243}$$

$$+ \frac{1}{2} \sum_{i=1}^{\eta(T)-\eta(t)} \sum_{j=1}^{\eta(T)-\eta(t)} \gamma_i\gamma_j\Phi(x_{\eta(t)-1},t)\{\Lambda^2(t,T) - \Lambda^2(t,x_{\eta(T)-1})\} \tag{B.244}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \Bigg[\sum_{b=0}^{\eta(t)-2} \Lambda_{x_k}(x_{b+1},x_{k+1})y_k(x_{b+1}) + \Lambda_{x_k}(t,x_{k+1})y_k(t) \tag{B.245}$$

$$+ \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{k-b} \sum_{j=1}^{k-b} \gamma_i\gamma_j\Phi(x_b,x_{b+1})\{\Lambda^2(x_{b+1},x_{k+1}) - \Lambda^2(x_{b+1},x_k)\} \tag{B.246}$$

$$+ \frac{1}{2} \sum_{i=1}^{(k-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i\gamma_j\Phi(x_{\eta(t)-1},t)\{\Lambda^2(t,x_{k+1}) - \Lambda^2(t,x_k)\}\Bigg] \tag{B.247}$$

$$= \int_t^T f(0,u)du \tag{B.248}$$

$$+ \sum_{b=0}^{\eta(t)-2} \Lambda_{x_{\eta(T)-1}}(x_{b+1},T)y_{\eta(T)-1}(x_{b+1}) \tag{B.249}$$

$$+ \Lambda_{x_{\eta(T)-1}}(t,T)y_{\eta(T)-1}(t) \tag{B.250}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \sum_{b=0}^{\eta(t)-2} \Lambda_{x_k}(x_{b+1},x_{k+1})y_k(x_{b+1}) \tag{B.251}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \Lambda_{x_k}(t,x_{k+1})y_k(t) \tag{B.252}$$

$$+ \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{\eta(T)-1-b} \sum_{j=1}^{\eta(T)-1-b} \gamma_i\gamma_j\Phi(x_b,x_{b+1})\{\Lambda^2(x_{b+1},T) - \Lambda^2(x_{b+1},x_{\eta(T)-1})\} \tag{B.253}$$

$$+ \frac{1}{2} \sum_{i=1}^{\eta(T)-\eta(t)} \sum_{j=1}^{\eta(T)-\eta(t)} \gamma_i\gamma_j\Phi(x_{\eta(t)-1},t)\{\Lambda^2(t,T) - \Lambda^2(t,x_{\eta(T)-1})\} \tag{B.254}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{k-b} \sum_{j=1}^{k-b} \gamma_i\gamma_j\Phi(x_b,x_{b+1})\{\Lambda^2(x_{b+1},x_{k+1}) - \Lambda^2(x_{b+1},x_k)\} \tag{B.255}$$

$$+ \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{i=1}^{(k-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i\gamma_j\Phi(x_{\eta(t)-1},t)\{\Lambda^2(t,x_{k+1}) - \Lambda^2(t,x_k)\} \tag{B.256}$$

Therefore, bond price:

$$B(t,T) = \exp\Big(- \int_t^T f(t,u)du \Big) \tag{B.257}$$

$$= \frac{B(0,T)}{B(0,t)} \exp\Big(- \sum_{b=0}^{\eta(t)-2} \Lambda_{x_{\eta(T)-1}}(x_{b+1},T) y_{\eta(T)-1}(x_{b+1}) \tag{B.258}$$

$$- \Lambda_{x_{\eta(T)-1}}(t,T) y_{\eta(T)-1}(t) \tag{B.259}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \sum_{b=0}^{\eta(t)-2} \Lambda_{x_k}(x_{b+1},x_{k+1}) y_k(x_{b+1}) \tag{B.260}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \Lambda_{x_k}(t,x_{k+1}) y_k(t) \tag{B.261}$$

$$- \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{\eta(T)-1-b} \sum_{j=1}^{\eta(T)-1-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \{ \Lambda^2(x_{b+1},T) - \Lambda^2(x_{b+1}, x_{\eta(T)-1}) \} \tag{B.262}$$

$$- \frac{1}{2} \sum_{i=1}^{\eta(T)-\eta(t)} \sum_{j=1}^{\eta(T)-\eta(t)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \{ \Lambda^2(t,T) - \Lambda^2(t, x_{\eta(T)-1}) \} \tag{B.263}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{b=0}^{\eta(t)-2} \sum_{i=1}^{k-b} \sum_{j=1}^{k-b} \gamma_i \gamma_j \Phi(x_b, x_{b+1}) \{ \Lambda^2(x_{b+1}, x_{k+1}) - \Lambda^2(x_{b+1}, x_k) \} \tag{B.264}$$

$$- \sum_{k=\eta(t)}^{\eta(T)-2} \frac{1}{2} \sum_{i=1}^{(k-\eta(t)+1)} \sum_{j=1}^{(a-\eta(t)+1)} \gamma_i \gamma_j \Phi(x_{\eta(t)-1}, t) \{ \Lambda^2(t, x_{k+1}) - \Lambda^2(t, x_k) \} \Big) \tag{B.265}$$